Article

Visual-Based Person Detection for Search-and-Rescue with UAS: Humans vs. Machine Learning Algorithm

Sven Gotovac ^{1,*}, Danijel Zelenika ², Željko Marušić ³ and Dunja Božić-Štulić ¹

¹ Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, 21000 Split, Croatia; dgotovac@fesb.hr

² Faculty of Mechanical Engineering, Computing and Electrical Engineering, University of Mostar, 88000 Mostar, Bosnia and Herzegovina; danijel.zelenika@fsre.sum.ba

³ Faculty of Science and Education, University of Mostar, 88000 Mostar, Bosnia and Herzegovina; zeljko.marusic@fpmoz.sum.ba

* Correspondence: sven.gotovac@fesb.hr; Tel.: +385-91-4305850

Received: 31 August 2020; Accepted: 8 October 2020; Published: 10 October 2020

Abstract: Unmanned Aircraft Systems (UASs) have been recognized as an important resource in search-and-rescue (SAR) missions and, as such, have been used by the Croatian Mountain Search and Rescue (CMRS) service for over seven years. The UAS scans and photographs the terrain. The high-resolution images are afterwards analyzed by SAR members to detect missing persons or to find some usable trace. It is a drawn out, tiresome process prone to human error. To facilitate and speed up mission image processing and increase detection accuracy, we have developed several image-processing algorithms. The latest are convolutional neural network (CNN)-based. CNNs were trained on a specially developed image database, named HERIDAL. Although these algorithms achieve excellent recall, the efficiency of the algorithm in actual SAR missions and its comparison with expert detection must be investigated. A series of mission simulations are planned and recorded for this purpose. They are processed and labelled by a developed algorithm. A web application was developed by which experts analyzed raw and processed mission images. The algorithm achieved better recall compared to an expert, but the experts achieved better accuracy when they analyzed images that were already processed and labelled.

Keywords: search and rescue (SAR); Unmanned Aircraft System (UAS); Deep CNN; aerial image processing; CNN; expert to CNN comparison

1. Introduction

Unmanned Aircraft Systems (UASs) have been recognized as a valuable new resource in practically all emergency services. According to [1], “up to 2017, at least 59 individuals have been rescued by drones from life-threatening conditions in 18 separate incidents around the globe”. In Croatia, besides the army and police, emergency services such as firefighters and the Croatian Mountain Rescue Service (CMRS) make intensive use of UASs. CMRS is a leader in the use of this technology in emergency situations. They have been using UAS technology intensively for the last six years [2]. At first, Unmanned Aerial Vehicles (UAV) were used for a terrain scouting and fast searching of areas of interest. Within a short time, this technology has been fully adopted and is now used intensively in all phases of SAR missions and on all types of terrain. In 2018, CMRS established a special UAS department. A new internal training and licensing program for UAS pilots has been developed and authorized by regulatory authorities. Forty-nine CMRS pilots successfully finished the program and have become licensed pilots. Currently, CMRS in total operates 40 UASs. By August 2020 alone, CMRS had detected and thereby rescued five people using drones.

The use of UASs, despite its evident advantages, has opened additional requirements. Besides the licensed pilots, the CMRS team also needs experienced personnel to process the information collected by the UAS. This primarily refers to the image processing acquired by UASs with a goal either of detecting lost persons or of finding some trace that may be helpful for the mission. How demanding this task is can be shown by the example of the search mission for the Pole Lukasz Dariusz, who became lost on 31 July 2020 on the mountain Biokovo. The SAR mission lasted for intensive 10 days, unfortunately without success. Over 340 rescuers and 6 drones were engaged in it. In the search procedure, drones take pictures of the terrain every 4 s, and within one flight they generate approximately 300 high-resolution images. More than 2000 images are generated per search day. The images are uploaded to the Cloud and analyzed by the CMSR personnel, namely people in the base, at home, and in the field. It is important to note that the processing of the aerial images of the Mediterranean karst landscape is very demanding due to the colors and shadows, as well as the number of different objects. The target objects are relatively small and often camouflaged within the environment, making detection a challenging and demanding task. Complexities of the images are shown in Figure 1. Empirically, it has been determined that the person who analyzes images needs 5 to 45 s to process a single image, depending on the complexity of the composition. These numbers indicate that image analyses require many personnel as well as a relatively long processing time, and both present a considerable problem.

Regarding this issue, a research group from the University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), Croatia, has been intensively working on the development of image processing algorithms for person detection in the aerial images taken from drones in the Mediterranean karst landscape for the SAR missions. More than 10 years ago, a mean-shift-based algorithm was developed and implemented [3]. This algorithm achieved a satisfactory level of detection and, by suggesting potential suspect locations in images, significantly simplified and sped up the visual inspection process. The drawback of this approach was its relatively long processing time as well as the relatively significant number of false detections that were counterproductive to the visual inspection process. In the next step, to overcome these disadvantages, we developed a completely new algorithm based on a visual attention algorithm that detects the salient or most prominent segments in the image [4]. The detected segments are afterwards processed/classified by fine-tuned convolutional neural networks (CNNs) to select the regions most likely to contain a person. We established a special database called HERIDAL for CNN training and testing purposes. The HERIDAL database contains over 68,750 image patches of people in wilderness viewed from an aerial perspective, as well as 500 labelled, full-size 4000×3000 -pixel real-world images which all contain at least one person for testing purposes. In parallel at that time, we were working on the solution to this problem with the state-of-the-art Faster R-CNN algorithm [5]. Although Faster R-CNN has shown excellent results in many complex computer vision detection problems, it expressed the problem of detecting smaller objects, which has been the subject of numerous scientific studies [6–8]. Faster R-CNN has been trained on full-size images from the HERIDAL database. A comparison of the developed methods was made on the HERIDAL image database, and the following results were obtained [4]: the mean-shift algorithm achieved 74.7% recall and 18.7% precision, Faster R-CNN 85.0% recall and 58.1% precision, while saliency + CNN 88.9% recall and 34.8% precision. Due to the great complexity of the images, in which it is difficult even for an experienced SAR operator to detect a person in the image, it could be concluded that the obtained results are extraordinary. This indicates that the proposed algorithms have great potential to be successfully applied within the SAR missions.

Figure 1. An example of the complexity of aerial images and enlarged parts of where people are located.

