

TOWARDS AN AUTOMATED FLOOD AREA EXTRACTION FROM HIGH RESOLUTION SATELLITE IMAGES

R.G.C.J. Kapilaratne ^{1,*}, S. Kaneta ¹

¹ Asia Air Survey Co. Ltd., 215-0004, Asao Ku, Kawazaki City, Kanagawa, Japan - (jie.kapilaratne, sni.kaneta)@ajiko.co.jp

KEY WORDS: Disaster mapping, Remote Sensing, Optical Satellite Images, Deep learning, Convolutional Neural Networks, Emergency Response Systems

ABSTRACT:

Flooding is considered as one of the most devastated natural disasters due to its adverse effect on human lives as well as economy. Since more population concentrate towards flood prone areas and frequent occurrence of flood events due to global climate change, there is an urgent need in remote sensing community for faster and reliable inundation mapping technologies to increase the preparedness of population and reduce the catastrophic impact. With the recent advancement in remote sensing technologies and integration capability of deep learning algorithms with remote sensing data makes faster mapping of large area is feasible. Therefore, this study attempted to explore a faster and low cost solution for flood area extraction by integrating convolution neural networks (CNNs) with high resolution (1.5m) SPOT satellite images. By consider the system requirement as a measure of cost, capabilities (speed and accuracy) of a deeper (ResNet101) and a shallower (MobileNetV2) CNNs on flood mapping were examined and compared. The models were trained and tested with satellite images captured during several flood events occurred in Japan. It is observed from the results that ResNet101 obtained better flood area mapping accuracy than MobileNetV2. Whereas, MobileNetV2 is having much higher capabilities in faster mapping in 0.3 s/ km² with a competitive accuracy and minimal system requirements than ResNet101.

1. INTRODUCTION

1.1 Background

During the past decade, world's water related disasters became frequent and severe as an adverse effect of changing pattern in global climate. According to annual report of HELP global report on water and Disaster 2019, year 2018 was marked highest number of water-related disasters in all parts of the world. Around 6,500 people lost their lives by water-related disasters among them majority are corresponds to flood events. Although, the global efforts and advancement of science and technology towards disaster mitigation approaches exacerbated, the amount of economic and human losses couldn't controlled as expected due to the increase in concentration of humans towards the vulnerable areas for flooding. Particular in Japan, population is more likely to concentrate on flood-prone areas besides the fact that geography and the climate of japan is more susceptible for frequent extreme hydro-meteorological events. According to infrastructure development institute of Japan and Japanese river association, about 49% of the population and 75 % real estates of the country are located in alluvial plains exposed to flood risk. Therefore, development of faster mapping of disaster events such as flooding is vital for authorities for their emergency response and proper action plans towards the protection of human lives and assets. Additionally, timely damage assessment is also crucial for better relief work planning as well.

With the advancement of the satellite imagining technologies during the last few decades and capabilities of commercial satellites to capture images on demand within hours, remote sensing rewarded as highly demanding surveying option for disaster mapping in near real time.

Existing studies on this regard, were mostly focused on manual methods such as change detection from pre and post disaster event using image algebra (band differencing, band rationing),

post classification comparison and object-based change detection (Amit and Aoki, 2018). However, the mapping performance of aforementioned methods vary upon several factors including study area, water characteristics, environmental and atmospheric conditions etc. (Goffi et al., 2020). Most importantly, such manual method requires longer time and unable integrate with an emergency response systems in the contexts of time, accuracy and automation capability. Thus, remote sensing community is always committed to develop technologies for better performance in near real time. Where, researchers found the best possible solution was to introduce neural networks, the basis of Deep Learning (DL) into such remote sensing applications (Ma et al., 2019, Chen et al., 2017; Zhu et al., 2017) due to its capability to produce accurate results for a larger area within almost no time in comparison to traditional methods.

Various DL approaches including Convolutional neural networks (CNN) and recurrent neural network (RNN) widely used for various remote sensing applications such as image classification, segmentation and object detection. Among them CNN have attained surprising dominance in remote sensing community due to its impressive results in challenging applications such as image classification and segmentations. Moreover, its image handling capability and integration ability of automatic feature extraction attracted majority of the researchers towards DL from conventional image processing technologies (Yang and Cervone, 2019).

1.2 CNN for Image Classification and Segmentation

CNNs are type of feed-forward artificial neural network made up of layers which includes learnable parameters including weights and biases (Gebrehiwot et al., 2019). Many studies have been carried out in CNN based image segmentation methods such as fully convolutional network (FCN), Seg-Net, U-Net etc. However, most of the remote sensing applications were used FCN based models as fully connected neural layers were replaced

by convolution neural layers to preserve the 2-D structure of the satellite images.

There are great number of works focused on FCN based approaches for image classification and segmentation. Among them works carried out by Castelluccio et al., (2015); Fu et al.,(2017) , Marmanis et al.,(2016) and Ireland et al., (2015) obtained promising results with about 80%- 90% of class accuracy. The common finding with aforementioned studies and other researches on similar direction is deeper convolutional networks corresponds to better results. It is highly agreed with the existing literature about FCN that higher feature extraction level can be achieved by going deeper network architectures.

