Convolutional Neural Network for Road Network Detections Using Sentinel 2A

Bayu Yanuargi
Master of Informatics Engineering
AMIKOM Yogyakarta University
Yogyakarta, Indonesia

Ema Utami
Master of Informatics Engineering
AMIKOM Yogyakarta University
Yogyakarta, Indonesia

Kusnawi
Master of Informatics Engineering
AMIKOM Yogyakarta University
Yogyakarta, Indonesia

Abstract:- Road network data is critical information that used by government for development planning and by the services provider such as transportation and logistic company to deliver their services and prices calculations. Use of sentinel 2A satellite imagery data will solve the road map update cost issue since this data is free to use. The only problem on sentinel 2A is only about the medial spatial resolution that only able to detect 10 meters object. Using GPS data for the ground truth data will help to create the road masking data that can be used for the training. The result of the combination between these two data on Convolutional Neural Network are satisfied enough with accuracies score is 99% and soft dice error only 0.5%.

Keywords:- *Convolutional Neural Network (CNN), Sentinel, GPS, U-Net, Deep Learning.*

I. INTRODUCTION

Accurate and up to date road network data is very critical to ensure the development of road infrastructure, urban planning, navigation/routing planning and other purposes of use are exactly precise. In order the road network data can keep up to date, the updating should be conducted frequently, especially in the big city such as Jakarta.

There are two methods can be used for the road network update, first is traditional survey by visiting directly to the field and second using remote sensing approach. Traditional surveying methods are laborious and time consuming, that's why the considerable development of remote sensing technology in recent year has opened the way to automatic road detection application supporting a wide coverage area [5].

Remote sensing using deep learning technology and high-resolution remote sensing images provided opportunity to help the user for the road network updating. The applications of geo-information updating, urban transportation planning, and autonomous vehicle operation all heavily rely on the extraction of roads from high-resolution satellite photos, a fundamental study in earth observation and remote sensing. Furthermore, precise road maps can help with scene comprehension by supplying background information for recognizing structures, crops, and many other surface features. Road extraction remains a difficult task despite decades of attention due to complex ground information, including the shadows cast by buildings, the shade of trees, moving vehicles, and structures that resemble roads. Therefore, poor connectivity and erroneous recognition may be problems with extracted [7].

This research will use Sentinel 2A satellite imagery and GPS data to train the model using U-Net algorithm. The goal of this research is how the GPS data can be a good or proper training data for the road data extraction from medium resolution of Satellite Imagery. We expecting that the extraction from Sentinel 2A will be optimized, since Sentinel 2A is the free sources product that can be used by another researcher with a low budget.

Convolutional Neural Network (CNN) model has been performed well on the high resolutions satellite imagery with missing error only 5.12% with redundancy of error is only 0.35% and the accuracies is 93.61% [10]. This research will try to extract the road network data using CNN algorithm to know how accurate is the algorithm to extract the road network data from the medium resolution satellite imagery and how it can be used to gather the new road that haven't available in the current RBI versions.

II. LITERATURE REVIEW

Road network data may be extracted from high resolution remotely sensed data because of recent developments in machine vision, which also make it possible to automate the mapping process. By using feature mapping approaches, the hybrid configurations on a dataset combining ground-truth road masks and aerial information from typical regions of Spanish territory obtained maximum IoU and F1 scores of 0.5790 and 0.7120, respectively, and a minimum loss of 0.4985. When compared to the original design created from scratch, the best hybrid model (using SEResNeXt50 as the backbone network and U-Net as the segmentation architecture) saw performance metrics improve by around 3.5% [2].

In another study using data from Google Earth at a resolution of 60 cm/pixel and road network graphs from the Open Street Map (OSM) covering the core of 37 cities in six different countries. The road tracing model used in this study is similar to the Road Tracer for building accurate road maps from satellite imagery. The iterative search algorithm used is CNN to get a road network map. The study compared Road Tracer and Road Tracer-M with an increase of 10.4% in F1 scores and 8.7% in IoU,relative increase in F1 38.6% and IoU 51.1% [7].

