A Review on Integrating Photogrammetry and Deep Learning for Semantic Segmentation Applications

Jorge Gálvez [a], Heriberto Zapata [a], Omar Avalos [a], Nayeli Pérez-Padilla [a], Victor García [a]

[a]Universidad de Guadalajara, CUCEI, Guadalajara, Jal, Mexico.
Av. Revolución 1500, Guadalajara, Jal, México
{jorge.galvez, omar.aalvarez, nayeli.perez, victor.ggutierrez}@academicos.udg.mx
heriberto.zapata7939@alumnos.udg.mx

Abstract. The monitoring of large areas is essential to assess their conditions and patterns of change and support decision-making from which different ways are look for ways and optimize these problems and techniques such as photogrammetry are used to assess changes in large areas and, in conjunction with semantic segmentation, are important to extract the necessary information and to provide classification analysis at the pixel level for an accurate assessment of irregular shapes The reference ground sample distance (GSD) method translates pixels into a unit of measurement, facilitating accurate image scale calculations and precise measurements. This method explores these strategies using deep learning, potentially improving accuracy and simplifying segmentation. It also looks at different works that talk about how segmentation and photogrammetry can be used in different areas, including their methods and results, to find the best ways to do things, adapt, and lead to new developments in the field.

Keywords: Semantic segmentation, Deep Learning, Photogrammetry, Ground Sample Distance.

1 Introduction.

Analyzing the conditions of large areas of buildings, forests, and vegetation is one of the most significant challenges due to the stress they suffer from the consequences of climate change [1]. Therefore, it is essential to constantly record the measurements of the areas to understand their situation and patterns of change and support decision-making regarding their care [2].

Obtain accurate measurements of objects, plants, and cities when looking at large, hard-to-reach areas, especially when government-protected areas and regulatory requirements are involved. Different techniques are used, from traditional methods to modern remote sensing tools. For example, aerial photogrammetry and satellite remote sensing are used to measure changes over large areas. Landsat and Sentinel are two tools that collect data on large areas of vegetation. Light Detection and Ranging (LIDAR)

https://doi.org/10.61728/AE20254926

technology is one of the most accurate tools for estimating biomass and forest density [3]. Geographic Information Systems (GIS) allow the integration of geospatial data obtained from different sources, facilitating the delineation and analysis of large land areas using software[4].

There are many works proposed using semantic segmentation and photogrammetry in different areas that talk about the improvement of mapping in segmentation, using UAV photogrammetry data (RGB, DSM, VDVI) focused on building and propose the combination of these methods to obtain very accurate results [5], methods that focus on improving the extraction of geometric information [6]; and these studies demonstrate the effectiveness of integrating semantic segmentation with photogrammetry for various applications, including building care, image analysis of historic building facades, and applied in architecture. [7] In addition, different applications of photogrammetry for environmental investigations and the impact that can be obtained by combining LiDAR photogrammetry and improved spatial resolution and accuracy and the possible investigations for understanding ecological processes and hazard prevention [8]. However, some of these techniques have disadvantages, such as their cost due to sensors and cameras [9].

The main objective is to understand how these techniques can be used in conjunction with deep learning and other methods, focusing on the advantages of their use, the methodology used, and proposals, as well as how it has improved computational cost and accuracy [10] . But first, we will analyze why to apply these semantic segmentation and photogrammetry techniques.

2 Semantic segmentation and photogrammetry

Currently, semantic segmentation is essential for understanding image details and provides pixel-level classification by assigning a label (e.g., “tree,” “object,” “dog”). This technique helps us to interpret the data, understand how to represent it, correctly differentiate objects of various classes, and work with the information within the images to ensure consistency between pixel labels [11]. This segmentation technique is commonly used in computer vision and aerial image analysis, as it allows for accurate analysis of images containing irregular shapes, something that other segmentation techniques have more difficulty with. However, one of the challenges it faces is recognizing the characteristics of different factors, such as image scenes, lighting conditions, and data quality.

These semantic segmentation and photogrammetry techniques information to be extracted from photographs or digital images. In other words, it is used to determine the shape, dimensions, and position of objects and terrain by analyzing photographs. It is also used in various fields because of its accuracy and speed, considering the problems where it is applied for disaster prevention and management and used for geography, landscape description, coastal studies, archaeology, etc. [12], driven by advances in computer vision and image processing data, provides a cost-effective solution for accurate ecosystem monitoring and land use planning.