Despite the excellent results achieved by these algorithms, it is important to evaluate how effective they are in actual SAR missions; therefore, it is necessary to compare the efficiency of the proposed algorithms and the SAR operator. In actual SAR missions, the operator analyzes hundreds of images. With time, the operator's attention decreases and the probability of oversight increases. This is not a problem for the algorithm. On the other hand, an experienced operator will probably spot a lost

person or a potential clue in a new environment for which the algorithm may not be adequately trained. It is also important to keep in mind that the operator needs up to a half-minute to process the image, while the algorithm does so in less than a second. Furthermore, distributed processing across multiple GPUs makes mission processing time negligible. The importance of time in SAR missions does not need to be especially emphasized. Therefore, to evaluate the effectiveness of this technology in SAR missions, we have conducted additional research.

Based on the experience of CMSR, statistics of previous SAR missions, as well as statistics collected and published by the SAR services of other countries [9], simulations of real situations encountered in SAR missions were designed. Typical terrains where SAR missions are conducted were defined, as well as clothes worn by people who have been lost, poses and places in which such people could be found, or where they could be hidden. All persons simulating lost persons were marked by a GPS locator for subsequent position labelling. The search mission simulation was performed with typical flight settings, a flight altitude Above Ground Level (AGL) of 50 m, and an image overlap of 40%. It was also considered that the simulations be carried out when certain weather conditions are met (very sunny, cloudy, etc.) as well as at certain times of the day (morning, noon, early evening). From the numerous simulations of SAR missions, the ones that best illustrate real situations were selected and added to the HERIDAL image database. Selected missions were processed by developed algorithms with the goal of further statistical evaluation, i.e., recall and precision calculation. It is important to highlight that, in real SAR missions, a large number of images contain no person on them. Those missions processed by the specially developed algorithm based on CNN with labelled region proposals, as well as non-processed missions, were added to the HERIDAL image database.

The final goal of this research is to estimate and compare the reliability as well as the efficiency of detecting a lost person by SAR staff and by the developed image-processing/artificial-intelligence algorithm. The remainder of the paper is organized as follows. Section 2 provides a brief review of the relevant literature. Section 3 describes the theoretical basis, particularities, and the procedure of SAR mission implementation. This is the basis for planning mission simulations to be used in the research. The algorithms used for person detection are presented in Section 4. Section 5 presents the experiment and the results of the study as well as discussion regarding the detection recall and precision between the expert and the proposed algorithms. This paper ends with the conclusions and proposals for further research.

2. Related Work

Person detection from an aerial perspective is a very complex computer vision task due to the wide range of possible object appearances caused by different poses, clothing color, lighting, and background [10]. For years, researchers have developed different approaches that can identify objects [11–13]. Standard object detection involves relatively large objects with sharp edges and corners. In UAV imagery, the objects are small, and we face image instability because the platform is moving. All these issues make this task very difficult. The standard approach for person detection generally addresses large object sizes within an image together with texture and shape information. One of the approaches that has been proposed is bimodal systems. A bimodal system is composed of thermal and optical images taken from UAVs to improve detection rate. Anna Gaszczak and Breckon [14] used both thermal and visible imagery to develop a real-time person and vehicle detection system by fusing the two image sources. Rudol and Doherty [15] also used thermal and visible imagery to find persons in different poses on the ground in video sequences. The first step is to identify high-temperature regions corresponding to human body silhouettes. After detection of corresponding regions, they analyze the visible spectrum using a cascade of boosted classifiers working with Haar-like features. In [16] the authors proposed a real-time model for the detection swimmers in open water using unsupervised learning. Detecting humans from aerial images in [17] was managed by pyramidal feature extraction of SSD for human detection and action recognition, while in [18] the authors proposed a model for human detection based on color and depth data.

Transfer learning is a powerful machine learning technique that enables us to reuse an already trained neural network on our own dataset. This technique is particularly useful in situations where we do not have enough data; for example, remote sense images are a kind of data that are complicated to collect. In the literature there were attempts to overcome this problem by using transfer learning. The authors in [19] proposed a model for semantic segmentation of remotely sensed images. The main problem was a lack of real-world images, so they used domain specific transfer learning. First, they trained the proposed model on a large dataset and then they used that model and re-trained it on a smaller dataset. The authors reported great performances of this technique. In [20] transfer learning was used to accelerate the classification of remote sensing data. The authors reported that using transfer learning on larger and generic natural images datasets can outperform using transfer learning on small remotely sensed datasets. In [21] the authors used transfer learning in a region-based CNN called a Double Multi-scale Feature Pyramid Network (DMFPN) where they used inherent multi-scale pyramidal features and combined them with low-resolution features and high-resolution features using transfer learning. The authors in [22] proposed a deep neural network model for the classification of SAR images that does not require a large labeled dataset by using transfer learning. As we can see from previous papers, transfer learning has shown great potential in overcoming a lack of data for training deep learning models and represents a step towards strong machine learning. Mediterranean regions are characterized by very hot summers, and since the search missions are during the summer period, thermal cameras are not suitable for this kind of problem. In the previous study, throughout the IPSAR project, we used optical cameras for image acquisition. In one paper [3], a method based primarily on the mean-shift segmentation algorithm was proposed. After segmentation is tuned for small segments, the heuristic rules approach is used, for example, using the sizes of segments and clusters to make decisions. The mean-shift algorithm was selected primarily because it had demonstrated good results regarding stability and segmentation quality. The algorithm is divided into two-stage mean-shift segmentation in order to reduce the high computational requirements and the quadratic computational complexity of the algorithm. This approach resulted in only a minor loss of accuracy. In [23], the authors used the aforementioned detection model to conduct performance comparisons of the system on compressive-sensing-reconstructed images and original images, focusing primarily on image quality and information exchange. In [24], the authors tried different approaches, applying and analyzing various salient detection algorithms to detect lost persons. In [4], the authors proposed a two-stage algorithm approach based on salient detection and convolution neural networks. This approach showed promising results, but the false alarm rate was reported as an issue.

Labelled-image databases are usually used for the training and testing of deep neural networks and were also applied in our research. The research community has recognized the importance of the impact of the label errors (label noise) in training datasets on the model accuracy and have introduced works attempting to understand noisy training labels [25].

In [26], the authors aimed to obtain a better understanding of the similarities and differences in overt classification behavior—and thus, very likely, computation—between Deep Neural Networks (DNN) and human vision.

Our research addressed a completely different problem: the influence of the algorithm on the decision of the expert as well as the comparison of algorithm and expert accuracy. Such research has been conducted especially in bioinformatics as well as in medicine in computer-aided diagnostics.

