Article

Deep-Learning-Based Automatic Sinkhole Recognition: Application to the Eastern Dead Sea

Osama Alrabayah [1,*,†], **Danu Caus** [2,3,4], **Robert Alban Watson** [5], **Hanna Z. Schulten** [5], **Tobias Weigel** [2,3,4], **Lars Rüpke** [1] **and Djamil Al-Halbouni** [6]

1   Helmholtz Centre for Ocean Research—GEOMAR, 24148 Kiel, Germany; lruepke@geomar.de
2   German Climate Computing Centre, 20146 Hamburg, Germany; caus@dkrz.de (D.C.); weigel@dkrz.de (T.W.)
3   Helmholtz Centre Hereon, 21502 Geesthacht, Germany
4   Helmholtz AI, Germany
5   School of Earth Sciences, University College Dublin, Belfield, D04 V1W8 Dublin, Ireland;
    hanna.schulten@rwth-aachen.de (H.Z.S.)
6   Institute for Earth System Science and Remote Sensing, University of Leipzig, 04103 Leipzig, Germany;
    dhalbouni@uni-leipzig.de
*   Correspondence: alrabayah@bwl.uni-kiel.de
†   Current address: Department for Technology Management, Christian-Albrechts-University at Kiel,
    24118 Kiel, Germany.

Abstract: Sinkholes can cause significant damage to infrastructures, agriculture, and endanger lives in active karst regions like the Dead Sea's eastern shore at Ghor Al-Haditha. The common sinkhole mapping methods often require costly high-resolution data and manual, time-consuming expert analysis. This study introduces an efficient deep learning model designed to improve sinkhole mapping using accessible satellite imagery, which could enhance management practices related to sinkholes and other geohazards in evaporite karst regions. The developed AI system is centered around the U-Net architecture. The model was initially trained on a high-resolution drone dataset (0.1 m GSD, phase I), covering 250 sinkhole instances. Subsequently, it was additionally fine-tuned on a larger dataset from a Pleiades Neo satellite image (0.3 m GSD, phase II) with 1038 instances. The training process involved an automated image-processing workflow and strategic layer freezing and unfreezing to adapt the model to different input scales and resolutions. We show the usefulness of initial layer features learned on drone data, for the coarser, more readily-available satellite inputs. The validation revealed high detection accuracy for sinkholes, with phase I achieving a recall of 96.79% and an F1 score of 97.08%, and phase II reaching a recall of 92.06% and an F1 score of 91.23%. These results confirm the model's accuracy and its capability to maintain high performance across varying resolutions. Our findings highlight the potential of using RGB visual bands for sinkhole detection across different karst environments. This approach provides a scalable, cost-effective solution for continuous mapping, monitoring, and risk mitigation related to sinkhole hazards. The developed system is not limited only to sinkholes however, and can be naturally extended to other geohazards as well. Moreover, since it currently uses U-Net as a backbone, the system can be extended to incorporate super-resolution techniques, leveraging U-Net based latent diffusion models to address the smaller-scale, ambiguous geo-structures that are often found in geoscientific data.

Keywords: deep learning; computer vision; CNN; U-Net; segmentation; automatic recognition; geohazards; subsidence; sinkholes; dead sea

1. Introduction

Subsidence is a worldwide phenomenon of vertical ground settlement, either due to natural or anthropogenic reasons [1]. A special form of subsidence is the appearance of enclosed depressions, so-called sinkholes, as a morphological landscape expression of karstified rock in the underground. Well-known examples of sinkholes are located

in Florida, Turkey, Germany, China, Spain, at the Dead Sea and in many other karst environments worldwide (see e.g., [1,2]).

Sinkholes are a remarkable natural hazard, with the potential to cause extensive damage to the environment, infrastructure, and human life (see [3–5]). Thorough and accurate mapping of sinkholes is important to identify patterns and monitor sinkhole activities, communicate necessary steps to prevent or mitigate their damage and it also contributes to the creation of comprehensive sinkhole inventories. Researchers such as Galve et al. [6], Galve et al. [7], Sevil and Gutiérrez [8], and Gutiérrez [9] have highlighted the importance of these sinkhole maps in supporting decision-making processes for land use, development projects, and hazard preparedness in areas susceptible to sinkhole formation.

Over time, the methods used to map sinkhole outlines have remarkably evolved, reflecting an influential shift in the types of the data and technological approaches used. Initially, the dominant methods were on-site field assessments and geophysical surveys, along with manual inspection of topographic maps and stereo-images (e.g., [10,11]). An alternative source of baseline data are Digital Elevation Models (DEMs), which can be derived from both passive remote sensing data, such as aerial photography and satellite imagery (e.g., [12,13]) or from active remote sensing sources such as airborne laser scans (LiDAR, [14]) or radar, e.g., the Shuttle Radar Topography Mission (SRTM, [15]). In the last 40 years, the increased availability of such DEMs has facilitated development of many automated methods of depression mapping, often performed within a Geographic Information System (GIS). These methods tend to leverage the geometric properties of sinkholes to identify them, typically by simulating the inundation of the terrain model with water (Figure 1). Despite the undoubted increase in efficiency and objectivity afforded by these approaches, they require up-to-date DEMs of sufficient resolution to be obtained every time a sinkhole inventory is to be updated, which can be extremely costly and time-consuming. Furthermore, the mathematical generalization of sinkhole geometry necessary to apply these 'top-down' approaches can be inflexible, and can result in large numbers of depressions being missed, especially in the case of subtle and shallow sinkhole morphologies, or nested systems. Please refer to [1] (Sect. 6.3.7) for a thorough review of sinkhole mapping techniques.

In the last 20 years however, the combined advances in image processing and statistical optimization have facilitated an explosion in 'data-driven' automatic image classification, e.g., training deep, convolutional neural network architectures to recognise data patterns and other methods (e.g., [16–20]). After manually labelling a subset of the total population of studied features, training data is extracted and turned into image patches of fixed size. The model then identifies common patterns among the training data, such as edges of various size and orientation for example. This is completed in several stages/layers, with the dimensions of input data being reduced between stages and more abstract and complex patterns identified at each stage. Analysing the statistical relationships between patterns, the model is then able to generalize and classify data it has not seen before. The exact architecture of a model can be adjusted according to the specific task at hand. The intrinsic nature of these models allows them to be far more scalable and efficient than explicit categorisation models, such as the approaches presented in Figure 1. Such machine learning frameworks have proved to be applicable to the detection and mapping of sinkholes (Table 1), and their versatility and adaptability has been shown across many other object mapping applications.

While previous studies have significantly advanced the field of sinkhole detection using machine learning models, a review of these works reveals specific methodological challenges. For instance, limitations in relation to data availability (especially high-resolution elevation models), the need for manual verification to ensure accuracy, a lack of testing outside of limestone karst regions, and resolution limitations that may not fully capture the diverse geometrical characteristics of sinkhole instances, have been noted across different approaches (see Table 1 for detailed limitations of each referenced study). These limitations

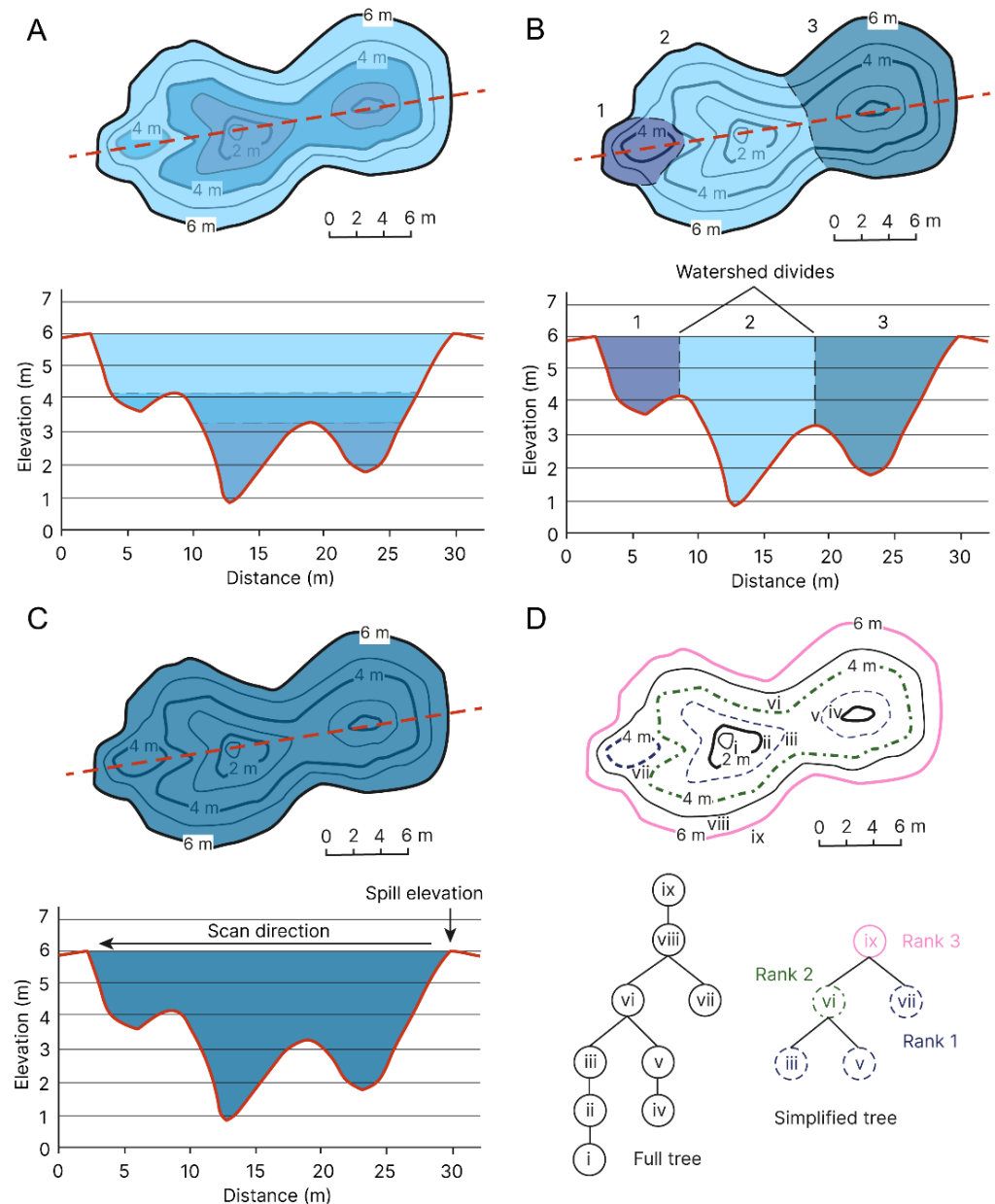underline the importance of developing more adaptable and efficient methodologies for sinkhole detection.

Figure 1. Schematic representations of different 'top-down' algorithmic methods of delineating sinkholes. (**A**) method of O'Callaghan and Mark, [21], which maps the depressions according to simulated stratification of water within them. Adapted with permission from [21], 2024, Elsevier. (**B**) the 'D8' method of Jenson and Domingue [22], which uses a moving window to map the watersheds within the depression. This method and that shown in (**A**) become very computationally intensive with high-resolution data. (**C**) the 'priority fill' method of Wang and Liu, [23], which is able to simulate filling of the entire compound depression in one pass of processing. This method offers an improvement in run-time of a factor of 30 on (**B**), but is not able to capture the internal complexity of the compound depression. Adapted with permission from [23], 2024, Taylor & Francis. (**D**) The 'contour tree' method developed by Wu et al. [24,25], which builds on the 'priority fill' method to produce a graph ('tree') of contours within the compound depression, allowing nested depressions to be identified and labelled by their rank. This allows for more accurate automated updating of depression location and morphometric databases. The method has since been further refined for efficient computation (see [26,27]). Adapted with permission from [25], 2024, Elsevier.

