

Classification of semantic segmentation using fully convolutional networks based unmanned aerial vehicle application

Shouket Abdulrahman Ahmed¹, Hazry Desa¹, Abadal-Salam T. Hussain^{1,2}

¹Centre of Excellence for Unmanned Aerial Systems (COEUAS), Universiti Malaysia Perlis, Perlis, Malaysia

²Department of Medical Instrumentation Engineering Techniques, Al-Kitab University, Kirkuk, Iraq

Article Info

Article history:

Received Jul 1, 2022

Revised Sep 27, 2022

Accepted Oct 27, 2022

Keywords:

Artificial neural network

Classification

Convolutional neural network

Deep neural networks network

Semantic segmentation

Unmanned aerial vehicle

ABSTRACT

The classification of semantic segmentation-based unmanned aerial vehicle (UAV) application based on the datasets used in this work and the necessary data preprocessing steps for the optimization and implementation of the models are also involved. The optimization of the various models was done using the evaluation metrics and loss functions because deep neural networks (DNNs) are just about writing a cost function and its subsequent optimization. convolutional neural network (CNN) is a common type of artificial neural network (ANN) that has found application in numerous tasks, such as image and video recognition, image classification, recommender systems, financial time series, medical image analysis, and natural language processing. CNN is developed to automatically and adaptively learn spatial feature hierarchies via backpropagation using numerous building blocks, such as pooling, convolution, and fully connected layers. The result of identification was excellent. The image segmentation was detected and comprehend the actual components of an image down to the pixel level. The result created an entire image segmentation masks with instances using the new label editor in the label box.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Shouket Abdulrahman Ahmed

Centre of Excellence for Unmanned Aerial Systems (COEUAS), Universiti Malaysia Perlis

Jalan Kangar-Alor Setar, 01000 Kangar, Perlis, Malaysia

Email: shouketunimap@gmail.com

1. INTRODUCTION

These are models whose architecture is made up of layers arranged in stacks. The models of each stage learn a useful pattern by successively filtering the input data. Figure 1 depicts a basic 4-layer neural network for the classification of handwriting from the Modified National Institute of Standards and Technology (MNIST) dataset [1]. The four components of a system for training a neural network include: i) Layers; ii) Input data; iii) Loss function; and iv) Optimizer.

Furthermore, neural networks are trained to identify the correct weight values that will specify the transformations to be performed on the incoming data [2]. Figure 2 depicts the network layers parameterization process. The setup of these weights is challenging to optimize for varied tasks because there can be millions of parameters that are interdependent. A loss function is utilized to generate a weight configuration because it measures the extent of performance of the layer representations from the network compared to the predicted output. The first stage of the backpropagation (BP) algorithm is the determination of the final loss value [3], followed by computation of each parameter's contribution in the computed loss value from the top layer to the bottom layer [4]. The loss function derivative can be determined at a given

training point by identifying the required adjustments to minimize the loss and optimize the model to a given evaluation metric known as stochastic gradient descent [5], [6].

The Figure 3 shows the final representation of a neural network training loop. The weights are randomly initialized (using a random initializer) in this training loop, which results in a high loss score because the initial weights are likely unsuitable for capturing the data patterns in a usable manner. The network, on the other hand, adjusts its weights with each training batch and gradually improves until it has a small loss [7].

