



Person identification with aerial imagery using SegNet based semantic segmentation

Rajeswari Manickam¹ · Sathesh Kumar Rajan¹ · Chidambaranathan Subramanian² · Arnold Xavi¹ · Golden Julie Eanoch³ · Harold Robinson Yesudhas⁴

Received: 15 July 2020 / Accepted: 25 August 2020 / Published online: 1 September 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In recent days, people in remote area suffer a lot due to variety of natural calamities such as flooding, earthquake and so on. It has been noted that people used to stay in top portions of their house when there is a flooding issue. Hence, it is very difficult for the rescue team to identify the location of a person by looking at the parts of a person such as hands, legs and partial image of a face using the existing approaches. In this proposed approach, an idea for detecting person when there is only parts such as legs, hands are visible from remote, wild or non-urban areas with the help of UAV-Unmanned Aerial Vehicle has been suggested. Detecting person and identifying the location from the image tends to be a difficult process due to very small and camouflaged objects in the images collected. In this approach, Semantic Segmentation using deep learning approach has been applied in order to detect a person. SegNet- Segmentation Network is the network architecture used in the process of semantically segment the image according to each pixel, hence identifying person is easy. The main objective of this proposed model is that, sometime UAV image may contain partial person images, like legs, hand, etc., that could not be identified by existing approaches were being recognized and identified successfully. This model is trained and tested using HERIDAL dataset. Over 70% images were trained and 30% images were used for testing. This enhanced deep learning model named as Semantic SegNet model achieved an accuracy of 91.04%. This proposed Semantic SegNet model has been compared with existing approaches such as VGG16, GoogleNet and ResNet- Residual neural Network for the same set of trained and tested images. Comparison table declared that this proposed Semantic SegNet Model outperformed other existing models.

Keywords Deep learning · Person detection · UAV · Semantic segmentation · SegNet

Introduction

In 2018, severe flood affected Kerala due to heavy rainfall which caused over 2 lakh people to loss their life. In that

occasion, people stayed in top portion of their houses. It is observed that, with the help of rescue mission team, shifting the trapped people into safest camps using helicopters and boats. Rescue mission officers received the information about

Communicated by: H. Babaie

✉ Golden Julie Eanoch
goldenjuliephd@gmail.com

Rajeswari Manickam
rajeswarim@sahridaya.ac.in

Sathesh Kumar Rajan
satheshpkd@gmail.com

Chidambaranathan Subramanian
scharan2009@gmail.com

Arnold Xavi
arnoldxavi.t@gmail.com

Harold Robinson Yesudhas
yhrobinphd@gmail.com

- ¹ Department of Computer Science and Engineering, Sahridaya College of Engineering and Technology, Kodakara, Kerala, India
- ² Master of Computer Applications Department, St. Xavier's College, Palayamkottai, Tamil Nadu, India
- ³ Department of Computer Science and Engineering, Anna University Regional Campus, Tirunelveli, India
- ⁴ School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

the place where the people get trapped by the phone calls. Information has been provided from the trapped people or using social media videos. Also, there was no electricity and internet due to the endless rain which made the trapped people finding difficult to communicate with the rescue team. Identifying the person who got trapped in various part of the house was hectic and time consuming without proper communication facilities. In some places, it is found that normal people couldn't directly go without the help of helicopters. In 2019, Malapuram, Kavalapara affected with sliding of land which requires finding the person very quickly to save their life for the resume mission team.

In this proposed approach, an idea for detecting person from remote, wild or non-urban areas with the help of UAV has been suggested even though only the parts of the person images are available (Harold Robinson and Golden Julie 2019). The proposed model uses Semantic segmentation method with SegNet architecture for person detection. Semantic segmentation is defined as the method used for segmentation where image is segmented to multiple classes (example: Person, Animal, Background) and each class is assigned with different colors. SegNet is the network architecture which is used in the process of semantically segment the image according to each pixel (Joshua Bapu et al. 2019). The proposed model includes Collecting UAV image set which is

to be labeled pixel wise using Semantic segmentation. SegNet architecture is constructed for training and testing the network. Also, User interface is to be built for deployment to computer applications.

There may be a chance for flooding in the coming years as 2018 & 2019; it is desirable to find out suitable methodology to overcome such issues. It is found that many dams in Kerala might have a chance to be broken; this proposed method will be helpful for the researchers and rescue mission team to identify and rescue the needy people.

The main objective is to collect various UAV image dataset for person identification and detection. Collected images need to be processed (Rudol and Doherty 2008) which involves image restoration and image enhancements. There are plenty of image processing algorithms available in order to perform image restoration and image enhancement. Many factors should be considered to apply image processing algorithms such as image resolution, scaling of an image. In recent days, many advanced image classification techniques such as Artificial Neural Network, fuzzy-set and so on. This work focuses on per-pixel classification approach in order to train whether it is a human or not. This is achieved by inputting an image that contains only parts of a human such as legs, hands and so on. Labeling is performed in such a way that

