Evaluation of Depth Anything Models for Satellite-Derived Bathymetry

Esra Günaydın [1*], İrem Yakar [1], Tolga Bakırman [2], Mahmut Oğuz Selbesoğlu [1]

[1] ITU, Department of Geomatics Engineering, 34469 Maslak Istanbul, Türkiye - esgunaydin, yakari, selbesoglu@itu.edu.tr
[2] YTU, Department of Geomatics Engineering, 34220 Esenler Istanbul, Türkiye - bakirman@yildiz.edu.tr

Keywords: Satellite Imagery, Deep Learning, Bathymetry, Monocular Depth Estimation, Depth Anything

Abstract

The emergence of foundation models has driven major advancements in computer vision and natural language processing, primarily due to their strong zero-shot and few-shot capabilities powered by large-scale, diverse datasets. While earlier approaches used supervised datasets, their limited scene diversity did not perform well in unseen environments. To overcome these limitations, recent works have leveraged unlabeled monocular images, which can be automatically labeled using pre-trained models. One model can be shown as Depth Anything, which demonstrated robust zero-shot performance across diverse scenarios, with Depth Anything V2 further improving accuracy. In this study, the performance of Depth Anything V1 and V2 models was evaluated in satellite-derived bathymetry using Sentinel 2 satellite imagery. The accuracy of these predicted depth maps was evaluated by comparing them with bathymetric data obtained from the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information (NCEI) as the ground truth. The results show that the correlation between Depth Anything V1 predictions and NOAA NCEI data was 56.69%, while the correlation for Depth Anything V2 reached 84.54%. The predicted depth maps were also scaled to obtain Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The RMSE and MAE values for Depth Anything V1 are 0.4135 m and 0.34 m, respectively, while the RMSE and MAE values for V2 are 0.2681 m and 0.2089 m, respectively. This improvement shows the capability of Depth Anything V2 in estimating underwater terrain from monocular satellite imagery, which also demonstrates its potential for cost-effective bathymetric mapping in remote sensing applications. In addition to deep learning-based approaches applied in the test area, a satellite-derived depth map was also generated using the classical band ratio method. Compared with reference bathymetric data, the correlation coefficient, RMSE, and MAE were found to be 38.20%, 0.4639m, and 0.3746m, respectively.

1. Introduction

The bathymetry is essential in application areas such as marine navigation, harbors, submarine pipelines, and cable laying. Bathymetry surveys generally use traditional methods such as the acoustic reflection measurement technique for data collection. Sensor technology has advanced further with developments such as depth profilers, current profilers, bio-optical sensors, and increased data collection accuracy (Ashpaq, 2021). However, the high cost and time-consuming nature of these traditional methods has increased the tendency towards alternative data sources in recent years; in this context, satellite-derived bathymetry (SDB) methods stand out as a remarkable solution. There are different SDB studies conducted in the literature. Kim et al. (2024) applied physics-based algorithms to overcome the limitation that only a few known values are available from a multispectral sensor. The physics-based algorithm is applied to Landsat Operational Land Imager (OLI) images in shallow coastal areas of Guam, Key West, and Puerto Rico. SDB depths are compared to airborne lidar depths, yielding a Root Mean Square Error (RMSE), typically less than 2 m at depths down to 30 m.

Uzakara et al. (2024) aimed to extract depth information using freely accessible Sentinel-2 multispectral images. Marine topography was determined using regression analysis with samples taken from reference data. Reference data were adjusted with changes in coastlines, because coastal changes were used as a parameter for these modifications. Jiang and Rutherfurd (2024) emphasized that bathymetry changes in large rivers are useful for flood planning, navigation, and channel change prediction. This study investigated whether historical bathymetry data of large rivers can be recovered with easily collected free satellite images and limited cross-sectional surveys usually available in less affluent countries. The method was tested on the heavily regulated Han River in central China. The depth derived from Landsat-8 and Sentinel-2 images could accurately match in situ depths in cross-sections to depths of 13.9 m, which is possible in areas where suspended sediment density is low due to dam regulation. Figliomeni and Parente (2024) investigated more classical methods and compared three different methods to estimate shallow water depth from Sentinel-2 satellite images in the Gulf of Pozzuoli, Italy. The evaluated methods were: Band Ratio Method (BRM), Third Degree Polynomial Method (3DPM), Principal Component Analysis Method (PCAM). Kalkan et al. (2021) aimed to investigate the effectiveness of k-Nearest Neighbor regression to derive bathymetry from Landsat 8 Operational Land Imager (OLI) data.