CNN technology, one of the most effective deep learning techniques in recent years, excels at picture segmentation and image recognition. It is also frequently used to extract roads from satellite data and, in general, produces reliable findings. However, at the moment, the extraction of roads based on CNN still necessitates a lot of manual preparation work. To achieve extraction, a significant number of samples can be marked, necessitating a lengthy drawing cycle and a high production cost. The GPS trajectories of a floating car are used as a training set (GPSTasST) in a new method for creating CNN sample sets that extracts multilayer urban highways from high resolution remote sensing data. This method rasterizes the GPS trajectories of floating car into a raster map and uses the processed raster map to label the satellite image to obtain a road extraction sample set. CNN can extract roads from remote sensing imagery by learning the training set. The results show that the method achieves a harmonic mean of precision and recall higher than road extraction method from single data source while eliminating the manual labelling work, which shows the effectiveness of this work [4].

The use of the CNN algorithm for pothole detection results in a fairly high accuracy, which is 92.8%. The CNN algorithm is used to classify the images generated by the dash camera. The first stage is the classification of images using feed forward. The second stage is the learning stage with the back propagation method. Using the wrapping method and cropping to concentrate on the classification object Using feed forward and reverse propagation methods is the next step in training. The feed forward method with updated weights and biases is then used in the classification stage. The GPS will determine the position coordinates if a hole is found and communicate those coordinates to the database server [1].

III. METHOD

A. *Data Pre-processing*

The data used in this study is Sentinel 2 Satellite Imagery data for the DKI Jakarta area and its surroundings. Sentinel-2 is a satellite launched by a collaboration between The European Commission and the European Space Agency in the Global Monitoring for Environment and Security (GMES) program. This satellite was launched to monitor the condition of the earth's surface, so that it is able to provide information on the current condition of the earth from space for environmental and security applications.

Sentinel-2 images can be obtained through the United States Geological Survey (USGS) Earth Explorer web (http://earthexplorer.usgs.gov), Sentinels Scientific Data Hub (https://scihub.copernicus.eu/), or the Earth Observing System. (https://lv.eosda.com/). The important parameters for this data is cloud cover should be less than 10% to ensure all object can be detected properly.

Based on the figure 1 below, the processing of the satellite imagery post download are to use only the 10 meters band and combine three band of 4,3,2 to make the panchromatic image of red green blue. Since the original coverages of the data is 110 x 100 km, we cropped the image only on the research area with final size is 20 x 10 km.

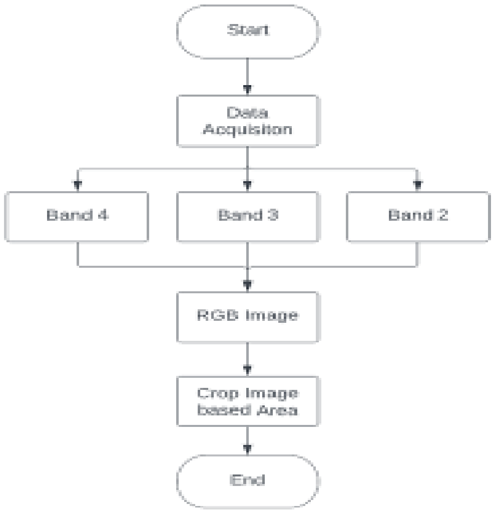

Fig 1. Satellite Imagery Pre-Processing

B. *Convolutional Neural Network (CNN)*

A deep learning algorithm called the convolutional neural network (CNN) is made to analyse two-dimensional data. CNN is frequently used to analyse and find features in images. In contrast to the neural network's typical input of 1-dimensional array data, the CNN technique uses 2-dimensional input data. Similar to generic neural networks, CNN is made up of a large number of neurons with weight, bias, and activation functions. Below Figure 2 is the steps of CNN.