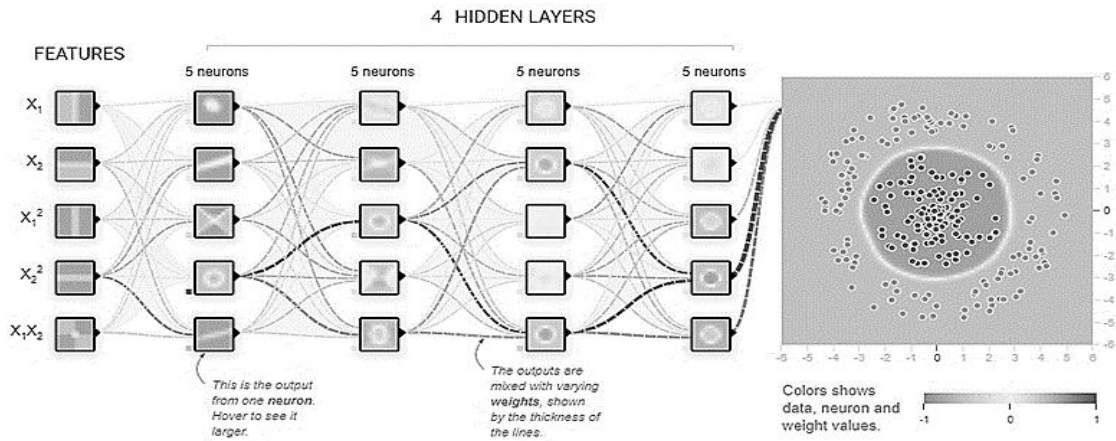


Figure 1. A neural network with four layers

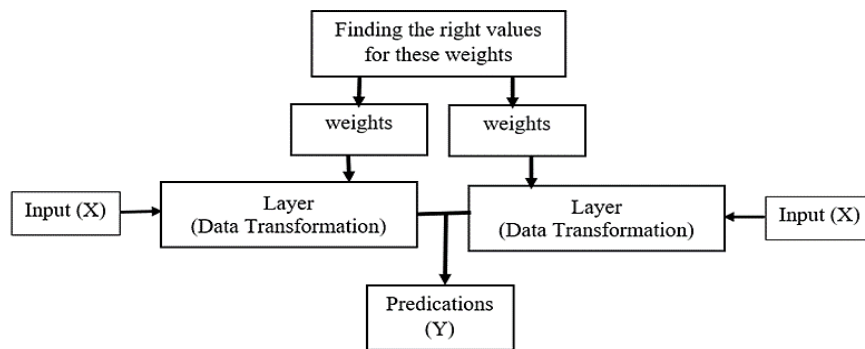


Figure 2. Weight values-based parameterization of DL layers

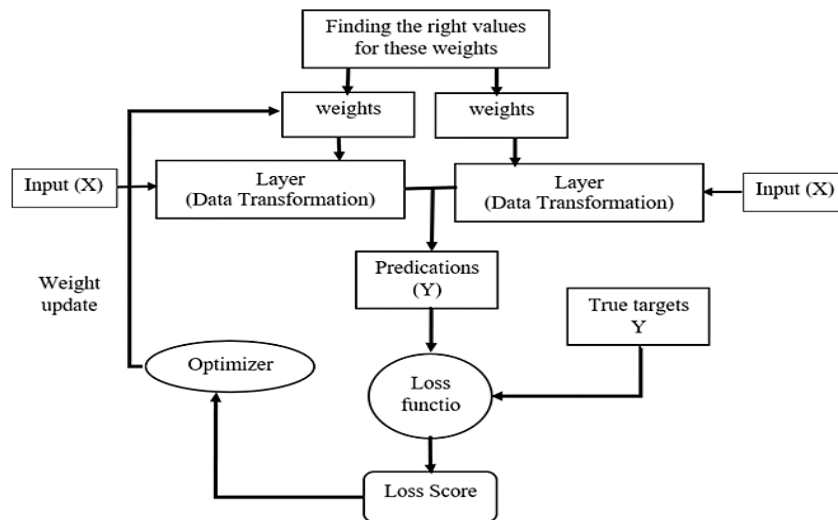


Figure 3. Training loop for a neural network

2. METHOD

2.1. Dense layers

Traditionally, densely linked layers, also known as completely connected layers [8], [9], have been used for datasets other than images, in which a form of connection exists between the neurons from one layer to the other. These dense layers [10], [11], on the other hand, can only learn patterns that fall within their input feature space, while each convolution filter is capable of learning patterns locally (kernel space or region of interest) [12]–[14]. This implies the possibility of fragmenting input images into edges and textures for easy learning; they are also more useful for classification rather than global patterns [15].

2.2. Convolutional neural networks (CNNs)

CNNs are a sort of neural network that is largely utilized in deep learning (DL) for tasks in computer vision [16]–[18]. CNNs are similar to traditional neural networks, except that rather than generic matrix multiplication [19], [20], CNNs depend on a convolution kernel that is found in one or more of the CNN network layers [21], [22]. CNNs can be implemented using image tensors as input because they can extract the relevant image features for classification or differentiation, and then output classifications [23], [24]. CNNs have found wide application because they can use relevant filters to learn translation-invariant patterns and spatial hierarchies. Translation invariant patterns learning implies that the network can identify similar patterns wherever else in the image after learning it in one region. This ensures image processing efficiency owing to the need to learn just few training samples that can be generalized [25].