With the mentioned, it is necessary to extract the information, as Linder says, considering some geometrical principles: camera, focal length, height mentioned in Fig.1 this relationship is used to obtain equation 1, corresponding ground sample distance (GSD), which allows them to convert pixels to any unit of measurement, thus applying the techniques previously exposed. In this case, from the decimal metric system, the centimeter-pixel ratio is applied, which represents the actual distance on the ground corresponding to the width or length of a pixel in the digital image, allowing accurate calculation of the image scale and reliable measurements. This technique depends on several parameters. The size of the pixel in the image and the characteristics of the camera, such as the resolution, the focal length defined by the camera sensor according to the manufacturer, and the flight altitude defined in the flight plan, are related to ensure the accuracy of the GSD on the ground [13].

In this case, we will focus on Ground Sample Distance (GSD), which allows the conversion of pixels to any unit of measurement, thus applying the techniques described above. In this case, the centimeter-to-pixel ratio is applied using a photogrammetric technique called Ground Sample Distance (GSD). This technique represents the actual distance on the ground corresponding to the width or length of a pixel in the digital image, allowing the image scale to be accurately calculated and reliable measurements to be made. The calculation of the GSD depends on several parameters. The GSD relates the pixel size in the image to camera characteristics such as resolution and focal length as defined by the camera sensor according to the manufacturer, and flight altitude as defined in the flight plan to ensure accuracy on the ground.

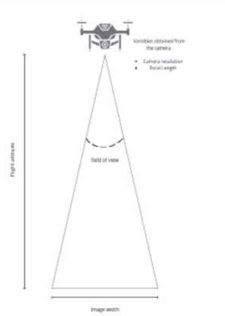

Fig.1 Representation of variables to obtain the GSD

$$GSD = \frac{Pixel\ size \times Altitude}{Focal\ Lenght} \quad (1)$$

It should be noted that the maximum possible GSD for aerial surveys depends on several factors, such as camera exposure interval, blur due to image motion, and altitude. The higher the altitude, the lower the accuracy. It also depends on what the flight is being conducted with. For example, analysis shows that a minimum GSD of 0.86 m is required to detect most aircraft accurately, and that performance decreases at coarser resolutions, indicating that finer GSDs improve the detection of smaller targets [15]

In the next section, show some work that proposes the use of segmentation and photogrammetry in different areas of work and its application, while highlighting the importance of the ground sample distance for the dataset, the methodology, and the characteristics that must meet the images used in the proposed work. It also discusses how these techniques are used to optimize and combine these techniques in different areas in conjunction with deep learning, which facilitates the segmentation[16] and has also improved the accuracy [17].

3 Geospatial mapping and analysis

3.1 Guidelines to Compare Semantic Segmentation Maps at Different Resolutions

The use of GSD and semantic segmentation to evaluate the accuracy of models at different spatial resolutions Ayala (2024)[18] proposes to compare maps by integrating pixel-wise metrics at both scene and region levels. The objective is to establish an evaluation process that includes different spatial resolutions and enhances comprehension of the different GSD values applied to building and pool detection problems. This shows that it is possible to optimize the balance between the computational resources and the accuracy of the results.

3.2 Generalized Sparse Convolutional Neural Networks for Semantic Segmentation of Point Clouds Derived from Tri-Stereo Satellite Imagery

Bachhofner (2020)[19] develops a neural network architecture based on U-Net that emphasizes applying point clouds derived from 0.7 m resolution (GSD) images obtained from photogrammetry to approach semantic segmentation using deep learning techniques. The methodology consists of the generation of digital surface models (DSM) and the application of experiments that adjust, median weighting of classes to mitigate information imbalance. Ultimately, the research results indicate that color aggregation optimizes class performance and increases the overall accuracy of the sparse convolutional network. The best performance was achieved with FCN-8s network, and for road classification, a decision tree achieved slightly better accuracy.

4 Environment

4.1 VEPL Dataset: A Vegetation Encroachment in Power Line Corridors Dataset for Semantic Segmentation of Drone Aerial Orthomosaics

Cano-Solis 2023 [20] proposes a method based on the combination of UAV imagery and deep learning for semantic segmentation of vegetation to automate the detection and facilitate the maintenance of power lines, using photogrammetry to generate high-resolution orthomosaics with a Ground Sample Distance (GSD) between 0.9 and 1.3 cm, and methods to increase the number of available examples performed by geometric and spectral augmentation on the original dataset. As a result, this approach facilitates the accurate identification of vegetation in power corridors, allowing for more economical and near real-time monitoring compared to techniques relying on satellite imagery or LiDAR.