In [27], the authors used CNNs trained with a low number of Epifluorescence Microscopy (EFM) images representing biofilms of different bacterial compositions and compared the performance of CNNs versus human experts in correctly classifying new images. Obtained results indicated that neural networks achieved an accuracy of 92.8% compared to 51.5% by human experts.

In [28], the authors compared the diagnostic performance of an artificial intelligence deep learning system with that of expert neuro-ophthalmologists in classifying optic disc appearance. They first trained and validated the CNN on 14,341 ocular fundus photographs from 19 international centers. The experiment was conducted on 800 new fundus photographs. The results of the CNN classification

were compared with those of two expert neuro-ophthalmologists who independently reviewed the same randomly presented images without clinical information. The experiment showed that this deep CNN system's performance at classifying optic disc abnormalities was at least as good as that of the experts. Similar research is presented in this paper with the significant difference that here it was necessary to design and carry out simulation of the missions on which the research will be conducted. The number of respondents is also much higher than in similar research, and thus the reliability of the experiment is better.

3. Design and Realization of Test Missions

The quality of this research significantly depends on the quality of the simulated missions on which measurements and result comparisons are carried out. Therefore, it is necessary to plan, conduct, and record test missions that are as close as possible to typical actual search missions. In this sense, as a first step, it is necessary to carry out analyses of the real missions in the Republic of Croatia as well as all around the world. This includes knowledge of the procedures and activities within the search mission as well as knowledge of the particularities related to the missing person. It is particularly important that the set of test missions contains as many realistic scenes as possible within the karst terrain with dense low vegetation, bushes, piles of stones, fissures, and sinkholes and also that it be conducted at the appropriate time of day with appropriate weather conditions and illumination. The simulated missing person must wear typical clothes, be set in an appropriate pose, and be placed within that terrain as realistically as is in real missions.

3.1. Characteristics of the Missing Person

Similar analyses were performed during the design of the HERIDAL image database. These data and analyses were extended to include the data needed to plan SAR mission simulations. The required data could be collected from statistics of the Croatian SAR team as well as from the literature. The book [9] represents the starting point of SAR theory, in which the author analyzed and interpreted data obtained from the International Search and Rescue Incident Database (ISRID), which contains 50,692 SAR cases. Accordingly, subject types can be classified by age, mental status, and activity into a dozen broad categories like children (various age groups), older demented people, irrational, mentally challenged, and despondent people, hunters, hikers, and climbers.

Each category is characterized by specific behavior patterns that can manifest when a person is lost and that can help predict the missing person's potential location and pose in the environment. Children frequently become lost because they are exploring or adventuring, take shortcuts, or are involved in fantasy play. When lost, they use a trail/road-following strategy or may hide intentionally because they are scared, sometimes simply sulking, but also to avoid punishment. Unfortunately, they are often dressed in clothes that fit well with the colors of the environment (green, brown, dark blue, or white) which makes them difficult to spot. They also have a much smaller projection and footprint in images compared to adults. Therefore, it is usually very difficult to distinguish them from natural artefacts, especially in low lighting as well as in high-contrast conditions.

Older people are usually found near a road or path where they become stuck in a bush. They often become tired and seek shelter, most often from the sun in summer and from rain and wind in winter. Demented people usually want to return to the locations where they were in their youth, the hamlets where they lived, the meadows where they kept cattle, or the forests where they collected wood. Like all older people, they often become tired or scared and seek shelter or just become stuck in a bush. They are commonly dressed in dark colors (black or gray), and therefore when they are sitting or lying in shadow, it is very hard to detect them.

Climbers and mountaineers are the categories that are most often the subject of searches in our areas. They often travel considerable distances to ascend prominent peaks or to climb rocks. The biggest risks for these two categories are an underestimation of the terrain difficulty and the time required to complete the climb, as well as weather conditions (especially during summer). It is

not uncommon for them to be completely unequipped. Overstrain, trauma, and injuries from falls are very frequent. Experienced hikers are dressed in brightly colored clothing and therefore easy to detect. Unfortunately, many amateurs are dressed in white clothes, which makes them difficult to spot on rocky terrain or in camouflage clothing, which makes them difficult to detect on bushy terrain. They often experience accidents and severe trauma, in which they remain immobile in various lying or sitting positions. Hunters behave similarly to mountaineers. The most common reason they need help is an injury, fall, or gunshot wound. Unfortunately, they are usually dressed in camouflage equipment, which makes them difficult to detect in aerial images.

Statistics data about the location where certain types of missing persons have been found are shown in Table 1.

Table 1. Statistical data about subject type and location where missing persons have been found.

Subject Type	Structure (Houses, Cottages, Huts, Barns)	Road, Drainage, Linear	Brush, Scrub, Woods	Field	Rock	Other
Children (7–9)	29	38	15	6	1	11
Demented	20	36	23	14	0	7
Climbers	0	27	9	9	27	28
Hikers	13	50	9	14	4	10
Hunters	8	55	14	0	2	21

3.2. SAR Mission Procedure

The activities of SAR teams are carried out according to strictly defined procedures. These procedures regulate aspects such as the size of the search team, its operational readiness and equipment, their location within the terrain, and how the terrain will be searched.

After receiving a call for help, the regional CRMS headquarters collect initial data regarding the missing person and the area of disappearance. The degree of urgency is determined based on various criteria. Priority is given to children and the elderly, as well as persons suffering from some diseases. Poor weather conditions, demanding terrain or environment, and a higher number of missing or injured persons contribute to an increase in the urgency of the reaction.

When the call is evaluated to require intervention, it is forwarded to the local CMSR team, which gathers members to conduct the search. A search manager is appointed, and the location of the base camp where all participants will be gathered is determined. Choosing a camp site is very important because it must be close to the search area, possibly with good mobile signal coverage and relatively isolated from visitors. The base camp is built around command vehicles that have the necessary ICT equipment. Improvised command vehicles are commonly used at first until a real one arrives.

In the meantime, detailed information about the subject is collected. Information such as gender, age, description of physiognomy, various interests, hobbies, time and the place of last sighting, appearance and description of clothing, health, and mental condition are just some of the valuable pieces of information that can help searchers to categorize a lost person into certain categories. Separate interviews are conducted with family members, friends, acquaintances, and people who last saw the missing person.

After the base camp is established, in accordance with the characteristics of the terrain and the lost person type, the area around the point where the lost person was last seen (Point Last Seen; PLS) is searched. The goal is to find the missing person as soon as possible or, if this is not the case, to find traces that would direct further search. UASs are regularly used in this initial phase of the search mission.