Spatial hierarchies show how the network's initial levels can learn few local patterns that can be generalized into larger patterns made up of these few patterns in the subsequent layers. This enhances the capability of the network to acquire abstract and complex visual concepts. It also allows these networks to produce predictions of greater accuracy rather than just vectorizing a complicated image with heavy reliance on pixels. To track these key properties and interactions between features, CNNs must be used in any image classification operation [26].

2.3. Convolution

Convolution is a procedure that combines two functions of a real-valued parameter to create a new function. Let $w(a)$ represent the weighting function that prioritize recent measurements, $s(t)$ represent the time-based output estimate function, and $x(a)$ represent the time-based input position function. Based on these defined functions, the generic equation for convolution is shown in (1) and Figure 4 shows visual representation of dimensional convolution [27].

$$s(t) = \int x(a)w(t - a)da \tag{1}$$

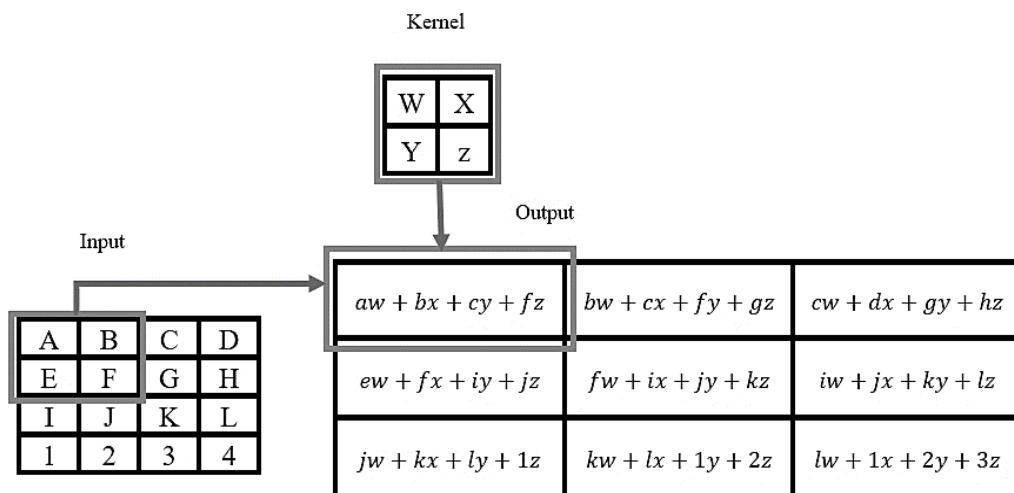


Figure 4. Visual representation of dimensional convolution

The $x(a)$ is the input function, $w(a)$ is the kernel, and $s(t)$ is the feature map in CNNs. In CNNs, convolutions are done on 2-D tensors made up of the input images' height, width, and color channels [28]. These operations take patches from the input and apply transformations to create a feature map with varying

depths as an output. The output feature map depth, which indicates the number of filters built by the layer and each of which can encode a different input data feature, is one of the two key criteria for creating these convolutions, coupled with the size of the extracted patches [29].

The computation of 2D convolutions is done by moving a square window of defined width and height across all input feature map pixels. For images, the input feature map has three dimensions which are height, breadth, and color bands (corresponding to red, green, and blue). When these feature maps are subjected to 2D convolution, 2-D patches of the surrounding features are created. Then, these patches are turned into a 1-D vector that reflects the output depth using a convolution kernel. After that, the vectors are spatially reconstructed into a 2-D output map that corresponds to all the input map's locations. Figure 4 depicts the process of this convolution [30].

2.4. Strides

The stride of the convolution is a characteristic that can determine the size of the output in CNNs. Stride is a convolution operation parameter that specifies the distance between the derived patches from the input feature map. A stride of 2 implies that the output feature map's width and height are downsized by a factor of 2 without padding as shown in Figure 5 [31].

