Information Sciences Letters
*An International Journal*

http://dx.doi.org/10.18576/isl/120643

# On the Application of Data Clustering Algorithm used in Information Retrieval for Satellite Imagery Segmentation

*Ahmed NourEldeen[1,*], Yasser Fouad[2], Mohamed E. Wahed[3], and Mohamed S. Metwally[1]*

[1]Department of Mathematics, Faculty of Science, Suez University, Suez, Egypt
[2]Faculty of Computers and Informatics, Suez University, Suez, Egypt
[3]Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

Received: 1 Apr. 2023; Revised: 16 May 2023; Accepted: 20 May 2023.
Published online: 1 Jun. 2023.

**Abstract:** This study proposes an automated technique for segmenting satellite imagery using unsupervised learning. Autoencoders, a type of neural network, are employed for dimensionality reduction and feature extraction. The study evaluates different segmentation architectures and encoders and identifies the best performing combination as the DeepLabv3+ architecture with a ResNet-152 encoder. This approach achieves high performance scores across multiple metrics and can be beneficial in various fields, including agriculture, land use monitoring, and disaster response.

**Keywords:** Convolutional neural networks (CNN); Data clustering; Land cover; Semantic segmentation; Satellite Imagery.

## 1 Introduction

Accurate land cover data is vital for various applications, including natural resource management, urban planning, and natural hazard assessment and mitigation [1-3]. In remote sensing, land cover classification and change detection are crucial tasks that have gained significant attention in recent decades. This focus is due to the increasing availability of remote sensing data and advancements in computer power and machine learning approaches [4,5], enabling large-scale autonomous land cover identification. Furthermore, remote sensing data can be integrated with other geospatial data, such as topographic maps, weather models, and demographic information, to generate hazard models and risk assessments. These models and assessments provide decision-makers with valuable information to develop mitigation and preparedness plans and to make informed decisions during and after a natural disaster.

Image segmentation is a fundamental vision problem with a long research history. Traditionally, this problem has been investigated in the unsupervised learning, which involves producing a pixelwise prediction to segment an image into coherent clusters that correspond to objects in the image. Classical computer vision has several well-known techniques for this problem, including normalized cuts [6, 7], methods based on Markov random fields [8], mean shift [9], hierarchical methods [10], and many more.

With the recent progress of deep learning in computer vision, there has been renewed interest in the image segmentation problems. Most recent research in this field has concentrated on semantic segmentation [11-16], an image segmentation task with supervision. Usually, these techniques use a fully convolutional networks are one example to produce pixel-level predictions, and then supervised training methods are utilized to determine filters that can segment new images.

One popular recent method is the U-Net architecture [14], which is a fully convolutional network that has produced remarkable results in the biomedical image domain. However, existing semantic segmentation methods demand a large amount of pixel-level labeled data, which can be challenging to acquire for new domains. In light of the segmentation problem's significance in various fields and the shortage of labeled training data, we revisit the unsupervised image segmentation problem and incorporate recent semantic segmentation concepts. To this end, we use introduce a hyper approach, which connects two fully convolutional network (FCN), similar to the U-Net, to create an autoencoder. The initial FCN encodes the input image into a k-way soft segmentation, while the second FCN does the opposite operation, transforming the segmentation layer into a restored image.

## 2 Related Work

Currently, most research on classifying land cover based on multispectral remote sensing data relies on supervised machine learning (ML) methods. When remote sensing images are classified pixel-by-pixel, this process is called image

[*]Corresponding author e-mail: ahmednour_cs@yahoo.com

segmentation. ML approaches can be classified according to their input data, namely pixel-based, spatial, and sequence-based approaches. Pixel-based methods, which classify individual pixels based on their corresponding spectral data, are commonly used for land cover classification from multispectral data. Examples of pixel-based ML models include Random Forest [17,18], support vector machines [19], and self-organizing maps [20,21]. However, pix-el-based approaches have a significant disadvantage in that they overlook spatial patterns, including important information about the classification task. This limitation is particularly relevant in land cover classification, where land cover classes, such as farmland or water bodies, often encompass coherent areas that are larger than one pixel. Consequently, pixel-based ML methods cannot leverage correlations between neighboring pixels to improve classification performance.

