Intelligent System for Building Separation on a Semantically Segmented Map

Volodymyr Hnatushenko[a], Vadym Zhernovyi[b] , Iryna Udovyk[a] and Olga Shevtsova[a]

[a] *Dnipro University of Technology, Dmytra Yavornytskoho av., 19, Dnipro, 49005, Ukraine*
[b] *Oles Honchar Dnipro National University, Gagarina av.,72, Dnipro, 49010, Ukraine*

Abstract

Terabytes of very high-resolution satellite imagery data are sent to land stations every day and only 5% of this information is used which raises a demand in automation of image processing routines. Semantic maps become especially popular for a wide range of analysis challenges like surveillance, vegetation monitoring, change detection, etc. Nowadays, deep learning approach to image processing suggests a very flexible and configurable tool for different needs – semantic segmentation included. With the use of deep learning, it is possible to extract unique features from data and adapt model and algorithms for specific data to achieve the best results possible. In current work the algorithm to instance-like segmentation is suggested. This algorithm is applied to a modified semantic segmentation neural network in order to work with separate instances of different land objects. There are other networks which already perform instance segmentation like Mask RCNN. However, often semantic segmentation networks provide better detection results regarding accuracy and a possibility to work with detected objects is crucial. Separated instances can be used in various calculations and measurements such as a size of these objects, distances, etc. In addition to the semantic segmentation neural network, an approach is suggested to approximate measurements of such essential physical parameters of land objects as perimeters, square areas and building density using knowledge of spatial resolution characteristics of the ultra-high-resolution remote sensing imagery used in current work as a source of data for the training dataset. The results of suggested methods can be applied to countless areas such as urban planning, built-up analysis, traffic control, etc. The solution is flexible and can be additionally adjusted for different needs which is discussed in our future research.

Keywords

Remote sensing, image, deep learning, semantic segmentation, masks, measurements.

1. Introduction

The satellite imagery is based on the complex process of converting solar energy reflected from the earth's surface and electromagnetic pulses, which are recorded digitally. Until a decade ago, access to satellite data was limited, and only military, large corporations, government agencies, and some scientific institutions could obtain such information. Now terabytes of satellite data are available to everyone. Every day we can see how our planet looks like with the help of satellite imagery. Remotely sensed images permit accurate mapping of land cover and can assist the planning and coordination of global change.

IntelITSIS'2021: 2nd International Workshop on Intelligent Information Technologies and Systems of Information Security, March 24–26, 2021, Khmelnytskyi, Ukraine

EMAIL: vvgnat@ukr.net (V. Hnatushenko); vadim.zhernovoy@gmail.com (V. Zhernovyi); udovik.im@gmail.com (I. Udovyk); shevtsova.o.s@nmu.one (O. Shevtsova)
ORCID: 0000-0003-3140-3788 (V. Hnatushenko); 0000-0002-0599-7992 (V. Zhernovyi); 0000-0002-5190-841X (I. Udovyk); 0000-0002-6421-8127 (O. Shevtsova)

Recently, semantic segmentation of land objects becomes extremely popular in remote sensing applications and systems. Such segmented maps have a lot of applications in different areas such as urban planning [1], agricultural applications [2], traffic estimation and monitoring as on land as well as in water [3], etc. Two categories of approaches are usually considered when solving semantic segmentation problems – definitive feature-based algorithms such as described in [2, 4] or stochastic deep learning approaches which are heavily relied on deep convolutional neural networks [5-7].

Considering the fact that feature-based hand-crafted algorithms and other machine learning approaches without neural networks may be successfully applied to solve certain satellite imagery processing problems, deep learning still remains more promising in the long run [8-10]. The advances in deep learning neural networks of direct propagation are the alternation of convolutional and max-pooling layers [11], topped with several fully connected or sparsely connected layers, according to followed by the final layer of classification. Training is usually done without any spontaneous pre-training. GPU-based approaches have won many image recognition competitions, including the IJCNN 2011 Traffic Sign Recognition Competition, [12] the Neural Structure Segmentation Competition in the electron microscopy stack. (Segmentation of neuronal structures in EM stacks challenge) ISBI 2012 [13], ImageNet [14] and others. Such guided methods of deep learning also became the first artificial image recognizers to achieve in some tasks efficiency comparable to human [15].