On the other hand, deep learning (DL)-based techniques have started to be widely used in SDB applications. Recently, the main reason for the widespread use of deep learning methods in the fields of earth observation and remote sensing is the capacity of Convolutional Neural Networks (CNNs) technology to successfully perform satellite image processing and feature analysis (Al Najar et al., 2021). DL is a field of machine learning algorithms that has seen great development in the last decade and has attracted attention by exhibiting strong capabilities in different domains (Goodfellow et al. 2016). One of the most successful applications of DL so far has been in computer vision applications in areas such as image classification, processing, and generation (Simonyan and Zisserman 2014). A natural extension of this application includes the use of DL in remote sensing through automatic processing of satellite images. Many recent studies have used DL methods in the field of satellite image classification or segmentation for different applications (Liu et al. 2017; Iglovikov et al. 2017). These approaches usually use deep learning to identify features in satellite images (Al Najar et al., 2021). Sonogashira et al. (2020) proposed a deep learning-based super-resolution method to accelerate the detailed mapping of the ocean floor. Obtaining high-resolution bathymetric maps is time-consuming and costly, as it normally requires depth measurements at many points on the sea surface. To overcome this problem, gridded bathymetric data is treated as digital images, and an image processing-based superresolution technique is applied to produce high-resolution maps from low-resolution data. The model is trained with deep learning architecture to automatically learn the geometric properties of the ocean floor. Al Najar et al. (2021) developed and compared two innovative deep learning based methods for bathymetry estimation which is critical for coastal development and risk management. Due to the cost and complexity of traditional echo-sounding methods, bathymetry estimation was performed using wave signatures from satellite images and physical inversion models. It was shown that the deep learning approaches presented in the study can estimate ocean depths with high accuracy in different simulation scenarios, which stands out as a new, effective and applicable method for bathymetry estimation.

In recent years, transformer-based deep learning models have also been used in SDB studies. Lv et al. (2025) introduces a new deep learning model called BathyFormer, which was developed to make bathymetric mapping in shallow waters near the shore faster and more cost-effective. This model, which is based on visual transformers and encoders, estimates water depths from high-resolution multispectral satellite images. Bathymetric data from CUDEM (Continuously Updated Digital Elevation Model) was used as training data. The model produced highly successful estimates with RMSEs of 0.55–0.73 m at depths of 2–5 m by analyzing spectral signatures and spatial patterns in satellite images. The system, which was tested in non-training areas in the Chesapeake Bay, offers a scalable and economical solution for areas such as coastal management, environmental monitoring and maritime. In addition, the advent of foundation models is highly impactful on the fields of natural language processing and computer vision, which leads to a significant transformation in these areas (Bommasani et al., 2021). Remarkable zero-shot and few-shot capabilities across a wide range of tasks were achieved (Touvron et al., 2023; Radford et al., 2021). However, in the domain of Monocular Depth Estimation (MDE), a critical task for applications such as autonomous driving, robotics (Wofk et al., 2019), and virtual or augmented reality, such foundation models have not yet been fully realized. This is primarily due to the immense challenge of collecting and annotating datasets with millions of depth-labeled images. On the other hand, the depth prediction using a single image was tried in earlier methods, such as MiDaS, using various labeled datasets. However, those models did not perform well in new or unfamiliar situations. The main reason for this situation is the scarcity of labeled datasets in different scenes. Therefore, the models could not generalize well to images very different from the training examples (Yang et. al., 2024). Yang et al. (2024) demonstrated that leveraging unlabeled data and semantic priors enhances monocular depth estimation, which is Depth Anything V1. Subsequently, the improved version, Depth Anything V2, built through architectural enhancements and large-scale data utilization, was officially released on June 14, 2024. It demonstrates substantial improvements over Depth Anything V1 in terms of fine-detail preservation and robustness. Compared to Stable Diffusion-based approaches, it benefits from faster inference speed, a smaller parameter footprint, and greater depth accuracy (Yang et al., 2024).