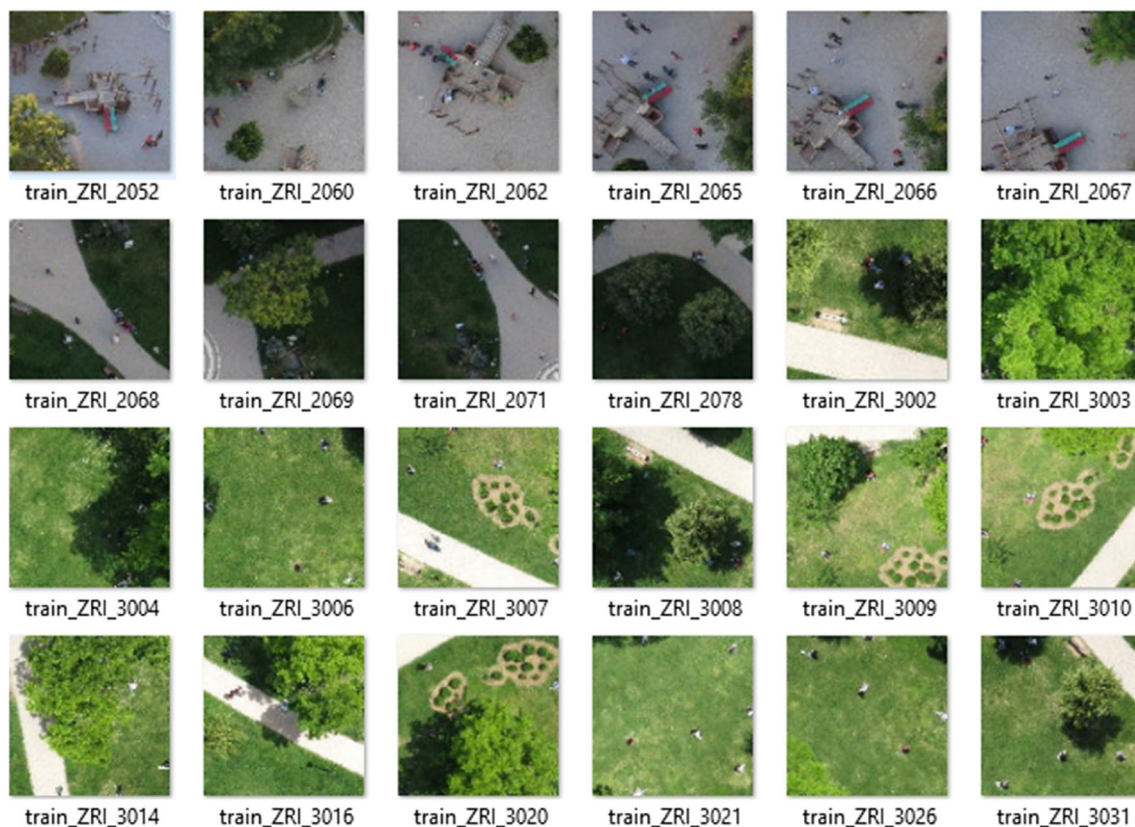


Fig. 1 Example Dataset used for person detection

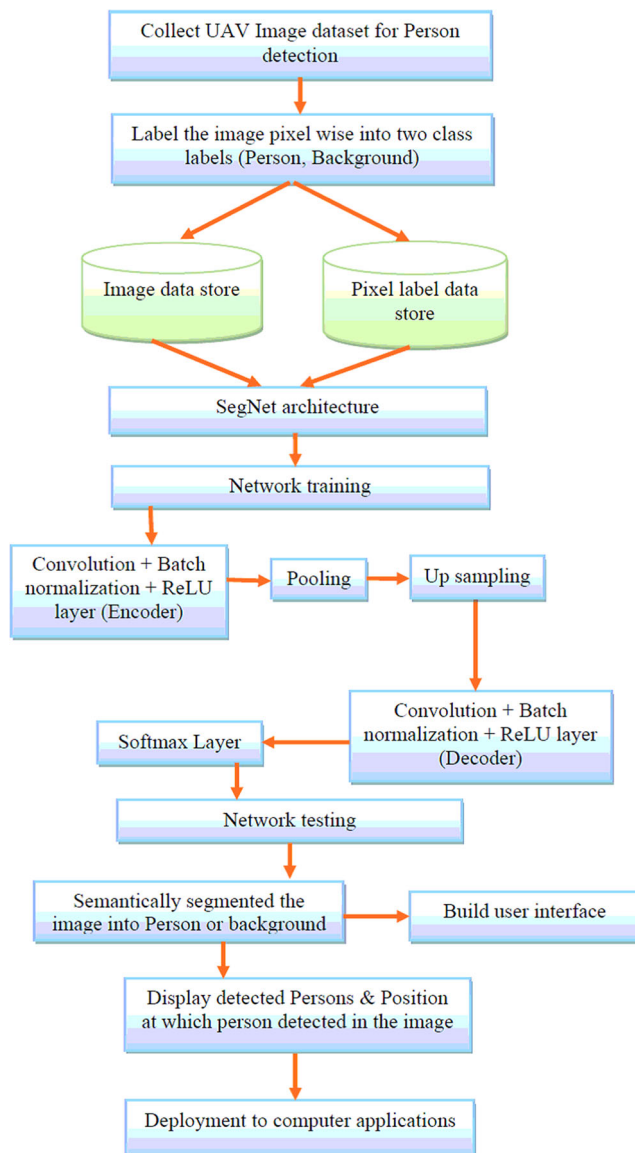


Fig. 2 System Architecture for Person detection

assigning human/ parts in one color and all the other things such as table, sofa and other materials in different color.

Semantic pixel-wise segmentation (Badrinarayanan et al. 2017) involves classifying pixels accurately to predict the class probabilities such as performing labeling operation. In this approach, a deep convolutional neural network applicable for segmentation has been used for object classification. Various UAV image are applied to this semantic SegNet as input and they have been trained in order to separate human parts and other non-living organism with different colors. This is achieved by using convolutional encoder-decoder along with batch normalization, pooling, up sampling and softmax operation. Accuracy of identifying a person is carried out and compared with existing approaches for the same UAV image dataset. Semantic segmentation is done on the images in order

to detect whether it is a person or any other things. User interface is built after performing semantic segmentation. At last, detected person and the position at which person detected in the image has been displayed.

The contribution of the paper is

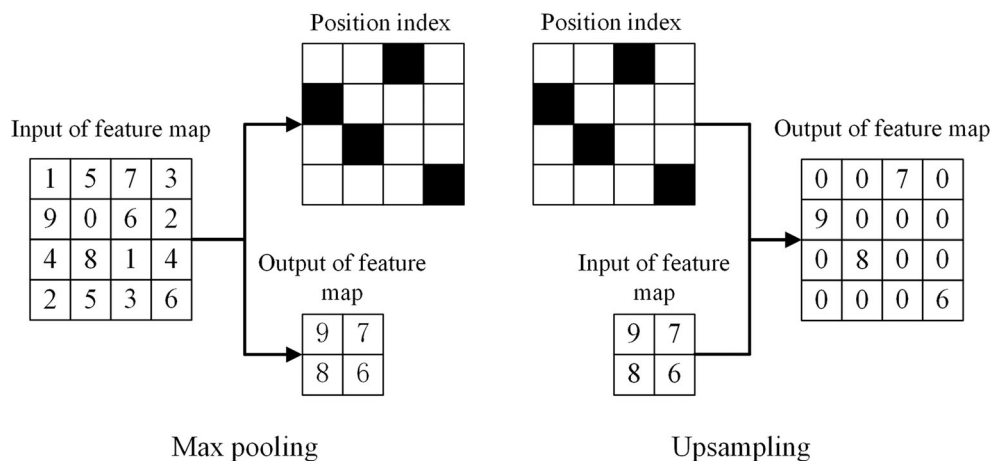
- Heavy rain causes Flooding in many of the countries which threatened most of the people's lives. Helping the needed people who are trapped in various parts of the houses would be identified easily by the rescue mission team.
- There is a threat that heavy rains increase the water level which makes the rescue team to find out the person by looking at the partial parts of a human such as upper body part or lower body parts.
- Due to appropriate working of quarries and cutting tree in hilly area causes landslide in most parts of the countries. Person detection will be very helpful during this situation.
- The most common hazard in forest is forests fire which affects many nearby people to adopt safety measures. This proposed work would be useful when people get trapped in the forest.
- There are many hairpin bends across the world which causes frequent accidents. This makes many persons being fallen down and very difficult to find out. This proposed work would be applied to detect the needy person easily.