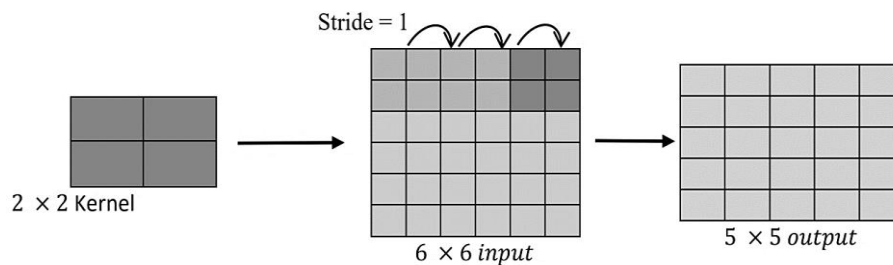


Figure 5. Strides

2.5. Semantic segmentation

Each pixel in an image is given a class label using the image classification approach known as semantic segmentation. For aerial imagery, semantic segmentation for use in a feature extractor is carried out to locate pertinent areas of an image that may point to invariance between seasons and times of the day. Even though fully convolutional networks (FCNs) can be used for this task, precise/accurate image segmentation and localizations learning necessitates a large amount of data. Due to the lengthy time required to manually label satellite images with many classifications, there are not many datasets with extensive pixel-by-pixel labels that are publicly available. Hence, FCNs and U-Nets are commonly employed in aerial imagery segmentation to address both issues [32].

2.6. Fully convolutional networks (FCN)

The advent of a variant of the CNN known as the FCN represents a big advancement in picture segmentation. The difference between FCN and the traditional CNN is that in the CNN, the completely linked terminal layers are transformed into convolution layers, resulting in the creation of a nonlinear filter for each output vector layer in the network. As a result, the completed network can function on inputs of any size and produce outputs with the same spatial dimensions. Hence, the classification network can generate a heatmap of the selected item class. The addition of layers and a spatial loss to the network results in an efficient scheme for end-to-end dense learning. Figure 6 depicts an example of this transformation [33].

FCNs are not only more flexible because they can take a variety of input image sizes, but they've also been shown to be more efficient for learning dense predictions thanks to in-network up-sampling. The FCN may also keep track of the input's spatial information, which is important for semantic segmentation because it requires both classification and localization. Although an FCN can accept any size input image, non-padded convolutions are used to decrease the output resolution. These were created to keep filter sizes modest and reduce the computational demands. The outcome is a coarse output with a size reduction equal to the pixel stride of the output units' receptive field [34], [35].

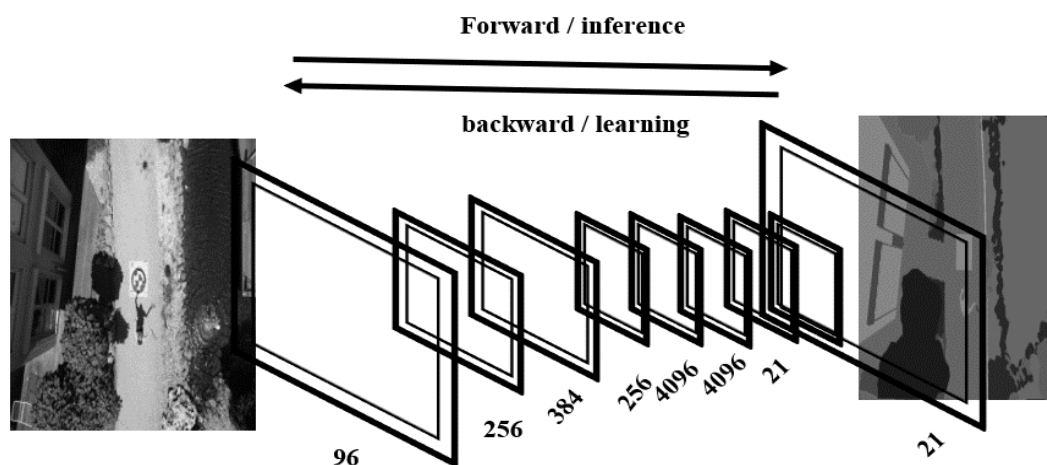


Figure 6. FCN making dense predictions for per-pixel tasks such as semantic segmentation