4.2 Deep Learning Application for Fracture Segmentation Over Outcrop Images From UAV-Based Digital Photogrammetry

Fractures affect properties of reservoir geologies, so their characterization is crucial for fluid flow modeling in 2021 Junior A[21] proposed that used images captured with drones and processed with digital photogrammetry, the automatic and accurate detection of fracture networks and validate two popular convolutional neural networks (CNNs), SegNet and U-Net, employed consist of image processing with segmentation techniques to identify fractures in outcrops and The results obtained show that both models achieved acceptable values presenting the U-Net presented closer segmented images than the SegNet model, with reduced learning time.

5 Town planning

5.1 Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks.

In the study in 2023, Hong D [8] proposes a method based on HighDAN, which aims to improve the generalization capability in semantic segmentation applied to cities. Of the main limitations are the traditional methods, which are usually designed for specific cases of cities and, therefore, have difficulties being applied in multiple regions. The methodology of the study is based on the use of hyperspectral, multispectral, and SAR images with a GSD of 10, allowing a better discrimination of classes. The results obtained in extensive experiments on C2Seg show that HighDAN outperforms other methods and demonstrate that it can be applied in different regions.

5.2 HAGDAVS: Height-Augmented Geo-Located Dataset for Detection and Semantic Segmentation of Vehicles in Drone Aerial Orthomosaics

This paper explains that semantic segmentation vehicle detection has applications in security, traffic and parking management, urban planning, logistics and transportation, among others. Ballesteros (2022) [22] proposes a method that integrates RGB spectral channel data with the Digital Surface Model (DSM) to obtain an accurate representation of the scene. The main idea is to improve the accuracy of deep learning models by incorporating additional height variables and Ground Sample Distance (GSD) information, which allows a more accurate differentiation between classes, and the methodology developed includes image acquisition and the generation of photogrammetric products. In addition, neural networks are used, and parameters are adjusted to optimize semantic segmentation and vehicle detection. The results of the study show that the fusion of RG-DSM data and the application of deep learning techniques allow for overcoming challenges such as occlusion and the presence of noise, providing a robust solution for vehicle detection, classification, and geo-localization.

5.3 Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic

In this work, Park (2022) [23]focuses on transportation infrastructure management and the objective is to eliminate vehicle interference in aerial images. A method based on photogrammetric UAVs is used, using images with a GSD of 0.94 cm/pixel, where deep learning and semantic segmentation techniques are applied to improve the quality of the orthomosaics. The objective is to remove these vehicles from each image individually, using segmentation with the Mask R-CNN model, and then reconstruct the removed areas. The methodology consists of preprocessing the images to avoid the appearance of undesired pixels, ensuring a correct joining of the photographs in the rest of the mosaic.

6 Civil Engineering and Architecture

6.1 Semi-autonomous inspection for concrete structures using digital models and a hybrid approach based on deep learning and photogrammetry

Mirzazade (2023) [24] presents a methodology that uses photogrammetry and deep learning to evaluate deterioration in concrete buildings, including bridges and tunnel linings. The main objective is to utilize photogrammetry to compute essential parameters such as Ground Sampling Distance (GSD) and Field of View (FOV), optimizing image resolution at varying distances. Additionally, convolutional neural networks (CNN) are employed for semantic segmentation and crack detection in the examined surfaces. The results show that this hybrid approach can assist human inspectors in detecting, locating, and quantifying damage in concrete infrastructure.

6.2 Semantic segmentation and photogrammetry of crowdsourced images to monitor historic facades

Lui Z 2022 [25] proposed that combining semantic segmentation and photogrammetry of crowdsourced images can be used to remotely monitor and measure changes in historic building facades, thus reducing the human and financial resources dedicated to on-site inspections comparing neural network architectures DeepLab family and DeeplabV3 integrating photogrammetry to build a model, allowing to measure the expansion of biological growth and experiments show that the selected model achieves an IoU of 66. 9% in the segmentation of seven classes, with specific values of 54.6% for vegetation in the one-stage model and 56.2% in the two-stage model and demonstrates that the combination of photogrammetry and semantic segmentation to reduce distortions in the images but with a high computational cost, especially in the optimization of the model and integration with photogrammetry.