Section 2 comprehends the literature survey; Section 3 describes the segmentation SegNet methodology, Section 4 discusses the results and Section 5 provides the conclusion and future work.

Literature survey

Deep neural network is the advanced technology that includes segmentation, object detection. A. Krizhevsky et al. (Krizhevsky et al. 2012) proposed AlexNet which is the classical image object recognition procedure network proposed in 2012 that contains 8 layers, five are convolutional layers and three are fully connected layers. AlexNet includes 60 M parameters and won the Image net contest in 2012. Its error rate is 15.3%. M.D. Zeiler et al. (Zeiler and Fergus 2013) proposed ZF Net in 2013. It achieved error rate of 14.8% which is less than the AlexNet. K. Simonyan et al. (Simonyan and Zisserman 2014) proposed VGGNet in 2014, it contains 16 layers. The number of parameters that VGGNet contains is 138 M. Its error rate was 7.3% that is less than the AlexNet. The size of the VGGNet is very large because it contains fully connected layers. C. Szegedy et al. (Szegedy et al. 2015) proposed GoogleNet which includes 22 layers. Its error rate is

Fig. 3 Pixel-wise Labeling



6.67% that is less than the VGGNet. GoogleNet includes 4 M parameters and won the ILSVRC competition in 2014.

K.He et al. (He et al. 2015) explained ResNet in the year 2015 and achieved error rate of 3.6% which was considered to be less than VGGNet and GoogleNet. It contains 152 layers. R.Girshick et al. (Girshick 2015) proposed Fast RCNN. In this method, regions of interest are pooled into a fixed-size feature map which is then mapped to a feature vector using fully connected layers. It includes two output vectors per RoI. One is softmax probabilities and another is per-class bounding box regression offset. It gives better performance than the original RCNN. S. Ren et al. (Ren et al. 2015) proposed Faster RCNN by adding some more modification on Fast RCNN. The method was dealing with RPN that is a fully convolutional. Advantage of Faster RCNN is that it gives better performance than fast RCNN. J.Redmon et al. (Redmon et al. 2015) proposed YoLo Detector. YoLo architecture used object detection as a regression. YoLo architecture gives better speed than the Faster RCNN. Drawback of YoLo architecture is that its accuracy is less when compared to Faster RCNN. H.

Namet al. (Nam and Han 2015) proposed MDNet which won the VOT2015 Challenge. Main part of MDNet is the layers which are shared and domain specific. Nonetheless, drawback of this method is that it demands to run CNN to extract features from images, make the system very slow. Paul Viola et al. (Viola et al. 2003) demonstrated an approach for finding humans who are walking. The approach combined both motion information and appearance information. The method achieved detection with low rate of false positive. Tian, Y., Luo et al. (Tian et al. 2015) elaborated a method that utilized extensive part detectors. The model was trained on weakly labeled data. The dataset used for that method was Caltech dataset. The model achieved a miss rate of 11.89%. Deepak Jaiswal et al. (Jaiswal and Kumar 2019) implemented a method for finding moving object with the help of GPUs in UAV videos. The detection process includes feature extraction, feature matching, image transformation, background subtraction, morphological processing, connected component labeling and finally detect the object from UAV videos.