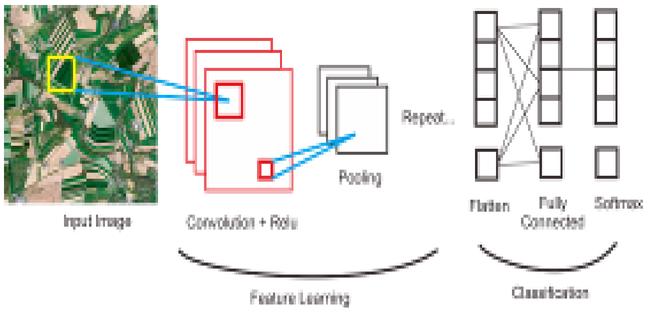

Fig 2. Convolutional Neural Network Steps

Convolutional, In CNN, Convolutional is a linear operation that involves multiplying a set of weights by the input. The convolution layer works by sliding the window on the input image. Each shift will be carried out a 'dot' operation between the input and the filter, resulting in a certain output value, as described in Figure 3 below.

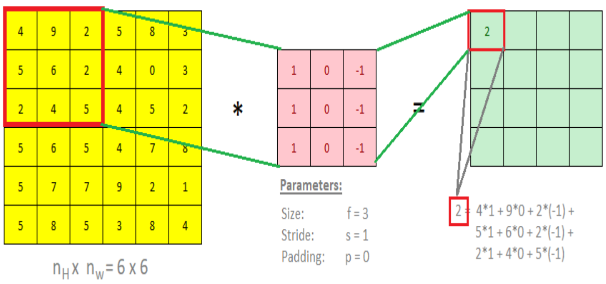

Fig 3. Convolutional Concept

Rectified Linear Unit (ReLu) is an operation to introduce nonlinearity and improve the representation of the model. The ReLU activation function is f(x) = max(0,x). The output value of the neuron can be expressed as 0 if the input is negative. If the input value is positive, then the output of the neuron is the value of the activation input.

Pooling or subsampling is a reduction in the size of the matrix. There are two types of pooling that are often used, namely average pooling and max pooling [8]. The purpose of using the pooling layer is to speed up the computational process. This can happen because after passing the pooling layer, fewer parameters need to be updated, so the risk of overfitting is minimal. Just like in the convolution layer, the pooling layer also has a filter with a certain size that will perform a sliding window process on the input matrix as on below figure 4.

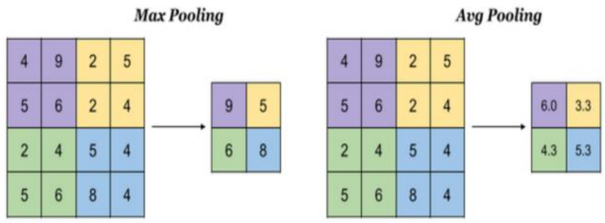

Fig. 4. Max pooling and average pooling concept

The flattening stage is changing from the matrix in the pooling layer to one column only (a single vector). Flatten serves to reshape a feature map from a multidimensional array into a vector [9]. This is necessary so that these values can be used as inputs to the fully connected layer. Dropout Regularization is a neural network regularization technique where some neurons will be randomly selected and not used in the training process. By removing a neuron, it means removing it temporarily from the existing network. The purpose of using this technique is to prevent overfitting. In addition, this technique can also speed up the learning process on the CNN architecture as a whole.

The fully connected layer is a feed forward neural network consisting of a hidden layer, activation function, output layer, and loss function. Fully connected layer is often used in multi-layer perceptron's with the aim of transforming the data dimensions so that the data can be classified linearly. The input to the fully connected layer is data from feature learning. Data from the feature learning process which is the input to this layer is data in vector form, which has previously been processed by flatten.

Softmax rule in the classification process is to calculate the probability of each target class against all existing target classes. The output probability range on softmax is a value from 0 to 1, and if all the probability values of the target class are added up, then the value will be equal to one. Softmax uses the exponential of a given input value and the sum of the exponential values of all values in the output.

C. Road Data

This study used GPS data captured by the Ride-Hailing Driver (GRAB) app in April 2019. These included 84,000 paths with a total of 88 million PDs (point data). Data were taken every second he depending on the driver's performance [3]. Figure 5 below shows an example of GPS data in Jakarta that used for this research.