Table 1. Overview of relevant studies which have applied Machine Learning (ML) and Deep Learning (DL) to detect sinkholes from remote sensing data.

Authors	Technique Used	Data Source	Key Insights on ML/DL Use	Study Limitations	Best Performance Metrics for Sinkhole Class
Lee et al. [28]	3D-Convolutional Neural Network (CNN)	Thermal images from drones, resolution: 640×480 pixels.	Demonstrated that a light-CNN algorithm can effectively be applied to thermal drone images for detecting artificial sinkholes.	Reliance on drone-based thermal imaging, risk of missing sinkholes due to drone speed and background patterns, and a benchmark dataset not fully representing sinkhole diversity, indicating a need for more varied data.	Precision: 87.9% Recall: 88.1%
Zhu and Pierskalla [29]	Random Forest Classifier	LiDAR data, average point spacing: 1 m, DEM cell size: 1.5 m.	One of the first studies to apply ML to the problem of sinkhole delineation from elevation data.	The model failed to transfer effectively between study areas (89% accuracy in detecting sinkholes in area 1 vs. 73% accuracy in area 2), elevation models of equivalent resolution are difficult and costly to obtain in more remote karst regions.	Precision: 84.71% Recall: 65.17%
Kang et al. [30]	Modified CNN architecture based on AlexNet (See Krizhevsky et al. [18]	Ground Penetrating Radar (GPR), original resolution: 50×50 pixels (B-scan), 50×13 pixels (C-scan), enhanced to 200×200 pixels.	Highlighted versatility of CNN architectures at sinkhole detection by applying them to GPR data.	Narrowly defined area of interest, so transferability untested, GPR data are difficult to obtain in more remote study areas.	(Original resolution) Precision: 88.26% Recall: 72.36%, (Enhanced resolution) Precision: 100% Recall: 100%,
Mihevc and Mihevc [31]	U-Net	LiDAR, DEM cell size: 1 m.	Proved U-Nets to be a highly scalable automatic approach to sinkhole detection—initially mapped > 470,000 sinkholes in Slovenia, and has now been applied to map > 400,000 sinkholes across the entire USA. See https://dolines.org accessedon 17 June 2024	Elevation models of equivalent resolution are currently unavailable in many karst regions, accuracy was not especially high (16% variation as compared to manual mapping for both sinkhole count and area), model performance relatively untested outside limestone karst areas.	Intersection over Union (IoU): 60.4% Dice Coefficient: 72.36%

Table 1. *Cont.*

Authors	Technique Used	Data Source	Key Insights on ML/DL Use	Study Limitations	Best Performance Metrics for Sinkhole Class
Nefeslioglu et al. [32]	Artificial Neural Network (ANN)	Satellite optical imagery and InSAR DEMs spatial resolution: 10 m.	Used ANNs for sinkhole susceptibility mapping and detection, confirming the value of ANN models in this field.	The accuracy of the used DEM, and its sensitivity to vegetation and land cover changes, may introduce errors in deformation mapping. This emphasizes the importance of the accuracy of sinkhole susceptibility assessments and deformation analyses.	Root Mean Square Error (RMSE): 45.1%
Rafique et al. [33]	U-Net	LiDAR DEMs, aerial imagery resolution: 1.524 m per pixel.	Integrated two types of raster data (optical imagery and elevation models) and their derivatives to improve U-Net performance in sinkhole detection, with good learning between US limestone karst areas.	Elevation models of equivalent resolution are currently unavailable in many karst regions, model performance relatively untested outside limestone karst areas. The study suggests that aerial images alone were not useful for sinkhole segmentation.	IoU: 45.38% Precision: 66.29%

Our research introduces an approach to sinkhole mapping, using RGB visual bands only as the data source. This method aims to overcome limitations highlighted in earlier studies by using a modified deep learning pipeline. Our main objective is to develop an automated model capable of accurately mapping the spatial distribution of sinkhole instances by analysing aerial images of different resolutions. This was supported by the use of labelled sinkhole data from the evaporite karst region of Ghor Al-Haditha, situated on the eastern shore of the Dead Sea. The project was performed in two distinct phases. In the **initial phase**, a U-Net model was trained, tested, and validated on a dataset of 250 instances derived from a high-resolution orthophoto captured in December 2016, featuring a ground sample distance (GSD) of 0.1 m. The **subsequent phase** involved transferring the model to a new dataset comprising 1038 instances, obtained from a Pleiades Neo satellite image (GSD of 0.3 m) from August 2022, capturing the same study area. The effectiveness of our algorithm was demonstrated through its accuracy in detecting sinkhole instances across both datasets, underlining the model's transferability and the feasibility of automating sinkhole mapping using readily available satellite images. The code is made publicly available at: **https://github.com/ducspe/sinkhole_geohazard_segmentation** (accessed on 17 June 2024).

2. Dead Sea Site Description and Sinkhole Evolution

Sinkhole formation in the Dead Sea region has intensified over the last 35 years, with an escalating occurrence of over 6000 sinkholes, which is closely associated with the rapid regression of the lake and associated shoreline migration [34–36]. Several conceptual models have been proposed on the origin and evolution of the sinkholes. Notably, different geoscientific methods have revealed that the underlying conditions for sinkhole formation vary from location to location at the Dead Sea shoreline. On the eastern shoreline, where our data subset stems from, and parts of the SW shoreline, physical subsurface erosion and chemical dissolution of evaporites, the general instability of non-evaporitic sedimentary materials and tectonic control, have all been suggested as underlying causes for this phenomenon [35,37–39]. For the majority of the western shoreline, however, tectonic control, a purely chemical dissolution of a massive salt layer and related salt-dissolution front migration have been suggested [40–44].

The selected site for this study, Ghor Al-Haditha, is situated on the south-eastern shore of the Dead Sea in Jordan (Figure 2). The site encompasses approximately 9.75 km^2 and lies 340 to 440 m below the mean sea level, bordered by the Dead Sea to the west and the Dead Sea highway to the east. Despite its small size, the site has a high density of sinkholes, with over 1000 sinkholes having formed from 1967 to 2017 [36]. The sinkholes at Ghor Al-Haditha are formed in three primary near-surface materials: unconsolidated to semi-consolidated lacustrine silty-clay carbonates, alluvial sand-gravel sediments, and rock salt with interleaved thin mud layers [45]. Sinkhole morphology is variable depending upon the mixture of these materials in which they are formed. Sinkholes formed primarily in alluvium and salt materials generally have a high depth to diameter ratio [36], indicating a collapse origin. On the other hand, sinkholes formed mostly in mud tend to be much wider and shallower. They are formed by surface sagging of overlying deposits, and are typically filled with large collapsed and inward-rotating chunks of sedimentary material [45,46]. Sinkhole clustering and coalescence into larger compound sinkholes and larger-scale karst depressions are common processes. The smallest features included in the dataset are ~3 m in diameter, while the largest are over 60 m. In addition to these enclosed depressions, several surface stream-channels fed by groundwater springs have formed in the former lakebed of the Dead Sea [37]. These canyons are characterised by steep bank slopes, and the springs feeding the streams often emerge within areas of subsidence and sinkhole formation.

Overall, despite the lack of solutional karst features, the dataset we have gathered encompasses a wide variety of evaporite sinkhole materials, morphologies and genesis types. The sparse vegetation and clear skies that are typical for the region amplify the visibility of sinkholes in aerial imagery. The dynamic topography, characterized by three

major wadi systems depositing alluvial fan deposits on the coastline [37], adds diversity to the training dataset. In this study, we aim to capitalize on these diverse characteristics to improve our models' performance in identifying sinkholes across different environments.
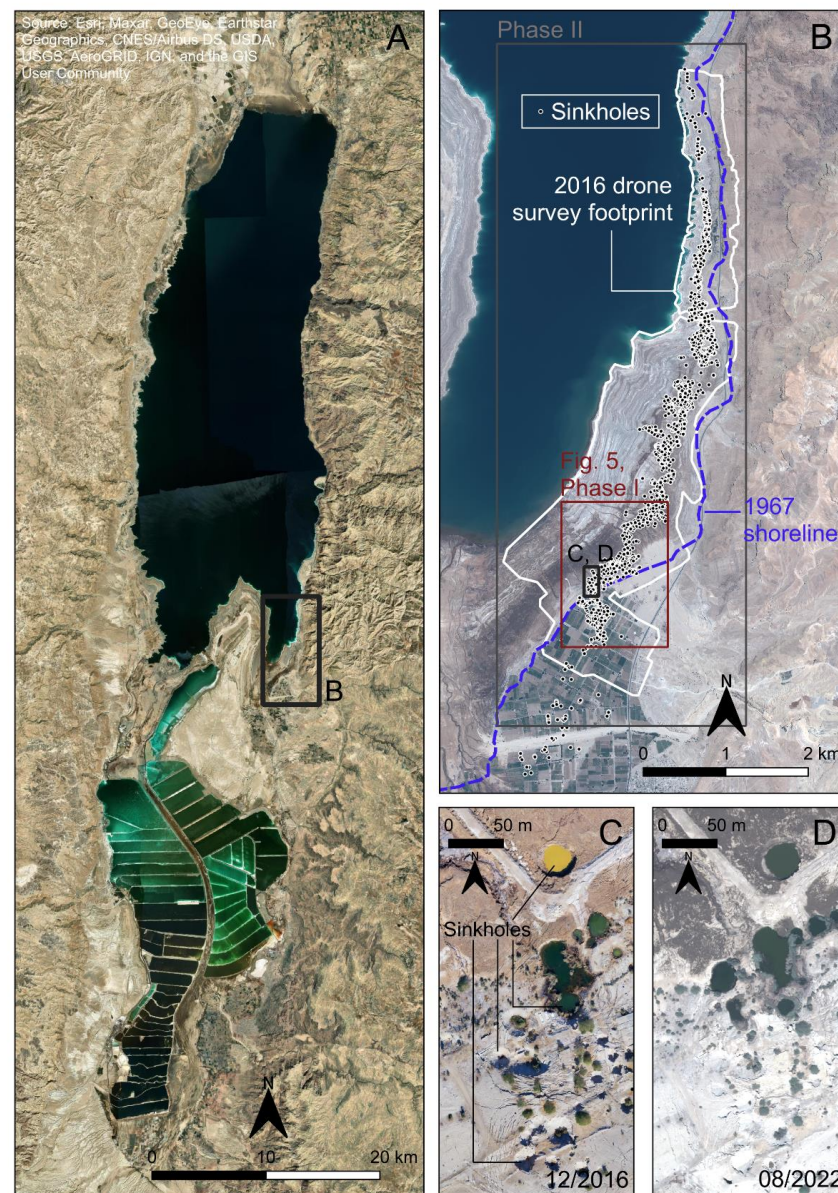

Figure 2. Overview of the study area. (**A**) ESRI satellite imagery of the Dead Sea. The location of part (**B**) is marked. (**B**) Pleiades 1-A satellite image from April 2018 of the Ghor Al-Haditha study area on the Dead Sea's eastern shore. The outline of data collected in the December 2016 drone survey, the extent of sinkhole formation across the study area and the position of the Dead Sea shoreline in 1967 are shown. Additionally, the areas covered by the datasets used for Phase I (Red) and Phase II (Grey) of our study are shown, as are the locations of parts (**C,D**), which depict sinkholes in both alluvium and mud materials as they appear in the 2016 structure-from-motion orthophoto and in Pleiades Neo satellite imagery from August 2022, respectively. Several new sinkholes have formed in 2022 as compared to 2016, and others have changed in shape and size.