Many unsupervised image segmentation methods rely on extracting features such as color, brightness, and texture from local patches, followed by clustering at the pixel level based on these features. Among these methods, the most commonly used ones are Felzenszwalb and Huttenlocher’s graph-based approach [22], Shi and Malik’s Normalized Cuts [6, 7], and Comaniciu and Meer’s Mean Shift [9]. Edge detection-based approaches, such as the one proposed by Arbelaez et al. [10,23] have also been shown to outperform classical methods. Additionally, a recent unified method for hierarchical im-age segmentation was proposed in [24].

One of the most used techniques in unsupervised feature learning is the encoder-decoder method, which has been widely studied and applied [25, 26]. The encoder function maps the input (such as an image patch) to a compressed feature representation, while the decoder function reconstructs the input from this lower-dimensional representation.

In this article, we have developed a hyper approach for Land cover classification. We used the encoder to maps the input to a dense pixelwise segmentation layer with the same spatial size instead of a low-dimensional space. The decoder then generates a reconstruction from this dense prediction layer.

## 3 Methodologies

### *3.1. Methodology*

Our proposed approach is using the autoencoders for satellite imagery segmentation technique for land cover classification.

We developed a new deep learning model by adopting segmentation mothed using the DeepLabv3+ [27] architecture as decoder and a ResNet-152 [28] encoder as in Figure 1.

The FCN architecture serves as the foundation for the DNN models employed in this study, which is made up of an encoder followed by a decoder [29]. While the decoder up samples the feature maps in the latent space to the original input size, the encoder collects features from the input image.

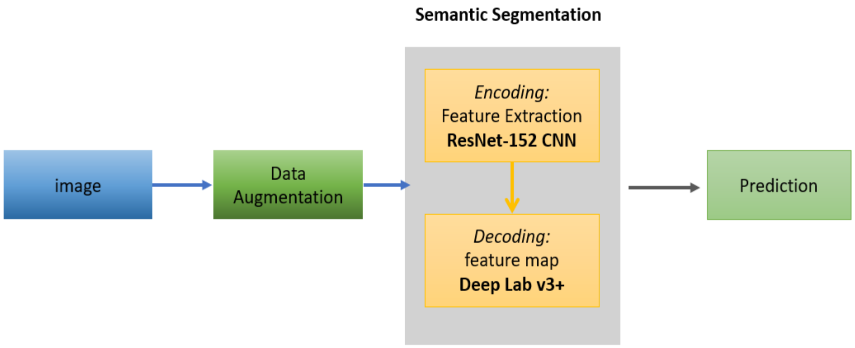
Semantic Segmentation
image
Data Augmentation
Encoding: Feature Extraction ResNet-152 CNN
Decoding: feature map Deep Lab v3+
Prediction

**Fig. 1:** Flowchart of the proposed model used deep neural network architecture which follows an encoder decoder structure.

### *3.2. Images dataset [30]*

In this study, three datasets were utilized, named “Aksu”, “Kestel”, and “Aksu + Kestel”. The “Aksu + Kestel” dataset is a mix of the two datasets, and the tiling approach used has a size of 512x512 pixels with 128-pixel overlaps. The overlap in the images is included not just to increase the number of patches, but also to boost the classifier's capacity to recognize the spatial continuity of the image. (i.e., contextual information) [29,31]. Each dataset's patches were split into training, validation, and test sets using a 70% ,20% and 10 %, respectively.

Figure 2 shows the sample of the images and their corresponding ground truth maps from the datasets. The optical images are presented in the first column, while the ground truth masks are presented in the second column. The image with different classes is shown.

## *3.3. Data Augmentation*

Basic image processing techniques such as flip, rotation, shift, and scale are used to augment the dataset and increase its volume. Additionally, we used a sampling technique to address the issue of under-represented classes, by over-sampling these samples to help the model focus more on them. Specifically, we used the "*compute_sample_weight*" method from the sklearn library to calculate weights, which were then fed into the PyTorch DataLoader as input [32].

## *3.4. Encoding for Feature Extraction*

The feature extraction process from the input images is handled by the encoder component, it is an important stage in data classification. For this purpose, ResNet-152[28], a Deep Residual Learning for Image Recognition architecture, is utilized due to its superior accuracy compared to other ResNet models [28]. Down sampling is conducted in the encoder component, while Atrous convolutions are used to convey the low-level information recovered from the image in the embedded space to the decoder component.