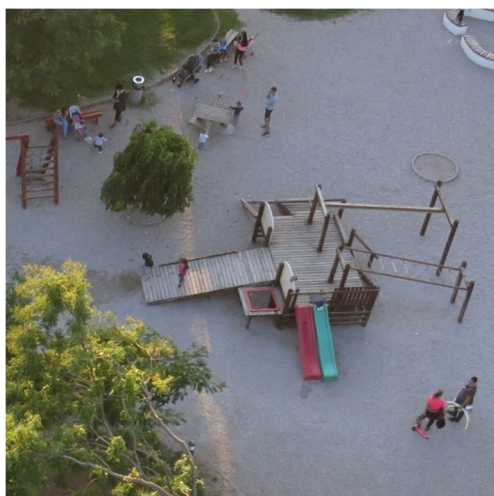


Fig. 4 Image without Labeling



Fig. 5 Image after Labeling

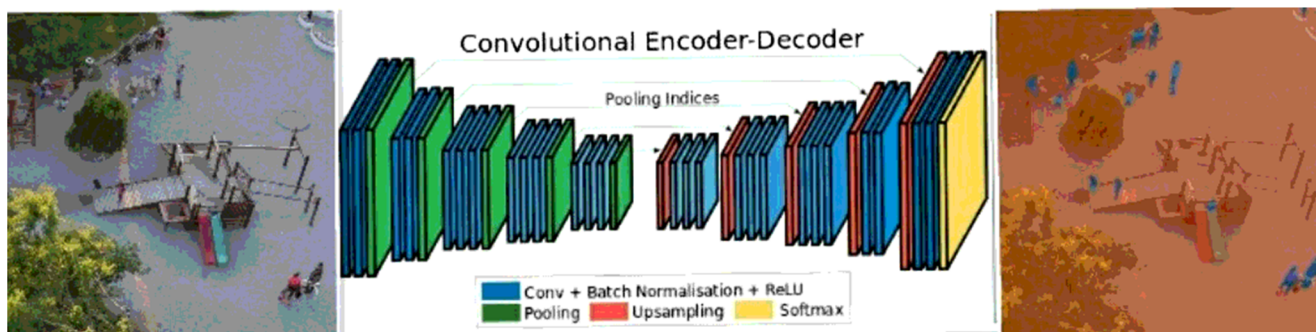


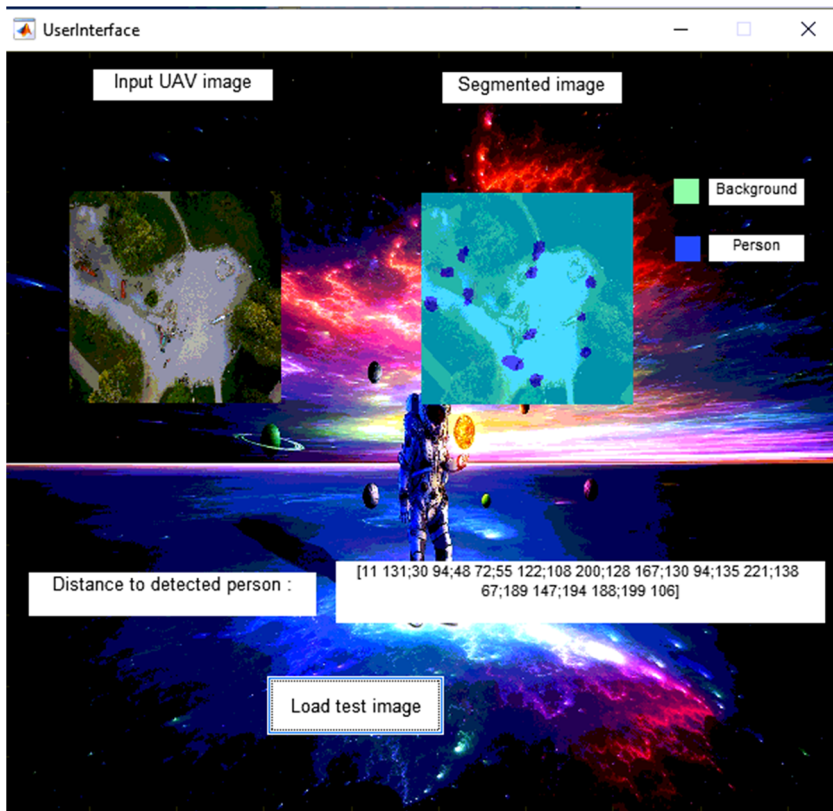
Fig. 6 Semantic SegNet architecture

Anna Gaszczak et al. (Anna Gaszczak and Breckon 2011) presented a method for detecting vehicles automatically. The approach contains cascaded Haar classifiers that are multiply trained. The method achieved accuracy of 70% for detecting human. Rudol P et al. (Rudol and Doherty 2008) proposed an approach for detecting the presence of person in video sequences with the help of thermal images and visible images. In hot summers using thermal cameras are not quite suitable is the main disadvantage of this method. Turić, H et al. (Turić et al. 2010) was selected primarily the mean-shift algorithm. Authors decided to modify the mean-shift algorithm. They applied two-stage mean shift segmentation from the base

which gives better result than previous one. Ming-Yu et al. (Chen et al. 2017) augment the state and art of RCNN algorithm and augmented RCNN improved 29.8% mean average precision over RCNN on detecting small object. Drawback of this method is that it is not applicable for the identification of big objects. Angelova et al. (Angelova et al. 2015) proposed an approach for identifying the pedestrian. The method cascades deep networks with fast features. It has achieved miss rate of 26.2%.

C.A.B.Baker et al. (Baker et al. 2016) stated the concept for searching after calamity occurred with the help of UAVs. They used Monte Carlo algorithm for the searching process.

Fig. 7 Person detection output



Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:19	39.50%	0.7677	0.0010
50	50	00:13:02	53.44%	0.6538	0.0010
100	100	00:25:52	75.77%	0.5378	0.0010
150	150	00:38:48	80.28%	0.4734	0.0010
200	200	00:51:41	81.55%	0.4322	0.0010
250	250	01:03:56	82.37%	0.4033	0.0010
300	300	01:17:11	82.69%	0.3824	0.0010
500	500	02:08:27	85.37%	0.3229	0.0010
550	550	02:21:04	86.07%	0.3125	0.0010
600	600	02:33:45	85.21%	0.3025	0.0010
650	650	02:46:54	86.69%	0.2899	0.0010
700	700	03:00:26	88.11%	0.2796	0.0010
750	750	03:13:16	88.70%	0.2717	0.0010
800	800	03:25:56	87.78%	0.2687	0.0010
850	850	03:39:52	88.95%	0.2581	0.0010
900	900	03:53:10	88.83%	0.2520	0.0010
950	950	04:06:03	88.25%	0.2494	0.0010
1000	1000	04:19:16	88.77%	0.2447	0.0010
1050	1050	04:32:55	88.97%	0.2407	0.0010
1100	1100	04:46:12	89.99%	0.2353	0.0010
1150	1150	04:59:16	89.74%	0.2327	0.0010
1200	1200	05:12:25	89.63%	0.2318	0.0010
1250	1250	05:25:32	90.36%	0.2238	0.0010
1300	1300	05:38:24	89.99%	0.2228	0.0010
1350	1350	05:51:51	90.54%	0.2174	0.0010
1400	1400	06:06:16	90.57%	0.2109	0.0010
1450	1450	06:20:13	90.59%	0.2134	0.0010
1500	1500	06:32:56	90.49%	0.2080	0.0010
1532	1532	06:40:52	91.04%	0.2036	0.0010

Fig. 8 Person detection verbose plot

Advantage of this method is that it gives 7% speed than former searching method. Vijay Badrinarayanan et al. (Badrinarayanan et al. 2017) proposed SegNet which includes two modules. The first is the encoder portion and the other one is the decoder portion. The number of trainable parameters is less (14.7 M) in the SegNet which is considered to be the advantage of this architecture. LichaoMou et al. (LichaoMou et al. n.d.) introduced a dataset called ERA. It includes 2864 videos. It is used for recognizing events from videos that are taken with the help of UAV. Hanno Hildmann et al.