Fig 5. GPS data in the research area

This study used GPS data captured by the Ride-Hailing Driver (GRAB) app in April 2019. These included 84,000 paths with a total of 88 million PDs (point data). Data were taken every second he depending on the driver's performance. Figure 8 below shows an example of GPS data in Jakarta that used for this research.

D. Data Training

To maximize training speed and power, we decomposed the Sentinel-2 mosaic and GPS grid into small chunks of data by creating a 256 × 256 cell grid within the training area. Using GPS data intersected with raster data, we created a reference road data set for training a U-Net Road recognition model. The Sentinel-2 image was then cropped to the chosen grid (256 × 256 grid) resulting in 2240 samples. Finally, we stacked the Sentinel-2 RGB bands with the updated raster GPS and split the 2240 chip samples into training, validation, and test datasets. Randomly split 80% of the chip samples for training and 20% evenly for validation and testing. 224 each. The data training and model building can be seen on figure 6 and table 1

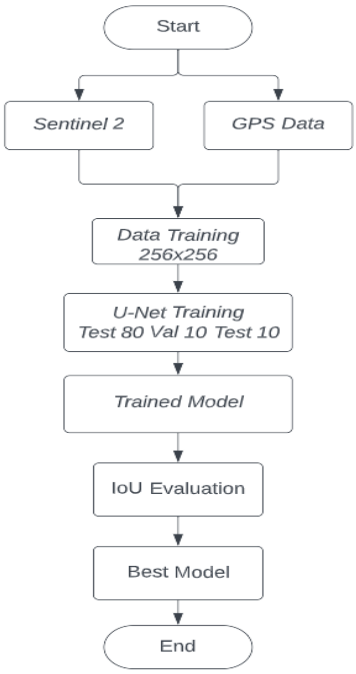

Fig 6. Data Training Flow

TABLE 1. TRAINING DATA.

Satellite Imagery	*GPS Label*
	
	
	

**Sample of data training*

E. U-Net For Road Detection

To ensure the model can detecting road object properly, the U-Net architecture layers and hyperparameter modified. First the input layer is using 256x256 data same with the output data with a zero-padding border around the image. Second is using Rectified Linear Unit Activation function (ReLU) to solve the linear problem during the training by give zero value if the input is less than zero. For the dropout layer we add value of 0.3 to avoid overfitting on our training data. In the end the dice loss function will replace the original loss function to minimize impact of imbalanced classes for the segmentation result. The process above can be seen on below figure 7.

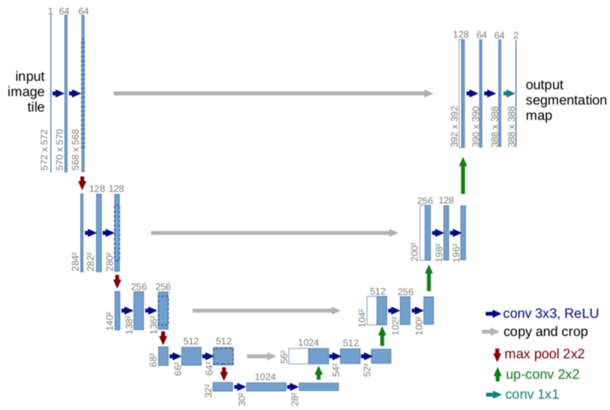

Fig 7. U-Nets Architecture

The U-Net architecture is built on top of Fully Convolutional Network (FCN) and modified in such a way as to produce better segmentation. Compared to FCN, the two main differences are (1) the U-NET is symmetrical and (2) the jump connection between the down sampling path and the up-sampling path applies the join operator instead of sum. This skip connection aims to provide local information to global information during up sampling. Because it is symmetrical, the network has a large number of feature maps in the up-sampling path, which makes it possible to transfer information. In comparison, the basic FCN architecture has only a limited number of class feature maps in its up-sampling path [6].