1.3 Challenges of CNN Integration with Remote Sensing Applications

One of the challenging issues of segmentation problems in remote sensing applications is that to balance the trade-off between strong down sampling (for better feature extraction with deeper networks) and accurate boundary localization. The most commonly used solution was to use hybrid FCN architecture (e.g. Sherrah, 2016). Although it improves the results this adds extra complexity to the model and leads to a longer training time and requires much computation resources. Most importantly, longer processing time accompanies with deeper networks highly matters when it comes to applications like disaster detection systems. Thus, there is a great necessity both in computer vision and remote sensing community to build and identify smaller and effective neural networks for balance the trade-off with minimum resources. Consequently, Iandola et al., (2016); Wu et al., (2015) and Wang et al., (2016) demonstrated effective small scale networks with competitive accuracy. However, the primary focus of these models were to minimize the overall size of the model to address the issue of resource restriction regardless of the model capabilities. In order to address both time and the resource issues, Howard et al., (2017) developed MobileNets network which can be used even in mobile devices without any advanced system properties. Yet, there is no or minimum attention has been given by the remote sensing community to investigate its potential for remote sensing application.

1.4 Motivation and Manuscript Organization

The motivation of this study was to examine and demonstrate the potential of a smaller network (MobileNet) for remote sensing applications with specific emphasis on flood area extraction from optical satellite data using image segmentation technique in order to integrate with an automatic disaster detection system. The performance of MobileNet for flood area extraction was compared with the readily used deep CNN network (Resnet101) which won the best FCN architecture with a larger margin at highly competitive computer vision challenge of ILSVRC 2015. This manuscript is organized as follows. The subsequent section consists of brief introduction and comprehensive comparison of MobileNet network with widely used Resnet101 network architecture followed by data and methodology sections. Thereafter, results and discussion sections can be found along with conclusion of the study.

2. MOBILENET ARCHITECTURE

MobileNet network is primarily built with depthwise separable convolutions (filters) know as factorized convolutions. These convolutions factorize standard convolutions into depth wise convolution and a 1x1 (point wise) convolution. Figure 1 is

illustrates the difference between typical convolution layer and a depth-wise separable convolution layer.

A regular convolution layer applies a convolution kernel (or filter) to all of the channels of an input image. It slides the kernel across the image and each step performs a weighted sum of the input pixels covered by the kernel across all input channels. But in the depth wise convolutions has two layers, one for filtering and the remaining is for combining. As of Howard et al., (2017) if a network uses 3x3 depth wise separable convolutions, 8~ 9 times computation cost reduction can be observed for a negligible reduction in accuracy. Similar observation has been experienced in this study as well. Further explanation on this regard will be discussed in results section of this manuscript.

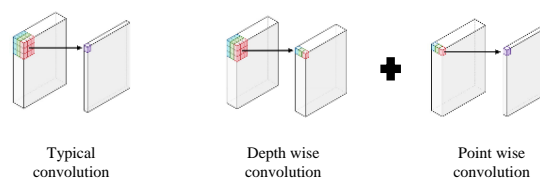


Figure 1. Comparison of typical convolution and depth-wise separable convolution

2.1 Comparison of Network Architectures of ResNet and MobileNet

The primary difference of the two networks can be found in the developmental objective of the two networks. The main focus of ResNet network development was not to computationally economical but to obtain the highest possible computational accuracy. Consequently, deepest network at present with 152 layers was designed. Since the network is much deeper it faced an issue of optimization. Furthermore, such deep networks having the trade-off of size and latency. Which alleviate such networks for being utilized in real world applications with computationally limited platforms such as autonomous vehicles, robotic visions and automated applications for near real time systems regardless of the higher computational accuracy. In contrast, MobileNet was designed to perform with competitive accuracy with minimal computation cost (Howard et al., 2017). Table 1 discusses model architecture difference in ResNet101 and MobileNet V2 (which was implemented in this study). For all the experimental purposes in this study we have used ResNet101dilated and MobileNetV2dilated as encoder and a common PPM+ deep supervision trick decoder for ResNet101 and MobileNetV2 models respectively.

Model Name	ResNet101	MobileNetV2
Number of feature maps	2048	320
Number of parameters (millions)	42.5	4.3
Encoder	ResNet101dilated	MobileNetV2dilated
Decoder	PPM + deep supervision trick	PPM + deep supervision trick

Table 1. Comparison of model architecture used in this study

For better use of the image features and to reduce the burden in design phase of the system, pre-trained version of the aforementioned models were fine-tuned with remote sensing

images. Several researchers have stated the advantages of using pre-trained networks despite using a network train from scratch. Among them Sherrah, 2016, one of the very first discussed successfully such applications in remote sensing images.

Both models were pre-trained with the densely annotated ADE20K dataset, which spans diverse annotations of scenes, objects, parts of objects and in some cases even parts of parts (Zhou et al., 2018). Further, Zhou et al., 2018, mentioned that pre-trained models with ADE20K data set is having high potential to segment wide variety of objects and scenes. Models were trained using images contained 10.5 object classes and 19.5 instances. Overall, 20000 images were used to train for 150 object categories.