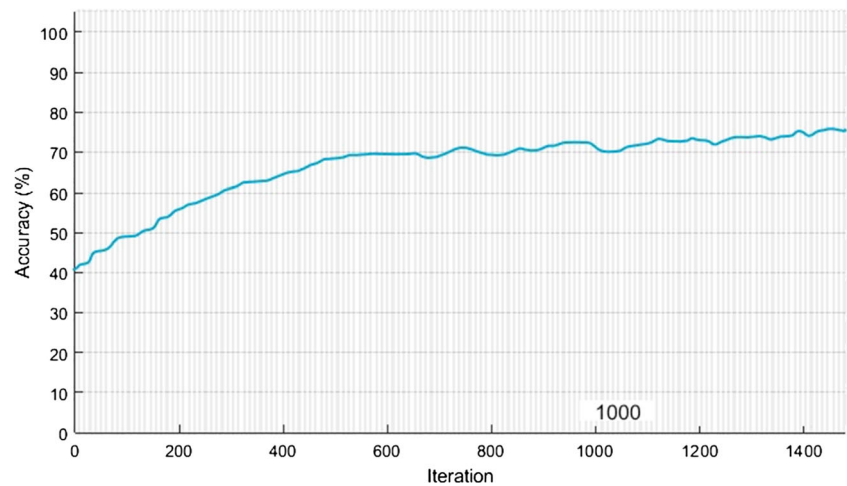
(Hildmann and Kovacs 2019) discussed a review based on application of UAVs in different social sectors. In that review they pointed out that UAV can be used for various applications such as safety of public people, disaster management, surveying for road making etc. Sven Gotovac et al. (Božić-Štulić et al. 2019) proposed a method for finding the presence of person in mediterranean and sub mediterranean landscapes with the help of UAV. The method used wavelet transform for saliency object detection and CNN for feature extraction. Model achieved accuracy of 88.9%. The main disadvantage of this method is that sometime image may contain partial person images like legs, hands, etc., that could not be detected by existing approaches. To overcome this drawback, our proposed model uses semantic segmentation and SegNet.

Najafzadeh & Zahiri (Najafzadeh and Zahiri 2015) elaborated neuro-fuzzy supported technique which basically applies GMDH dependent application to forecast flow release in linear compound channels. Zahiri & Najafzadeh (Zahiri and Najafzadeh 2018) have applied

Table 1 Person detection accuracy and loss comparison

Models	Accuracy	Loss
VGG 16	75.77%	0.5378
GoogleNet	80.28%	0.4734
ResNet	86.69%	0.2829
SegNet	91.04%	0.2036

Fig. 9 Person detection accuracy with VGG 16



many soft computing methods such as gene-expression programming (GEP), model tree (MT) and evolutionary polynomial regression (EPR) for predicting flow discharge in linear compound channels.

Methodology

In this proposed approach, we suggest an idea for detecting person from remote, wild or non-urban areas with the help of UAV. Main advantage of this proposed model is that, sometime UAV image may contain the partial person images, like legs, hand, etc., that could not be identified by existing approaches were being recognized and identified successfully. Fig. 1 shows the example dataset with various images containing partial parts of the person and other nonliving things. Fig. 2 represents the person detection architecture.

The proposed model uses Semantic segmentation method with SegNet architecture for person detection. Semantic segmentation is defined as the method used for segmentation where an image is segmented to multiple classes (example: Person, Background) and each class is assigned with different colors. SegNet is the network architecture which is used in the process of semantically segment the image according to each pixel. The proposed model includes six modules that are (1) Collect UAV image set, (2) Pixel wise labelling for Semantic segmentation, (3) SegNet architecture, (4) Training and

testing the network, (5) Build user interface, (6) Deployment to computer applications. The first four modules are included in semantically segmentation process.

Collect UAV Imageset

UAV is the latest method used in aerial imagery for searching purposes. Advantage of using UAV is that it is uncomplicated to use and very cheap rated. For taking images using this method, good resolution DSLR - Digital Single-Lens Reflex camera Cameras are attached with the UAV. At the same time of capturing the image also noted the location and height of the flight from the ground. It is important to set a constant image resolution. UAV image dataset used in this work is HERIDAL dataset. HERIDAL contains 68,750 aerial images that are mainly used for IPSAR - Institute of Professional Studies and Research projects. The proposed model uses images with dimensions of 500*500 for training and testing. For getting better result for the semantic segmentation process, the resolution and very high quality image is very important.

Pixel wise labeling for semantic segmentation

The proposed model has been implemented in MatLab2019a version. For the pixel wise labeling process, inMatLab there is a default option called image Labeler which is available for labeling the images.

Fig. 10 Person detection loss with VGG16

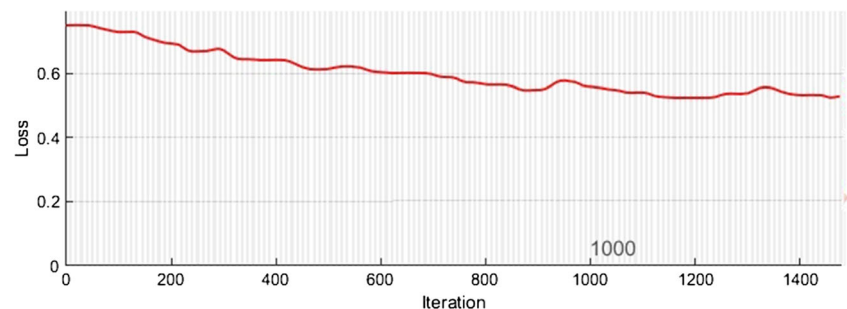
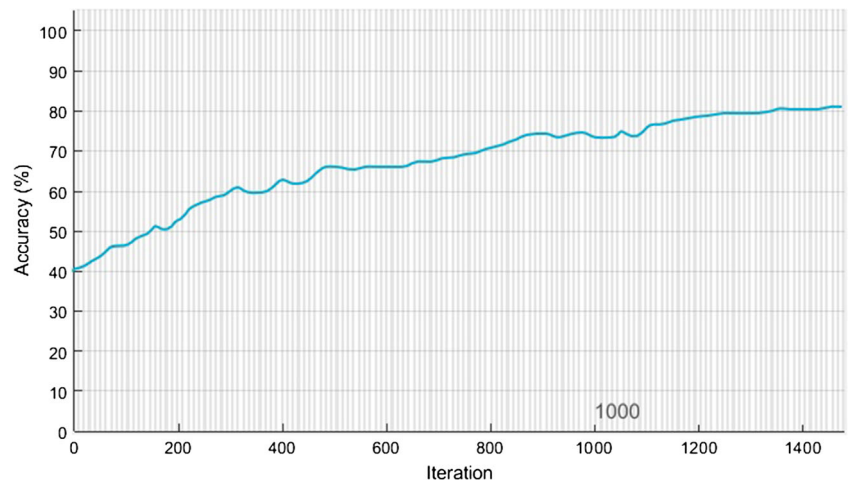