The deep learning applications in remotely sensed images are different from those in natural images. The remotely sensed images usually have more complicated and diverse patterns. Thanks to the strong ability of deep learning in feature representation, deep learning has been introduced into environmental remote sensing and applied in many aspects, including land cover mapping, environmental parameter retrieval, data fusion and downscaling, and information construction and prediction. More detailed applications of deep learning in environmental remote sensing are as follows [16].

Most deep learning solutions make use of neural network structures based on convolutional neural networks. Certain neural networks type suit better for different challenges – remote sensing is not an exclusion. One of the problems that exists even today for remote sensing imagery processing is that it is often required to be process in patches or tiles since some neural networks accept 3-channel images of certain resolution when most of remote sensing imagery may reach a resolution of more than ten thousand pixels per a dimension. Fully Convolutional Network (FCN) was invented to address the mentioned problem [17-19]. Using this neural network type, it is possible to generated segmented map of any size. Another popular approach for semantic segmentation is encoder-decoder architecture which results in generating a semantic map of the same resolution and dimensions as the original image. Later, more complex solutions were developed to advance generation of semantic maps namely SegNet [20], DCNN+CRF [21], SS-CNN [22] and others. Good review of neural network applications for remote sensing data is provided in [23]. One of the main reasons to choose one neural network architecture before others is to make use of both spectral and spatial information which is often provided for the most used satellite imagery. In most cases the results of deep learning solutions for remote sensing are applied successfully only to certain imagery type which it was implemented and tested with. However, there are exclusions where a proper combination of neural network architectures and parameters solved the problem of semantic segmentation for similar imagery from multiple different satellite vehicles. These problems and solutions are described in detail in [24]

In current research paper the modified Unet-like architecture is suggested for processing a very high-resolution hyperspectral WorldView-3 Imagery data. WorldView-3 Imagery is used to design a dataset for the neural network training. Additional layers are designed on a top of the neural network architecture to separate instances of detected land objects and perform land object density measurements algorithm.

2. Pre-processing

In current work WorldView-3 imagery is used as a source for a training and testing datasets. WorldView-3 is the high resolution satellite sensor operating at an altitude of 617 km. WorldView-3 satellite provides 31 cm panchromatic resolution, 1.24 m multispectral resolution, 3.7 m short wave infrared resolution (SWIR) and 30 m CAVIS resolution. The satellite has an average revisit time of <1 day and is capable of collecting up to 680,000 km2 per day [25].

In order to achieve the best results possible applying deep learning to satellite imagery, it is essential to consider the main characteristics of sensory systems that determine the suitability of the data to solve a problem, there are four types of distinction:

- spectral;
- spatial;
- radiometric;
- temporal.

Spectral resolution is the ability of a sensor system to register electromagnetic radiation of a specific frequency range, which is determined by the number of satellite channels, for example, the intervals of wavelengths of the electromagnetic spectrum to which the sensor is sensitive. WorldView-3 provides a wide range of options regarding spectral resolution including panchromatic images, multispectral visible and non-visible bands. Non-visible bands are short-wave infrared bands and are not used in current research.

Spatial resolution is the size of the smallest object on the earth's surface that differs in the image, that is, it is actually the physical size of a pixel. Currently, the best commercially available imagery has a spatial resolution of 30 cm – WorldView-3 satellite (excerpt from the Tvis website). This means that a 30 × 30 cm object will appear in the image as a single pixel. So, the objects, such as cars, will be noticeable in the picture and their color can be determined (if the picture is color), but smaller details (registration number, design features that help determine the make and model) will not be read in the picture [26]. These characteristics are the main ones which were taken to consideration when designing a dataset for the deep neural network training. Details on the dataset design are described in [27]. In order to improve a quality of a dataset further, additional image enhancement techniques can be applied, in current research [28] are used. There was an additional attempt of shadow detection algorithm [29] application to the dataset but it did not show any significant improvement for post-processing algorithm performance. Resulting spectral and spatial characteristics as well as amount of information in pixels are provided in a Table 1.

Table 1
WorldView-3 Imagery Spatial Characteristics

Type	Wavebands	Pixel resolution	Num. channels	Size
Grayscale	Panchromatic	0.31 m	1	16924 x 17020
8-band	Multispectral pansharpened	0.31 m	8	16924 x 17020
16-band	Multispectral	1.24 m	8	4255 x 4231
	Shor-wave infrared	7.5 m	8	670 x 688

Radiometric resolution is the number of possible encoded spectral luminance values in the data file for each spectral band indicated by the number of bits. It is determined by the number of gradations of color values, the corresponding transitions from the brightness of absolutely "black" to absolutely "white" and is expressed in the number of bits per pixel. For WorldView-3 this value is 16-bit which means that spectral luminance values for this imagery varies from 0 to 65535. For the use with the neural network these values are normalized between 0 and 1, 16-bit values (FP16). Among the best practices for training a Neural Network is to normalize your data to obtain a mean close to 0. Normalizing the data generally speeds up learning and leads to faster convergence [30].