IV. RESULT

Data training and validation using Google Colabs targeting to train dataset, getting a model and validating the training result. On the training process, we were used 50 epochs with the result as below table.

ISSN No:-2456-2165

TABLE 2. TRAINING RESULT

Epoch	Accuracy	Soft Dice Lost	Loss	Val Accuracy	Val dice loss	Val Loss
1	0.87805	0.27090	0.37232	0.88172	0.27360	0.36169
5	0.94258	0.10747	0.16497	0.89634	0.14253	0.29585
10	0.96868	0.05483	0.08334	0.89727	0.12303	0.32373
15	0.97880	0.03536	0.05481	0.89725	0.11394	0.36209
20	0.98411	0.02594	0.04084	0.89584	0.11274	0.41052
25	0.98787	0.01973	0.03151	0.89506	0.11106	0.45484
30	0.99150	0.01401	0.02272	0.89623	0.10846	0.49072
35	0.99404	0.01006	0.01631	0.89947	0.10375	0.57911
40	0.99538	0.00785	0.01294	0.89784	0.10495	0.68984
45	0.99629	0.00633	0.01050	0.89841	0.10374	0.82875
50	0.99719	0.00492	0.00815	0.89791	0.10411	0.72644

Based on training data with 50 epochs, can be seen that the training and validation shown very good accuracies nearly to 1. The training reaches the maximum numbers of 0.99 in the epoch 50, and the validation reach the highest score on epoch of 27 with 0.90 score. Based on Figure 7 and 8 below, it shows that the graph of the U-Net model does not show overfitting or underfitting conditions, but rather towards a convergent graph. Underfitting is an event where the model that has been formed is unable to see the logic behind the dataset, so the model cannot make predictions with accurate results on training data or data validation. Meanwhile, Overfitting occurs when the model is too focused on a particular training dataset, so it cannot make predictions correctly when given another, similar dataset [6]. In this study the number of epochs = 50, but due to limited equipment, researchers implemented an early stop monitor. Early stop monitor is a condition where the training phase is stopped before the entire epoch is completed. The condition applied is, if every 10 epochs there is a decrease in accuracy then the training phase will be stopped. The accuracy, dice loss and comparison between both can be seen on figure 8 – 10 below.