Fig. 11 Person detection accuracy with GoogleNet



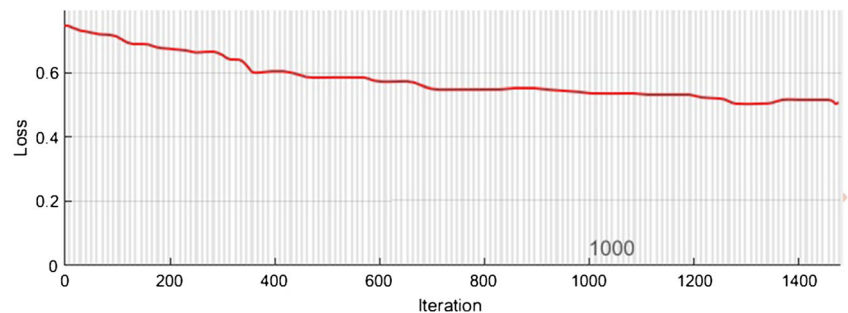
Load the image from dataset to image labeler in MatLab. Create 2 class labels named Person and Background using “define label” option. In this work, Person is assigned with blue color and Background is assigned with green color. In an image, all nonperson portions such as non-living organism and other animals is treated as Background. An image may contain partial body like lower body, upper body, hands, legs and so on is also labeled as Person. Labeling has been done for all the images which are taken for the training set and create a pixel label data store. The pixel-wise total cross-entropy for every pixel is computed using Eq. (1)

$$cross_{entropy} = -\frac{1}{pq} \times \sum_{i=1}^p \sum_{j=1}^q [\omega_{ij} \log \beta_{ij} + (1-\omega_{ij}) \log (1-\beta_{ij})] \quad (1)$$

where I and j are the pixel values, p is the value of height, q is the value of weight, ω_{ij} is the extra value, β_{ij} is the prediction value. The mean value of intersection over the union is computed for evaluating the index value to provide the segmentation process in Eq. (2)

$$mean_{IoU} = \frac{1}{pi} \sum_{i=1}^{pi} \frac{|\delta_i \cap \gamma_i|}{|\delta_i \cup \gamma_i|} \quad (2)$$

Fig. 12 Person detection loss with GoogleNet



The accuracy for the pixel is computed in Eq. (3)

$$Accuracy_{pixel} = \frac{\sum_{x=0}^k p_{i_{xx}}}{\sum_{x=0}^k \sum_{y=0}^k p_{i_{xy}}} \quad (3)$$

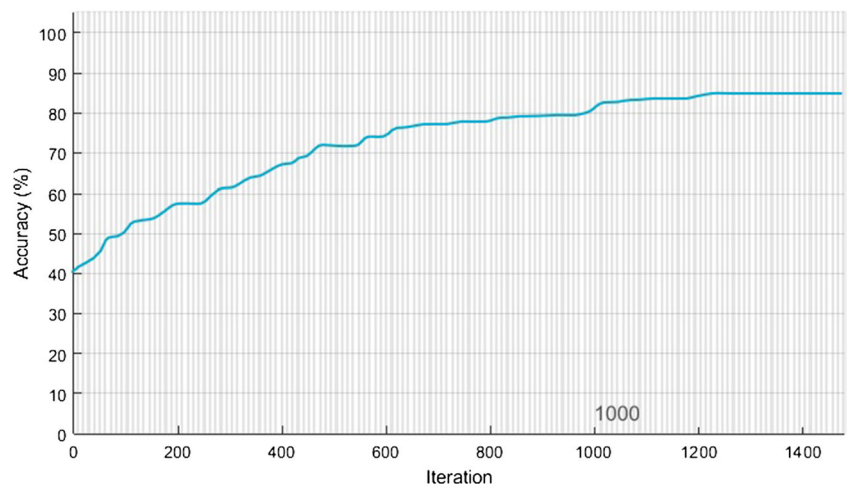
All images in the training set are made available by creating a pixel label data store which is depicted in Fig. 3.

For ex, Fig.4 depicts the example image before applying labeling and Fig.5 represents the labeled image as stated above.

SEgNET architecture

SegNet is the network architecture which is used in the process of semantically segment the image according to each pixel. The network mainly includes two modules. The first is the encoder portion and the other one is the decoder portion. Encoder part mainly contains convolution that is Batch normalized and rectified linear unit non linearity. Max pooling layer is included in the last portion of the encoder. Apply Maxpooling process with stride of 2 and window 2×2 . Perform subsampling function by factor 2 after performing Maxpooling. Using Maxpooling indices in the encoding sequence the sparse encoding due to Maxpooling process is upsampled in the decoder. Decoder part upsampled the encoder part output and given it to convolution layer, and finally it is gone through softmax classifier which classifies

Fig. 13 Person detection accuracy with ResNet



the image to two classes that is Person or Background. The Fig. 6 represents the diagram of SegNet architecture.

In the decoder portion, Maxpooling indices upsampled the lower resolution feature maps cause to cutting down the number of parameters. SegNet architecture is only convolutional. There are 13 convolutional layers in the encoder part of SegNet that is resembled to VGG16 architecture. In favor of maintaining feature maps of top resolution, avoided the fully connected layers. That is the reason why the number of trainable parameters is less (14.7 M) in the network. The decoder portion also includes 13 layers. Using stochastic gradient descent, the whole architecture trained end to end.

Training and testing THE NETWORK

In the process of training the network, image data store and the pixel label data store are imported to the SegNet with the size 500 × 500. Number of classes can be defined which includes the labeling of entire image into person, animal, background and so on. In this approach, number of classes has been assumed to be 2 such as Person and Background for our convenience. It is to set the training options such as Sgdm, Maximum epoch (2000), Initial learning rate (1e-3), Mini batch size (64), Sgdm is called the Stochastic Gradient Descent with

Momentum. Also, one epoch is defined as; it has to occur when the whole dataset is gone through network architecture in forward and backward direction in one single time.

Plot the training progress after the training process, and the particular plot is called verbose plot, which includes the detail such as epoch, iteration, time elapsed, Base learning rate, mini batch loss and mini batch accuracy. Read the test images from data store for testing and perform image overlay function. Test each image with the trained network and check whether the input image is properly segmented to two classes such as person or background.