3. RESULTS AND DISCUSSION

The unmanned aerial vehicle (UAV) system benefits from this model application especially with the limited resources of UAVs system. Fully autonomous UAVs need to understand their surrounding environments in detail. Always the surrounding objects are in 3D while token pictures were shown in 2D which is critical for high-level decision-making. For example, always drones need to understand their surrounding environment in real time; therefore, real-time semantic mapping is significant and worth exploring in this type of drone application. Such application and processes will be effective for most important factor of UAV which is the power resources, so semantic segmentation will reduce the time and used power to process video images or frames in order to reach a final decision [36].

4. CONCLUSION

Using FCN has been shown to be not only more flexible as images can be obtained with different input sizes, but also more efficient as the network uses upsampling to learn dense predictions increase. The input's spatial information can also be preserved using FCN, especially for semantic segmentation applications. This task requires both classification and localization, hence, any image size can be used as an input with FCN, even though the output resolution is negatively impacted by convolution without padding. Hence, these were introduced as a way of reducing the computational demands and filter sizes. The result is a coarse output that is reduced in size by a factor equivalent to the pixel pitch of the incoming field of the output unit.

ACKNOWLEDGEMENTS

The Authors declare that they have no conflict of interest. The authors are thankful for the Centre of Excellence for Unmanned Aerial Systems (COEUAS), Universiti Malaysia Perlis, Jalan Kangar-Alor Setar, 01000 Kangar, Perlis, MALAYSIA for their support and providing their labs and environments to do the testing.

REFERENCES




- [1] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.
- [2] M. Dian Bah, A. Hafiane, and R. Canals, "Deep learning with unsupervised data labeling for weed detection in line crops in UAV images," *Remote Sensing*, vol. 10, no. 11, 2018, doi: 10.3390/rs10111690.
- [3] G. S. Saragih, Z. Rustam, and J. E. Aurelia, "A hybrid model based on convolutional neural networks and fuzzy kernel K-medoids for lung cancer detection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, p. 126, 2021, doi: 10.11591/ijeecs.v24.i1.pp126-133.
- [4] J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community," *Journal of Applied Remote Sensing*, vol. 11, no. 04, p. 1, 2017, doi: 10.1117/1.jrs.11.042609.
- [5] L. J. Biffi *et al.*, "Article atss deep learning-based approach to detect apple fruits," *Remote Sensing*, vol. 13, no. 1, pp. 1–23, 2021, doi: 10.3390/rs13010054.
- [6] P. S. Bithas, E. T. Michailidis, N. Nomikos, D. Vouyioukas, and A. G. Kanatas, "A survey on machine-learning techniques for UAV-based communications," *Sensors (Switzerland)*, vol. 19, no. 23, 2019, doi: 10.3390/s19235170.
- [7] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep

- convolutional neural networks architectures,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1015–1027, 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.
- [8] W. Boonpook, Y. Tan, and B. Xu, “Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry,” *International Journal of Remote Sensing*, vol. 42, no. 1, pp. 1–19, 2021, doi: 10.1080/01431161.2020.1788742.
- [9] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12346 LNCS, pp. 213–229, 2020, doi: 10.1007/978-3-030-58452-8_13.
- [10] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11218 LNCS, pp. 139–156, 2018, doi: 10.1007/978-3-030-01264-9_9.
- [11] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, “Unsupervised learning of visual features by contrasting cluster assignments,” *Advances in Neural Information Processing Systems*, vol. 2020-December, 2020.
- [12] E. mehdi Cherrat, R. Alaoui, and H. Bouzahir, “A multimodal biometric identification system based on cascade advanced of fingerprint, fingervein and face images,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3, p. 1562, 2020, doi: 10.11591/ijeecs.v17.i3.pp1562-1570.
- [13] C. C. V. Da Silva, K. Nogueira, H. N. Oliveira, and J. A. D. Santos, “Towards open-set semantic segmentation of aerial images,” *2020 IEEE Latin American GRSS and ISPRS Remote Sensing Conference, LAGIRS 2020 - Proceedings*, pp. 16–21, 2020, doi: 10.1109/LAGIRS48042.2020.9165597.
- [14] D. C. de Oliveira and M. A. Wehrmeister, “Using deep learning and low-cost RGB and thermal cameras to detect pedestrians in aerial images captured by multicopter UAV,” *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072244.
- [15] H. A. Almurieb and E. S. Bhaya, “Best neural simultaneous approximation,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1584–1590, 2020, doi: 10.11591/ijeecs.v20.i3.pp1584-1590.
- [16] Y. Guan, Y. Han, and S. Liu, “Deep learning approaches for image classification techniques,” *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms, EEBDA 2022*, pp. 1132–1136, 2022, doi: 10.1109/EEBDA53927.2022.9744739.
- [17] C. M. Gevaert, C. Persello, F. Nex, and G. Vosselman, “A deep learning approach to DTM extraction from imagery using rule-based training labels,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 142, pp. 106–123, 2018, doi: 10.1016/j.isprsjprs.2018.06.001.
- [18] A. K. Abbas, S. Hamad, and N. A. Hamad, “Single line to ground fault detection and location in medium voltage distribution system network based on neural network,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 2, p. 621, 2021, doi: 10.11591/ijeecs.v23.i2.pp621-632.
- [19] C. M. Gevaert, C. Persello, R. Sliuzas, and G. Vosselman, “Monitoring household upgrading in unplanned settlements with unmanned aerial vehicles,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 90, 2020, doi: 10.1016/j.jag.2020.102117.
- [20] G. Ghiasi, T. Y. Lin, and Q. V. Le, “NAS-FPN: Learning scalable feature pyramid architecture for object detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 7029–7038, 2019, doi: 10.1109/CVPR.2019.00720.
- [21] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [22] N. Horning, E. Fleishman, P. J. Ersts, F. A. Fogarty, and M. Wohlfeil Zillig, “Mapping of land cover with open-source software and ultra-high-resolution imagery acquired with unmanned aerial vehicles,” *Remote Sensing in Ecology and Conservation*, vol. 6, no. 4, pp. 487–497, 2020, doi: 10.1002/rse2.144.
- [23] T. Onuean, A., Jung, H., & Kim, “Handwriting Thai digit recognition using convolution neural networks,” *Proceedings of the Korean Institute of Information and Commucation Sciences Conference, The Korea Institute of Information and Commucation Engineering*, pp. 15–17, 2021.
- [24] M. D. Hossain and D. Chen, “Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 115–134, 2019, doi: 10.1016/j.isprsjprs.2019.02.009.
- [25] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 5967–5976, 2017, doi: 10.1109/CVPR.2017.632.
- [26] M. M. Jasim, H. K. Al-Qaysi, and Y. Allbadi, “Reliability-based routing metric for UAVs networks,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, pp. 1771–1783, 2021, doi: 10.11591/ijeecs.v21.i3.pp1771-1783.
- [27] G. Jakovljevic, M. Govedarica, F. Alvarez-Taboada, and V. Pajic, “Accuracy assessment of deep learning based classification of LiDAR and UAV points clouds for DTM creation and flood risk mapping,” *Geosciences (Switzerland)*, vol. 9, no. 7, 2019, doi: 10.3390/geosciences9070323.
- [28] M. S. Bin Safaron *et al.*, “Directional cloverleaf antenna for unmanned aerial vehicle (UAV) application,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 2, pp. 773–779, 2019, doi: 10.11591/ijeecs.v14.i2.pp773-779.
- [29] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, “A survey: Deep learning for hyperspectral image classification with few labeled samples,” *Neurocomputing*, vol. 448, pp. 179–204, 2021, doi: 10.1016/j.neucom.2021.03.035.
- [30] N. Thirumaleshwar Hegde, V. I. George, C. G. Nayak, and K. Kumar, “Transition flight modeling and robust control of a VTOL unmanned quad tilt-rotor aerial vehicle,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 3, p. 1252, 2020, doi: 10.11591/ijeecs.v18.i3.pp1252-1261.
- [31] K. Kim and H. S. Lee, “Probabilistic anchor assignment with IoU prediction for object detection,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12370 LNCS, pp. 355–371, 2020, doi: 10.1007/978-3-030-58595-2_22.
- [32] S. K. Debnath *et al.*, “A review on graph search algorithms for optimal energy efficient path planning for an unmanned air vehicle,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 2, pp. 743–750, 2019, doi: 10.11591/ijeecs.v15.i2.pp743-749.
- [33] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, “Panoptic segmentation,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 9396–9405, 2019, doi: 10.1109/CVPR.2019.00963.
- [34] S. Ben Amarar and P. Zong, “3D path planning, routing algorithms and routing protocols for unmanned air vehicles: A review,” *Aircraft Engineering and Aerospace Technology*, vol. 91, no. 9, pp. 1245–1255, 2019, doi: 10.1108/AEAT-01-2019-0023.