Moreover, the capability of trained models with large number of object categories for remote sensing applications is rare or has not been documented. Thus, in this study authors attempted to explore the potentials and capabilities of such pre-trained networks for remote sensing applications.

2.2 Related Works

With the availability of high spatial and spectral resolution data containing complex details of land features, there was an urging need in developing an advanced mapping technologies with abilities to cater issues with conventional rule based technologies and potentials to handle data with larger complexities. More importantly when it comes to flood area mapping, rule based technologies tend to fail due to lack of abilities to handle high spectral variations of flood waters and mixing of spectral properties of land and flooded pixels (Feng et al., 2015; Sarker et al., 2019). Consequently, yet there is no successful regional scale work has been found in literature for flood area mapping with rule based technologies despite few successful local scale works (Ogashawara et al., 2013, Amarnath 2014, Haq et al., 2012).

Further, object-based mapping also failed in the context of flood area mapping as flood water spread along the other land use classes create inter-class spectral similarity alongside the intra-class spectral heterogeneity (Sarker et al., 2019). Such issues evolved flood mapping into a new era with deep neural networks (DNNs). With the recent advancements in computer peripheral for speed up computations such as graphical processing units (GPUs), DL achieved much attention not only in computer vision applications but also in its sub divisions such as remote sensing due to its high potential of integrating with near real time information dissemination systems.

Disaster area recognition is one of the key areas in remote sensing where the application of DL, reported an increase trend. The main cause for the observation was due to its adaptability and generalization nature in comparison to the conventional disaster mapping technologies. A handful of works can be found in recent literature on flood mapping using CNN methods. Those studies can be found mainly in two categories, depending whether the change detection technique was used or not. Amit et al., (2017) used pre and post event images for the extraction of disaster area from aerial images. Even though the authors argued the potential of the method to be implemented on disaster detection mechanism around the world, yet there are several limitations with the methodology which need to be addressed. The main drawback is the necessity of image pre-processing. As the colour variation of remotely sensed images controlled by many factors such as satellite sensor, spatial location, seasonality etc., regional application of the method will be costly. Moreover, one of the major drawbacks of change detection mechanisms is the

necessity of pre disaster image. As the occurrence of the disaster events are unpredictable in the contexts of location and time, requirement of pre-disaster image cease the usability of the method in an emergency response systems.

Sarker et al., (2019) and Nogueira et al., (2018) proposed new CNN network architectures for flood area mapping using low and high resolution satellite images respectively. Where the Sarker et al., (2019) used Landsat images while Nogueira et al., (2018) applied to images with ground resolution 3m. However, flood area extraction ability of the Sarker et al., (2019) need to be improved further for its applicability with emergency response systems. Whereas, the CNN model developed by Nogueira et al., (2018) apparently have high potential of being adopted for emergency response system however, the lack of information about generalisation ability and computational performance makes it difficult understand its appropriateness. Therefore, in this study authors tried to investigate on a low cost and accurate CNN based flood area extraction solution for high resolution optical satellite images which can be adopted for near real-time emergency warning system with a minimal system requirement.

3. DATA AND METHODOLOGY

The proposed system of this study consists of three distinct steps. During the first step training and validation data sets will be prepared. Thereafter, a model training and training accuracy estimation will be performed. Model validation will be evaluated as the second step. As for the final step, testing will be performed with each trained models. The overview of the overall methodology used in this study is shown in Figure 2.

3.1 Data

All the training images were taken by SPOT 6 and SPOT 7 satellite sensor during or just after the heavy rainfall events occurred in Joso, Kurashiki and Osaki cities from Ibaraki, Okayama and Miyagi prefectures of Japan respectively. Whereas, test images were observed at Fukushima and Saitama prefectures. Map of training data and test sites used in this study are shown in Figure 3. Further information about the data sets can be found at Table 2 and 3.

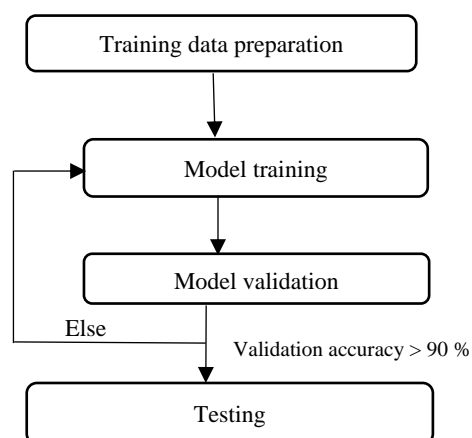


Figure 2. Overall methodology used in the study

3.1.1 Training Data Preparation: All the satellite images used in this study were pan sharpened in ERDAS Imaging 2016 environment using HCS resolution merge technique. The original resolution of panchromatic images and multispectral images are 1.5 m and 6.0 m respectively.