## *3.5 Decoding for Feature Map*

The decoder component of the deep neural network (DNN) upscales the feature maps in the latent space to their original image size. Through a benchmarking process, we identified the DeepLab v3+ [27] architecture as the most effective in generating densely predicted segmentation maps.

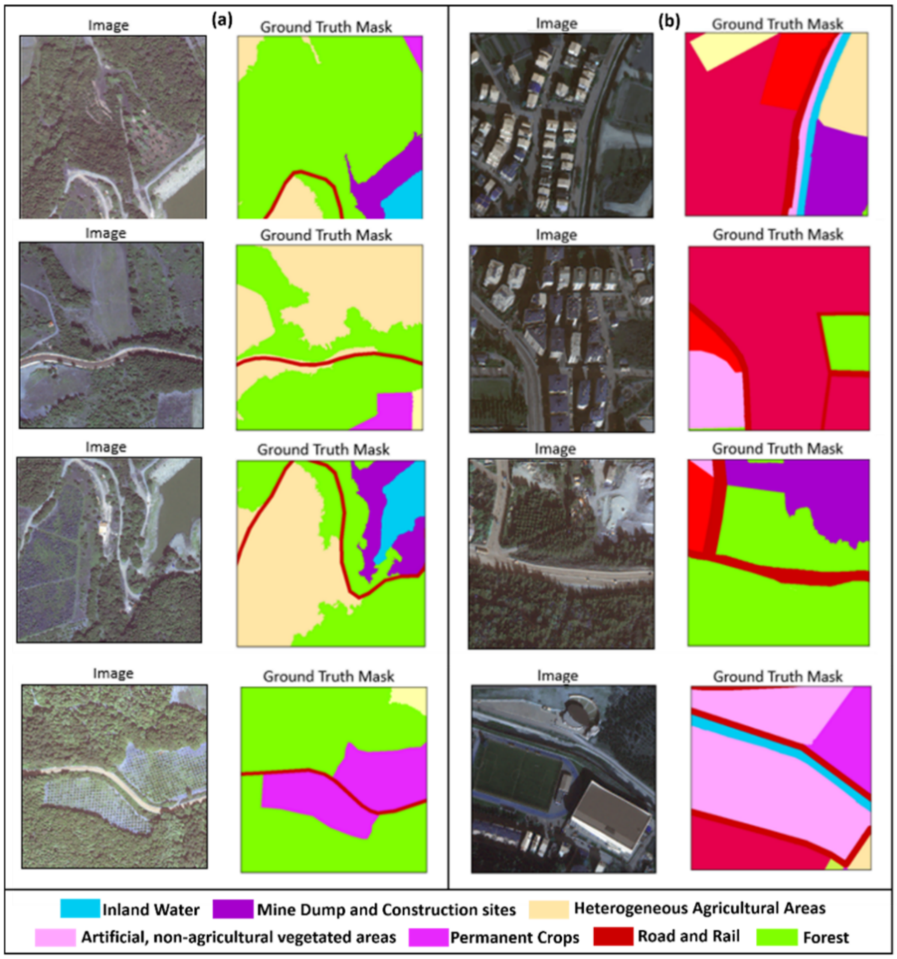
(a)
(b)
Image
Ground Truth Mask
Inland Water
Mine Dump and Construction sites
Heterogeneous Agricultural Areas
Artificial, non-agricultural vegetated areas
Permanent Crops
Road and Rail
Forest

**Fig. 2:** Example of images and their corresponding ground truth masks. Specifically, (a) displays sample patches extracted from the Aksu region, while (b) shows sample patches taken from the "Kestel "region [30]

### *3.6. Performance Evaluation Metrics*

The joint loss function used in this study combines the Dice loss and Focal loss to address the issue of disparity between classes [33, 34]. Equation (1) shows the constructed loss function, where the Dice loss is represented by the first term, and the Focal loss, which is weighted by a coefficient of 0.5, is represented by the second term. Both loss functions are effective in dealing with class imbalance. In the Dice loss function, pi and $gi$ denote the prediction equivalent pixel values and ground truth, respectively. The Focal loss function term $a_t$ is a weighted hyperparameter offset that scales the main term to address class disparity. The operator γ acts as a relaxation parameter that adjusts the priority assigned to correctly or incorrectly categorized samples.