While these models have typically been applied to close-range imagery, the primary objective of this study is to explore their effectiveness in SDB from optical satellite images, thereby extending their applicability to large-scale remote sensing scenarios. In this respect, Depth Anything V1 and V2 models were evaluated in terms of SDB. In this context, depth maps were created using these two versions of the model from the Sentinel-2 satellite imagery. Subsequently, the Bathymetric Attributed Grid (BAG) Mosaic dataset by the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information (NCEI) was used for the validation of the predicted depths.

2. Materials & Methods

2.1 Depth Anything Models

Depth Anything is a transformer-based monocular depth estimation model trained on a large-scale dataset comprising 1.5 million labeled and 62 million unlabeled images. To support cross-dataset learning where scale and shift may vary, the model employs an affine-invariant loss function, defined as:

$$L_l = \frac{1}{HW}\sum_{i=1}^{HW} \rho(d_i^*, d_i) \tag{1}$$

Here is d_i^* and d_i refer to ground truth and predicted depths, respectively. Additionally, $\rho(d_i^*, d_i)$ represents the affine-invariant mean absolute error loss, which measures the absolute difference between the scaled and shifted versions of the prediction and the ground truth (Yang et al., 2024). The training pipeline of Depth Anything V2 consists of three main stages. First, a teacher model based on DINOv2-G is trained exclusively on high-quality synthetic images. Second, this model is used to generate accurate pseudo-depth labels for large-scale, unlabeled real-world images. Finally, student models with various DINOv2 backbones (small, base, large, and giant) are trained on these pseudo-labeled real images to achieve robust generalization. Compared to V1, it produces more detailed and robust predictions, supported by large-scale pseudo-labeled real images. Offered in various parameter scales, the models outperform Stable Diffusion-based approaches in both accuracy and inference speed when fine-tuned for metric depth estimation (Yang et al., 2024).

2.2 Band Ratio Method

The Band Ratio Method (BRM) allows depth estimation by comparing the reflectance ratios between specific spectral bands with in-situ measurements. This method focuses on the blue and green bands, especially by taking advantage of the intrinsic optical properties of water. In clear waters, the reflectance intensity of the blue band tends to be higher than the green band. In coastal or turbid waters, the green band exhibits higher reflectance values due to the density of organic and inorganic substances in the water. These spectral differences provide important clues in terms of revealing bathymetric changes in the study area (Figliomeni and Parente, 2023). In this context, the Z (depth) value can be calculated using a regression formula based on band ratios (Stumpf et al., 2003):

$$Z = m_1 \frac{\ln(n*\rho_w(\lambda_i))}{\ln(n*\rho_w(\lambda_j))} - m_0 \tag{2}$$

Here,

- The coefficient m_1 is the constant used to scale the depth ratio.
- n is a value that is kept constant throughout the study area to ensure that the logarithm remains positive.
- m_0 is the depth offset representing the point $Z = 0$.
- ρ_w represents the magnitude of the radiation (reflectance value) reflected from the water surface.
- λ_i and λ_j represent two different spectral bands.