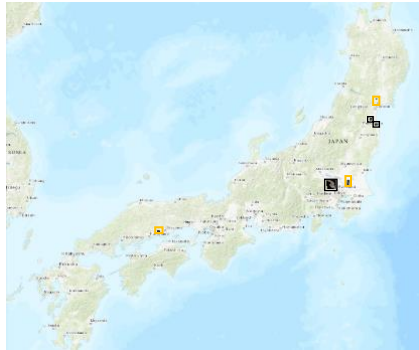


Figure 3. Location Map of training and test data sets. Training data sets bounded by yellow boxes © SPOT image AIRBUS DS 2018

Location	Joso	Osaki	Kurashiki
Satellite sensor	SPOT 6		SPOT 7
Resolution	1.5 m		
Colour composite	True colour :Red: band 1 Green: band 2 Blue: band 3		
Occurrence of event	2015/09/9-11		2018/07/5-5
Imaging date	2015/09/12	2015/09/11	2018/07/09
Sample size (pixels)	512 x 512		
Number of samples	1211		

Table 2. Training data sets used in the study

Location	Abukuma, Arakawa, Utagawa
Satellite sensor	SPOT 7
Resolution	1.5 m
Colour composite	True colour :Red: band 1 Green: band 2 Blue: band 3
Occurrence of event	2019/10/13-15
Imaging date	2019/10/13
Sample size (pixels)	512 x 512
Total area	693 Km ²

Table 3. Test data sets used in the study

The true colour composites were prepared from respective pan sharpened images. No other pre-processing steps were performed for the data sets used in this study. The data sets used to train the models consists with images obtained under different conditions (colour variation, spectral characteristics etc.). In order to avoid the possible negative effects on a smooth automating process of flood area extraction, manual intervention in image pre-processing steps such as colour variation balancing were not undertaken. Thus, model was trained using satellite images with minimal pre-processing exposure.

3.1.2 Annotation Data Preparation: Annotation masks used in this study contain labels from 1 to 3 representing no data, land (remainders excluding other two categories) and flooded area respectively. In order to improve the accuracy of training data preparation, true colour and false colour composites of both post disaster and pre disaster images along with NDWI image were used as shown in Figure 4. Frequently encountered issue during annotation data preparation was misclassification of empty paddy lands with paddy area covered with flooded water. To overcome aforementioned error, NDWI image was utilized along with pre and post disaster images since literature (e.g. Soltanian et al., 2019; Qiao et al., 2011 and Rokni et al., 2019) emphasized the advantages of using NDWI images towards flood area extraction. The expert was carefully compared all three images of each scene and prepared the annotation files in Adobe Photoshop CC 2019 environment. The annotation images used in this study was in grey scale.

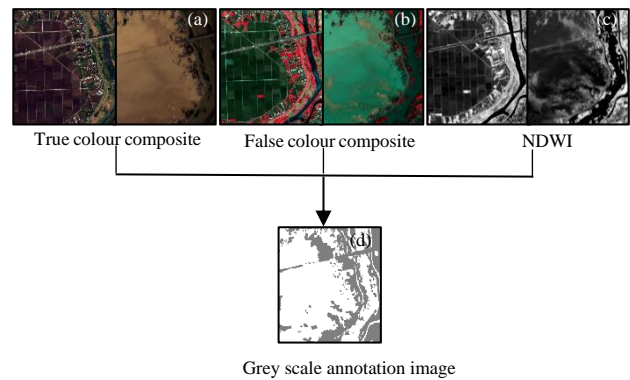


Figure 4. Annotation data preparation ((a) True colour composite of SPOT image (b) false colour composite (c) NDWI image (d) annotation image) © SPOT image AIRBUS DS 2018

3.1.3 Image Registration and Tilling: Consequently, as shown in Figure 5 image registration was carried out for the annotation file in ERDAS imaging 2016 environment with corresponding post-disaster image by matching the map model and spatial reference. Consequently, spatially referenced annotation files and image files were diced into 512x512 before start the training process. All the testing images were also tilled into same size with training images. Thereafter, 1/10 of annotation and respective post disaster image tiles were randomly separated for validation purposes.

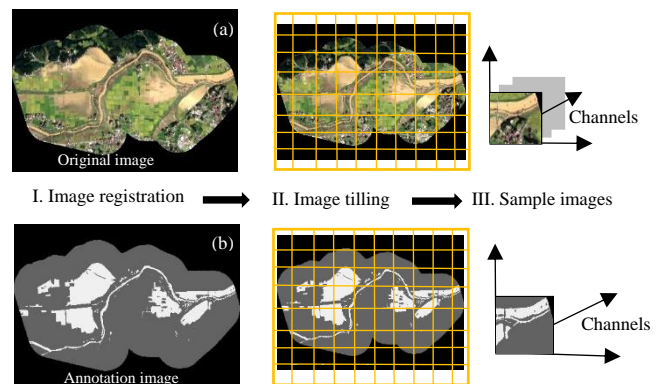


Figure 5. Training image sample preparations ((a) True colour composite of SPOT image (b) corresponding annotation image of Osaki Japan) © SPOT image AIRBUS DS 2018

3.2 Training Phase

This study adopted the sample program from Zhou et al., (2018) and used pre-trained version of Resnet101 and MobileNetV2 with ADE20K dataset. Two GPUs (graphical processing unites) were used for training process and only one GUP was used during testing phase. The experimental setup used in this study is summarized at Table 4. Post disaster image files and corresponding annotation files were used to train the networks. The ppm_deepsup was the common encoder for networks while resnet101dilated and mobilenetv2dilated were used as decoders for Resnet101 and MobileNetV2 respectively. 2048 feature channels were used in Resnet101 whereas, MobileNetv2 used 320 feature maps.