Build user interface

In this model, 70% of the images were trained and 30% of the images were tested which is to be made available for the end users. User interface is the only part that is visible to the outside person that contains a button such as load test image, input UAV image showing section and output segmented image, spot at which person found, Person and Background color representing section. For the calculation of position at which person is found, take the origin as (0,0) and set pixel distance = 0.5.

Fig. 14 Person detection loss with ResNet

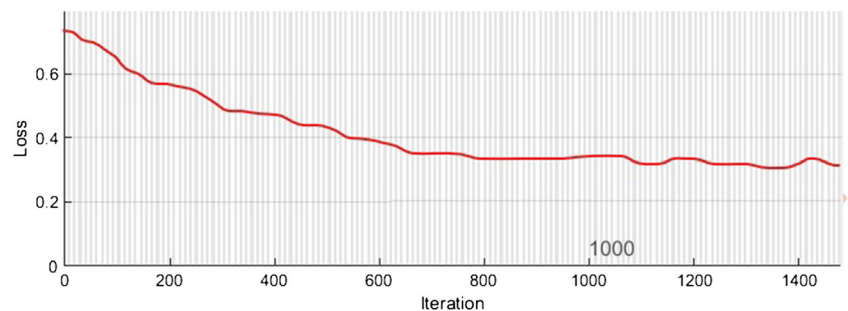
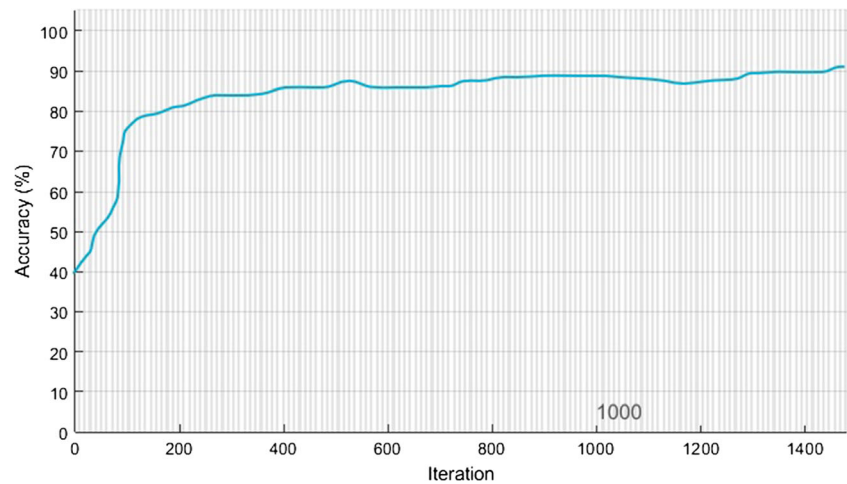


Fig. 15 Person detection accuracy using Semantic SegNetmodel



Deployment to computer applications

Using Application compiler in the MatLab 2019a, deployed the proposed model into a computer based application. Proposed system can run on any computer system that does not require MatLab to be installed using this deployment. Initially, import the created user interface file to the application compiler in the MatLab. Select the runtime included in the package which includes file size 754MB. It requires an installer to be generated that includes MatLab runtime installer. At last, customize the packaged application and its appearance. Fig. 7 shows the person detection output screen after deployment.

Results and discussions

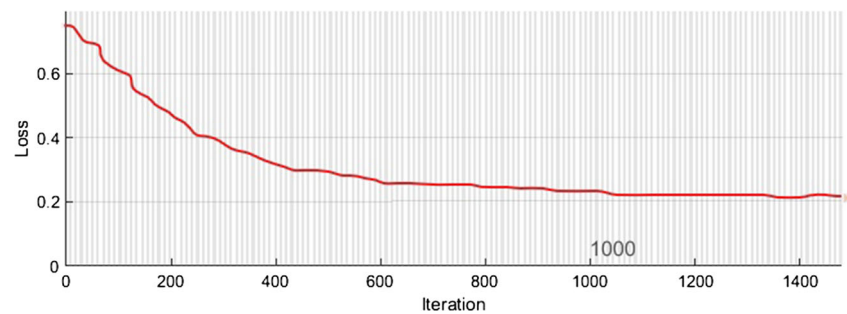
MatLab 2019a has been used for implementing this SegNet architecture which requires 13 min and 02 s to complete 50 iterations to achieve an accuracy of 53.44% and loss of 0.6538. For the entire training process, base learning rate achieved is 0.0010 with accuracy and loss calculated based on Mini batch. Completed 100 iterations at Epoch of 100 and achieved an accuracy of 75.77% and loss of 0.5378 after 25 min and 52 s. It achieved an accuracy of 81.55% and loss of 0.4322 at iteration of 200 at a time of 51 min and 41 s with Epoch of 200. It achieved accuracy of 82.69% and loss of

0.3824 with Epoch and iteration of 300 after 1 h 17 min and 11 s. It achieved accuracy of 85.37% and loss of 0.3229 at Epoch of 500 and iteration of 500 at a time of 2 h 8 min and 27 s. Completed 650 iterations at Epoch of 650 and achieved an accuracy of 86.69% and loss of 0.2899 after 2 h 46 min and 54 s. Completed 750 iterations at Epoch of 750 and achieved an accuracy of 88.70% and loss of 0.2717 after 3 h and 13 min and 16 s.

It achieved an accuracy of 88.95% and loss of 0.2581 at iteration of 850 at a time of 3 h 39 min and 52 s with Epoch of 850. After 4 h 32 min and 55 s it achieved an accuracy of 88.97% and loss of 0.2407 with Epoch and iteration of 1050. It achieved accuracy of 89.99% and loss of 0.2353 at Epoch of 1100 and iteration of 1100 at a time of 4 h 46 min and 12 s. For completing 1250 iterations it took 5 hours 25 min and 32 s, achieved accuracy of 90.36% and loss 0.2238 with Epoch of 1250. After 5 h and 51 min and 51 s completed 1350 iterations at Epoch of 1350 and achieved an accuracy of 90.54% and loss of 0.2174. It achieved accuracy of 90.57% and loss of 0.2109 at iteration of 1400 at a time of 6 h 6 min and 16 s with Epoch of 1400. After 6 h 20 min and 13 s it achieved accuracy of 90.59% and loss of 0.2134 with Epoch and iteration of 1450. Accuracy of this proposed system remains the same and fluctuates between 90.14% and 90.57% when the number of iterations gets increased.