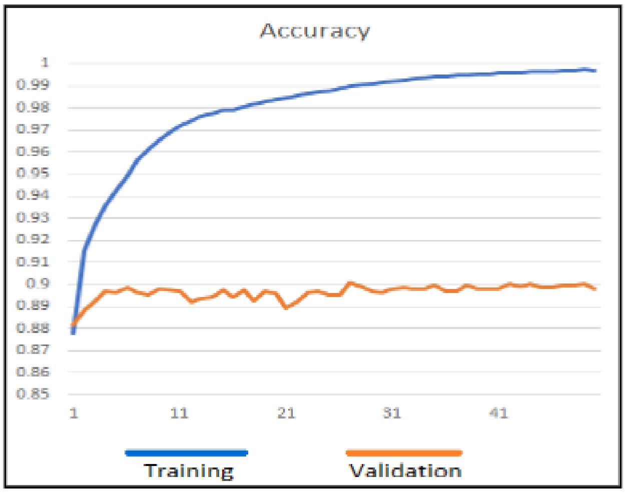

Fig 8. Training Accuracy Curve

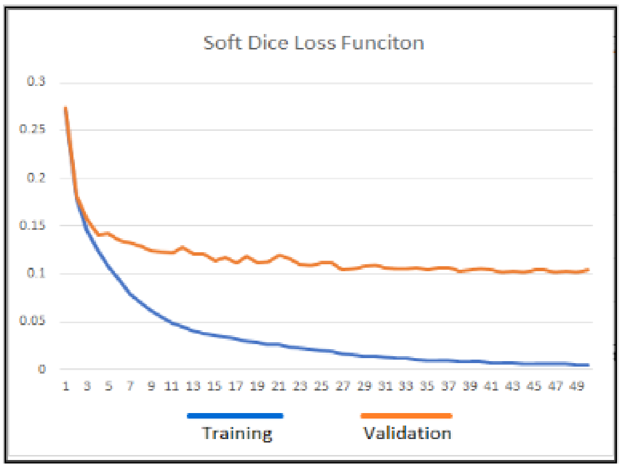

Fig 9. Soft Dice Loss Functions

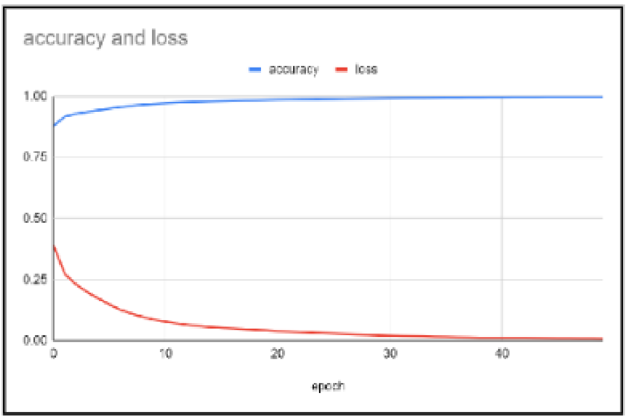

Fig 10. Accuracy vs Loss

Trained model applied to the datasets of Sentinel 2A. The prediction took 55 second to generate 2240 images. The model build based on the training on epoch 49 which have lesser Soft Dice Error on the data training 0.005 with accuracy of 0.997. The prediction using the best model produce quite good result that shown on below picture where the predicted mask able to detect the similar pattern as ground truth mask. The missing road on predicted mask mostly are smaller road class that not expected to be detected from sentinel 2A because of the limitation of spatial resolution of this data that only able to record 10 meters wide per pixel as we can see on below figure 11.