Sinkhole development in the Ghor Al-Haditha area has been rampant since 1986, with over 1000 sinkholes appearing between 1967–2017 [36], and more than 1000 further sinkholes forming between 2017–2024 [47]. The incessant formation of sinkholes has resulted in significant damage, disrupting infrastructure and affecting agriculture in the

area [36]. In response to these challenges, the Ministry of Energy and Mineral Resources of Jordan has commenced geologic and geophysical surveys in this area since the early 1990s aiming to understand the causes and consequences of the sinkhole formation (e.g., [48]).

As the Dead Sea evolves as a potential site for geotourism, the careful identification and mapping of sinkholes becomes very important [49,50]. This attempt goes beyond the immediate safety concerns, offering a route to revive local economies impacted by sinkhole formations. Enhancing our precision in detecting and tracking these formations supports the safety standards of the region for geotourism, protecting both visitors and local communities. Proactive identification can not only prevent significant economic losses from infrastructure damages, but also offers valuable insights into the future trajectory of sinkhole formation. A comprehensive grasp of current sinkhole patterns, therefore, becomes useful in developing informed prevention and response strategies, ensuring the Dead Sea's viability as a geotourism hotspot without compromising on safety or environmental sanctity [49,50].

Various studies, deploying geological, geophysical and hydrogeological surveys, remote sensing, and numerical simulation, have been undertaken at Ghor Al-Haditha to comprehend the spatio-temporal development and the mechanisms of sinkhole formation here (e.g., [51]). Through these studies, local authorities have been able to delineate areas susceptible to sinkhole threats more effectively. Furthermore, aerial images of different resolutions collected over the years by satellites, as well as balloon and drone surveys offer a chronological illustration of sinkhole evolution in the region [47]. Given the dynamic geology of the region, these aerial images form a large and diverse training dataset for our deep learning model.

3. Materials and Methods

3.1. Deep Learning Approach

We chose to frame the research problem of mapping and delineating sinkholes as an **instance segmentation** problem (Figure 3), enabling the classification of sinkhole instances at the pixel level. This approach offers advantages over simple classification or object detection methods by facilitating detailed spatio-temporal morphometric analysis and evolution monitoring of the mapped sinkholes instances. Moreover, it allows for clear delineation of 'redundant' and 'non-redundant' sinkholes (see Sevil and Gutiérrez [52] for a recent example of this). The model development and training process relied exclusively on the colour channels (RGB) present in the aerial data, as opposed to incorporating other channels like Digital Surface Models (DSMs), which might not be available for all regions. This methodology makes the model more applicable to a broader range of cases.
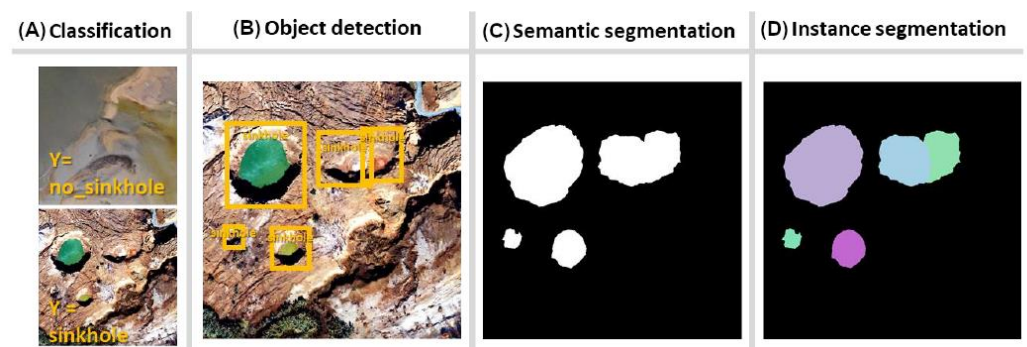

Figure 3. Relevant computer vision problems within which we could frame our task. We chose image segmentation in the end. (**A**) *Image classification*: an entire image is classified according to a label. (**B**) *Object detection*: the task of detecting instances of objects of a certain class within an image. (**C**) *Semantic segmentation:* label each pixel of an image with a corresponding class, i.e., per pixel classification (**D**) *Instance segmentation:* label each pixel of an image with a corresponding class and detect instances of objects of each class within an image.

3.2. Datasets and Annotation Process

The study was conducted in two phases, employing two different datasets: high-resolution (HR) drone orthophoto imagery for Phase 1 and low-resolution (LR) optical satellite imagery for Phase 2. The first dataset was compiled from a point cloud collection of high-resolution drone images, which were processed by using structure-from-motion photogrammetry (see [35], for an overview of this process), and the second dataset was generated from a single Pleiades Neo scene acquired in August 2022. Both datasets are taken from the same region shown in Figure 2 above. Notably, the satellite imagery covers a larger area that includes the region covered by the drone. The description of both datasets and the process of their annotation are elaborated below.

3.2.1. Dataset for Phase 1 (HR Drone Images)

For the initial phase of the study, a high-resolution dataset was employed that has been gathered in December 2016 through drone-based, close-range aerial surveys. This dataset comprises optical orthophoto mosaics with a resolution of 0.1 m/pixel, acquired via a 12 MP DJI Phantom 3 inbuilt camera at an altitude of around 100 m. The manual annotation process was directed by a digital surface model (DSM) which was devised by photogrammetric processing of the optical images. For a more detailed explanation regarding the creation of orthophoto mosaics and DSMs, refer to Al-Halbouni et al. [35] and Watson et al. [36]. The dataset was explicitly annotated for the purposes of this research project to train a deep learning model. In this phase, particular emphasis was placed on the precision and quality of the annotation process, prioritizing the accuracy of labelled data over its quantity.

The annotation process involved manually digitizing sinkhole extents within the ArcGIS Pro V. 2.9 software, employing various layers and tools within the software to guide the annotation process (Figure 4). In this way, we created a sinkhole instance segmentation mask image where each sinkhole was designated with a distinct colour (Figure 5). Expert knowledge of the distinction between an enclosed sinkhole and an open stream-channel sink has been incorporated at this stage. Finally, the mask image was exported as a TIFF RGB image of the same dimensions as the orthophoto image (12,633 × 15,062 × 3).
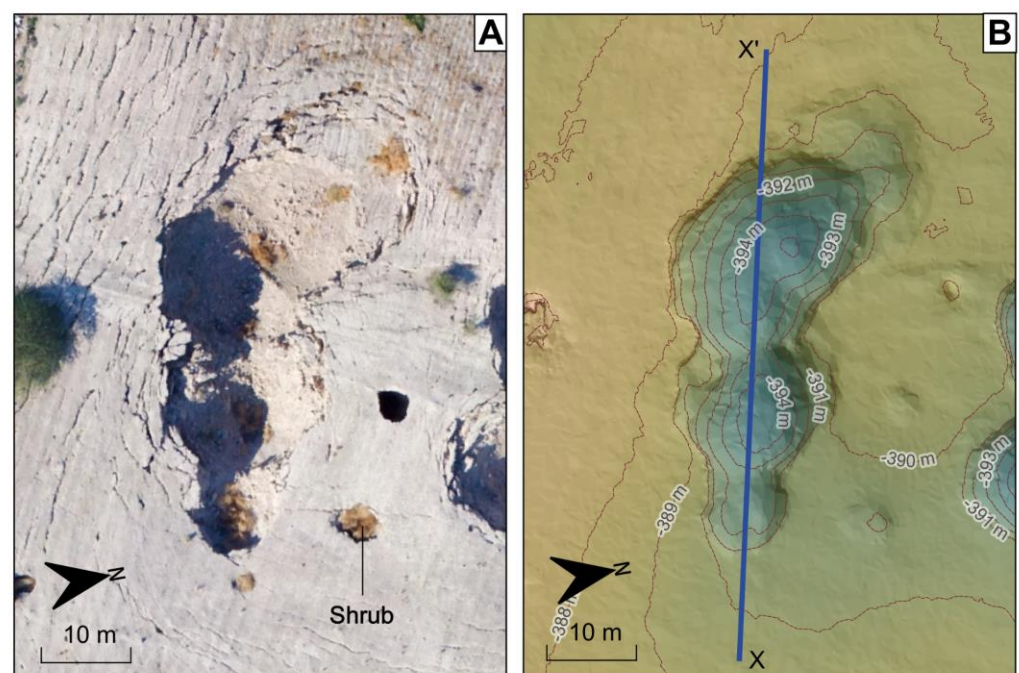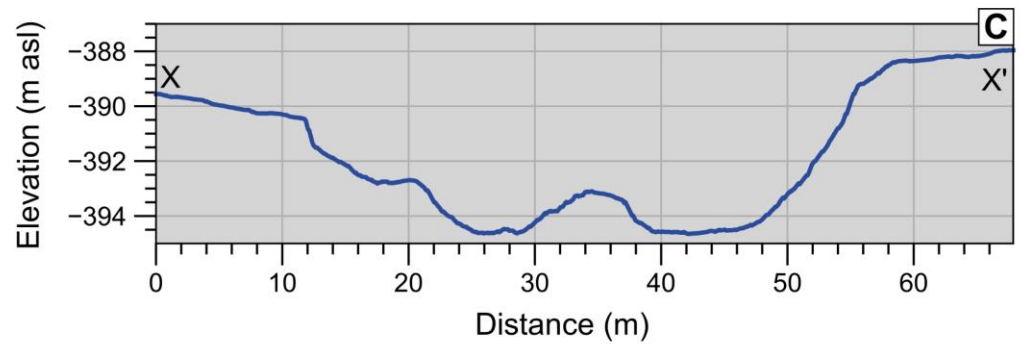