During the training process, number of batch per GPU , optimisation algorithm and learning rate for decoder and encoder were commonly set for both models as 2, SGD (Stochastic gradient descend) and 0.02 respectively. The training process was carried out for 30 epochs with 5000 iterations per each epoch.

Operating system	Ubuntu 18.04
Framework	Pytorch 1.3.1
Networks used	ResNet101, MobileNetV2
GPU	NVIDIA TITAN RTX VRAM:24GB Driver Version: 430.64 CUDA Version: 10.1

Table 4. Experimental setup used in the study

3.3 Validation and Testing Phase

Once the training process completed, trained models were validated with the validation data set. Models were trained for further epochs unless model reached more than 90% validation accuracy. The accuracy of the validation phase was estimated using Jaccard index or intersection over union (IOU) (Equation (7)). For each class, mean IOU and pixels wise accuracy (Equation (8)) were estimated. The estimated model weights of a certain epoch at which model gained more than 90 % accuracy used for inference.

693 km² area was available for testing purposes. However, annotation data was available only for the Abukuma area covering 67.19 km². Thus, quantitative analysis of test data was carried out only for the Abukuma area while qualitative analysis was conducted for whole area.

$$IoU = \frac{(true\ positives)}{(true\ positives + false\ positives + false\ negatives)} \quad (7)$$

$$pixel\ wise\ accuracy = \frac{correctly\ classified\ pixels}{labelled\ pixels} \quad (8)$$

$$F - score = 2 \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (9)$$

The accuracy assessment of testing phase was carried out mainly using recall and precision indexes which are commonly used in data science. For the tested categories in this study, precision and the F-score was estimated. It is observed that ResNet101model obtained more than 95% training accuracy and less than 0.01 loss at the 23rd epoch, while MobileNetV2 reached similar training accuracy at 29th epoch. Thus, evaluation was carried out using the weights at 30th epoch for both the models.

		Ground truth labels			Precision
		No data (1)	Land (2)	Flood area (3)	
Test result label	No data (1)	a	b	c	a/(a+b+c)
	Land (2)	d	e	f	e/(d+e+f)
	Flood area (3)	g	h	i	i/(g+h+i)
Recall		a/(a+d+g)	b/(b+e+h)	c/(c+f+i)	

Table 5. Accuracy assessment for test data set

4. RESULTS AND DISCUSSION

Results and discussion section is organized as follows. Subsection 4.1 discusses about the validation phase while other sections presents the results and discussion about test phase.

4.1 Training Time and Validation Accuracy

As mentioned in the previous section model training was carried out until validation accuracy obtained more than 90%. ResNet101 model obtained the validation accuracy more than 90% even with the estimated weights at 23th epoch. Whereas, MobileNetV2 was reached the same level of accuracy with 30th epoch weights (Figure 6). Table 6 summarizes the training time and validation accuracy of the models. As of the system capabilities used in this study, in average more than half hour training time difference was observed for an epoch. Overall, ResNet101 obtained above 91% accuracy whereas MobileNetV2 able to obtained above 90% validation accuracy. However, obtained difference in class IoU during model validation is smaller. Thus, in comparison to used training resources, MobileNetV2 obtained a competitive accuracy with a minimal system requirement.

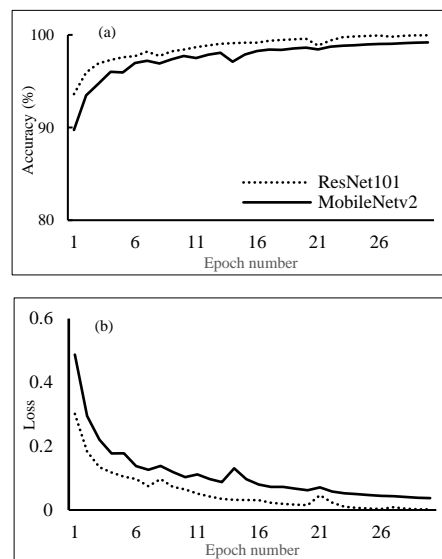


Figure 6. Comparison of training accuracy (6(a)) and loss (6(b)) of models

Model name	Training Time (h:mins)	Intersection over Union			Accuracy (%)
		No data	Land	Flood	
ResNet101	27 h 41m	0.9739	0.8445	0.8518	91.17
MobileNetV2	10 h 25 m	0.9825	0.8479	0.8699	90.17

Table 6. Accuracy comparison of models during validation

4.2 Test Results

As mentioned in section 3, qualitative and a quantitative analyses were carried out with test results. During the qualitative analysis test results were discussed comparing the original post disaster image. Whereas, quantitative analysis was carried out mainly on the following two contexts. (i) processing time and (ii) accuracy. According to the given explanation in subsection 3.3, accuracy was estimated. As illustrated in Figure 7, processing time was estimated from the time taken by tiling of original image up to creation of final geo-referenced mosaic image.