3. Case Study

In this study, Sentinel-2 bands 4, 3, and 2 were specifically utilized from an image of a selected test area in Tampa Bay. These bands correspond to red, green, and blue wavelengths and were chosen to generate true-color images. The spatial resolution of Sentinel-2 images is 10 m, which provides sufficient detail to generate depth maps of the study area. The satellite image selection was carried out between January 2025 and March 2025 with a cloud coverage below approximately 10%. The images were mosaiced, and a monocular image of the study area was obtained. Sentinel-2 images from the "COPERNICUS/S2_SR_HARMONIZED" collection in Google Earth Engine were utilized, where Level-2A surface reflectance data had been atmospherically corrected using the Sen2Cor algorithm. The Depth Anything V1 and V2 models were utilized to derive a depth map from this image. The generated depth maps range between 0-1 normalized values. Indeed, the high-resolution BAG Mosaic dataset provided by NOAA NCEI (NCEI Bathymetric Data Viewer, 2025), with an approximate cell size of 10 meters, was used as the reference bathymetry data for validation purposes. The depth maps obtained from Sentinel-2 satellite images were compared with this reference bathymetry data, and their accuracy was tested. NOAA BAG Mosaic dataset is a known source with its high resolution and accurate data, and therefore was selected as a suitable "ground truth" data to evaluate the reliability of the depth maps used in the study. The accuracy of the bathymetry data was used to show how close the depth maps derived from satellite images are to the ground truth, and the performance of each model was presented.

The cell size was selected as 1/3 arc-second (approximately 10 meters) in the BAG Mosaic dataset provided by NOAA. The aim was to be compatible with the spatial resolution of Sentinel-2 satellite images. Selecting the same resolution (10 meters) for NOAA data is important to increase the accuracy of depth maps and ensure comparability of both datasets, since the spatial resolution of Sentinel-2 images is 10 meters. This similar spatial resolution enables the verification and integration processes between satellite data and bathymetry data.

The selected study area lies off the coast of Tampa Bay, Florida (27.60°N, 82.66°W), where high-resolution NOAA BAG Mosaic data were available for bathymetric validation. Figure 1 (a) represents the study area, 1 (b) represents the bathymetric data, and 1 (c) shows the Sentinel 2 imagery used in the study.

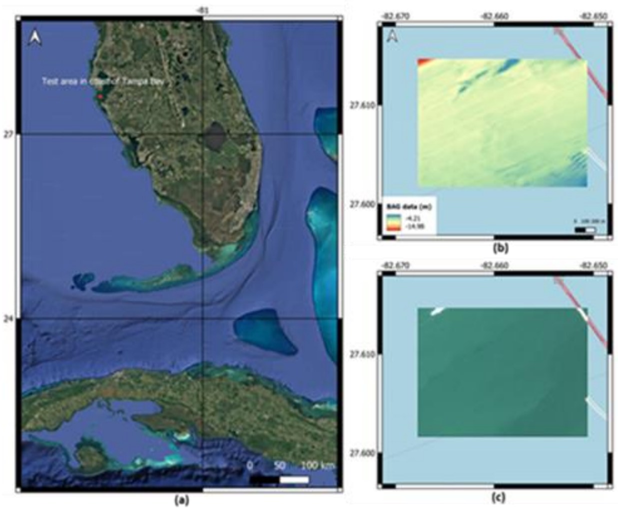

Figure 1. (a) Study area, (b) Bathymetric data, (c) Sentinel-2 imagery

The workflow of the study is shown in Figure 2.

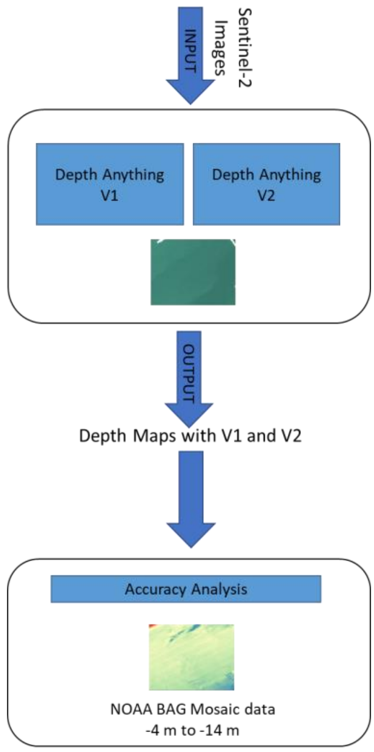

Figure 2. The workflow of the study

4. Results

The Pearson correlation coefficient (R) was calculated according to ground truth bathymetric data for Depth Anything V1 and V2. Regression analysis was used to examine the relationship between bathymetry data (ground truth) and model output and to evaluate the accuracy of the model's predictions. In this analysis, bathymetry data is taken as the independent variable, while the depth values predicted by the model are considered as the dependent variable. Regression analysis is used to determine the linear (or non-linear) relationship between the data sets. In this way, it is revealed how the values predicted by the model relate to the actual depths. The regression coefficients obtained by regression analysis show whether the model made the correct prediction and to what extent it is on the correct scale.