While the initial search of the terrain around the PLS point is being carried out, the search manager prepares a search plan. The area around the PLS is divided into zones according to terrain characteristics as well as the probability that a lost person could be found in it (Figure 2). Available

SAR resources, personnel, dogs, and UASs are allocated to each zone. Their task is to search a zone and report results to the search manager.

Figure 2. Division of the search area into zones according to the terrain characteristics and the lost person type.

For the area to be searched by the UAS, a flight plan is made using the appropriate mission planner software, while sometimes the drone is operated manually. The drone pilot monitors a video from the drone to detect a missing person or some trace. In parallel, high-resolution terrain images are taken for post-processing. After the drone completes the mission, these images are either inspected on-site at the base camp or sent over the mobile network to the Cloud, where they are processed by CMRS members. In both cases, there are drawbacks. In the SAR base camp, the operators do not have adequate equipment (large monitors, powerful computers), and they are usually not trained enough because they are local members. Of course, their tiredness should be also considered. A high-bandwidth mobile network is often not available at search locations, which is why the upload of high-resolution images to the Cloud is almost impossible from the base camp.

If the search achieves no result, then the zones that may not have been searched with a high degree of reliability due to terrain characteristics are detected and searched again but much more carefully. A new zoning that covers a wider area also must be constructed. The SAR mission procedure is presented in the flowchart in Figure 3.

Figure 3. The flowchart of the search-and-rescue (SAR) mission procedure.

3.3. Planning and Carrying Out SAR Mission Simulations

According to previous considerations, the research team at FESB has prepared and recorded a total of 36 SAR mission simulations. In the preparation phase of each simulation, the subject type and his or her pose was selected, the terrain was carefully analyzed, and accordingly, the locations where the missing persons were to be placed were selected. Depending on the selected location, weather conditions, and time of the day, the appropriate missing-person's clothing was chosen. Once the person was placed in the selected location, the search phase began. Using the flight planning software UgCS (<https://www.ugcs.com/>), the search area, flight parameters (drone altitude and speed), and points at which the shooting were to be performed were defined. These settings were uploaded to the drone that conducted the search mission. We were using DJI Phantom 3 Pro and MAVIC PRO drones. After the search was done, the images were reviewed on-site to determine their quality as well as the quality of the simulation. If necessary, modifications to the search scenario were made or a completely new scenario was generated, and the mission was resumed. Usually, four to five mission simulations were made at each location. Afterward, those that best suited the actual search mission were selected and added to the HERIDAL image database.

The preparation as well as the carrying out of the search mission are presented in Figure 4.

Figure 4. (a) Preparation of the drone flight for the search mission; (b) carrying out the search mission with Unmanned Aircraft Systems (UASs).

Among all the performed missions, only four were selected for this experiment. They are from the Herzegovinian region at locations where, according to the experience of the local SAR service, rescue missions are very common. Table 2 presents their basic information: number of persons, number of images with persons, number images without persons, the total number of images, and weight of the mission. The selected missions were from the areas of Stolac, Goranci, Jastrebnika, and Popovo polje. In the mission Goranci, the total number of persons was five within five different images. In missions Popovo polje and Stolac, one image contained more than one person. The total amount of images per mission ranged from 47 to 54 images. The number of persons per mission ranged from 5 to 11. Missions were also categorized according to terrain structure. The weight of the Stolac mission is “hard” because of the non-uniform terrain structure with rocks, low vegetation, and shadows. Pictures from the Jastrebnika and Popovo polje are mostly a uniform type of terrain with relatively dense vegetation, so the weight of these missions is “medium-hard”.

Table 2. Information about SAR missions used within this experiment; NH—number of persons, IWH—number of images without persons, INotH—images without persons.

Location	Mission	NH	IWH	INotH	Total	Weight
Goranci	GOR_1	5	5	42	47	medium
Jastrebnika	JAS_1	5	5	47	52	medium–hard
Popovo polje	POP_1	11	7	47	54	medium–hard
Stolac	STO_2	10	6	44	50	hard
Total	4	31	23	180	203	

4. Image Processing and Region Proposal Algorithm—ScoreMap to ROI

A special algorithm for the image processing and region proposal, ScoreMap to ROI has been developed. It is based on the CNN network. The detection process is divided into three parts. In the first part, features are extracted using a CNN network based on the VGG-16 architecture. In the second part, the regions of interest are calculated from the score map. In the final step, the best K regions are selected using the average Euclidean distance algorithm.

4.1. Feature Extraction

In the first part of the proposed algorithm, the VGG-16 network is used. The fully connected layer was removed and the 3×3 convolution layer (conv_6) was added after the (pool_5) layer. The prediction was performed using a 1×1 convolution layer with a sigmoidal activation function. This approach is similar to Faster-RCNN [22] architecture. The last convolutional layer uses a 512-d features vector. For an input image whose dimensions are 4000×3000 pixels, the size of the feature vector on the last convolutional layer is $125 \times 93 \times 512$. The CNN network produces three score maps, and the final score map is obtained by their sum as in Figure 5.

Figure 5. A convolution neural network (CNN) network used to extract features and predict the score map.

4.2. Proposing the Region of Interest

Regions of interest are extracted from the score map provided by the CNN network. The score map was converted to a binary map according to a threshold. After that, the algorithm for the detection contour was applied. For each contour, a minimum bounding box which contains all points in one

contour was calculated. The result of this operation is a list of regions of interest of different dimensions and shapes as shown in Figure 6. Each region is defined by (x_1, x_2, y_1, y_2) coordinates, where (x_1, y_1) is the upper-left point, and (x_2, y_2) is the lower-right point of the bound box.

Figure 6. Feature extraction with the ROI MaxPool layer. For each ROI the max pool operation is used to reduce feature space and then the average Euclidean distance to select the best features.

The process of extracting the region of interest is similar to the RPN algorithm in the Faster R-CNN [29] architecture. The RPN algorithm tries to draw a boundary box around each position on the feature maps of the last convolutional layer in the network. The results are many overlapping boundary boxes. To remove overlapping boundary boxes, Faster-RCNN employs a non-maxima suppression algorithm. In contrast, the proposed method merges all neighboring positions in score maps, where the network has detected the presence of a suspicious object in one area of interest. In this way, the operator performing the image inspection focuses on one area of interest where the network has detected the presence of a suspicious object.