Temporal characteristics are not considered since the change detection for the same territories in different time is out of scope in current research.

3. Neural network

For years, Unet architecture remain popular choice in many areas of research where semantic segmentation is required. Originally Unet was designed for biomedical image segmentation [31].

However, today Unet is applied successfully in other areas of knowledge including remote sensing [10, 32-34].

The main focus of this research is post-processing of Unet segmentation results so additional tuning was done to the neural network used in current work in order to aid separation of the whole mask to separated instances and measurements.

Our neural network consists of:

- Input layer of size 512 x 512
- 5 encoder blocks (Fig. 1)
- 1 extra convolution block (Fig. 1)
- 5 decoder blocks (Fig. 1)
- Output sigmoid activation layer.

Unlike most Unet-like architecture applications, dropout is not used in current work – in conducted experiments different dropout layers combination didn't show any improvements for post-processing algorithm.

Training is run on 4255 training and 759 validation samples. Random hue, horizontal flipping and height-width shifting are applied as augmentation for the training dataset. Hue delta value of 0.1 is chosen. Such augmentation is not applied to a validation dataset.

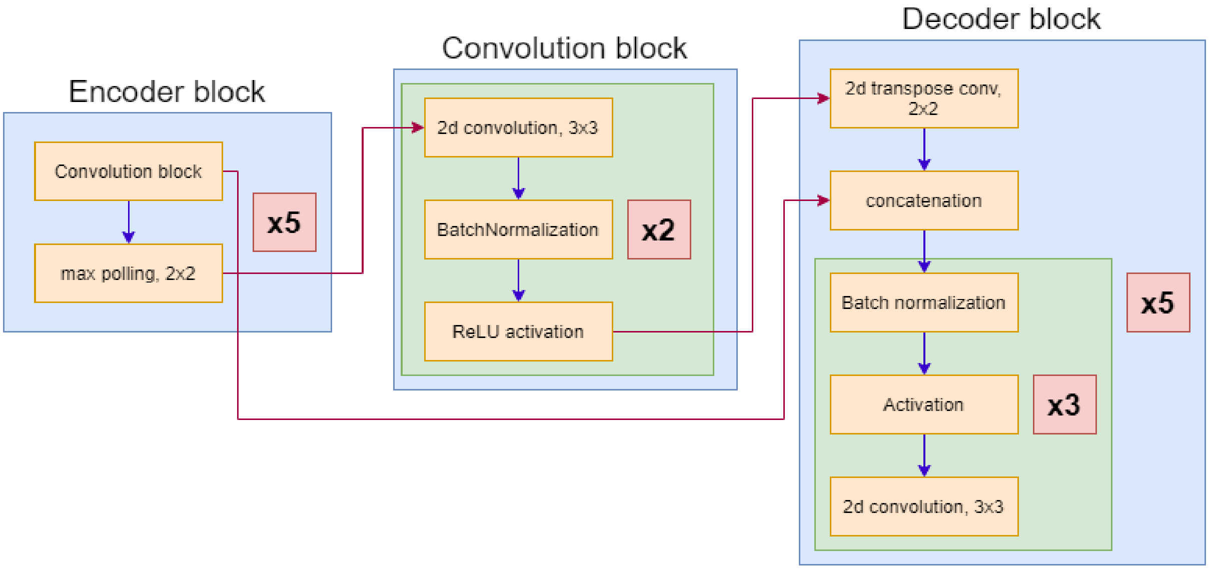

Figure 1: Unet backbone blocks

There are metrics specifically developed to adequately measure deep learning solutions performance. Custom metrics and loss functions were developed in current work for better representation of both - the neural network performance and the post-processing algorithm performance. Since the goal of the article to achieve good instance segmentation, it was decided to apply a modifier dice coefficient (F1 score)

$$Dice = \frac{1}{N}\sum_{i=0}^{n}\frac{2 * p(y_i) * y_i + 1}{p(y_i) + y_i + 1}, \tag{1}$$

where $p(y_i)$ is a predicted mask and y_i is a ground truth annotated mask available from the training data.