The relationship between Depth Anything V1 and bathymetry is modeled by the following linear regression function:

$$1.953 * \text{Model Output} - 8.932 \tag{3}$$

While the relationship between Depth Anything V1 and bathymetry data is defined by:

$$3.509 * \text{Model Output} - 10.179 \tag{4}$$

The results are presented in Figure 3 (a) and (b) for Depth Anything V1 and V2, respectively.

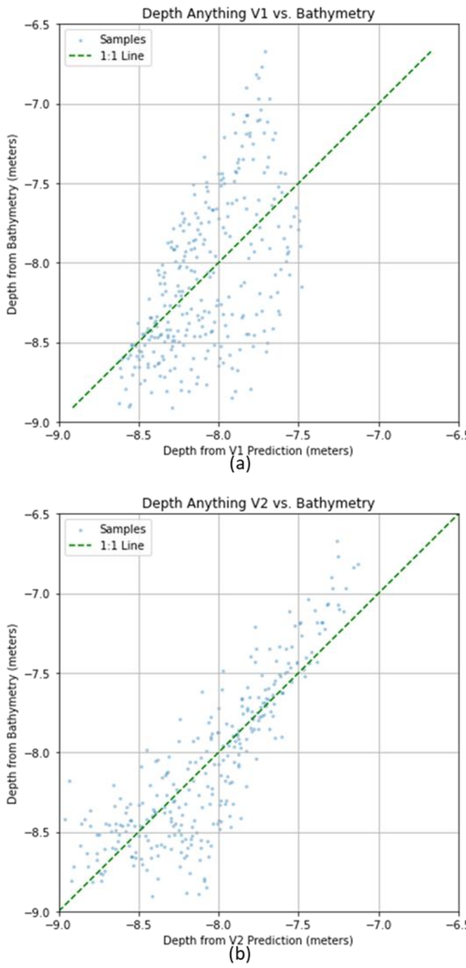

Figure 3. Scatter plots of bathymetric data and (a) Depth Anything V1 prediction, (b) Depth Anything V2 prediction