6.3 Towards Mesh-Based Deep Learning for Semantic Segmentation in Photogrammetry

In this work, Knott (2021)[6] applies MeshCNN, a deep learning model specifically designed for 3D triangular meshes, to the field of photogrammetry and incorporates photometric features, and the experiments performed show that the inclusion of color information significantly improves the model performance in segmentation tasks. Finally, considering the computational cost and complexity required to optimize the model configuration, performance could be significantly improved by introducing a high-quality benchmark.

6.4 Photogrammetric Techniques and Image Segmentation via Machine Learning as Supporting Tools in Paving Asphalt Mixtures Studies

This work focuses on the study of the formulation asphalt mixtures for the construction of roads in Piemonte (2022) [26], proposing a method based on photogrammetry to obtain information, taking advantage of photogrammetric techniques to analyze displacements and textures in asphalt samples, guaranteeing a GSD of 0.1 mm, and using Machine Learning for automatic segmentation of images, achieving less than 2% error in the classification of the classes of interest. It is important to note that although the use of ML algorithms for semantic segmentation is mentioned.

7 Conclusions

As previously mentioned, it is evident that numerous applications currently exist in different areas, each utilizing photogrammetry and semantic segmentation while proposing distinct methodologies and avenues for enhancement. Some implicitly indicate that the application of Ground Sample Distance (GSD) to the dataset is a critical variable for assessing the value of a pixel within the image and the precision of object representation. One of the primary obstacles in implementing photogrammetric approaches is the necessity for specialized sensors, which may require distinct instruments for each task, such as customized sensors or specialist cameras for projects. Should the goal be continuously monitored, depending solely on these methods for measurements may prove intricate and expensive. However, if the goal is to compute regions within images, a standard camera can be used by translating pixels into a specific unit of measurement. Using these strategies and considering the camera's characteristics, this methodology can be applied to different contexts, ensuring its relevance across different areas, identifying best practices, finding innovative solutions to their needs, and inspiring new advances in the different areas.

References

[1] I. P. on C. C. (IPCC), "Changing State of the Climate System," *Climate Change 2021 – The Physical Science Basis*, pp. 287–422, Jun. 2023, doi: 10.1017/9781009157896.004.

[2] S. Fritz *et al.*, "A comparison of global agricultural monitoring systems and current gaps," *Agric Syst*, vol. 168, pp. 258–272, Jan. 2019, doi: 10.1016/J.AGSY.2018.05.010.

[3] K. Wołk and M. S. Tatara, "A Review of Semantic Segmentation and Instance Segmentation Techniques in Forestry Using LiDAR and Imagery Data," *Electronics (Basel)*, vol. 13, no. 20, p. 4139, Oct. 2024, doi: 10.3390/electronics13204139.

[4] S. Laudares, J. B. Laudares, and M. P. Libório, "Geographic Information Systems Used as a Practical Way of Teaching Mathematics," *Journal of Geographic Information System*, vol. 08, no. 05, pp. 608–617, 2016, doi: 10.4236/jgis.2016.85050.

[5] W. Boonpook, Y. Tan, and B. Xu, "Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry," *Int J Remote Sens*, vol. 42, no. 1, pp. 1–19, Jan. 2021, doi: 10.1080/01431161.2020.1788742.

[6] M. Knott and R. Groenendijk, "TOWARDS MESH-BASED DEEP LEARNING for SEMANTIC SEGMENTATION in PHOTOGRAMMETRY," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Copernicus GmbH, Jun. 2021, pp. 59–66. doi: 10.5194/isprs-annals-V-2-2021-59-2021.

[7] A. Song and Y. Kim, "Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote

sensing potsdam and cityscape datasets," *ISPRS Int J Geoinf*, vol. 9, no. 10, Oct. 2020, doi: 10.3390/ijgi9100601.

[8] M. Schirrmann, A. Hamdorf, A. Garz, A. Ustyuzhanin, and K. H. Dammer, "Estimating wheat biomass by combining image clustering with crop height," *Comput Electron Agric*, vol. 121, pp. 374–384, Feb. 2016, doi: 10.1016/j.compag.2016.01.007.

[9] A. Faridhouseini, A. Mianabadi, M. Bannayan, and A. Alizadeh, "Lidar remote sensing for forestry and terrestrial applications," *International Journal of Applied Environmental Sciences*, vol. 6, pp. 103–118, Jan. 2011.