Figure 4. *Cont.*

Figure 4. The different layers that were used to guide the annotation process for the training dataset in Phase I. The sinkhole cluster shown here is the same as that highlighted in Figure 5C–F. (**A**) RGB orthophoto mosaic. (**B**) DSM data visualized as a hill-shaded relief map. Contour lines generated from the DSM data with an interval of 1 m were also used. (**C**) Elevation profile generated within ArcGIS Pro (V. 2.9) along the axis of a sinkhole cluster. The tool was used in special cases to find the exact edges, especially the edges between compound (merged) sinkholes, as presented in the image.
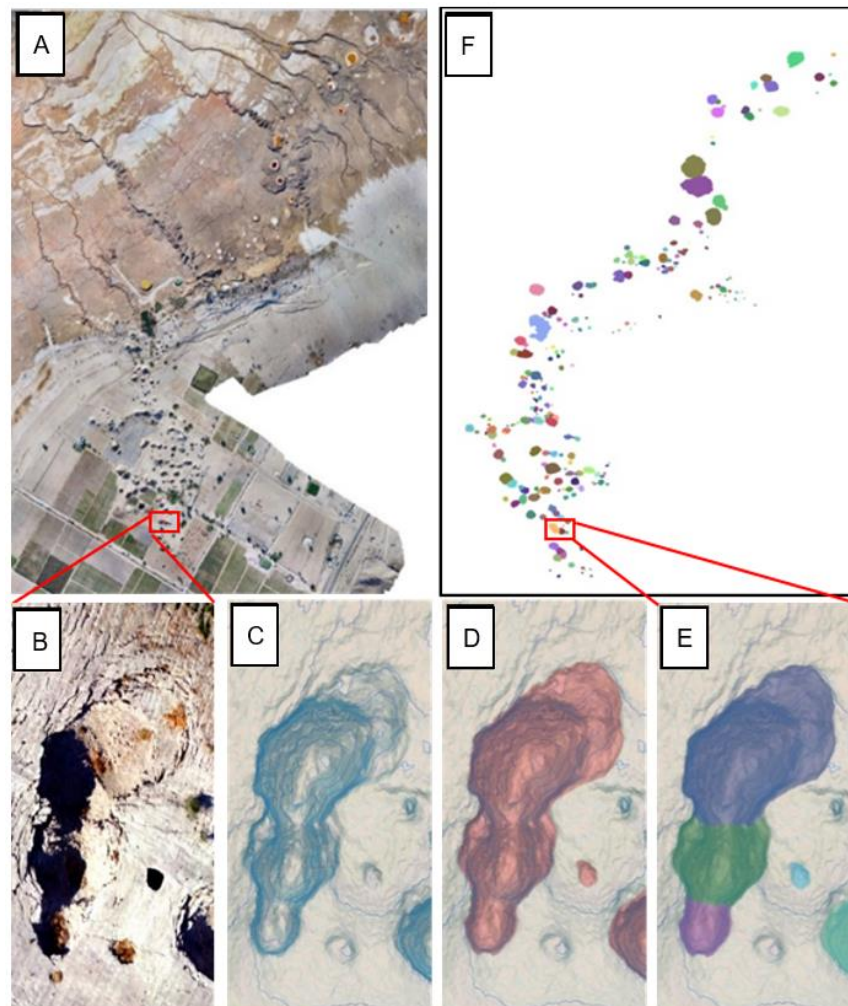
Figure 5. Generating the sinkhole instance segmentation mask image. (**A**) The selected area from the drone image for training sample generation. (**B**) Several depicted sinkholes. Note the 3 compound sinkhole instances. (**C**) Using different layers to guide the annotation process. (**D**) Different polygons were manually drawn for each sinkhole instance with precise edges. (**E**) Converted polygons to a raster layer where each sinkhole is presented using a different colour. (**F**) TIFF mask image with all the sinkholes in the selected area.

3.2.2. Dataset for Phase 2 (LR Satellite Images)

In the subsequent phase of our research, the focus shifted towards exploring the potential of transfer learning [53] to enrich the versatility of our model. To this end, we adapted the model that was initially tailored for drone data, to suit satellite imagery. This transition leveraged a dataset curated from a collection of pre-existing datasets, originating from research studies conducted on satellite images from the year 2022. The images stem from Pleiades Neo satellite with a resolution of 0.3 m/pixel, acquired in August 2022 and pansharpened. The annotation process for sinkhole instances was guided by the central points of these sinkhole instances present within the original dataset. Utilizing the capabilities of ArcGIS Pro, we meticulously mapped the extent of each sinkhole as polygons.

The annotation of the satellite images was enhanced using the streaming tool on ArcGIS software—a convenient feature allowing users to craft polygons reflecting the computer mouse movements. Several defining characteristics of the sinkholes assisted in the annotation process. These included pronounced shadowing typically observed in the southern corner, noticeable alterations in texture and colouration, a discernible bright salt layer, and occasionally, water accumulation at the sinkholes' depocentres, as depicted in Figure 6. Annotation limitations primarily stemmed from the lower resolution of satellite images—in comparison with the drone case—and the absence of elevation data to guide the process.
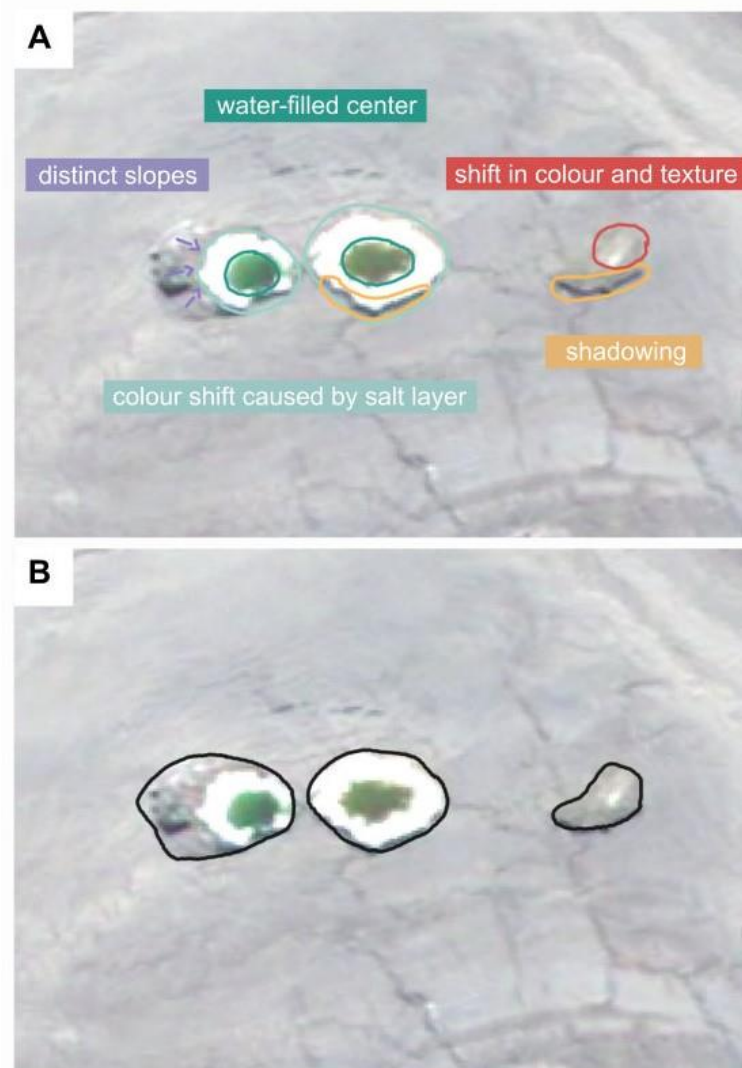

Figure 6. (**A**) Defining features of sinkhole outlines. (**B**) Mapping of sinkholes in the area.

3.2.3. Annotation Special Cases

During the annotation process, we encountered a few unique scenarios. For instance, where vegetation obscured parts of a sinkhole, making the borders not entirely visible, an estimation method was employed for the high-resolution drone dataset. In such cases, sinkholes that were predominantly concealed by vegetation were not mapped. On the other hand, in the low-resolution satellite dataset, the annotator resorted to satellite images from previous years to estimate the borders of obscured sinkholes. This situation was not frequent, affecting only approximately 1 to 10 sinkholes per image, due to the sparse vegetation in the Dead Sea region.

Another unique case involved compound (merged) sinkholes. These were treated differently between the two datasets: in the high-resolution dataset, each sinkhole within a compound structure received a separate annotation, while in the low-resolution dataset, compound sinkholes were consistently mapped as a single unit. The subsequent section will focus on the data preprocessing and training methodology for the deep learning model.

3.3. Deep Learning Model Architecture

The choice of an appropriate CNN architecture is important in achieving the objectives of our sinkhole recognition project. In this study, our aim is to identify individual instances of sinkholes, a task known as instance segmentation. This poses a challenge, particularly given the constraints of our limited dataset. To address this, we selected the U-Net architecture [54], which despite its typical association with semantic segmentation, presents a viable solution for our requirements. The U-Net architecture was deliberately chosen for several reasons:

- **Simplifying intermediary steps**: U-Net generates semantic segmentation maps that serve as simplifying intermediary steps in our pipeline, followed by post-processing operations like connected-component labelling (CCL) [55] to generate the instance segmentation map. This two-step approach reduces the complexity of the problem, allowing for more accurate segmentation despite limited data.
- **Adaptability to limited datasets**: U-Net is particularly adept at handling limited datasets due to its efficient structure. The fully convolutional nature of U-Net allows it to perform well even with relatively small amounts of training data, which was crucial considering the limited number of annotated sinkhole instances available for our study.
- **Multiscale feature extraction**: U-Net's architecture, with its encoder-decoder structure and skip connections, allows it to capture multiscale features effectively [56]. This is advantageous for detailed sinkhole identification, as it enables the network to retain high-resolution information, while also learning more abstract representations at the same time. In certain scenarios, the skip connections can also help manage class imbalance challenges, commonly encountered in image segmentation tasks, as they facilitate the retention of high-resolution information, important for accurately depicting smaller-scale, minority classes [54].
- **Scalability**: U-Net is known for its scalability and efficiency in processing large datasets. Even if the datasets grow substantially with more drone and satellite data being accumulated over time, U-Net's fully convolutional architecture can keep pace with the increased scale and is amenable to efficient parallel processing in hardware. The fully convolutional nature, also allows to address various input sizes seamlessly, such as the ones we experimented with: 128×128 and 256×256, as well as other shapes that may arise in the future due to our focus on multi-resolution aspects.
- **Strategic goals**: Additionally, the U-Net architecture fits well within our strategic goal of developing a multi-scale, multi-resolution sinkhole detection system. Given the potential for future integration of super-resolution techniques via latent diffusion models such as SR3, which is a U-Net-based super-resolution diffusion model [57], U-Net provides a robust foundation that can be expanded upon. It acts therefore as a backbone and allows to connect heterogeneous components, i.e., segmentation

and super-resolution modules, in a consistent manner. Adding such super-resolution techniques can help improve the detection of tightly spaced geological features, for example, around merged sinkholes' edge areas, something we encountered issues with in this work, and would like to address next.

We also considered other networks, such as Mask R-CNN [58] and Cascade R-CNN [59]. These architectures are specifically designed for instance segmentation and could potentially handle the task end-to-end. However, they are naturally more computationally intensive to be able to isolate the instances as well, as opposed to simply segment semantically. This is to be expected, since instance segmentation is generally a harder task than semantic segmentation. Furthermore, we do not have enough depictions of the same sinkholes from various perspectives to be able to train them properly.

We would also like to mention the recent advancements in segmentation foundational models, such as Segment Anything [60], that generally make use of the Vision Transformer technology/ViT [61]. Unfortunately, they do not work well with geological data, possibly because of the statistics of the data distributions they were originally trained on, amongst other things. Strategically however, we also did not choose a transformer architecture on our end, because it is attention-based and is therefore much more data-demanding than traditional convolution-based models.

Given the data constraints, U-Net was more suitable for our needs, allowing us to reduce the complexity of the task first by doing semantic segmentation, and then performing instance segmentation as a second step, building therefore on the abstraction principle, i.e., the intermediary maps provided by the U-Net. In addition to this, U-Net is also more universal, allowing us to reuse its latent space embeddings in a consistent manner, namely, in the super-resolution extension we are planning via U-Net based diffusion models that would hopefully address the lower edge segmentation scores we are currently facing.

Considering all the above points, we specifically chose the U-Net architecture because it fits our expectations in terms of intended use-case, computational complexity, and consistency with future development plans and features that we intend to try out and possibly incorporate into the broader geohazard detection system.