It was successfully used in [35] to represent instance segmentation results. Original dice coefficient is modified by adding 1 for intersection and union parts of the equation to prevent division by zero. Additionally, a dice loss function is used for training

$$loss_{dice} = 1 - Dice, \tag{2}$$

where *Dice* is defined in (1). Dice loss function is a metric for measuring overlap for ground truth and segmented masks. Dice loss metric is very flexible and could be additionally optimized which may improve results more [36], but in current work such optimization is not investigated.

Unfortunately, for the neural network architecture using $loss_{dice}$ led to overfitting of the model despite a popular neural network architecture and a relatively big dataset. In order to overcome this problem an improvement was implemented for the loss function which helped in another problem using similar neural network model [37]. The solution is to define a more complex loss

$$loss = \mathrm{loss_{dice}} + loss_{bce}, \tag{3}$$

where $loss_{bce}$ is a binary cross entropy loss or log loss function which is defined as

$$loss_{bce} = -\frac{1}{N}\sum_{i=0}^{n} \mathrm{y_i} * \log(\mathrm{p(y_i)}) + (1 - \mathrm{y_i}) * \log(1 - \mathrm{p(y_i)}), \tag{4}$$

where $p(y_i)$ is a predicted mask and y_i is a ground truth annotated mask available from training data.

Adam function was chosen as an optimizer. Additional standard metrics such as accuracy, precision and recall were also calculated for secondary analysis of the results.

Mixed precision technique was applied in order to improve training speed of relatively big model (10 million parameters) in the limited training environment. Mixed precision technique is described in detail in [38].

Final results of training are mentioned in Table 2.

Table 2
Unet training results

Metric	Value
Accuracy	0.9834
Precision	0.9217
Recall	0.7641
F1 Score	0.8355

Such results are in line with ones in similar works [10] which have a better recall rate and around the same precision.

Preliminary research showed that more training and dataset cleanup are the main contributors in a better recall.

4. Post-processing

The main goal of current article is post-processing which is implemented as top layers for the Unet-like backbone. This post-processing is capable of multiple sequential steps:

- Instance separation
- Semantic labeling
- Measurements of land objects
- Building density

4.1. Instance separation

Output of Unet is pixel-wise grayscale values in a range of 0..1 which represent a degree of how much a pixel belong to certain class. In order to distinguish detected masks properly from the background, a threshold must be implemented so every pixel can be separated by this value.

Most research use a middle value of 0.5. However, for the suggested approach another value performs visual better. This value is obtained by application of Otsu's method [39] to the segmentation results which is an automatic image thresholding technique used for classification of all pixels into two classes – foreground and background which results in a binarized image (Fig. 2). According to the Otsu method, the optimal threshold for binarization reaches the minimization of the weighted sum of variances within each cluster or, on the other hand, the maximum sum of the interclass variances.

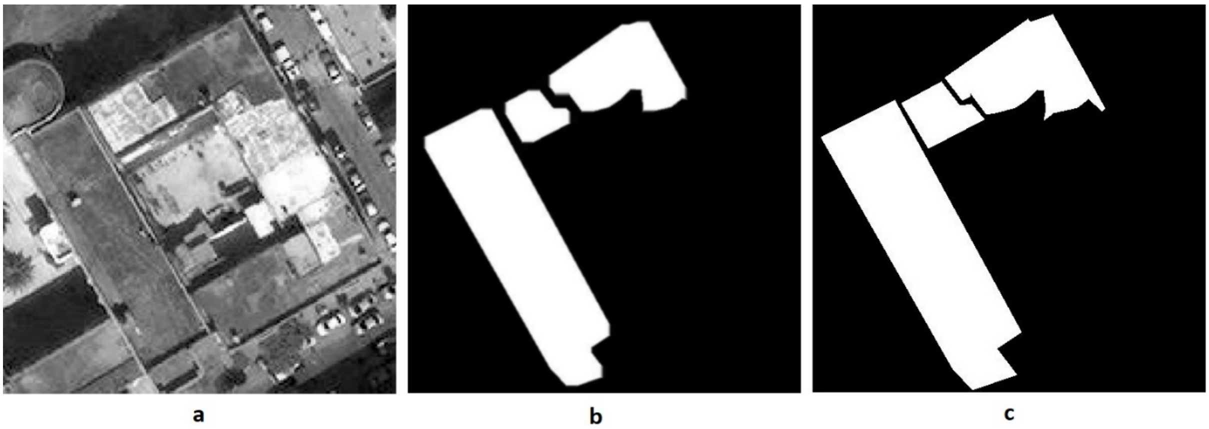

Figure 2: Results of applying Otsu's thresholding algorithm, original image (a), segmented image (b), post-processed image (c)