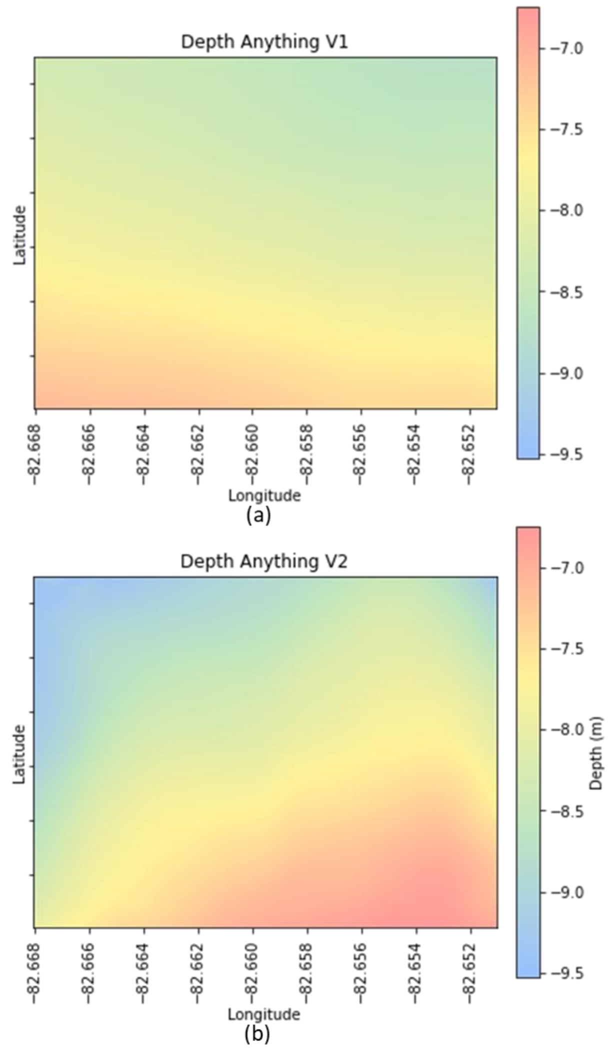

Figure 4. The depth maps generated with (a) Depth Anything V1 and (b) Depth Anything V2

The accuracy analysis was conducted at depths ranging from approximately 7-8 meters, in a region located in the central part of the study area, which covers approximately 0.7 km². The correlation between Depth Anything V1 and bathymetric data was found to be 56.69%, while the correlation between V2 and bathymetric data was found to be 84.54%. The predicted depth maps were also scaled to obtain RMSE and Mean Absolute Error (MAE). The RMSE and MAE values for Depth Anything V1 are 0.4135 m and 0.3400 m, respectively, while the RMSE and MAE values for V2 are 0.2681 m and 0.2089 m, respectively. This improvement shows the capability of Depth Anything V2 in estimating underwater terrain from monocular satellite imagery, which also demonstrates its potential for cost-effective bathymetric mapping in remote sensing applications. The depth maps generated with Depth Anything V1 and V2 are shown in Figure 4, respectively.

The first depth map (Depth Anything V1) shows the depth changes in large areas with smooth and continuous transitions. This map is very useful for understanding general trends and large-scale topographic structure. This indicates that the model estimates a surface with less detail but more stable and continuous. The second depth map (Depth Anything V2) contains much higher frequency details and reveals small changes in the surface more clearly. This map provides a more detailed view by emphasizing fine structural differences and local depth variations.

In addition to deep learning-based approaches applied in the test area, a satellite-derived depth map was also generated using the classical band ratio method. Compared with reference bathymetric data, the correlation coefficient, RMSE, and MAE were found to be 38.20%, 0.4639 m, and 0.3746 m, respectively. Table 1 presents the statistical evaluation results for three different methods.

Method	R	RMSE (m)	MAE (m)
Band Ratio	38.20%	0.4639	0.3746
Depth Anything V1	56.69%	0.4135	0.3400
Depth Anything V2	84.54%	0.2681	0.2089

Table 1. R, RMSE, and MAE values of three different methods

5. Conclusion

In this study, the performance of Depth Anything V1 and V2 models in the prediction of depths is evaluated using the Sentinel-2 satellite image. Known for their advanced zero-shot and few-shot capabilities, these models provide an effective method for extracting depth maps, especially from unlabeled monocular images. The results show that Depth Anything V1 provides 56.69% correlation with NOAA NCEI data, while Depth Anything V2 increases this value to 84.54%, demonstrating that the model provides a significant improvement in terms of correlation. In addition, the depth maps created by these models are measured with RMSE and MAE values, and are determined as 0.4135 m and 0.3400 m for V1, and 0.2681 m and 0.2089 m for V2, respectively. This improvement shows that Depth Anything V2 has further improved its ability to predict depth from monocular satellite images and offers great potential for cost-effective bathymetry mapping in remote sensing applications. The results were compared with classical BRM, and it was inferred that DL-based approaches show superior accuracy compared with classical methods.

In conclusion, this study highlights the effectiveness of deep learning-based approaches in the process of making bathymetry predictions from satellite images. The increased accuracy provided by Depth Anything V2 offers a potentially cost-effective method in areas such as underwater mapping and environmental monitoring. Especially in developing countries, it becomes possible to extract bathymetry with such open-source, low-cost satellite data as an alternative to high-cost field studies and commercial satellite images. Furthermore, this approach has great potential for shallow water and coastal regions, where continuous and accurate bathymetric data is essential for erosion monitoring, navigation, and coastal ecosystem management. Environmental management and disaster prevention strategies can be substantially improved by being able to monitor depth variations in these vulnerable locations. For future studies, it is recommended to further develop these models for larger geographic areas, diverse environmental conditions, and greater depth ranges in satellite-based bathymetry applications. Furthermore, this approach could potentially be used to freshwater environments like lakes and reservoirs, where bathymetric mapping is equally necessary for pollution prevention, ecosystem evaluation, and water resource management.