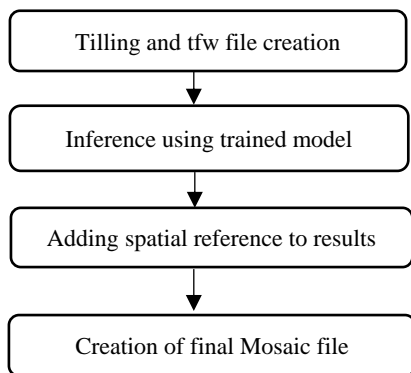


Figure 7 Processing steps of test phase of the model

4.2.1 Processing Time Estimation:

Tiling and tfw file creation: The test images were diced into 512 x 512 size before feed them to the trained model. Due to possible occurrence of header information loss during testing, spatial reference need to be added before creation of mosaic image.

Inference: Thereafter, tiled images were added to trained models for inference.

Geo-referencing and Mosaic image creation: Subsequently, obtained result images were geo-referenced using corresponding tfw files. Finally, Mosaic images were created using spatially referenced image files.

	Area (km ²)	ResNet101 (seconds)	MobileNetV2 (seconds)
Arakawa	553.6	367.07	152.59
Abukumagawa	67.2	67.13	22.52
Utawaga	73.0	69.58	23.80

Table 7. Comparison of processing time during testing

As per the obtained results summarized in Table 7, to process an area of 1 km² MobileNetV2 required in average 0.3 seconds for all 03 test sites while ResNet101 need 0.7 s/km². Overall, MobileNetV2 probably save about 7 minutes time during a processing of 1000 km² area. Saving 7 minutes computation time might not have much value unless it is an integrated system developed for dissemination emergency warnings during a flood event. In such situation, earlier information retrieval will lead to save invaluable human lives as well as economic assets. Though it is extremely important to save time, accurate flood area extraction also matters.

Flood Area Extraction Accuracy: Due to non-availability of ground truth data of Arakawa and Utawaga areas, flood area extraction accuracy was estimated only for Abukumagawa region.

4.2.2 Flood Area Extraction Accuracy: Due to non-availability of ground truth data of Arakawa and Utawaga areas, flood area extraction accuracy was estimated only for Abukumagawa region. Figure 8 represents the extracted inundated areas by ResNet101 and MobileNetV2 models.

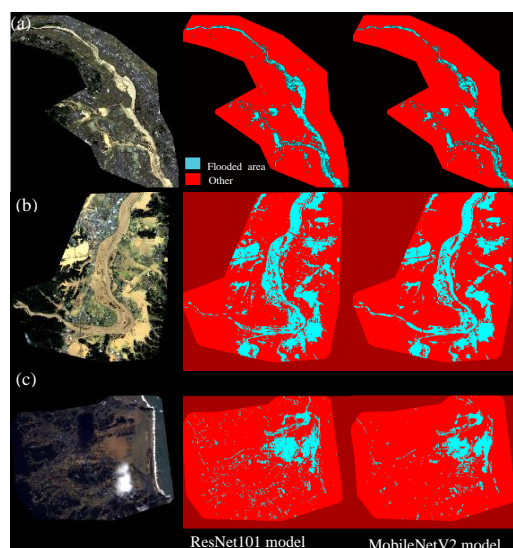


Figure 8. Comparison of flood area extraction (a). Arakawa (b). Abukumagawa (c). Utawaga © SPOT image AIRBUS DS 2019

Overall, it is obvious that the false extraction and mis-extraction of flooded areas are comparatively low for all three testes areas. However, false extraction of flooded areas are relatively high in Utawaga region than other areas. Such observation is more likely to correlate with the larger paddy coverage of Utawaga region than other two areas. Both models were more or less equally failed to extract inundated areas covered with less amount flood water regardless of their overall success in mapping flooded areas with greater accuracy.

Inundated area extraction capability of two models were quantitatively analysed using Abukumagawa ground truth data. The post disaster image and the corresponding annotation image data are shown in Figure 9. During the analysis it is found that average percentage difference of extracted flood area between ResNet101 model and MobileNetV2 models is 1.6% for Abukumagawa region. Thus, the deeper ResNet101 obtained higher flood area extraction capability for the tested 03 sites. However, qualitatively it can be observed that there are several areas where MobileNetV2 shown better performance than ResNet101 in flood area mapping. The estimated precision recall and *F-score* estimations for flood area extraction are summarised at Table 8.