Another algorithm is implemented to separate instances from the whole mask. This technique involves feature-based analysis which distinguish arrays of pixels using a centrosymmetric filter structure. Such approach helps to keep together pixels that belong to one territory but consists of multiple objects of search (Fig. 3).

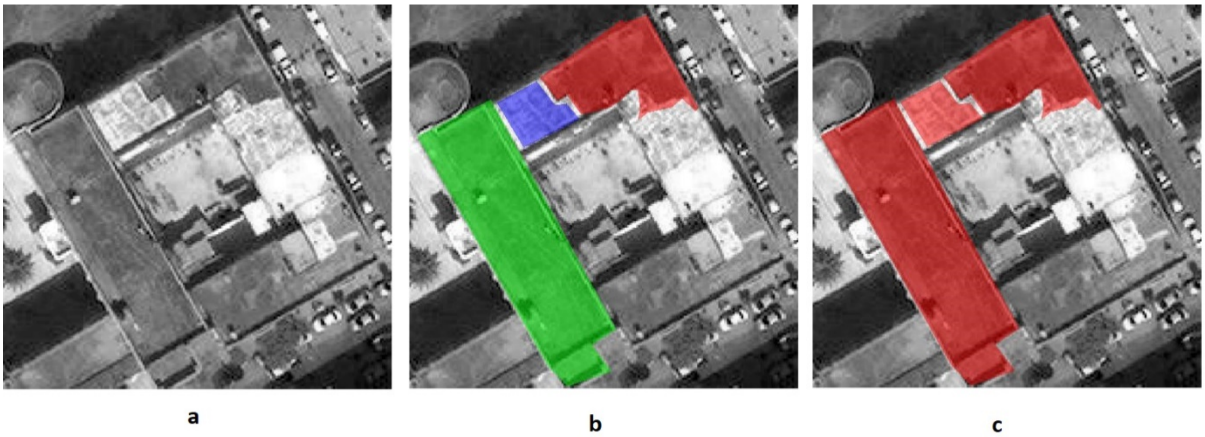

Figure 3: Post-processing algorithm: original image (a) is segmented by 3 split areas (b) but classified as a single instance (c)

Additional semantic search is performed after separation to obtain coordinates and dimensions of all objects found. These coordinates and dimensions are used for measurements and density calculations.

4.2. Land objects measurements

This part of post-processing is considered as the simplest in terms of computation complexity. It is based on knowledge of structure of the satellite imagery. It is known that a length of one pixel is 31 cm which gives an opportunity to calculate physical parameters of detected objects.

First, the minimal fitting rectangle is calculated for one instance. This rectangle is built with the largest vertical and horizontal diameters (Fig. 4).

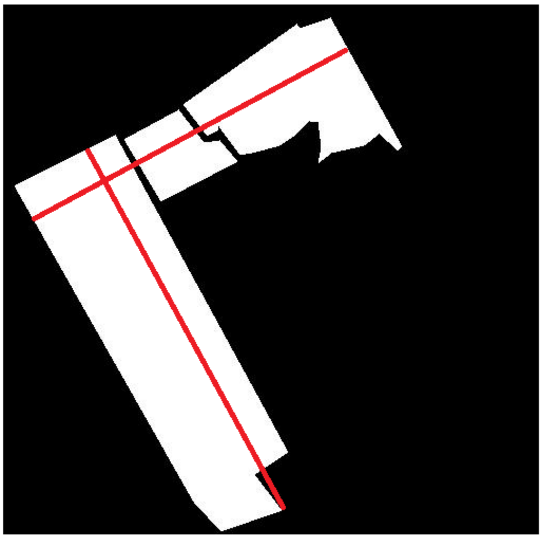

Figure 4: Maximum vertical and horizontal diameters based on minimal fitting a rectangle

Using calculated diameter land objects physical parameters are assumed – perimeter (5), square area (6).

$$Perimeter = 2 * (vmax + hmax), \tag{5}$$

where *hmax* is a maximum horizontal diameter assumed from the minimal fitting rectangle, *vmax* – vertical.

$$SqArea = vmax * hmax, \tag{6}$$

where *hmax* and *vmax* are maximum horizontal and vertical diameters correspondingly.