$$L = \frac{2\sum_i^N pigi}{\sum_i^N p_i^2 + \sum_i^N g_i^2} + (-a_t(1-p_t)0^{\gamma} \log\log p_t) * 0.5 \tag{1}$$

In addition to qualitative analysis, standard evaluation metrics have been utilized to evaluate the effectiveness of the created classifiers. The quantitative metrics employed in this paper include Intersection over Union (IoU), Precision, Recall, F1 score, and Accuracy values, which are computed from the confusion matrix according to False Negative (FN), True Negative (TN), False Positive (FP), and True Positive (TP). These metrics are represented in Equations (2)–(6):

$$IoU = \frac{TP}{TP + TF + TN} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{F1} - \text{score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{6}$$

## 4 Implementation and Discussion of Results

### *4.1. Experimental Setup*

The machine used for the experiments was an Intel Core i7 CPU, 32 GB RAM with an onboard NVIDIA GeForce GTX1070. The training processes had a time constraint of 100 epochs with early stop technique. With a value of 0.9, the Adam optimization algorithm used. The Python programming language is used to implement all the routines in the Pytorch library.

### *4.2. Evaluation*

We used the primary challenge metric to evaluate the suggested model like IoU, F-1 Score, Precision and Recall for comparing six well-known deep neural network architectures to get the best segmentation model, which are:

- DeepLab v3+ [35],
- (PAN) Pyramid Attention Network [36,37],
- U-Net ++ [38],
- (FPNs) Feature Pyramid Networks [39],
- Linknet [40], and
- (PSP-Net) Pyramid Scene Parsing Network [36].

**Table 1**: Comparison of different architectures' segmentation outcomes (bold text denotes the best-performing configuration).

| Architecture | IoU | F-1 Score | Precision | Recall |
|---|---|---|---|---|
| **DeepLabv3+(ours)** | **90.34** | **95.07** | **95.02** | **95.15** |
| DeepLabv3+ [30] | 89.46 | 94.35 | 94.25 | 94.49 |

| | | | | |
|---|---|---|---|---|
| PAN | 82.78 | 90.37 | 90.34 | 90.47 |
| U-Net ++ | 81.54 | 89.54 | 89.63 | 89.45 |
| FPN | 76.45 | 86.39 | 86.39 | 86.38 |
| Linknet | 74.75 | 84.99 | 84.95 | 85.04 |
| PSPNet | 71.20 | 82.44 | 82.44 | 82.45 |

**Table 2:** Encoder comparison utilizing the DeepLabv3+ segmentation architecture (bold text denotes the best-performing configuration).

| **Architecture** | **Parameters** | **IoU** | **F-1 Score** | **Precision** | **Recall** |
|---|---|---|---|---|---|
| **ResNet152 (ours)** | **58M** | **90.34** | **95.07** | **95.02** | **95.15** |
| ResNeXt50 | 22M | 89.46 | 94.34 | 94.25 | 94.49 |
| ResNet50 | 23M | 87.32 | 93.08 | 92.99 | 93.16 |
| DPN68 | 11M | 80.83 | 88.61 | 88.61 | 88.61 |
| MobileNet | 2M | 79.07 | 88.09 | 88.15 | 88.02 |
| Efficientnet-b0 | 4M | 79.94 | 88.48 | 88.42 | 88.55 |
| Efficientnet-b1 | 6M | 82.64 | 90.24 | 90.16 | 90.32 |
| Efficientnet-b2 | 7M | 83.36 | 90.58 | 90.52 | 90.64 |

### *4.3. Comparison with Similar works*

Table 3 compares our model to the most recent approach in terms of F1-score, Precision, and Recall; our model exceeded the most recent approaches. As a result of the combination of DeepLab v3+ and ResNet-152, F1 Score of 95.07 %.

**Table 3**: Comparative F1 score, Precision, and Recall between our approach and the newest approach

| **Model** | **Encoder** | **Decoder** | **F-1 Score** | **Precision** | **Recall** |
|---|---|---|---|---|---|
| **Ours** | **ResNet152** | **DeepLabv3+** | **95.07** | **95.02** | **95.15** |
| Sertel, etc. [30] | ResNeXt50 | DeepLabv3+ | 94.35 | 94.25 | 94.49 |

Our study explored the effectiveness of using DeepLab v3+ and ResNet-152 for land cover clustering. Our results demonstrate that this approach is more accurate than other clustering methods previously used in remote sensing. One possible reason for the high accuracy of our approach is the ability of DeepLab v3+ to capture fine-grained details in the imagery. This is due to its powerful encoder-decoder architecture that utilizes dilated convolutional layers and Atrous spatial pyramid pooling. This allows for the extraction of high-level features at different scales, improving the quality of segmentation results.