In this way, a significantly smaller number of region proposals are obtained than with the RPN algorithm in the Faster-RCNN architecture. However, even this approach still proposes a significant number of regions that do not contain a subject of interest. Too many regions can be counterproductive in the search and rescue mission because, in this, way valuable time is spent on inspecting regions that do not contain humans. It is therefore necessary to find an adequate way to filter the regions and leave only the proposals that are most likely to contain humans. An algorithm for calculating the average Euclidean distance is used for this task.

4.3. Filtering the Regions of Interest

The filtering of ROIS is the most significant part of the algorithm because it is necessary to determine whether the region of interest contains a human or not. If an ROI contains people, it is retained and if not it is rejected. An algorithm for calculating the average Euclidean distance (AED) is in charge of performing this task. The input data in the AED algorithm is a list of regions of interest with a feature fixed size. In this case, for each ROI, a $1 \times 1 \times 512$ feature vector is used for each ROI. The ROI MaxPool layer is responsible for extracting this feature vector. For each region of interest whose dimensions are $H \times W \times 512$, where H and W are height and width of the ROI, a max pool operation is performed, with filter of size 3 and step 2. The reduced feature map is transformed into a 1-D vector with $1 \times 1 \times 512$ features. In the last step, the best feature is selected by the AED algorithm. The procedure for extracting features is shown in Figure 6.

The AED algorithm is used for two purposes, to select the best features vector inside each region of interest as in Figure 6, and also for the selection of TOP-K ROIs as in Figure 7, where K is the number

of proposals. The input in the AED is a 1-D array of feature vectors, where each feature vector is the size of $1 \times 1 \times 512$. The AED algorithm for each region provides a score, which indicates how different that region is from the others. The AED algorithm first calculates the distances between all ROI pairs, according to the Equation (2). The matrix A in Equation (1) is the result of these operations. Each element in the matrix indicates the difference between the two pairs of ROIs. The matrix A is symmetrical because the difference between ROI-A and ROI-B is the same as the difference between ROI-B and ROI-A. Then the average Euclidean distance is calculated for each ROI according to the Equation (3), where M is the total number of ROIs. This operation can be done either by rows or by columns since A is a symmetric matrix. In the last step, the TOP-K of the best ROIs is selected.

$$A = (a_{xy}) \quad (1)$$

$$a_{xy} = \sqrt{\sum_{i=1}^{512} (x_i - y_i)^2} \quad (2)$$

$$F = [f_y], f_y = \frac{1}{M} \sum_{v \in V} a_{vy} \quad (3)$$

Figure 7. ScoreMap to ROI.

This part of the algorithm is responsible for filtering the ROIs according to the result from the AED algorithm. The algorithm rejects all ROIs that contain an object from the nature and retains only ROIs that do not correspond to the natural environment.

5. Experiments and Results

5.1. Experiment

In the experiment, offering the participant four selected missions for analysis was planned, as described in the previous section. To compare the ScoreMap to ROI algorithm to the expert, but also to evaluate the impact on the detection of image labelling by the ScoreMap to ROI algorithm, the participant was offered two non-labelled and two labelled missions for analysis. Each participant received a different combination of labelled and non-labelled missions. During the mission processing, all the activities of the participant were recorded, i.e., the processing time of each image and location of the proposal where the participant assumes that a person is.

For purposes of the research, a special web-based application for comparing the success of detection of participants against the image-processing algorithm as well as evaluating algorithm efficiency was developed. The core applications are based on Python, Django Rest Framework, and React JavaScript. The application can be accessed at URL: <http://34.65.81.113/> or at the website: <http://ipsar.fesb.unist.hr/HERIDAL%20database.html>

When the participant logs in to the system, a list of missions is displayed on the screen, as shown in Figure 8. Here the participant can view brief instructions on how to complete this survey and to select the mission to be analyzed. Labelled and non-labelled missions are marked differently.

Figure 8. List of missions of the developed web application for detection accuracy comparison between the expert and the ScoreMap to ROI algorithm.

Each mission contained 40–50 pictures and 3–5 pictures with people. The list of images is located on the left side of the screen, as in Figure 9. The image currently being processed is highlighted. The participant can return to any image at any time. The largest part of the screen is occupied by the image on which the current inspections are performed. On the right side is a list of proposals selected with the assumption that a person is in it. At the top of the screen are controls for zoom, so the participant can view suspicious parts of the image in detail. Selecting (tagging) possible areas with persons is done with a double-click. If the participant assumes that a selection is wrong, he or she can delete it. Images where the selection has been done receive a special tag so that the participant can return to it at any time and make an additional check. When the mission inspection has been completed, the results as well as all the time needed to analyze every single image are stored in the database for later analyses.

Figure 10 presents an image of a labelled mission with the region proposals in which the ScoreMap to ROI algorithm estimated that a person was located.

5.2. Results

The research involved 48 respondents. Each participant inspected two missions with the support of the ScoreMap to ROI algorithm, and two missions without support.

Figure 9. Main screen of the developed web application for detection accuracy comparison between the expert and the ScoreMap to ROI algorithm.

Figure 10. Main screen of the developed web application for detection accuracy comparison between the expert and the ScoreMap to ROI algorithm when the labelled mission is processed.

To capture the central tendency of the data (precision, recall, and time needed to inspect the image) a mean value and median of one were used. The mean value was calculated after removing outliers—observations that lie outside of the interval

$$\left[Q1 - \frac{3}{2}(Q3 - Q1), Q3 - \frac{3}{2}(Q3 - Q1) \right], \quad (4)$$

where Q1 stands for lower quartile 2 and Q3 stands upper quartile 3 of the sample—from the sample since some data were significantly smaller than we expected (maybe some individuals were not dedicated enough to inspect the images with care). The 95% confidence interval for the mean value (calculated without outliers) is

$$\left[M - t \frac{s}{n}, M + t \frac{s}{n} \right] \quad (5)$$

where M is mean value, s is standard deviation, n is sample length, and t is the 97.5% quantile of the Student's t -distribution with $n - 1$ degree of freedom. The percentage of correct inspection (recognition of people) is denoted with p , and the 95% confidence interval is

$$\left[p - z \sqrt{\frac{p(1-p)}{n}}, p + z \sqrt{\frac{p(1-p)}{n}} \right] \quad (6)$$

where $z = 1.96$ is the 97.5% quantile of the standard normal distribution.