Building density is calculated using (7) as a percentage of all pixels identified as land objects to a total number of image pixels. Though these calculations only assume the physical sizes of land object because AI system cannot be 100% accurate, the formula also considers a correction by using precision and recall values as a factor.

$$BD = \frac{precision}{recall} * \frac{\sum_{i=0}^{m} p(y_i)}{\sum_{j=0}^{n} x_j}, \tag{7}$$

where, *m* is a number of segmented pixels, *n* is a total number of image pixels, $p(y_i)$ is a predicted pixel, x_j is an image pixel. *Precision* and *recall* are corresponding neural network metrics obtained during validation stage of the training.

5. Experiment

Experiments were conducted in multiple ways – for a neural network part, the post-processing and for the whole system. The necessity of such conditions is justified by modularity and comparability of each part to similar approaches for pre-processing, neural network processing and postprocessing.

For the neural network accuracy, precision, recall and F1-score are calculated and compared to similar Unet-like neural network solutions for remote sensing semantic segmentation. Unet-like architectures are chosen for comparison since we are not suggesting a brand-new neural network architecture to compare it with wider range of neural network architectures – the main focus of the research is the post-processing part. The point of such comparison is to demonstrate that the suggested Unet-backbone is not worse than existing similar models but still optimized for the needs of the post-processing module. The suggested architecture (SA) is compared to the original Unet, HSFA-Unet [9], Refined Unet [10], Stacked Unets [40, 41].

All mentioned neural networks are ran on a custom dataset used for development and testing in the current work [27]. The result of neural networks testing is provided in Table 3.

Table 3
Building calculation results

Neural network	Accuracy	Precision	Recall	F1-score
SA	**0.9834**	0.9217	0.7641	0.8355
Unet	0.8911	**0.9316**	**0.7923**	**0.8563**
HSFA-Unet	0.9831	0.8832	0.7373	0.8036
Refined Unet	0.7712	0.6909	0.7601	0.7238
Stacked Unets	0.8989	0.8877	0.7878	0.8347

Traditional Unet is slightly better in terms of general performance but the suggested Unet-like architecture is superior considering application of post-processing routine because of much higher accuracy than for the original Unet.

When an optimal neural network architecture was determined, another experiment was conducted to test the main part of the research – the port-processing method for measurements of land objects. Thus, calculations for the building on figures 2-4 mentioned in Table 4.

Table 4
Building calculation results

Metric	Value
Maximum horizontal diameter, m	105.4
Maximum vertical diameter, m	121.8
Perimeter, m	454.4
Area, m2	12837
Density, %	23

Further experiments showed that the density calculation for the whole scene instead of separate tiles decreases approximately by 10%.

6. Results and discussion

In current research, the approach for end-to-end AI pipeline is suggested including pre-processing, neural network modeling and post-processing.

Pre-processing stage is heavily relied on the results of previous work [27]. These results suggest a complete approach for dataset development for solving remote sensing problems and used in current research with minimal changes which include additional augmentation and image enhancement techniques in order to improve performance of the neural network processing and post-processing.

The second part of the solution keeps the novelty of Unet for Remote sensing by suggesting another approach to configure this architecture for solving many different problems and challenge including instance segmentation and land object measurements which it is not originally purposed for. Custom metrics and loss functions are developed which complement the post-processing and are highly suggested for using when solving land measurements tasks such as urban planning, etc. This section suggests all the required information to successfully conduct described experiment.

The post-processing is the main part of the research. All the previous work regarding the neural network configuration and custom metrics development are done to compliment post-processing. The post-processing workflow solves a very applied task fully automated – no interaction with other systems or operator is needed.

All suggested mechanisms are flexible and interchangeable. These results can be used to conduct experiments in other areas of research (i.e. Healthcare) and with other neural network architecture.

The developed approach can be improved further in multiple way:

- Increasing a number of data and its clean up
- Optimizing dice loss function
- Fine-tuning of neural network or replacing with another one

- Investigating and adding factors to density calculation mechanism such as counting vegetation and other objects

All results are currently applied to the only class of objects – building. The other classes of objects are planned to add to dataset in future. Since the source if data remains the same – WorldView-3 imagery, the developed approach will demonstrate the same performance for new classes of objects which may be trees, vehicles, etc.