To ensure that our system effectively identifies each sinkhole instance, we integrated a post-clustering algorithm into our methodology. However, this method faced limitations in differentiating compound sinkholes specifically, often classifying them as single instances. To address this, we experimented with adding a third class in the segmentation process to represent the edges between merged sinkhole instances. The idea behind is that we will attempt to predict where the in-between sinkhole edges are, subtract those pixels such that the sinkholes are separated first, and then apply the clustering algorithm on the separated sinkholes.

Our preprocessing phase included a multistep procedure to incorporate this edge class effectively. Initially, we employed a customized Sobel filter to detect edges between compound sinkholes in the mask image. The formula used for the filtered edge image was:

$$I = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2},$$

where I is the intensity of the pixels in the edge image and f is the 2D function depicting the original RGB label image. We apply this formula on the label image to detect where there is a sharp transition between the pixel values of one sinkhole and another sinkhole. This effectively means we detect the boundary between 2 merged sinkholes. The pixel intensity of this boundary is the magnitude of the label image gradient, and the components of this gradient are the derivatives/sharp transitions in the x (horizontal) and y (vertical) directions of the image. Note that we apply this formula efficiently, such that we do not detect edges between the sinkholes and the black background, but rather only between merged sinkholes. We do this by scanning the image to see where black background is present and ignoring those patches, i.e., not applying the formula there. Once we have the

thin edges computed in this manner, we apply dilation, a morphological computer vision operation, to dilate the edges to a certain extent. A dilated edge has the interpretation of a region of uncertainty, encoding the ambiguity, even for experts, regarding the question: where exactly does one sinkhole end and the other one begin? The model then will have the chance to encode the class uncertainties in its final layer, and become more or less uncertain depending on the different data examples it sees. The dilated edges are finally overlaid onto the binary label image to create a 3-class label image: 'Sinkhole' class, 'Background' or 'Non-sinkhole' class, and 'Edge' class between sinkholes (Figure 7). Subsequently, this finalized label image is ready to be patched and used to train the U-Net.
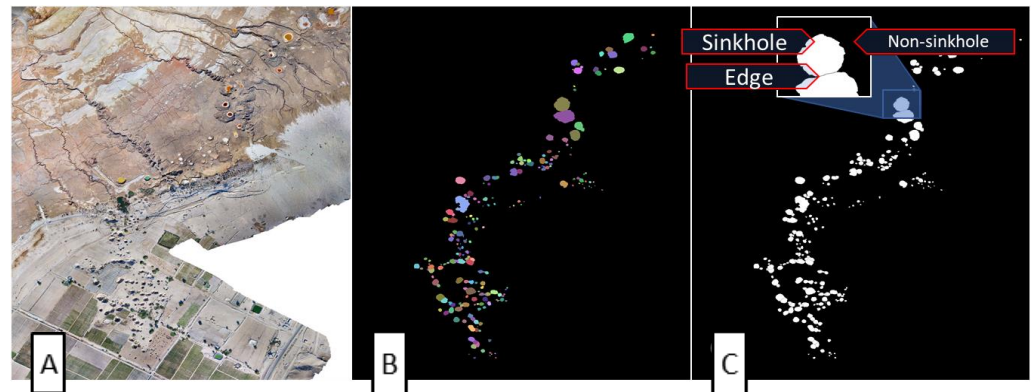
Figure 7. (**A**) Drone RGB image for the research area, (**B**) Sinkhole instance segmentation label image as created for the drone image case, and (**C**) The derived 3-class label image.

In the final phase of data preparation, we segmented the large images from the original orthophoto and the associated 3-class binary label into smaller, equally-sized tiles using a sliding window method with a 50% shifting/pixel overlap. This overlapped tiling allowed us to generate more data and hence facilitated the training of a more accurate model. The labelled images were then divided into training, validation, and testing sets in an 80:10:10 ratio, ensuring a comprehensive evaluation of the model's performance.

Our adaptation of U-Net was further refined to address the class imbalance challenge, a common issue in image segmentation. We employed data augmentation techniques and additionally, also experimented with specialized loss functions to see if they help balance the representation of different classes. Figure 8 illustrates the developed methodology for sinkhole instance segmentation.

The image tiles in our study are processed through a U-Net implemented in PyTorch Lightning, beginning with a double convolution block to extract basic features such as edges of different angles. This initial block consists of a convolutional layer, batch normalization and ReLU nonlinearity, repeated twice, where batch normalization helps to decouple the convolutional layers for better convergence during training. Following this initial stage, the U-Net architecture includes four down-sampling stages in the encoder, each comprising a max-pooling layer and a double convolution. This structure progressively learns more abstract features, with input/output channel tuples increasing from (64, 128); (128, 256); (256, 512) and finally to (512, 1024) throughout the stages. The encoder's compressive path is mirrored by an expanding decoding path with four up-sampling stages, each consisting of a transposed 2D convolution, followed by a double convolution block. The channel tuples in these stages reverse the encoder's pattern, decreasing from (1024, 512) to (128, 64). In the last layer, adapted for our ternary segmentation task, a 2D convolution layer aggregates 64 channels into three: one for the background, one for edges between sinkholes, and one for the sinkholes themselves. We kept the skip connections between the encoder and decoder, allowing unimpeded information flow across, enabling the decoder to access detailed information from the encoder. Our U-Net model is flexible to input sizes, but for this study, we focused on 128×128 image patches.
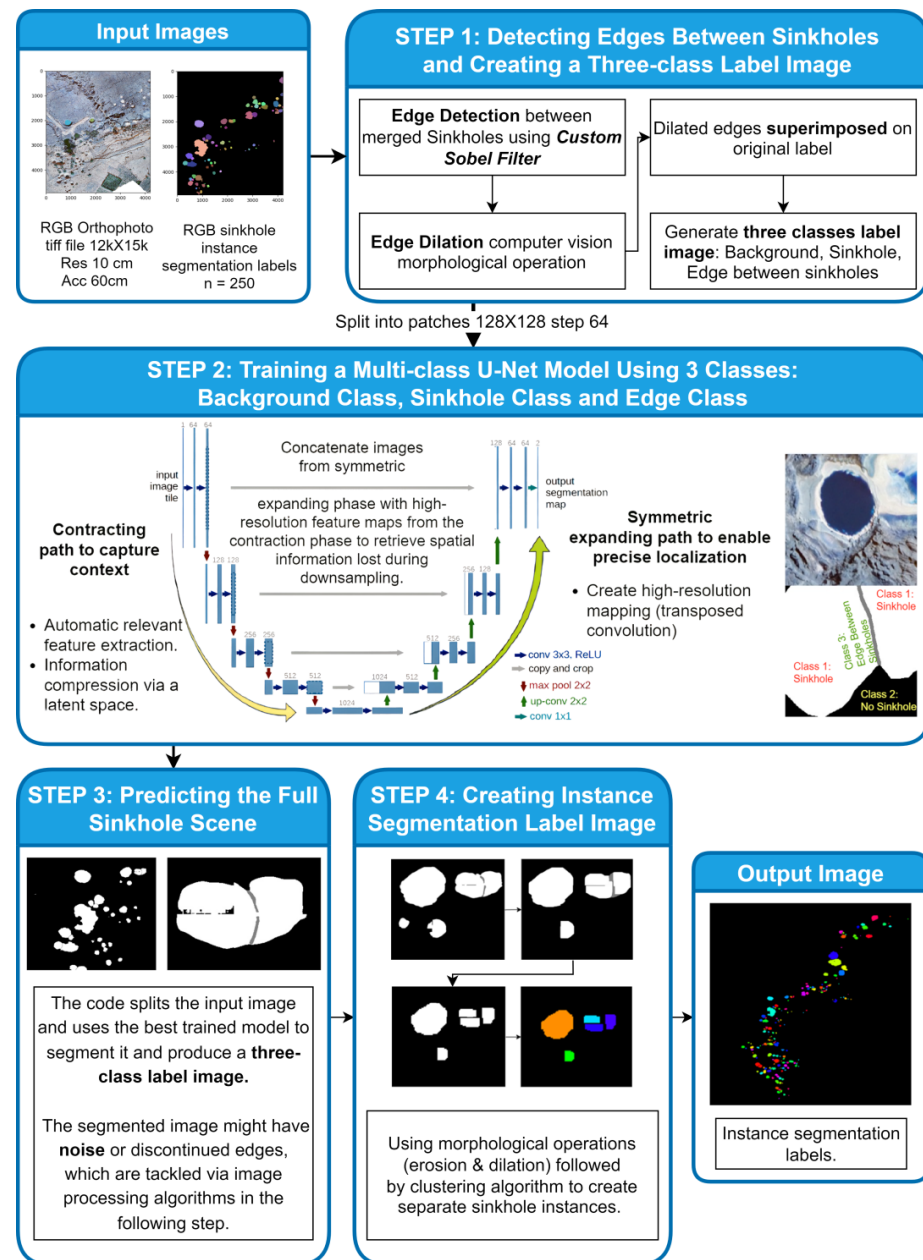
Figure 8. Overview of the sinkhole instance segmentation pipeline used in phase I of the study. This diagram illustrates the multi-stage process used to train a multi-class U-Net model, adapted from Ronneberger et al. [54]. The workflow begins with pre-processing the mask image (STEP 1) to detect edges between sinkholes, transforming the original two-class mask (Background and Sinkhole) into a three-class mask (Background, Sinkhole, and Edge Class). The input RGB orthophoto and the generated three-class mask are then used to train the multi-class U-Net model (STEP 2). The best-trained model is then applied to segment the full orthophoto, generating a semantically segmented mask (STEP 3). This mask undergoes a post-processing step (STEP 4) to generate the final instance segmentation mask image.

3.4. Transition from Higher- to Lower-Resolution Satellite Imagery

Our research pivots on the use of high-resolution drone images in the first phase. These images, owing to their level of detail, allowed for intricate mapping, annotation, and sinkhole detection. Training our deep learning models on this dataset ensured a robust understanding of sinkhole morphologies, their varied appearances across different terrains, and the intricate details that separate them from the surrounding landscape.

In the second phase, our research confronted a particular challenge for the field: How do we leverage the knowledge acquired from high-resolution images when faced with lower resolution satellite data? For this part, we turned to satellite images from the year 2022, which inherently lack the details present in drone samples. The transition involved several key modifications, which will be listed below.

3.4.1. Addressing Combined Sinkholes

In the first phase, employing high-resolution drone images, the distinction between combined (merged) sinkholes was prioritized for various reasons. Foremost, a clear understanding of individual sinkhole boundaries is pivotal for advanced sinkholes hazard mitigation and monitoring efforts. This demarcation helps in comprehending sinkhole merging patterns, useful for nuanced decision-making within sinkhole management activities. Training with the additional 'Edge' class broadens the model's exposure and is an extra step towards generalization. Delineation between merging sinkholes helps more accurate tracking and offers insights into sinkhole growth and future potential developments.

However, as the study transitioned to low-resolution satellite images in the second phase, adjustments were imperative. The reduced granularity of these images constrains the discernment of boundaries between closely clustered sinkholes. Thus, recognizing them as a unified instance became more accurate and avoided data extraction errors. This approach better aligns with practical scenarios where the overarching objective is to identify a broader hazardous area rather than discrete sinkholes. Also, given the guidance of annotations for satellite images through centre points from the high-resolution dataset, an attempt to define boundaries in clustered sinkholes could jeopardize annotation consistency. Lastly, considering the limited number of combined sinkholes in the first place, recognizing them as singular instances alleviated the data imbalance issue, ensuring a more adequate dataset for model training.