Another factor contributing to the accuracy of our approach is the use of ResNet-152 as the backbone network. ResNet-152 is a deep residual network architecture that is widely used for image classification tasks. Its use in our approach allows for the efficient ex-traction of high-level features, improving the performance of the clustering process.

Our results demonstrate that our approach is not only accurate but also efficient. The use of DeepLab v3+ and ResNet-152 allows for the processing of large-scale remote sensing data in a short amount of time. This is particularly important for apps like natural resource management and urban planning, where accurate and timely information is critical.

## 5 Conclusions

In conclusion, our study provides strong evidence that the use of DeepLab v3+ and ResNet-152 is an effective approach for land cover clustering in remote sensing. The high accuracy and efficiency of this approach make it a promising tool for a wide range of applications, including natural hazard assessment and mitigation, urban planning, and natural resource management.

Based on these findings, future research directions can be explored to further enhance and expand the application of this approach. Some potential avenues for future investigation include:

- Improving the scalability: Investigate methods to optimize the DeepLab v3+ and ResNet-152 approach to handle large-scale satellite imagery datasets. This can involve exploring distributed computing techniques or parallel processing algorithms to enable efficient analysis of high-resolution imagery.
- Incorporating multi-temporal analysis: Extend the approach to incorporate temporal information by integrating time-series satellite imagery. This would enable the detection and monitoring of dynamic changes in land cover patterns over time, providing valuable insights for long-term environmental monitoring and climate change studies.
- Fusion with other data sources: Investigate the fusion of satellite imagery with other geospatial data sources, such

as aerial photography, LiDAR, or socioeconomic data. By combining multiple data modalities, it is possible to enhance the accuracy and richness of land cover clustering results and enable more comprehensive analysis.

- Transfer learning and domain adaptation: Explore the transferability of the DeepLab v3+ and ResNet-152 model to different geographic regions or domains. Investigate techniques for domain adaptation to effectively utilize pre-trained models in areas with limited labeled data, thereby improving the generalizability and applicability of the approach.
- Incorporating uncertainty estimation: Develop methods to estimate and quantify the uncertainty associated with the land cover clustering results. This would provide decision-makers and stakeholders with a better understanding of the reliability and confidence levels of the obtained clusters, supporting informed decision-making in various applications.

By addressing these research directions, we can further advance the capabilities of land cover classification in remote sensing, enabling its broader adoption and unlocking its potential in addressing critical environmental and societal challenges.

## Conflicts of Interest

The authors declare that there is no conflict regarding the publication of this paper.

## References

[1] Green, K.; Kempka, D.; Lackey, L. Using remote sensing to detect and monitor land-cover and land-use change. *Photogramm. Eng. Remote Sens*. 60, 331–337, (1994).

[2] Loveland, T.; Sohl, T.; Stehman, S.; Gallant, A.; Sayler, K.; Napton, D. A Strategy for Estimating the Rates of Recent United States Land-Cover Changes. *Photogramm. Eng. Remote Sens.* 68, 1091–1099 (2002).

[3] Yuan, F.; Sawaya, K.E.; Loeffelholz, B.C.; Bauer, M.E. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote Sens. Environ*. 98, 317–328 (2005).

[4] Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36 (2012).

[5] Riese, F.M.; Keller, S. Supervised, Semi-Supervised, and Unsupervised Learning for Hyperspectral Regression. *In Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*; Prasad, S., Chanussot, J., Eds.; Springer International Publishing: Cham, Switzerland, Chapter 7, pp. 187–232 (2020).

[6] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. *In Computer Vision and Pattern Recognition,* 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 1124–1131. IEEE, (2005).

[7] J. Shi and J. Malik. Normalized cuts and image segmentation. IEEE *Transactions on pattern analysis and machine ceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1916–1922, (2013).

[8] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE *transactions on medical imaging*, 20(1):45–57, (2001).