After statistical data pre-processing, the analyses were carried out. We adopted two well-known measurements to properly evaluate the separate stages and the overall detection model: Recall and Precision.

$$\text{Recall} = \frac{TP}{TP + FN}, \text{ Precision} = \frac{TP}{TP + FP} \quad (7)$$

Recall is the number of true positives relative to the sum of the true positives and the false negatives. Recall represents the percentage of people correctly detected among all the candidate regions that should have been detected as people. Here, true positive (TP) is the number of correctly detected persons, and false negatives (FN) is the number of mis-detected persons (detection failures). Recall also represents the detection rate. Precision represents the percentage of correctly detected people divided by the total detected people. It includes false positives (FP), detected objects that are false alarms or are incorrectly detected as people.

The developed ScoreMap to ROI algorithm was applied to all simulated missions. The results for the missions to be used in the further experiments are shown in Table 3.

Table 3. ScoreMap to ROI algorithm results.

Location	Mission	TP	FP	FN	Recall
Goranci	GOR_1	5	84	0	100.0%
Jastrebnika	JAS_1	5	73	0	100.0%
Popovo polje	POP_1	10	95	1	90.9%
Stolac	STO_2	8	118	2	80.0%
Total		28	370	3	90.3%

Some false negative as well as false positive examples are presented in Figures 11 and 12, respectively.

Figure 11. False negative examples for the ScoreMap to ROI algorithm: (a) The person is in occlusion, (b) The person is dressed in grey clothes and black pants and it is very difficult to distinguish him or her from the environment, (c) The person is in occlusion, and only parts of his clothing and arms are clearly visible.

Figure 12. False-positive examples for the ScoreMap to ROI algorithm: (a,c) The algorithm, due to the shadow, made a wrong conclusion, (b) The stone and a blue plastic bag create an object that is almost impossible to distinguish from a human.

Figure 13 shows the detection results of the participants on the missions with the support of the ScoreMap to ROI algorithm.

Figure 13. The results of participants in missions with the ScoreMap to ROI support.

Figure 14 shows the detection results on missions without support.

Figure 14. The results of participants in missions without support.

Figure 15 shows a comparison of missions with the ScoreMap to ROI support and missions without support in terms of recall and precision.

Figure 15. Comparing missions with the ScoreMap to ROI support and missions without support in recall and precision.

Figure 16 shows the relationship between the average time spent inspecting a single mission with the ScoreMap to ROI support and missions without support.

Figure 16. Average inspection time for each mission type.

Figure 17 shows the overall detection results: The mission with the ScoreMap to ROI algorithm, with the ScoreMap to ROI support, and without support.

Figure 17. Results comparison between missions: The mission with the ScoreMap to ROI (algorithm alone), with the ScoreMap to ROI support, and without support.

Figure 18 presents images where the participants achieved the worst recall in missions (mission Stolac 2) with the ScoreMap to ROI support. The persons labelled in red color are those that are not

detected by the ScoreMap to ROI algorithm while the detected persons are labelled in yellow color. The recall of the least detected persons achieved by the analyses of those images with and without the ScoreMap to ROI support is presented in Figure 19.

Figure 18. The most difficult examples in missions (mission Stolac 2) with the ScoreMap to ROI support.

Figure 19. Comparison of the most difficult examples (the least detected persons) in missions (mission Stolac 2) with the ScoreMap to ROI support and the same examples in missions without support.

Images analyzed in missions without algorithm support were where the worst recall was achieved and are presented in Figure 20. The recall of the least detected persons achieved by the analyses of those images with and without the ScoreMap to ROI support is presented in Figure 21.

Figure 20. The most difficult examples in missions (mission Popovo polje 1) without support.

Figure 21. Comparison of the most difficult examples (the least detected persons) in missions (mission Popovo polje 1) without and with the ScoreMap to ROI support.

Finally, Figure 22 shows the results achieved by qualified SAR experts compared to the average SAR member.

Figure 22. Comparison of the recall achieved by the qualified SAR experts compared to the average SAR member.

6. Discussion

An analysis of the results and discussions are presented in this section.

The results obtained by the developed ScoreMap to ROI algorithm are the basis for further analyses. They are presented in Table 3. The recall of the algorithm was about 90%. Although the algorithm managed to detect almost all the people, it generated many false alarms. The best results were achieved on the Goranci and Jastrebnika missions, where the recall was 100%. A slightly worse result was achieved on the Popovo polje mission, where only one person was not detected. False negative examples are presented in Figure 11, indicating how difficult it is in some situations to detect people in aerial images. A person from Figure 11a is in occlusion, and only parts of his clothing and arms are clearly visible. In Figure 11b, a person is dressed in grey clothes and black pants and it is very difficult to distinguish him or her from the environment, while in Figure 11c, the person is also in occlusion.

False positive examples presented in Figure 12 show how it is difficult or sometimes even impossible to distinguish objects within an image. In Figure 12a,c, the algorithm, due to the shadow, made a wrong conclusion. In Figure 12b, stone and a blue plastic bag create an object that is almost impossible to distinguish from a human.

It follows the analysis and the discussion of the results obtained by the experiment presented in the previous chapter. This is the main contribution of this paper.

According to the Figure 13, the best recall was achieved on the missions of Goranci (84%) and Jastrebnika (83%). In the Popovo polje and Stolac missions, participants achieved a slightly worse recall. The recall in Popovo polje was 72% and 63% in Stolac. Participants achieved the worst recall on the Stolac mission as did the ScoreMap to ROI algorithm, which is to be expected since the experts assessed this mission as very difficult, as shown in Table 2. In general, the detection results are consistent with the complexity of the missions given in Table 2.

We have participants who achieved a very high recall of 100% in a relatively short time of inspection. On the other hand, some of the participants spent a lot of time on inspection, and they achieved a quite poor recall of less the 20%. With these results, we cannot confirm the fact that participants will have better recall if they spend more time on the inspection. These results are expected and indicate that the reliability of the detection is strongly influenced by the experience of the observer.

Analysis of Figure 14 shows that the best recall, 76%, was achieved on the Stolac missions. These results were not expected, as the experts assessed this mission to be very hard due to the structure of the terrain. This is contrary to the detection results of the same missions with the support of ScoreMap to ROI algorithm, where the worst recall was achieved. This result may be explained by the fact that people mostly focused on the ScoreMap to ROI proposals and did not focus enough on other parts of the image.

In the missions without support, the worst recall, 54%, was on the Popovo polje mission. In general, the detection results were not quite consistent with the complexity of the missions estimated in Table 2.

As in the previous case, it is not possible to relate the relationship between the time spent on image inspection and the recall of detection, and the same conclusion holds.