7. Conclusions

The suggested end-to-end approach has been proven to provide promising results in processing multiple types of very high resolution satellite imagery data. Current paper demonstrates good quality of processing for WorldView-3 imagery. The obtained results lead to conclusion that the methods which are suggested in this research paper are suitable for non-RGB images of very high resolution such as satellite imagery data. Even though the approach and methods are applied to WorldView-3 imagery data in current work, it is not limited and can be used in similar satellite imagery for another satellite vehicle such Landsat or Sentinel.

Application of the suggested methods to different remote sensing imagery is possible due to flexibility deep learning tools provide and all aspects of adaptation and optimization for the use with different imagery is covered in previous sections.

Another important aspect of work is that it demonstrates an application of deep learning tools on not popular open-source remote sensing images (such as Landsat 8, GeoEye-1 or Sentinel-2) but a commercial one -WorldView-3, which is of better resolution and quality and the least covered with research in compression to government satellite vehicles. Furthermore, the better resolution and quality and informational content of WorldView-3 imagery may impact solutions using deep learning in multiple ways – better and worse due to neural networks specifics. The latter increases importance of covering with research the ‘unpopular’ commercial satellite imagery. Commercial remote sensing imagery is very important in terms of application in different fields of knowledge due to its usually better technical quality than public satellites, as well as commercial imagery provides coverage of the land more frequently which may be crucial for change detection and treating any sorts of humanitarian crisis.

8. References

[1] D.M. Hordiiuk and V.V. Hnatushenko, Neural network and local laplace filter methods applied to very high resolution remote sensing imagery in urban damage detection, 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF), Lviv, 2017, pp. 363-366, doi: 10.1109/YSF.2017.8126648.

[2] V. Hnatushenko, P. Kogut, M. Uvarov, On Satellite Image Segmentation via Piecewise Constant Approximation of Selective Smoothed Target Mapping, Applied Mathematics and Computation, Vol.389, 2020, Id 125615, 26p, doi.org/10.1016/j.amc.2020.125615.

[3] D. Hordiiuk, I. Oliinyk, V. Hnatushenko, K. Maksymov, Semantic Segmentation for Ships Detection from Satellite Imagery. 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO). doi:10.1109/elnano.2019.8783822.

[4] Zhu, Xiao Xiang, et al, Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[5] D. Mozgovoy, V. Hnatushenko, and V. Vasyliev, Accuracy evaluation of automated object recognition using multispectral aerial images and neural network, Proc. SPIE 10806, Tenth International Conference on Digital Image Processing (ICDIP 2018), 108060H (9 August 2018); https://doi.org/10.1117/12.2502905.

[6] Zhang Liangpei, Lefei Zhang, and Bo Du, Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine 4.2 (2016) 22-40.

[7] Ma Lei, et al, Deep learning in remote sensing applications: A meta-analysis and review. ISPRS journal of photogrammetry and remote sensing 152 (2019) 166-177.

[8] Yuan, Qiangqiang, et al, Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment 241 (2020) 111716.

[9] He Nanjun, Leyuan Fang, and Antonio Plaza, Hybrid first and second order attention Unet for building segmentation in remote sensing images. Science China Information Sciences 63.4 (2020) 1-12.

[10] L. Jiao, L. Huo, C. Hu and P. Tang, Refined UNet: UNet-Based Refinement Network for Cloud and Shadow Precise Segmentation. Remote Sens. 2020, 12, 2001. https://doi.org/10.3390/rs12122001.

[11] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, J. Schmidhuber, Flexible, High Performance Convolutional Neural Networks for Image Classification. International Joint Conference on Artificial Intelligence (IJCAI-2011, Barcelona) (2011).

[12] Zhang, Jianming, et al, Lightweight deep network for traffic sign classification. Annals of Telecommunications 75.7 (2020) 369-379.

[13] D. Ciresan, A. Giusti, L. Gambardella, J. Schmidhuber, Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. In Advances in Neural Information Processing Systems (NIPS 2012), Lake Tahoe (2012).

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada, (2012).

[15] Agostinelli, Forest, Michael R. Anderson, and Honglak Lee, Adaptive multi-column deep neural networks with application to robust image denoising. Advances in Neural Information Processing Systems (2013).