[9] D. Comaneci and P. M. M. Shift. A robust approach toward feature space analysis. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 24(5), (2002).

[10] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. IEEE *transactions on pattern analysis and machine intelligence*, 33(5):898–916, (2011).

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, (2015).

[12] A.Chaurasia and E. Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. *arXiv preprint* arXiv:1707.03718, (2017).

[13] A.Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint* arXiv:1606.02147, (2016).

[14] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, (2015).

[15] P. Kr¨ahenb¨uhl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *In Advances in neural information processing systems*, pages 109–117,( 2011).

[16] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE *International Conference on Computer Vision*, pages 1529–1537, (2015).

[17] Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett*. 27, 294–300 (2006).

[18] Keller, S.; Braun, A.C.; Hinz, S.; Weinmann, M. Investigation of the impact of dimensionality reduction and feature selection on the classification of hyperspectral EnMAP data. *In Proceedings of the 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Los Angeles, CA, USA, 21–24; pp. 1–6 August (2016).

[19] Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. IEEE *Trans. Geosci. Remote Sens*. 42, 1778–1790 (2004).

[20] Riese, F.M. Development and Applications of Machine Learning Methods for Hyperspectral Data. Ph.D. Thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, (2020).

[21] Riese, F.M.; Keller, S.; Hinz, S. Supervised and Semi-Supervised Self-Organizing Maps for Regression and Classification Focusing on Hyperspectral Data. *Remote Sens.* 12, 7(2020).

[22] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graphbased image segmentation. *International journal of computer vision*, 59(2):167–181, (2004).

[23] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. *In Computer Vision and Pattern Recognition*, 2009. CVPR 2009. IEEE Conference on, pages 2294–2301. IEEE, (2009).

[24] J. Pont-Tuset, P. Arbelaez, J. T. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping for image segmentation and object proposal generation. IEEE *transactions on pattern analysis and machine intelligence*, 39(1):128– 140, (2017).

[25] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, (2006).

[26] F. J. Huang, Y.-L. Boureau, Y. LeCun, et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition. pages 1–8,( 2007).

[27] Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Computer Vision—ECCV; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; *Springer International Publishing*: Cham, Switzerland, Volume 11211, pp. 833–851, ISBN 978-3-030-01233-5 (2018).

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV] 10 (2015).

[29] Ekim, B.; Sertel, E.; Kabadayı, M.E. Automatic Road extraction from historical maps using deep learning techniques: A regional case study of Turkey in a German WorldWar II Map. *Int. J. Geo-Inf*. 10, 492 (2021).

[30] Sertel, E.; Ekim, B.; Ettehadi Osgouei, P.; Kabadayi, M.E. Land Use and Land Cover Mapping Using Deep Learning Based Segmentation Approaches and VHR Worldview-3 Images. *Remote Sens*. 14, 4558 (2022).

[31] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *In Proceedings of the Medical Image Computing and Computer-Assisted Intervention*—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J.,Wells,W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, pp. 234–241(2015).

[32] Sklearn Package. Available online: https://scikit-learn.org/stable/about.html#citing-scikit-learn (accessed on 1 February 2022).

[33] Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised dice overlap as a deep learning loss

function for highly unbalanced segmentations. arXiv:1707.03237 *(2017)*.

[34] Mulyanto, M.; Faisal, M.; Prakosa, S.W.; Leu, J.-S. Effectiveness of focal loss for minority classification in network intrusion detection systems. Symmetry, 13, 4 (2021),.

[35] Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Computer Vision—ECCV; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; *Springer International Publishing*: Cham, Switzerland, Volume 11211, pp. 833–851, ISBN 978-3-030-01233-5 (2018).

[36] Zhao, H.; Shi, J.; Qi, X.;Wang, X.; Jia, J. Pyramid scene parsing network., arXiv:1612.01105 (2017).

[37] Li, H.; Xiong, P.; An, J.;Wang, L. Pyramid attention network for semantic segmentation. *arXiv (2018), arXiv:1805.10180*.

[38] Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. *arXiv:1807.10165 (2018)*.

[39] Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. *arXiv:1612.03144 (2017),*.

[40] Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 *IEEE Visual Communications and Image Processing (VCIP), St. Petersburg*, FL, USA, 10–13, pp. 1–4 December (2017).