3.4.2. Modifications in Data Pre- and Post-Processing

Transitioning from high-resolution drone to low-resolution satellite images required some pre-processing steps to be modified. For the satellite imagery, histogram equalization was applied to enhance image contrast, and additionally mean subtraction was further completed to centre the pixel values, optimizing it for transfer learning. An important difference to reiterate is that the drone-based pre-processing method put an emphasis on identifying combined sinkhole boundaries, employing edge detection and dilation techniques to label transitions between sinkholes. In contrast, the satellite case, aligned with the decision to consider combined sinkholes as singular entities, omitted these steps, accommodating the lower resolution limitations and the goal for robust annotations. Thus, only two classes were used in this approach: 'Sinkhole' class, and 'Background' or 'Non-sinkhole' class.

3.4.3. Transfer Learning and Freezing of Certain U-Net Layers for the Satellite Case

In transitioning to satellite imagery, we applied transfer learning by initializing our satellite experiment models with the best weights obtained from the drone experiments and continuing training with satellite data. This approach involved strategic decisions on which layers to freeze and which to fine-tune. The key scenarios were as follows:

Freezing Initial Encoder Layers: By freezing the early layers, we took advantage of the recognition capability of basic features, e.g., patterns and textures with various angles learned in the drone training phase. We assume that these fundamental features are generally transferable and useful across different datasets.

Freezing Half of the Encoder Layers: This strategy extends beyond basic features, transferring more complex feature combinations learned from the drone data. We assume that the effectiveness of this method varies, as these complex features may or may not be as relevant for satellite data.

Freezing the Entire Encoder: Here, only the decoder was fine-tuned. We anticipated potential limitations since the encoder's ability to adapt to the complex, special features of the satellite dataset was restricted.

Unfreezing the Entire Encoder: This scenario entailed training on satellite data with all layers of the U-Net, including both the encoder and decoder. This approach allows for comprehensive fine-tuning using the new data, benefiting from the efficient starting point provided by the drone-trained weights. A good starting point for the weights also ensures quicker convergence to an optimal set of weights for the satellite dataset case. Although this method allows the model to learn new features from the satellite data, it may lead to some loss of previously learned information from the drone data. However, with this partial loss of information we gain also the benefit of adapting more flexibly to the new datasets, taking advantage at the same time of good initialization points. We can minimize this partial loss to some extent by choosing a more gradual re-training process, with smaller learning rates. Comparing the results from the 'Unfreezing the Entire Encoder' experiment to the other experiments provides valuable empirical insights into the trade-offs between potential risks and gained benefits of this approach.

3.5. Model Evaluation

The model's performance and accuracy in detecting and segmenting sinkholes from satellite and drone images were evaluated using multiple performance metrics. Considering the safety risks associated with undetected sinkholes (**False Negatives**) and the potential costs of monitoring **False Positives**, the metrics prioritized minimizing false negatives over false positives, i.e., we penalize more the cases of not detecting a 'Sinkhole' class. We computed the following metrics: model accuracy, specificity, per class precision, recall, F1 score, i.e., dice score. (refer to Table 2). Below are some brief definitions:

Confusion Matrix: A table used to describe the performance of a classification model by comparing the predicted class for each data instance to its actual class label [62].

True Positives (TP): These are pixels correctly identified as belonging to the target class. For the 'Sinkhole' class, it represents the number of pixels that are correctly identified as 'Sinkhole' in the prediction, while also classified as 'Sinkhole' in the ground truth.

True Negatives (TN): In our multi-class segmentation context, *TN* for a specific class refers to pixels that are correctly identified as not belonging to that class. To calculate it, we assume all pixels not involved in *TP*, *FP* and *FN* for a class are *TN*s.

False Positives (FP): These are pixels incorrectly labelled as belonging to the target class. For the 'Sinkhole' class, it represents the number of pixels that do not actually belong to a sinkhole, but are predicted as such.

False Negatives (FN): These are pixels that belong to the target class, but are not identified as such. For the 'Sinkhole' class, it represents pixels that are truly part of a sinkhole, but missed (i.e., predicted as either 'Background' or 'Edge').

Specificity: measure of the model's ability to correctly identify true negatives (*TN*), i.e., correctly predict the absence of a condition. It is calculated as:

$$Specificity = TN/(TN + FP)$$

Recall (also known as **Sensitivity**): represents the model's ability to correctly identify all actual instances of a specific class. It is the percentage of correctly predicted class pixels out of the total existing pixels of that class. For the 'Sinkhole' class, it is calculated as:

$$Recall_{sinkhole} = TP_{sinkhole}/(TP_{sinkhole} + FN_{sinkhole})$$

Precision: the percentage of correctly predicted class pixels out of all pixels predicted as the class of interest. For the 'Sinkhole' class, it is calculated as:

$$Precision_{sinkhole} = TP_{sinkhole}/(TP_{sinkhole} + FP_{sinkhole})$$

F1 Score: Harmonic mean of precision and recall for each class. F1 score is used to find an equilibrium between the reliability of positive predictions and the model's ability to detect positives. For the 'Sinkhole' class, it is calculated as:

$$F1_{sinkhole} = 2 \times (Precision_{sinkhole} \times Recall_{sinkhole})/(Precision_{sinkhole} + Recall_{sinkhole})$$

Accuracy: The proportion of correctly identified pixels for a specific class (both *TP* and *TN*) relative to the total number of pixels in the image. It is calculated as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN).$$

Table 2. Performance metrics for the developed models.

(A) Phase I—Train Models with Drone Images Exploring Different Loss Functions					
Focal Loss with Gamma = 1					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	97.477	96.222	99.937	96.846	99.751
Edge	17.117	17.244	99.996	17.18	99.989
Background	99.909	99.944	96.622	99.927	99.803
Focal Loss with Gamma = 2					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	97.302	96.587	99.932	96.943	99.764
Edge	16.962	15.433	99.997	16.161	99.99
Background	99.915	99.941	96.86	99.928	99.804
Focal Loss with Gamma = 5					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	97.242	96.372	99.93	96.805	99.752
Edge	16.498	15.63	99.997	16.052	99.989
Background	99.913	99.94	96.783	99.927	99.798
Weighted CE					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	89.547	90.642	99.73	90.091	99.273
Edge	7.809	10.79	99.995	9.06	99.987
Background	99.806	99.708	92.138	99.757	99.243
Non-weighted CE					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	97.364	96.791	99.933	97.077	99.776
Edge	18.31	17.244	99.997	17.761	99.99
Background	99.921	99.942	97.077	99.932	99.81
(B) Phase II—Fine-Tune the Best Drone Model using Satellite Images					
Freezing Initial Encoder Layers					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	90.415	92.055	99.973	91.228	99.95
Background	99.959	99.977	88.948	99.968	99.937
Freezing Half of the Encoder Layers					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	88.668	91.341	99.967	89.985	99.943
Background	99.956	99.969	88.203	99.963	99.925
Freezing the Entire Encoder					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	85.088	82.049	99.96	83.541	99.909
Background	99.928	99.952	80.636	99.94	99.88
Unfreezing the Entire Encoder					
	Precision	**Recall**	**Specificity**	**F1 Score**	**Accuracy**
Sinkhole	90.119	91.898	99.972	91	99.949
Background	99.959	99.975	88.843	99.967	99.934

4. Results

4.1. Experiment Setup

We used PyTorch and PyTorch Lightning as our frameworks of choice for all the experiments and took advantage of the scaled training capabilities they provide over multiple compute nodes and GPUs. Microsoft's Neural Network Intelligence (NNI) tool was utilized to explore the search space for different hyperparameters, such as batch size and learning rate, thus maximizing model performance on the available dataset. We can highlight the details of our **best experiment setup**: **a batch size of 64, 1000 epochs** (In deep learning, an **epoch** refers to one complete cycle through the entire training dataset during the model's training process. We cycle through several epochs to complete the training, i.e., tune the network weights) **of training using the Adam optimizer** (The **Adam optimizer** is an algorithm for optimizing neural networks, combining the advantages of AdaGrad and RMSProp to adjust learning rates based on recent gradient changes, enhancing the efficiency and speed of training) [17] **and a learning rate of 0.0003.** All convolutional layers were set to have a 3×3 kernel (In a convolutional neural network (CNN), a **kernel** is a small matrix used to apply a filter across an input image to extract features such as edges and textures by performing convolution operations) size, and padding was enabled to ensure that the output feature maps maintain the same spatial dimensions as the input. To increase the training data volume and enhance the network's generalization performance, data augmentation techniques such as rotation, horizontal and vertical flipping of the image patches were employed using the 'albumentations' library.

We made use of three distinct loss functions: non-weighted cross-entropy, weighted cross-entropy, and focal loss. We deliberately chose non-weighted cross-entropy to begin with, because the task of semantic segmentation can be viewed as a per-pixel multiclass classification task, where each pixel decides on a categorical label. Cross-entropy loss is therefore a desirable choice, especially for such a multi-class scenario. The alternative would have been dice loss, but since this one suffers from gradient instabilities, we decided in favour of nonweighted cross-entropy, with the added constraint of evaluating on the dice metric instead. Later on, when faced with data imbalance issues, we chose two more functions that would put more focus on the minority classes, while at the same time being natural extensions of the parent loss function, i.e., the nonweighted cross-entropy. The first choice was naturally weighted cross-entropy, which puts a fixed/static attention on the minority pixels. Subsequently, we chose a more dynamic/adaptive attention using the focal loss, to contrast it with the fixed scenario of weighted cross-entropy. Focal loss offers a more soft, gradual/adaptive attention on the minority pixels throughout the training procedure as evidenced by Lin et al. [63].

- **Non-Weighted Cross-Entropy**

  Calculated as:

  $$CE = -y\,log(p) - (1-y)log(1-p)$$

where y is the ground truth and p the predicted probability for class with label 1. This loss function measures the disparity between predictions and actual values, treating all classes equally.

- **Weighted Cross-Entropy**

  Changes the standard cross-entropy by introducing weights for classes:

  $$WCE = -w_1\,y\,log(p) - w_0\,(1-y)log(1-p)$$

with w being the weight assigned to each class. This approach gives higher importance to underrepresented classes. We assigned weights inversely proportional to the number of pixels in each class, i.e., $w = \frac{1}{counted\ number\ of\ pixels\ of\ a\ particular\ class}$.

- **Focal Loss**

Expressed as:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma log(p_t)$$

where α_t and γ are hyperparameters, and p_t is the model's estimated probability for the class with label t. Focal loss dynamically focuses on challenging misclassifications during training, adapting to problematic cases in a more responsive manner than weighted cross-entropy. This adaptability may allow for a quicker and more efficient training process, as it prioritizes difficult-to-classify instances on the fly. Note that Focal Loss is a generalization of the non-weighted cross-entropy loss, with $\gamma = 0$ and $\alpha_t = 1$ for all classes.