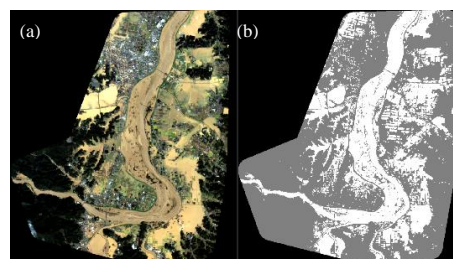


Figure 9. Post disaster image (a) and annotation image (b) of Abukumagawa © SPOT image AIRBUS DS 2019

Model	ResNet101	MobileNetV2
Precision (%)	81.5	80.1
Recall (%)	80.1	79.6
F-score	0.80	0.79

Table 8. Accuracy comparison of flood area mapping

The observed slightly higher precision and recall values of ResNet101, demonstrates the lesser false extraction and mis-extraction of inundated area mapping capability than MobileNetV2. Overall, both models were reported higher mis-extraction than false extraction which can be observed with Figure 8 as well. However, in comparison to MobileNetV2, comparatively higher precision values than recall of ResNet101 indicates the higher mis-extraction than false extraction tendency of the model. The differences of extracted flooded areas in Arakawa and Utawaga with two models were also less than 1%. This indicates further that the both models having almost similar capability to map the inundated areas. However, when it comes to the mapping of slightly inundated areas with lesser flood water (e.g. circled area in yellow of Figure 10) was not up to the desired level and need to be further improved.

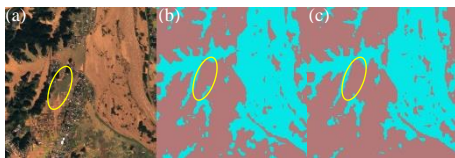


Figure 10. Inundated area mapping with (a) ResNet101 (b) MobileNetV2 models © SPOT image AIRBUS DS 2019

The possible fact caused such inability regardless of the size of the model, might be the spectral characteristics difference in training and test data sets at lightly inundated areas. Further, this study used quite larger sample size at training phase in order to save processing time at testing, this might lead to knowledge loss about such areas by the network. The similar works in literature (11*11 (Sarker et al., 2019) and 32*32 (Amit and Aoki, 2017)) used extremely smaller samples for training. Hence, authors will test such possibilities in future for the possible improvements on final results while preserving the processing speed.

Moreover, this study did not conduct any image pre-processing tasks (e.g. Amit and Aoki, 2017) to eliminate the effects of possible colour variations caused due to differences in weather conditions of the imaging date. Lack of such normalizations during training data preparations may also contributed negatively to the final results. Nevertheless, at the current stage of the study able to use minimal training samples covering only three prefectures of Japan. As explained at the Section 1 and 2 automation flood mapping become extremely difficult due to larger variation in spectral characteristics of flood water, hence it is obvious that networks need to be trained with more number of training samples covering different regional inundation events. Moreover, it should be noted that there is an inevitable addition of human error factor at the preparation of annotation images due to unavailability of ground truths for corresponding flood events at the time of this manuscript preparation.

By comparing the obtained higher accuracy at validation phase by the trained models and comparatively lesser performance at testing phase also suggested the necessity of model training with various spectral characteristics of flood water. Overall, obtained competitive accuracy of flood area mapping with minimum training samples using original forms of the satellite data conveys the potential of improved flood area extraction with better training.

At the time of this manuscript preparation, there is no existing study which estimated the CNN's capability of flood area mapping with high resolution (> 1.5 m) optical satellite images. However, Amit and Aoki, (2017) and Gebrehiwot et al., (2019) carried out similar study with Aerial image data and obtained more than 90% accuracy whereas Sarker et al., (2019) reached more than 76% precision in flood area mapping in his study with low resolution Landsat satellite data. Nogueira et al., was attained maximum of 84% IOU accuracy for his inundation area mapping with high resolution (3m orthorectified) satellite data at the testing phase. However, none of the studies were discussed about processing speed of the networks which is a crucial measure for emergency response systems. Consequently, it is clear that this study provides a greater insight about CNNs networks and their capabilities to integrate with emergency response systems.

Considering the processing time along with the obtained accuracy of the models it can be concluded that both models having high potentials for inundation area mapping. Further improvements are necessary for accurate flood mapping before integrating to an emergency warning system. At the current stage of the study, ResNet101 slightly ahead in flood mapping than MobileNetV2. However, MobileNetV2 is having high potential of adopting to an emergency warning system due to its competitive accuracy along with less system requirements and lesser processing time than ResNet101 in flood mapping.

5. CONCLUSIONS

More than half of the world population concentrated into flood prone areas. Particular in Japan about 49% of the population and 75% of the real state are located in alluvial plains under inundation risk. Hence, faster and reliable flood mapping is vital for support emergency response planning. With the recent development in remote sensing imaging technologies and DL networks this become feasible.

Therefore, this study explored the capabilities of deeper and shallower CNN networks for precise inundation mapping from high resolution optical satellite data in a shorter time. This study used SPOT satellite images captured from 03 prefectures of Japan at flood events occurred in year 2015 and 2018 for training data preparation and tested with images obtained during Typhoon-19 in year 2019. The tested model architectures of this study are ResNet101 which the deepest network at present and MobileNetV2, one of the shallowest network which was developed to use even in mobile devices with minimum system requirement.