According to the Figure 15, the participants achieved slightly better recall in missions with the ScoreMap to ROI support compared to missions without support, as presented in Figure 12. Recall in missions with the ScoreMap to ROI support was 72%, and in missions without support, it was 66%. The average precision in missions with the ScoreMap to ROI support was 46%, and in missions without support, it was 53%. Participants made fewer mistakes if they performed the inspection without support, but they had a worse recall, which in this application is more important.

From Figure 16 we can conclude that users with the Score Maps to ROI performed the inspection much faster. The average inspection time for ScoreMap to ROI missions was 1110 s, and for missions without support it was 1495 s. Inspection time was about 30% faster in the missions with the ScoreMap to ROI support. Two conclusions could be drawn. One is that due to the already pre-processed images, their analysis is carried out faster, which is positive. On the other hand, the downside is that apparently the participants conducted the analysis with less attention and in some cases superficially.

Analysis of Figure 17 shows that the best recall achieved by the ScoreMap to ROI algorithm was 90%, while the recall achieved with the ScoreMap to ROI support was much worse, 72%. This indicates that the algorithm can recognize very difficult cases, which could be easily overlooked by a less experienced observer. This is a confirmation that in real SAR missions, where images are often analyzed by less experienced searchers, mainly volunteers, the proposed algorithm could significantly contribute to the reliability of detection. Participants analyzing the missions without support achieved the worst recall of only 66%. This is a verification that the algorithm contributes a lot to the detection of people in aerial images.

However, the fact that the participants did not recognize particular labelled persons in the picture requires additional analysis. Therefore, the labelled images with the lowest recall, presented in Figure 18, were additionally carefully examined. The ScoreMap to ROI algorithm did not detect person S2 in the image STO_2033 as well as person S3 in the image STO_2007, labelled in red color, while the others S1, S4 and S5, labelled in yellow color, were detected by the algorithm. As expected, the worst recall was achieved in the images with the ScoreMap to ROI support where the persons were not detected by the algorithm. Moreover, these persons had a much worse recall compared to those without the ScoreMap to ROI support. The results are presented in Figure 19. For example, the recall of the person S2 was only 19%, while inspection of the same image without the ScoreMap to ROI support was 44%. This could be explained by the fact that people are most focused on the algorithm proposals and do not pay enough attention to the rest of the image. This image contains two humans, S1 and S2. The human S1 was recognized by the ScoreMap to ROI algorithm and therefore has much better recall.

Furthermore, the results obtained for person S5 in the image STO_2017 are very interesting. This person had a very low recall in the mission with the ScoreMap to ROI support of 50%, although it was detected by the algorithm. The explanation is that the person is at the very edge of the image. If the participant is not concentrated enough, the label could be easily omitted. In addition, it is natural during the analyses to be focused on the center of the image. On the other hand, in missions without the ScoreMap to ROI support, participants are more concentrated and analyze the image in more detail. This is verified by the data regarding the average time spent on image analysis. The participants spent up to three times as much time analyzing those images without the ScoreMap to ROI support.

Regarding images presented in Figure 20, it is important to emphasize that the ScoreMap to ROI algorithm successfully detected the persons in all of the images. It is interesting to comment that the person P2 in image POP_1012 had the worst recall of only 12%, but in other images the recall was not much better. For the algorithm, person detection in these examples is not a problem, since all of them are in the open field and in the poses for which the algorithm has been trained. On the other hand, for the person who analyzes these images, this is a difficult visual inspection task because the people in these pictures are wearing black or white clothes and in some way blend with the environment, so it is very difficult to distinguish them from shadows or stones. This provides more evidence of how

useful such an algorithm could be in SAR missions. Moreover, the recall in all these images analyzed with the ScoreMap to ROI support is relatively low. Additional analysis of these results shows that, for the images where two persons are present and labelled, the person who analyzes these images often detects only one person. This could be also explained by the lack of concentration as well as the wrong assumption that only one person is present in the image. This is also an explanation for Figure 13, where the participants did not detect all the labelled persons.

Figure 22, as expected, shows a large difference in the quality of detection achieved by the qualified SAR experts compared to the average SAR member. Furthermore, although image processing and artificial intelligence algorithms increase accuracy and reduce detection time, these analyses indicate that it is very important to work on the education of SAR members who analyze these images.

Finally, this research has shown that the application of image processing algorithms and artificial intelligence for human detection on aerial images taken in SAR missions achieves better results compared to the average SAR searcher, thus increasing the likelihood of finding a missing person.

7. Conclusions

This paper addressed the problem of comparing the accuracy of human detection in aerial images taken by UAVs in SAR missions between an algorithm based on deep neural networks and an SAR expert. For the purpose of the research, test search missions were planned in accordance with the experiences and statistics of the CMRS service as well as the data available in the world literature. Thirty-six test missions were recorded and added to the already-existing HERIDAL image database, where over 68,750 image patches of people in wilderness viewed from an aerial perspective, as well as 500 labelled, full-size 4000×3000 -pixel real-world images that all contain at least one person are stored. All planned, completed, and recorded missions were processed by a specially developed image-processing algorithm based on deep CNN, named ScoreMap to ROI. The algorithm is described in this paper. A web application has been developed with which experts can analyze recorded missions. Four missions were offered for analysis, two of which were processed and labelled with the ScoreMap to ROI algorithm, while the other two were unmarked. Forty-nine experts analyzed proposed missions, and all data regarding processing were stored in the database. Analyses of the obtained results were performed, and they indicate that the ScoreMap to ROI algorithm achieves better recall compared to the average observer. An observer who processed already-labelled missions using the algorithm also achieved better recall and precision compared to an observer who processed the missions that were the same but unlabeled. This experiment proved the effectiveness of image processing algorithms as support to SAR missions.

In future research, the focus will be on the modification and implementation of the developed algorithm on edge devices in real time. In this way, the processing will be performed on a UAV, and parts of the image detected as regions of interest can be transmitted in real time to the operator for inspection. This would significantly increase search efficiency.

Author Contributions: Conceptualization, S.G. and D.B.-Š.; methodology, S.G.; software, D.Z.; validation, S.G., D.Z., Ž.M. and D.B.-Š.; formal analysis, S.G. and D.Z.; investigation, S.G. and D.Z.; resources, D.B.-Š. and Ž.M.; data curation, S.G. and D.Z.; writing—original draft preparation, S.G.; writing—review and editing, S.G., D.Z., Ž.M. and D.B.-Š.; visualization, S.G., D.Z., Ž.M. and D.B.-Š.; supervision, S.G.; project administration, S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by project “Prototype of an Intelligent System for Search and Rescue”, grant number KK.01.2.1.01.0075 funded by European Regional Development Fund.