Progress monitoring involved computing the validation loss, and dice score (see Figure 9) after each epoch. The model checkpoint was updated each time the validation dice score improved. We used early-stopping to prevent overfitting, i.e., not to learn by heart the train dataset, which is usually correlated with an inability to generalize to other third-party datasets. More concretely, if the validation score did not improve after a tolerance threshold of 10 epochs, then we stopped training. The best model was evaluated quantitatively on the test dataset by calculating the test metrics, as well as qualitatively by storing the prediction maps for visual inspection (see Appendix A).
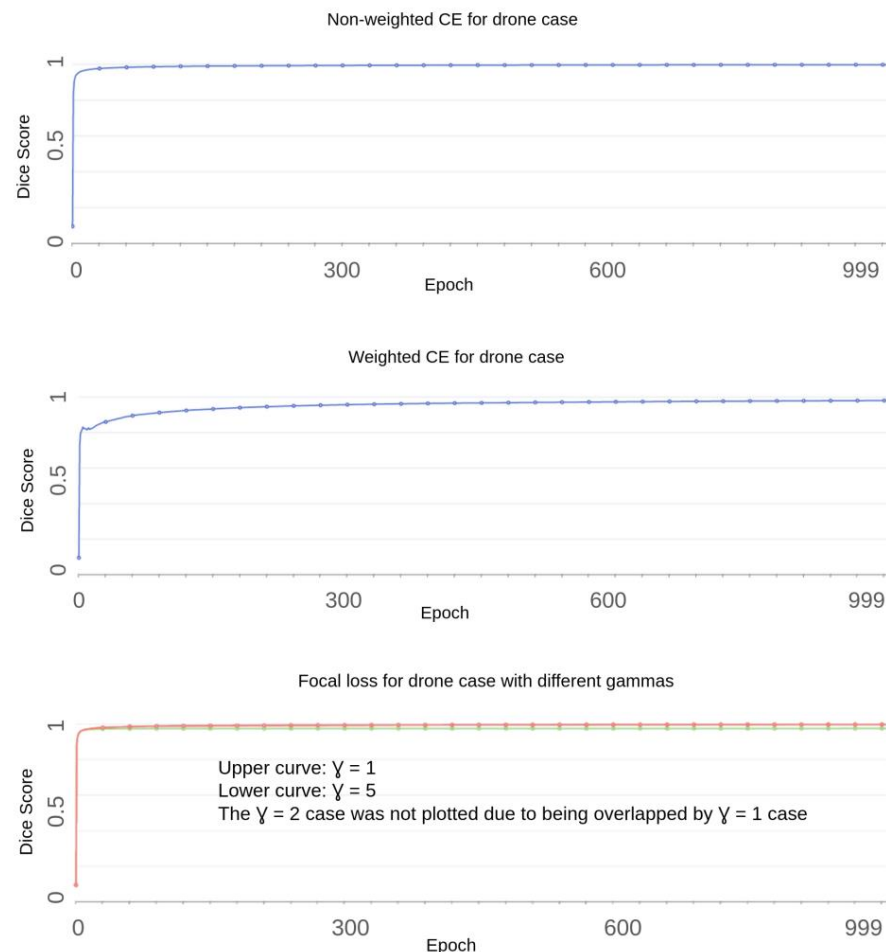
Figure 9. Model performance for Phase I as judged by the average dice score.

4.2. Performance of the Model

The initial results were obtained by training the model using the drone image dataset. Table 2 provides the outcomes from the various experiments performed. These results underscored the importance of choosing and tuning the appropriate loss functions, primarily due to the imbalanced nature of the dataset. To monitor the minority classes, i.e., sinkholes and edges, per-class metrics were reported. Regarding the accuracy metric, we would like

to point out that because of the very big number of background pixels, the 99% mark was consistently reached.

4.3. Performance Analysis across Datasets

In this section, we highlight how the model performed across the high-resolution drone and lower-resolution satellite imagery (See Table 2). This includes looking at the **recall** for the 'Sinkhole' class for accurate sinkhole detection (maximizing TP) and lowering risk by reducing missed detections (minimizing FN). In addition, we will consider the **F1** Score which serves as a key metric for balancing precision and recall, indicating the model's overall effectiveness in identifying sinkholes. Moreover, given the class imbalance, specificity and accuracy might be less indicative of model performance for sinkhole detection compared to recall and precision.

4.3.1. Phase I—Trained with Drone Images

The experiments demonstrate notable consistency in achieving high precision and recall for the 'Sinkhole' class across different loss functions, with the precision for 'Sinkhole' remaining above 89% across all experiments. The highest sinkhole recall, achieved using non-weighted CE, stands at 96.79%, alongside an F1 score of 97.08%. The 'Edge' class, on the other hand, reveals significantly lower precision and recall across all experiments, barely reaching the highest recall of 17.24% with both non-weighted CE and focal loss (Gamma = 1), and an F1 score of 17.761 achieved through non-weighted CE. Meanwhile, the 'Background' class consistently exhibits high performance, which can be attributed to its majority representation in the dataset (See Table 2A).

4.3.2. Phase II—Trained with Satellite Images

In the second phase of our research, we aimed to maintain the model's performance in identifying the 'Sinkhole' class on lower resolution satellite images via transfer learning.

Our experiments revealed a consistent performance in detecting and delineating sinkholes, where 'Freezing Initial Encoder Layers' was the most effective strategy, achieving a recall of 92.055% and an F1 score of 91.228%. This was closely followed by 'Unfreezing the Entire Encoder', then 'Freezing Half of the Encoder Layers', and finally, 'Freezing the Entire Encoder' showing the least effectiveness. In all experiments, the precision and F1 Score for the 'Sinkhole' class remained high, above 85.088% and 83.541%, respectively (See Table 2B). It is worth noting that the competitive results of 'Unfreezing the Entire Encoder' indicate that the model is capable of adapting to new data, flexibly altering all the drone experiment weights in a holistic manner, while still maintaining a reasonable level of performance.

5. Discussion

This work demonstrates the capability of our implemented U-Net-based pipeline to accurately detect sinkholes. The system's effectiveness in segmenting separate sinkhole instances with accurate detection of their boundaries was particularly evident when trained with high-resolution drone images. Scholars have highlighted the importance of high-resolution data in enhancing the accuracy of geological hazard detection [64] and our results confirm this perspective, yet we also show that considerable accuracy can be maintained even with lower resolution images through techniques like transfer learning. Throughout both deployment phases—from high-resolution drone to lower-resolution satellite imagery—the model maintained consistent high precision and F1 scores for the 'Sinkhole' class (Table 2). Such consistency under varied imaging conditions and resolutions is critical in minimizing potential risks associated with inaccurate sinkhole detection [3,65]. Our findings confirm the scalability and capability of the U-Net architecture to effectively detect sinkholes from aerial data, as previously noted by Mihevc and Mihevc [31]. In addition to the quantitative evaluations, we carried out qualitative assessments throughout visual inspection of prediction maps (see Appendices A and B), because sometimes just looking at the numbers may not convey the full picture.

5.1. Challenges in Sinkhole Edge Detection

One of the major challenges we faced was the accurate segmentation of edges between merged sinkholes, primarily due to class imbalance and the less distinct nature of these features compared to the sinkholes themselves. This difficulty aligns with the findings from Kang et al. [30], who noted the challenges in detecting sinkholes within narrowly defined areas and diverse datasets, and also echoes concerns noted by Nefeslioglu et al. [32] about the complexities involved in distinguishing closely spaced geological features due to overlapping characteristics.

Edges are inherently more challenging to detect than other classes, because they represent thin, often ambiguous transitions between distinct sinkholes, which can be difficult for the model to learn and generalize. In addition, this class contains significantly fewer training samples compared to the 'Sinkhole' and 'Background' classes, which constitutes an imbalance that affects the model's ability to accurately learn its characteristics. This imbalance is reflected in the lower recall and F1 scores for the 'Edge' class, as the model tends to have a bias towards the more prevalent classes.

One other aspect we can point out contributing to low 'Edge' recall and F1 scores, is that there is a very high inter-class similarity between sinkhole pixels and edge pixels. One can say that an edge pixel is actually a sinkhole pixel that belongs to several sinkholes simultaneously. Given that the sinkhole class can be regarded as subsuming the edge class in some sense, it is understandable that the neural network has difficulties predicting the edge class in particular.

Yet another important point is the fact that we apply the computer vision morphological operation of dilation as a pre-processing step to the thin edges derived via customized Sobel edge detection. This is accomplished to mark the transition regions between two merged sinkholes and emulate the natural uncertainty that even human experts experience when demarcating the exact boundaries where one sinkhole ends and the other one begins. The dilation, however, does have also a confusing effect for the CNN, because naturally some pixels from the sinkhole class are put in the edge set. Hence, the CNN will encounter some ambiguity stemming from occasional data mixing related to dilation. Nevertheless, we do believe in keeping the dilation step to encode uncertainty and let the network reduce this uncertainty in a data-driven manner, while we scale and accumulate more data over time, from new sources. And to mitigate the data mixing issue, the dilation operation can be coupled with an additional fuzzy logic block, where we label pixels probabilistically, e.g., a certain pixel is 70% sinkhole-like, and 30% close to a 'pure' edge. We would like to pursue this direction, as it resembles human intuition, amongst other things. Currently, we are forcing the CNN to draw a hard distinction between two ambiguous classes, when in fact we may benefit from a softer decision-making between the two.

Despite these challenges, the model showed a degree of success in edge classification, laying initial groundwork for further improvement in future studies. On the other hand, the 'Background' class demonstrated high performance, facilitated by its majority representation in the dataset. The ease of classifying 'Background' pixels also contributes to the model's overall effectiveness in distinguishing salient sinkhole features from their surroundings, which is critical for generating accurate sinkhole maps.

5.2. Handling Class Imbalance

Class imbalance in our dataset posed a significant challenge, affecting the model's ability to learn from less represented classes such as 'Sinkhole' and 'Edge'. The literature confirms that the effectiveness of machine learning models in environmental applications is heavily reliant on the balance and representation of classes within the training data [31,64]. Recognizing this, we have adopted strategies such as using weighted cross-entropy and focal losses. Both of these loss functions narrow the attention of the model in the initial cycles of the training towards the minority classes, either in a more fixed/static manner in the case of weighted cross-entropy loss, or more adaptively in the case of focal loss, as presented by Lin et al. [63]. However, adapting these strategies in practice proved to be

more challenging than expected and surprisingly: non-weighted cross-entropy provided better results for the 'Sinkhole' and 'Edge' minority classes with the least amount of energy (Table 2). We would like to note that so far, we took a principled approach, and in the case of weighted cross-entropy for example, we made the class weights inversely proportional to the number of class pixels available. But technically, these class weights can be searched more empirically, i.e., by brute force within a broader search space. The search space can also be extended for the gamma parameter in the case of focal loss as well. Therefore, we generally assume that by expanding the search space for the class-weights in the case of weighted cross-entropy loss; and gamma hyperparameter in the case of focal loss we might reach better local minima for our model. However, this of course comes at a cost of much more training resources and GPU time. For efficiency reasons therefore, we would like to expand the search space for these alternative loss function hyper-parameters once we gather more data for the edge class either from new sources, or from super-resolution techniques. We assume that applying our above-mentioned loss functions is indeed promising, provided that one reaches a data quantity threshold, i.e., one has a critical mass of samples available. In our case, the sinkhole pixels are very underrepresented with respect to the background, and the edge pixels are extremely underrepresented. Hence, we would like to collect especially more edge samples in the future and increase this type of data pool in particular.

Considering this, as well as the points mentioned in Section 5.1, dealing with class imbalance is a complex matter, requiring not only expanding the search space for the hyperparameters of our alternate loss functions, but also broadening the data pool for the minority classes, as well as better curating these samples to minimize data mixing, and employing fuzzy logic to encode uncertainty and enable soft decision-making, rather than hard demarcation of naturally ambiguous classes.