Proposed model achieved Mini batch accuracy of 91.04% and mini batch loss of 0.2036 at Epoch of 1532 and iteration

Fig. 16 Person detection loss using Semantic SegNetmodel



of 1532 with in a time of 6 h 40 min and 52 s training process. This model is tested with various test images in the data store. Then, test image is semantically segmented into Person or Background. If a test image includes a person or partial body parts of a person, then the person is identified successfully and the position at which the person present is also shown in the output user interface. The Fig.8 represents the person detection verbose plot.

Comparison with VGG 16, GOOGLNET, RESNET

Table 1 represents the person detection accuracy comparison of the proposed model with VGG 16, GoogleNet, ResNet models. VGG 16 model achieved an accuracy of 75.77% with loss of 0.5378 which is shown in Figs. 9 and 10. GoogleNet model achieved an accuracy of 80.28% with loss of 0.4734 which is represented in Figs. 11 and 12. ResNet model achieved an accuracy of 86.69% with loss of 0.2829 which is shown in Figs. 13 and 14.

Proposed model Semantic SegNet - achieved an accuracy of 91.04% with loss of 0.2036. It shows that our proposed model achieved better accuracy and loss than the VGG 16, GoogleNet and ResNet models which are shown in Figs. 15 and 16.

Conclusion

In recent days, we are facing many natural calamities such as flooding, earthquake and so on. It is very difficult to rescue the needy people without proper location information even though we have many recent methodologies and techniques. Most of the techniques recommended rescuing the trapped people when the entire image of a person has been taken using UAV techniques. The proposed model used Semantic segmentation method with SegNet architecture for person detection. Semantic segmentation has been applied for segmenting the entire image into multiple classes such as Person, Background, and Animal and so on. In this technique, each class has been assigned with different colors after performing semantic segmentation on the input image according to each pixel. This model has been trained and tested using HERIDAL dataset for which over 70% images were trained and 30% images were used for testing. This work attempted to reduce the search time of identifying the people who got trapped in some of the places though only the parts of a person is visible in the images. This enhanced deep learning model achieved an accuracy of 91.04%. Another challenging task with respect to natural calamities is to detect a person during night time. In future, Semantic SegNet model will be extended to detect person at night vision.

References

- Angelova, A, Krizhevsky, A, Vanhoucke, V, Ogale, A and Ferguson, D (2015). Real-time pedestrian detection with deep network cascades. In Proceedings of BMVC 2015
- Anna Gaszczak, JH and Breckon, Toby P (2011). Real-time people and vehicle detection from UAV imagery (Vol. 7878, pp. 7878–7878-13)
- Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. PAMI
- CAB Baker, S Ramchurn, WT Teacy, NR Jennings (2016). Planning search and rescue missions for UAV teams, proceedings of the twenty-second European conference on artificial intelligence, Netherlands, IOS Press
- Božić-Štulić D, Marušić Ž, Gotovac S (2019) Deep learning approach in aerial imagery for supporting land search and rescue missions. Int J Comput Vis 127:1256–1278
- Chenyi Chen, Ming-Yu Liu, Oncel Tuzel, Jianxiong Xiao (2017). R-CNN for Small Object Detection. Springer International Publishing AG, (pp. 214–230)
- R Girshick (2015). Fast r-cnn. In International Conference on Computer Vision (ICCV)
- Harold Robinson Y, Golden Julie E (2019) SMR: A Synchronized Multipath Rebroadcasting Mechanism for Improving the Quality of Conversational Video Service. Wirel Pers Commun 104(3): 1149–1173
- K. He, X. Zhang, S. Ren, and J. Sun (2015). Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385
- Hanno Hildmann and Ernő Kovacs (2019). Review: using unmanned aerial vehicles (UAVs) as Mobile sensing platforms (MSPs) for disaster response, Civil Security and Public Safety.Drones
- Deepak Jaiswal ,Praveen Kumar (2019). Real-time implementation of moving object detection in UAV videos using GPUs. J Real-Time Image Proc
- J Joshua Bapu, D Jemi Florinabel, Y Harold Robinson, E Golden Julie, Raghvendra Kumar, Vo Truong Nhu Ngoc, Le Hoang Son, Tran Manh Tuan, Cu Nguyen Giap (2019). “Adaptive convolutional neural network using N-gram for spatial object recognition”, Earth Science Informatics, pp 1–16
- A Krizhevsky, I Sutskever, and GE Hinton (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105
- LichaoMou, YuanshengHua, Pu Jin, and Xiao Xiang Zhu. ERA: A Dataset and Deep Learning Benchmark for Event Recognition in Aerial Videos
- Mohammad Najafzadeh and Abdolreza Zahiri (2015). Neuro-Fuzzy GMDH-Based Evolutionary Algorithms to Predict Flow Discharge in Straight Compound Channels. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001185](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001185), 20, 04015035
- H Nam and B Han (2015). Learning multi-domain convolutional neural networks for visual tracking. arXiv preprint arXiv:1510.07945
- J Redmon, S Divvala, R Girshick, and A Farhadi (2015). You only look once: Unified, real-time object detection. arXiv preprint arXiv: 1506.02640
- S Ren, K He, RB Girshick, and J Sun (2015). Faster R-CNN: towards real-time object detection with region proposal networks. CoRR, abs/1506.01497
- Rudol, P and Doherty, P (2008). Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery. In 2008 IEEE aerospace conference (pp. 1–8)
- K Simonyan and A Zisserman (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
- C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich (2015). Going deeper with

- convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9
- Tian, Y, Luo, P, Wang, X and Tang, X (2015). Deep learning strong parts for pedestrian detection. In: 2015 IEEE international conference on computer vision (ICCV) (pp. 1904–1912)
- Turić H, Dujmić H, Papić V (2010) Two-stage segmentation of aerial images for search and rescue. *Information Technology and Control* 39:138–145
- Viola, P, Jones, MJ and Snow, D (2003). Detecting pedestrians using patterns of motion and appearance. In proceedings ninth IEEE international conference on computer vision (Vol. 2, pp. 734–741)
- Abdolreza Zahiri and Mohammad Najafzadeh (2018). Optimized expressions to evaluate the flow discharge in main channels and floodplains using evolutionary computing and model classification, *International Journal of River Basin Management*, Volume 16, - Issue 1
- Matthew D. Zeiler and Rob Fergus (2013). Visualizing and Understanding Convolutional Network

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.