[10] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *Int J Multimed Inf Retr*, vol. 7, no. 2, pp. 87–93, Jun. 2018, doi: 10.1007/s13735-017-0141-z.

[11] S. Hao, Y. Zhou, and Y. Guo, "A Brief Survey on Semantic Segmentation with Deep Learning," *Neurocomputing*, vol. 406, pp. 302–321, Sep. 2020, doi: 10.1016/j.neucom.2019.11.118.

[12] W. Linder, *Digital photogrammetry: A practical course, fourth edition*, 4th ed. Berlin/Heidelberg, Germany: Springer Berlin Heidelberg, 2016. doi: 10.1007/978-3-662-50463-5.

[13] B. Felipe-García, D. Hernández-López, and J. L. Lerma, "Analysis of the ground sample distance on large photogrammetric surveys", doi: 10.1007/s12518-012-0084-2.

[14] A. S. Peng, A. Turkmen, B. Eickhoff, M. Finta, and P. Gerads, "Design of a Ground Sampling Distance Graphical User Interface for an Unmanned Aerial Vehicle System," in *2019 53rd Annual Conference on Information Sciences and Systems, CISS 2019*, Institute of Electrical and Electronics Engineers Inc., Apr. 2019. doi: 10.1109/CISS.2019.8692786.

[15] M. Ciolino and W. Maddox, "From Pixels to Planes: Minimum Ground Sample Distance for Aircraft," *ArXiv*, vol. abs/2412.02137, 2024, doi: https://doi.org/10.48550/arXiv.2412.02137.

[16] J. M. Junior *et al.*, "Integration Of Photogrammetry and Deep Learning In Earth Observation Applications," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 574–577. doi: 10.1109/IGARSS47720.2021.9554811.

[17] B. Neupane, T. Horanont, and J. Aryal, "Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis," Feb. 02, 2021, *MDPI AG*. doi: 10.3390/rs13040808.

[18] C. Ayala, C. Aranda, and M. Galar, "Guidelines to Compare Semantic Segmentation Maps at Different Resolutions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024, doi: 10.1109/TGRS.2024.3369310.

[19] S. Bachhofner *et al.*, "Generalized sparse convolutional neural networks for semantic segmentation of point clouds derived from tri-stereo satellite imagery," *Remote Sens (Basel)*, vol. 12, no. 8, Apr. 2020, doi: 10.3390/RS12081289.

[20] M. Cano-Solis, J. R. Ballesteros, and J. W. Branch-Bedoya, "VEPL Dataset: A Vegetation Encroachment in Power Line Corridors Dataset for Semantic

Segmentation of Drone Aerial Orthomosaics," *Data (Basel)*, vol. 8, no. 8, Aug. 2023, doi: 10.3390/data8080128.

[21] A. M. Junior *et al.*, "DEEP LEARNING APPLICATION FOR FRACTURE SEGMENTATION OVER OUTCROP IMAGES FROM UAV-BASED DIGITAL PHOTOGRAMMETRY," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 4692–4695. doi: 10.1109/IGARSS47720.2021.9553232.

[22] J. R. Ballesteros, G. Sanchez-Torres, and J. W. Branch-Bedoya, "HAGDAVS: Height-Augmented Geo-Located Dataset for Detection and Semantic Segmentation of Vehicles in Drone Aerial Orthomosaics," *Data Descriptor*, vol. 7, no. 4, Apr. 2022, doi: 10.3390/data7040050.

[23] J. Park, Y. K. Cho, and S. Kim, "Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic," *International Journal of Applied Earth Observation and Geoinformation*, vol. 115, Dec. 2022, doi: 10.1016/j.jag.2022.103111.

[24] A. Mirzazade, C. Popescu, J. Gonzalez-Libreros, T. Blanksvärd, B. Täljsten, and G. Sas, "Semi-autonomous inspection for concrete structures using digital models and a hybrid approach based on deep learning and photogrammetry," *J Civ Struct Health Monit*, vol. 13, no. 8, pp. 1633–1652, Dec. 2023, doi: 10.1007/s13349-023-00680-x.

[25] Z. Liu *et al.*, "Semantic segmentation and photogrammetry of crowdsourced images to monitor historic facades," *Herit Sci*, vol. 10, no. 1, Dec. 2022, doi: 10.1186/s40494-022-00664-y.