[16] Yuan, Qiangqiang, H. Shen, T. Li, Zhi-wei Li, Shuwen Li, Yun Jiang, Hongzhang Xu, W. Tan, Q. Yang, Jiwen Wang, Jianhao Gao and Liangpei Zhang, Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment 241 (2020) 111716.

[17] Fu, Gang, et al, Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sensing 9.5 (2017) 498.

[18] Maggiori, Emmanuel, et al, Fully convolutional neural networks for remote sensing image classification. 2016 IEEE international geoscience and remote sensing symposium (IGARSS). IEEE, 2016.

[19] Sun, Weiwei, and Ruisheng Wang, Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. IEEE Geoscience and Remote Sensing Letters 15.3 (2018) 474-478.

[20] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39.12 (2017) 2481-2495.

[21] Papandreou, George, et al, Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. Proceedings of the IEEE international conference on computer vision (2015).

[22] Zhang, Mengmeng, Wei Li, and Qian Du, Diverse region-based CNN for hyperspectral image classification. IEEE Transactions on Image Processing 27.6 (2018) 2623-2634.

[23] M. Y. Saifi, J. Singla, Nikita, Deep Learning based Framework for Semantic Segmentation of Satellite Images. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc48092.2020.iccmc-00069

[24] E. Saralioglu, O. Gungor, Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network. Geocarto International, 1–21, (2020). doi:10.1080/10106049.2020.1734871

[25] Satimagingcorp. WorldView-3 Satellite Sensor | Satellite Imaging Corp. (2016). URL: https://www.satimagingcorp.com/satellite-sensors/worldview-3/

[26] Schowengerdt R., Remote sensing: models and methods for image processing, New York: Academic Press. 2007. p.560.

[27] V. Hnatushenko, and V. Zhernovyi, Complex Approach of High-Resolution Multispectral Data Engineering for Deep Neural Network Processing. In: Lytvynenko V., Babichev S., Wójcik W.,

Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing, (2020) vol 1020. Springer, Cham. https://doi.org/10.1007/978-3-030-26474-1_46.

[28] V.J. Kashtan, V.V. Hnatushenko and Y.I. Shedlovska, Processing technology of multispectral remote sensing images, 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF), Lviv, 2017, pp. 355-358, doi: 10.1109/YSF.2017.8126647.

[29] Y.I. Shedlovska and V.V. Hnatushenko, Shadow detection and removal using a shadow formation model, 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2016, pp. 187-190, doi: 10.1109/DSMP.2016.7583537.

[30] T. Stöttner (2019, May 16), Why Data should be Normalized before Training a Neural Network. Medium. URL: https://towardsdatascience.com/why-data-should-be-normalized-before-training-a-neural-network-c626b7f66c7d.

[31] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham (2015).

[32] He Nanjun, Leyuan Fang, and Antonio Plaza, Hybrid first and second order attention Unet for building segmentation in remote sensing images. Science China Information Sciences 63.4 (2020) 1-12.

[33] Sun Shuting, et al. "L-UNet: An LSTM Network for Remote Sensing Image Change Detection" IEEE Geoscience and Remote Sensing Letters (2020). doi: 10.1109/LGRS.2020.3041530.

[34] Cao Kaili and Xiaoli Zhang, An improved res-unet model for tree species classification using airborne high-resolution images. Remote Sensing 2020; 12(7): 1128. https://doi.org/10.3390/rs12071128.

[35] V. Hnatushenko and V. Zhernovyi, Method of Improving Instance Segmentation for Very High Resolution Remote Sensing Imagery Using Deep Learning. In: Babichev S., Peleshko D., Vynokurova O. (eds). Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science, vol. 1158. Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_21.

[36] Milletari Fausto, Nassir Navab and Seyed-Ahmad Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV). IEEE, 2016.

[37] Carvana Image Masking Challenge | Kaggle. (2015). URL: Kaggle. https://www.kaggle.com/c/carvana-image-masking-challenge/discussion/40199

[38] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia & H. Wu, Mixed precision training. arXiv preprint arXiv:1710.03740, 2017.

[39] Nobuyuki Otsu, A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[40] A. Ghosh, M. Ehrlich, S. Shah, L. Davis, & R. Chellappa, Stacked U-Nets for Ground Material Segmentation in Remote Sensing Imagery. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.252-256. doi:10.1109/cvprw.2018.00047.

[41] X. Yuan, J. Shi, & L. Gu, A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. Expert Systems with Applications, 2020, 114417. doi:10.1016/j.eswa.2020.114417.