Data Availability

Sentinel-2 imagery used in this study was obtained from the Google Earth Engine data catalog (https://code.earthengine.google.com/).

References

Al Najar, M., Thoumyre, G., Bergsma, E. W., Almar, R., Benshila, R., Wilson, D. G., 2023. Satellite derived bathymetry using deep learning. *Machine Learning*, 1-24.

Ashphaq, M., Srivastava, P. K., Mitra, D., 2021. Review of near-shore satellite derived bathymetry: Classification and account of five decades of coastal bathymetry research. *Journal of Ocean Engineering and Science*, 6(4), 340-359.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint*, , arXiv:2108.07258.

Figliomeni, F. G., Parente, C., 2023. Bathymetry from satellite images: A proposal for adapting the band ratio approach to IKONOS data. *Applied Geomatics*, 15(3), 565-581.

Figliomeni, F. G., Parente, C., 2024. Methods for Satellite Derived Bathymetry from Sentinel-2 images: a comparison. The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 48, 165-173.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. Cambridge, MA, USA, *MIT Press*.

Iglovikov, V., Mushinskiy, S., Osin, V., 2017. Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint*, , arXiv:170606169

Jiang, H., Rutherfurd, I., 2024. Potential to use free satellite imagery to retrieve the past bathymetry of large rivers. *Remote Sensing Applications: Society and Environment*, 33, 101133.

Kalkan, B., Bakirman, T., Bayram, B., 2021. Satellite derived bathymetry with machine learning: A case study for California. Paper presented at the International Symposium on Applied Geoinformatics (ISAG 2021), Istanbul, Turkey.

Kim, M., Danielson, J., Storlazzi, C., Park, S., 2024. Physics-based satellite-derived bathymetry (SDB) using Landsat OLI images. *Remote Sensing*, 16(5), 843.

Liu, Q., Hang, R., Song, H., & Li, Z., 2017. Learning multiscale deep features for high-resolution satellite image scene classification. IEEE *Transactions on Geoscience and Remote Sensing*, 56(1), 117–126.

Lv, Z., Herman, J., Brewer, E., Nunez, K., Runfola, D., 2025. BathyFormer: A Transformer-Based Deep Learning Method to Map Nearshore Bathymetry with High-Resolution Multispectral Satellite Imagery. *Remote Sensing*, 17(7), 1195.

NCEI Bathymetric Data Viewer, 2025. https://www.ncei.noaa.gov/maps/bathymetry/ (05 April 2025).

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sutskever, I., 2021. Learning transferable visual models from natural language supervision. In: *International conference on machine learning*, 8748-8763.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*, arXiv:1409.1556.

Sonogashira, M., Shonai, M., Iiyama, M., 2020. High-resolution bathymetry by deep-learning-based image superresolution. *Plos one*, 15(7), e0235487.

Stumpf RP, Holderied K, Sinclair M., 2003. Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnol Oceanogr* 48:547–556

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., Lample, G., 2023. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Uzakara, H., Demir, N., Karakış, S., 2024. Satellite-based bathymetry supported by extracted coastlines. PFG–Journal of Photogrammetry, *Remote Sensing and Geoinformation Science*, 92(4), 317-334.

Wofk, D., Ma, F., Yang, T. J., Karaman, S., Sze, V., 2019. Fastdepth: Fast monocular depth estimation on embedded systems. In: *2019 International Conference on Robotics and Automation (ICRA)*, 6101-6108.

Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. *arXiv preprint*. arXiv:2401.10891. https://doi.org/10.48550/arXiv.2401.10891

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth Anything V2. *Advances in Neural Information Processing Systems*, *37*, 21875–21911.