- [35] A. Kirillov, Y. Wu, K. He, and R. Girshick, "Pointrend: Image segmentation as rendering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 9796–9805, 2020, doi: 10.1109/CVPR42600.2020.00982.
- [36] Z. Li, J. Zhao, X. Zhou, S. Wei, P. Li, and F. Shuang, "RTSDM: A real-time semantic dense mapping system for UAVs," *Machines*, vol. 10, no. 4, 2022, doi: 10.3390/machines10040285.

BIOGRAPHIES OF AUTHORS






Shouket Abdulrahman Ahmed    is awarded M.Sc. in Electrical System engineering, University Malaysia Perlis (UniMAP), Perlis, Malaysia, (2016), thesis title is "Electrical Intelligence System Based on Microcontroller for Power Station". Currently, His expertise is in the areas of power electronics. He has published more than three journal articles and conferences. After that, he obtained a Bachelor's degree (full time), in electrical and electronics Engineering, Grad 3 from Thames Polytechnic, UK (1989). He can be contacted at email: shouketunimap@gmail.com.



Hazry Desa    received his Bachelor of Mechanical Engineering from Tokushima University, Japan. He was a Senior Design Mechanical Engineer in R&D Sony Technology Malaysia in 1997. In 2001 he was employed by Tamura Electronics Malaysia. He pursued his PhD at the Artificial Life and Robotics Laboratory, Oita University, Japan in 2003. In 2008, he was appointed as the Deputy Dean at the School of Mechatronic in Universiti Malaysia Perlis (UniMAP) and soon after was appointed as the Dean of Center for Communication Skills and Entrepreneurship. He was promoted to Associate Professor in 2010 and Professor in 2016 by UniMAP and consequently appointed as the first dean of the School of Business Innovation and Technopreneurship by UniMAP until Feb 2017. In March 2017 until 2019, he is appointed as the Dean for the Faculty of Engineering Technology. He is also the Director of the Autonomous System and Machine Vision Research Cluster (AutoMAV) since 2009, which then is upgraded to the Center of Excellence for Unmanned Aerial Systems (COEUAS) in 2013 until present. He has supervised a number of postgraduate students in either Master or Doctor of Philosophy levels. His research interests include mobile robot, unmanned systems and aerial robotics. A lot of achievements have been made through the years such as journal publications, patents and prototypes developments in UAS area. He is also registered with Engineering Council United Kingdom as Chartered Engineer (CEng), registered as a Professional Engineer (Ir.) with Board of Engineers Malaysia (BEM), registered as a Professional Technologist (Ts.) with Malaysia Board of Technologist (MBOT), senior member of IEEE and a member of IET. He can be contacted at email: hazry@unimap.edu.my.



Abadal-Salam T. Hussain    is currently a PhD faculty staff, Assistant Professor-Department head of Medical Instrumentations Technique Engineering, Alkitab University. Previously Faculty staff of School of Electrical System Engineering, University Malaysia Perlis (UniMAP), Malaysia (2013-2016). Before that, he was affiliated with the University of Sharjah, UAE (2004-2007) and "Sharjah Institute of Technology (SIT), Sharjah, UAE (2007-2011). Abadal also has a GCE from the University of London & a Bachelor of Engineering from the Royal Naval Engineering College, Manadon, Plymouth, UK in (1989). He also has a Master of Science awarded by the University of Science & Technology Oran Algeria in 2000 and a doctorate (PhD) in Mechatronic Engineering awarded by the University Malaysia Perlis (UniMAP) in 2013. He can be contacted at email: asth2233@gmail.com, asth3@yahoo.com.