Input Image	Ground Truth Mask	Predicted Mask

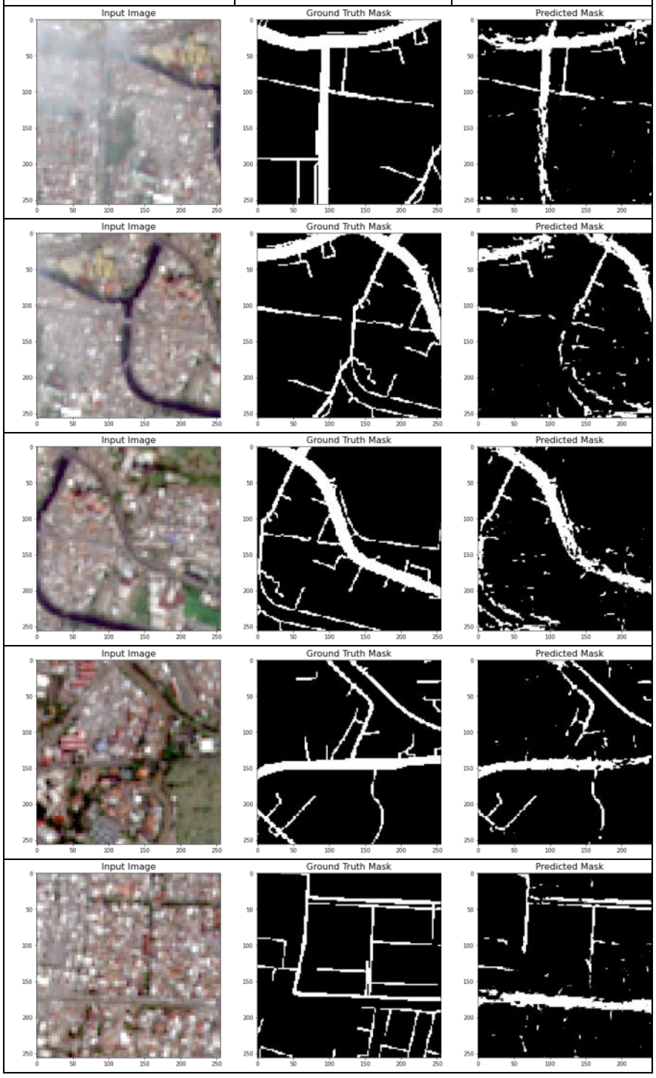

Fig 11. Comparison Between Sentinel 2A, Ground Truth Mask and Predicted Mask

V. CONCLUSION

Based on the training and validation, the model built with very high accuracies of 0.997 nearly to 1 and less soft dice loss is only 0.005 nearly zero. With this model accuracies and loss, the expectation to extract road data from medium resolution satellite imagery can be gained. Beside that the GPS data also provide quite good ground truth mask for the data training with low loss value and high quality of data training.

The combinations of GPS data and Sentinel 2A based on the prediction result is enough to get good quality road data with low cost. Comparing with high resolution imagery, Sentinel 2A and GPS data is good enough to be used for the 1:25.000 Road network map updating. For next development related with the combination between GPS data and medium resolution is how to optimizing the GPS not only analyzed using density, but optimized it by detecting the middle of road based on density and use that as the source of ground truth mask.

REFERENCES

[1]. A. D. Wiratmoko et al., "Design of Potholes Detection as Road’s Feasibility Data Information Using Convolutional Neural Network(CNN)," 2019 International Symposium on Electronics and Smart Devices (ISESD), 2019, pp. 1-5, doi: 10.1109/ISESD.2019.8909461.

[2]. C.-I. Cira, R. Alcarria, M.-Á. Manso-Callejo, and F. Serradilla, “A Deep Learning-Based Solution for Large-Scale Extraction of the Secondary Road Network from High-Resolution Aerial Orthoimagery,” Applied Sciences, vol. 10, no. 20, p. 7272, Oct. 2020, doi: 10.3390/app10207272.

[3]. Huang, X., Yin, Y., Lim, S., Wang, G., Hu, B., Varadarajan, J., Zheng, S., Bulusu, A., & Zimmermann, R, “Grab-Posisi: An Extensive Real-Life GPS Trajectory Dataset in Southeast Asia”, 2019. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility.

[4]. J. Zhang, Q. Hu, J. Li and M. Ai, "Learning From GPS Trajectories of Floating Car for CNN-Based Urban Road Extraction With High-Resolution Satellite Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 1836-1847, March 2021, doi: 10.1109/TGRS.2020.3003425.

[5]. Qianqian Zhang, Qingling Kong, Chao Zhang, Shucheng You, Hai Wei, Ruizhi Sun & Li Li ,“A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network”,2019, European Journal of Remote Sensing, 52:1, 572-582, DOI: 10.1080/22797254.2019.1694447

[6]. Suta, Ida & Sudarma, Made & Kumara, I Nyoman Satya. “Segmentasi Tumor Otak Berdasarkan Citra Magnetic Resonance Imaging Dengan Menggunakan Metode U-NET”, 2020. Majalah Ilmiah Teknologi Elektro. 19. 151. 10.24843/MITE.2020.v19i02.P05.

[7]. Y. Wei, K. Zhang and S. Ji, "Road Network Extraction from Satellite Images Using CNN Based Segmentation and Tracing," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 3923-3926, doi: 10.1109/IGARSS.2019.8898565.

[8]. Bejiga, M.B., Zeggada, A., Nouffidj, A., & Melgani, F. (2017). A Convolutional Neural Network Approach for Assisting Avalanche Search and Rescue Operations with UAV Imagery. Remote. Sens., 9, 100.

[9]. A. Chakravarty and J. Sivaswamy, "RACE-Net: A Recurrent Neural Network for Biomedical Image Segmentation," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 3, pp. 1151-1162, May 2019, doi: 10.1109/JBHI.2018.2852635.

[10]. Chen, Yuantao & Wang, Jin & Chen, Xi & Kumar, Arun & Yang, Kai & Cao, Zhouhong. (2019). Image Super-Resolution Algorithm Based on Dual-Channel Convolutional Neural Networks. Applied Sciences. 9. 2316. 10.3390/app9112316.