ResNet101 was able to extract flood area with 80.1% accuracy with processing speed of 0.7 s/km² while MobileNetV2 network with 79.6 % accuracy at 0.3 s/km² speed. Obtained results demonstrated that MobileNetV2 having a high potential of flood area mapping from optical satellite data with competitive accuracy. It is observed that both networks are obtained 90% accuracy in Jaccard index during the validation phase where it

less performed at the testing phase. This reflects the necessity of further training with different inundation conditions. Overall, it is found that in comparison to the existing studies of the field this study provides a better insight to scientific community towards the low cost automation solution for inundation mapping with high resolution satellite data.

ACKNOWLEDGEMENTS

Authors would like to acknowledge the financial support from Japan's Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Enhancement of National Resilience against Natural Disasters"

REFERENCES

- Amarnath, G., 2014: An algorithm for rapid flood inundation mapping from optical data using a reflectance differencing technique. *J. Flood Risk Manag.* 7, 239–250.
- Amit, S. N. K. B., Aoki, Y., 2017: Disaster detection from aerial imagery with convolutional neural network. *Knowledge Creation and Intelligent Computing (IES-KCIC)*, International Electronics Symposium on IEEE, 239–245.
- Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015: Land use classification in remote sensing images by convolutional neural networks. *arXiv*, arXiv:1508.00092.
- Chen, Y., Jiao, L., Li, Y., Zhao, J., 2017: Multilayer projective dictionary pair learning and sparse autoencoder for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.* 55 (12), 6683–6694.
- Feng, Q., Gong, J., Li, Y., 2015: Flood mapping based on multiple end member spectral mixer analysis and random forest classifier-the case study of Yuyao, China. *Remote Sens. Space sci.* 21, S37-S41.
- Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017: Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.*, 9, 498.
- Gebrehiwot A., Hashemi-Beni L., Thompson G., Kordjamshidi P., Langan T.E., 2019: Deep convolution neural network for flood extent mapping using unmanned aerial vehicles data. *Sensors.* 19, 1486.
- Goffi, A., Stroppiana D., Brivio P.A., Bordogna G., Boschetti M., 2020: Towards an automated approach to map flooded areas from Sentinel-2 data and soft integration of water spectral features. *Int J Appl Earth Obs Geoinformation* 84(2020)101951.
- Haq, M., Akhtar, M., Muhammad, S., Paras, S., Rahmatullah, J., 2012: Techniques of remote sensing and GIS for flood monitoring and damage assessment: A case study of Sindh province, Pakistan. *Egypt. J. Remote Sens. Space Sci.* 15, 135–141.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*.
- Iandola F. N., Moskewicz M. W., Ashraf K., Han S., Dally W. J., Keutzer K., 2016: Alexnet-level accuracy with 50x fewer parameters and 1mb model size. *arXiv:1602.07360*, 2016.
- Ireland, G., Volpi, M., Petropoulos, G.P., 2015: Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: A case study from a Mediterranean flood. *Remote Sens.*, 7, 3372–3399.
- Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson A.B., 2019: Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sens.* 152 (2019)166-177.
- Marmanis, D., Wegner, J.D., Galliani, S., Schindler, K., Datcu, M., Stilla, U., 2016: Semantic segmentation of aerial images with an ensemble of CNNs. *ISPRS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences*, Prague, Czech Republic.
- Nogueira, K., Fadel, S.G., Dourado, I.C., Werneck, R.D.O., Muñoz, J.A., Penatti, O.A., Calumby, R.T., Li, L.T., dos Santos, J.A. and Torres, R.D.S., 2018: Exploiting ConvNet diversity for flooding identification. *IEEE Geoscience and Remote Sensing Letters*, 15(9), pp.1446-1450.
- Ogashawara, I., Curtarelli, M., Ferreira, C., 2013: The use of optical remote sensing for mapping flooded areas. *J. Eng. Res. Appl.* 3, 1956–1960.
- Qiao, C., Luo, J., Sheng, Y., Shen, Z., Zhu, Z. Ming, D., 2012: An adaptive water extraction method from remote sensing image based on NDWI. *Journal of the Indian Society of Remote Sens.*, 40(3), pp.421-433.
- Rokni, K., Ahmad, A., Selamat, A. and Hazini, S., 2014: Water feature extraction and change detection using multitemporal Landsat imagery. *Remote sens.*, 6(5), pp.4173-4189.
- Sherrah, J., 2016: Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv:1606.02585*.
- Soltanian F.K., Abbasi M., Bakhtyari H.R.R., 2019: Flood monitoring using NDWI and mindwi spectral indices: a case study of Aghqala flood -2019, *ISPRS, XLII-4/W18*.
- Wang, Z., Liu, J., Li, J., Zhang, D.D., 2018: Multi-spectral water index: a native 10m multispectral water index for accurate water mapping on Sentinel-2. *Remote Sens. (Basel)* 10, 1643 2018.
- Wu, H., Adler, R.F., Tian, Y., Huffman, G.J., Li, H., Wang, J., 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* 50, 2693–2717.
- Yang L., Cervone, G., 2019: Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event. *Journal of soft computing*. <https://doi.org/10.1007/s00500-019-03878-8>.
- Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2018: Semantic Understanding of Scenes through ADE20K Dataset. *International Journal on Computer Vision (IJCV)*.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017: Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* 5 (4), 8–36.