Acknowledgments: We sincerely thank Croatian Mountain Search and Rescue Service for overall support in our research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alliance for Drone Innovation, Fact and Figures. Available online: <http://www.droneinnovation.org/facts-and-figures.html> (accessed on 14 February 2020).
2. Biočić, F.; Vuica, R. *Integrating Search and Rescue Drones into Airspace—Case Study—Croatian Mountain Rescue Service*; Version 1; EENA European Emergency Number Association: Brussels, Belgium, 2019.
3. Turić, H.; Dujmić, H.; Papić, V. Two-stage segmentation of aerial images for search and rescue. *Inf. Technol. Control* **2010**, *39*, 138–145.
4. Božić-Štulić, D.; Marušić, Ž.; Gotovac, S. Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions. *Int. J. Comput. Vis.* **2019**, *127*, 1256–1278.
5. Marušić, Ž.; Božić-Štulić, D.; Gotovac, S.; Marušić, T. Region proposal approach for human detection on aerial imagery, 2018. In Proceedings of the 3rd International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 26–29 June 2018.
6. Ren, Y.; Zhu, C.; Xiao, S. Small object detection in optical remote sensing images via modified faster R-CNN. *Appl. Sci.* **2018**, *8*, 813. [CrossRef]
7. Huang, J.; Shi, Y.; Gao, Y. Multi-scale faster-RCNN algorithm for small object detection. *Comput. Res. Develop.* **2019**, *56*, 319–327.
8. Cao, C.; Wang, B.; Zhang, W.; Zeng, X.; Yan, X.; Feng, Z.; Liu, Y.; Wu, Z. An Improved Faster R-CNN for Small Object Detection. *IEEE Access* **2019**, *7*. [CrossRef]
9. Koester, R. *Lost Person Behavior: A Search and Rescue Guide on Where to Look for Land, Air, and Water*; dbS Productions: Charlottesville, VA, USA, 2008. Available online: <https://books.google.hr/books?id=YQeSIAAACAAJ> (accessed on 14 February 2020).
10. Enzweiler, M.; Gavrilu, D.M. Monocular pedestrian detection: Survey and experiments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2179–2195. [CrossRef] [PubMed]
11. Viola, P.; Jones, M.J.; Snow, D. Detecting pedestrians using patterns of motion and appearance. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 2, pp. 734–741. [CrossRef]
12. Tian, Y.; Luo, P.; Wang, X.; Tang, X. Deep learning strong parts for pedestrian detection. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Las Condes, Chile, 11–18 December 2015; pp. 1904–1912.
13. Hosang, J.; Omran, M.; Benenson, R.; Schiele, B. Taking a deeper look at pedestrians. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
14. Gaszczak, A.; Toby, J.H.; Breckon, P. Real-time people and vehicle detection from UAV imagery. In *Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, San Francisco, CA, USA, 24–25 January 2011; International Society for Optics and Photonics: Bellingham, WA, USA, 2011; Volume 7878, p. 78780B. [CrossRef]
15. Rudol, P.; Doherty, P. Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery. In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; pp. 1–8. [CrossRef]
16. Lygouras, E.; Santavas, N.; Taitzoglou, A.; Tarchanidis, K.; Mitropoulos, A.; Gasteratos, A. Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations. *Sensors* **2019**, *19*, 3542. [CrossRef] [PubMed]
17. Mishra, B.; Garg, D.; Narang, P.; Mishra, V. Drone-surveillance for search and rescue in natural disaster. *Comput. Commun.* **2020**, *156*, 1–10. [CrossRef]
18. Al-Kaff, A.; Gómez-Silva, M.J.; Moreno, F.M.; de la Escalera, A.; Armingol, J.M. An appearance-based tracking algorithm for aerial search and rescue purposes. *Sensors* **2019**, *19*, 652. [CrossRef] [PubMed]
19. Panboonyuen, T.; Jitkajornwanich, K.; Lawawirojwong, S.; Srestasathien, P.; Vateekul, P. Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning. *Remote Sens.* **2019**, *11*, 83. [CrossRef]
20. Pires de Lima, R.; Marfurt, K. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sens.* **2020**, *12*, 86. [CrossRef]
21. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep transfer learning for few-shot SAR image classification. *Remote Sens.* **2019**, *11*, 1374. [CrossRef]

22. Zhang, X.; Zhu, K.; Chen, G.; Tan, X.; Zhang, L.; Dai, F.; Liao, P.; Gong, Y. Geospatial object detection on high resolution remote sensing imagery based on double multi-scale feature pyramid network. *Remote Sens.* **2019**, *11*, 755. [[CrossRef](#)]
23. Musić, J.; Orović, I.; Marasović, T.; Papić, V.; Stanković, S. Gradient compressive sensing for image data reduction in UAV based search and rescue in the wild. *Math. Probl. Eng.* **2016**. [[CrossRef](#)]
24. Gotovac, S.; Papić, V.; Marušić, Ž. Analysis of saliency object detection algorithms for search and rescue operations. In Proceedings of the 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Orlando, FL, USA, 25 May 2002; pp. 1–6. [[CrossRef](#)]
25. Karimi, D.; Dou, H.; Warfield, S.K.; Gholipour, A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med Image Anal.* **2020**, *65*. [[CrossRef](#)]
26. Wichmann, F.A.; Janssen, D.H.J.; Geirhos, R.; Aguilar, G.; Schutt, H.H.; Maertens, M.; Bethge, M. Methods and measurements to compare men against machines. In Proceedings of the IS&T International Symposium on Electronic Imaging 2017, Burlingame, CA, USA, 29 January–2 February 2017.
27. Buetti-Dinh, A.; Galli, V.; Bellenberg, S.; Ilie, O.; Herold, M.; Christel, S.; Boretska, M.; Pivkin, I.V.; Wilmes, P.; Sand, W.; et al. Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm composition. *Biotechnol. Rep.* **2019**. [[CrossRef](#)] [[PubMed](#)]
28. Biousse, V.; Newman, N.J.; Najjar, R.P.; Vasseneix, C.; Xu, X.; Ting, D.S.; Milea, L.B.; Hwang, J.; Kim, D.H.; Yang, H.K.; et al. Optic Disc Classification by Deep Learning versus Expert Neuro-Ophthalmologists. *Ann. Neurol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
29. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015; Volume 1, pp. 91–99.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).