5.3. Effectiveness of Transfer Learning

Our research highlights the model's adaptability across different resolutions and imaging conditions through strategic application of transfer learning [53]. This adaptability is important for practical applications, ensuring that the model can be deployed in various real-world scenarios with different data quality and resolutions [65]. A key to our success was the strategic freezing and unfreezing of specific layers within the U-Net architecture, which played an important role in achieving high precision and recall for the 'Sinkhole' class. Especially beneficial was the 'Freezing Initial Encoder Layers' approach. It capitalized on the fundamental features recognized from the high-resolution drone imagery, effectively transferring this knowledge to interpret the lower-resolution satellite images. Karpatne et al. [65] and Ma and Mei [64] further reinforce the importance of transfer learning for a wider applicability across fields and aerial data distributions.

5.4. Model Generalisability to Other Karst Environments

The adaptability of our model to different geological settings and multi-resolution scenarios broadens its applicability and utility for geohazard management. However, computer vision models for landform mapping can produce unexpected predictions when applied to a different geographical area than that where they were first trained [33]. This so-called 'out-of-distribution' phenomenon is one of the greatest challenges for machine learning mapping and requires considerable attention. This is especially true in karst environments, whose landscape configuration is highly variable. Unique environments can develop within very small areas, and their characteristics depend upon many factors, including the lithology and structure of the host rocks, the present and past climates which have prevailed in a given karst area, and the surface and subsurface hydrological conditions [1,66,67].

The karst environments on the shores of the Dead Sea have formed by dissolution and physical erosion of subsurface evaporite deposits, which are interlayered with poorly consolidated alluvial and lacustrine sediments [34,35,37,40]. As the climate is very arid,

there is very little surface water or vegetation present in the study area, meaning that there is not really an epikarst layer present. Dissolution therefore is almost absent as a surface process: collapse into subsurface voids is the primary mechanism of sinkhole formation, along with surface sagging across broader areas, with wide areas of subsidence and coalescence of sinkholes forming larger depressions [8,36]. The resulting landscape is one in which optical imagery allows clear delineation of sinkholes by the human eye (Figures 2C,D and 6), particularly with respect to the open sinks which form at the margins of stream-channel meanders (Figure 5A). Although sinkholes do have different morphologies, and thus different visual characteristics when formed in the alluvial fan deposits, as compared to the lacustrine mud deposits (cf. Figure 10, [35], and Figure 5, [36]), they can both be accurately delineated from optical imagery alone, by our model.

However, this may not necessarily be the case in solutional karst environments, where shadowing and colour gradients between sinkholes and the background image are far less pronounced. In such environments, a hill-shaded elevation model is likely to be more suitable as input data for classifying sinkholes [31]. Furthermore, the general absence of vegetation at Ghor Al-Haditha also lends itself to sinkhole detection from optical imagery.

There is considerable scope for applying our model to other karst environments, though further training and validation would be required to ensure accurate transferability. Fine-tuning the model would have to be carried out on additional datasets that capture the variety of sinkhole morphologies occurring in different geological and climatic settings, along with different vegetation covers and optical characteristics. For example, in a forested karst landscape, our approach would likely require significant adaptation, as vegetation would obscure the true land surface. For this case, it might be possible to incorporate LiDAR and multispectral data, which can be corrected to remove vegetation [68]. In urban environments, occlusion of sinkholes would present additional challenges, as the visual appearance and morphology of sinkholes will differ from natural cavities due to the influence of anthropogenic structures such as buildings and vehicles [69]. It may be anticipated that, as the number of recognised sinkhole occurrences in urban areas has increased substantially in recent decades [70], adaptation of our model to urban landscapes may be especially important. Adaptive learning methods can be used to allow the model to dynamically adjust to new data distributions and enhance its performance in different environments. Techniques such as domain adaptation and domain generalization can help the model learn invariant features that are relevant across various settings [71].

6. Conclusions

Our research, focusing on the identification and mapping of sinkholes in the evaporite karst at Ghor Al-Haditha on the eastern shore of the Dead Sea, demonstrated the effective use of a system designed for geological structure recognition and centred around the U-Net architecture. The research was carried out in two phases. Initially, the model was trained, validated, and tested using high-resolution drone-based orthophoto images (0.1 m GSD) captured in December 2016 and covering 250 different sinkholes (see Figure 5F). In the second phase, the model was fine-tuned and tested on a larger dataset with lower resolution from a Pleiades Neo satellite image (0.3 m GSD) covering 1038 different sinkholes.

The methodology highlights a strategic layer freezing and unfreezing during the training process, which supports the model's adaptability to different image resolutions. Our dual-phase approach has consistently returned high recall and F1 scores for the 'Sinkhole' class under various imaging conditions. Notably, the highest recall in Phase I was achieved using non-weighted CE, at 96.79%, alongside an F1 score of 97.08%. In Phase II, the 'Freezing Initial Encoder Layers' strategy achieved a recall of 92.06% and an F1 score of 91.23%, showing the robustness and effectiveness across input scales.

Furthermore, the deliberate use of RGB-only visual bands in aerial data—previously considered as not useful by some authors [33]—proved to be promising in our methodology. This broadens the model's applicability and enhances scalability due to more readily available data inputs.

The model tries to address the technical challenge of class imbalance via the use of more sophisticated loss functions, such as weighted cross-entropy and focal loss. However, further fine-tuning of class weights and gamma is necessary for these loss functions to enhance the results beyond the non-weighted cross-entropy baseline. This, however, should be completed with a larger and better curated dataset for the edge class in particular. Additionally, given that we applied dilation as a preprocessing step to encode transition region uncertainty within merged or coalesced sinkholes, we intend to pursue fuzzy logic as a means towards soft decision-making for 'sinkhole vs edge' classification, to accommodate the high inter-class similarity of these minority classes. Since the sinkhole sample-set can be regarded as subsuming the edge sample-set, this would be a natural and promising next step to follow.

The successful application of our model to delineate sinkhole instances at Ghor Al-Haditha is a crucial initial step towards the integration of automatic sinkhole detection within a geohazard monitoring system along the eastern shore of the Dead Sea. Our model's ability to utilize detailed local data and transfer this knowledge to less detailed, more broadly available data sources is an important distinction from previous work. This adaptability of the model to imagery of lower granularity allows deployment to study areas where sub-decimetre resolution drone imagery is not available. With proper training and calibration, the model could be tested in other karst settings, with different geological and climatic contexts. Investigating the integration of multimodal data sources, such as LiDAR and multispectral maps, could further enhance the robustness and accuracy of sinkhole detection models in areas with occluding vegetation, offering a more comprehensive understanding of geohazard dynamics.

In conclusion, while our model has shown promising results in sinkhole segmentation, the identified challenges partly align with findings from other studies, such as those by Mihevc and Mihevc [31] and Kang et al. [30], and indicate the need for future research and iterative advancements in the field. These efforts will hopefully fill current gaps, and enhance the scientific understanding and technological applications of machine learning and artificial intelligence in geosciences, thereby improving ecological monitoring and hazard mitigation.

Appendix A. Comparative Analysis of Results from Drone Imagery Training

The following figures show the comparative analysis of the results from the developed model trained on drone imagery with three different loss functions: non-weighted cross-entropy, weighted cross-entropy, and focal loss.

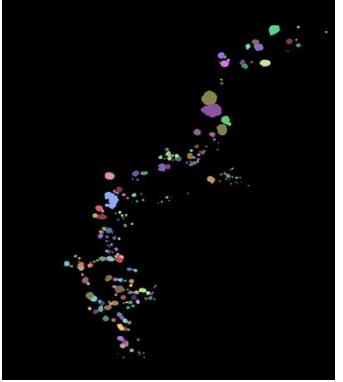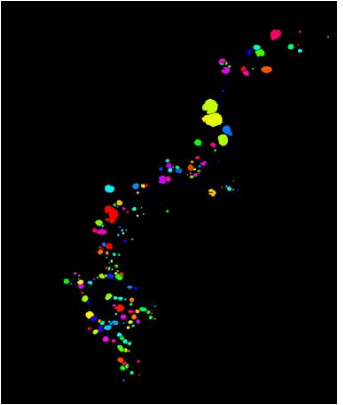**Table A1.** Semantic and Instance Segmentation Results (Phase I).
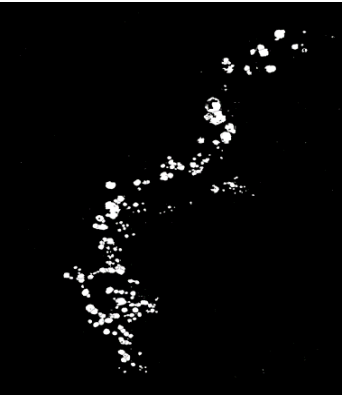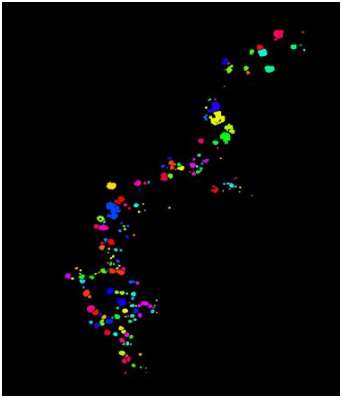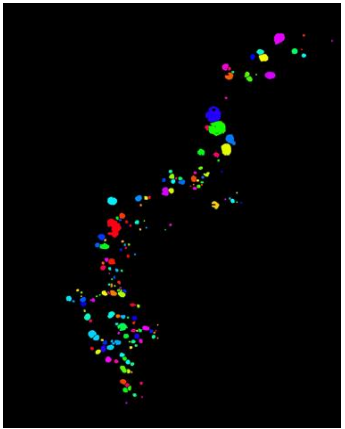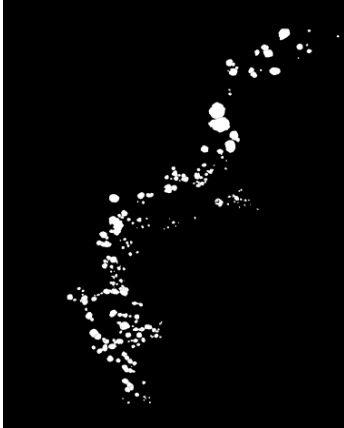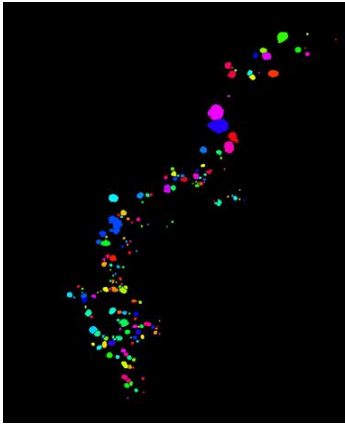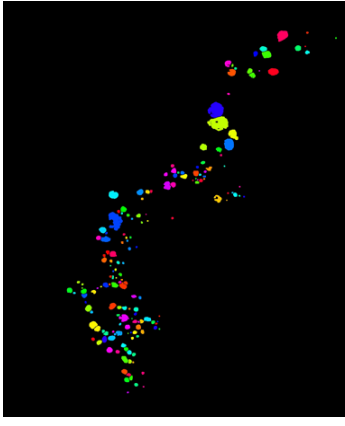
	Semantic Segmentation	Instance Segmentation
Ground truth		
Non-weighted CE		
Weighted CE		

Table A1. *Cont.*

	Semantic Segmentation	Instance Segmentation
Focal gamma 1		
Focal gamma 2		
Focal gamma 5		

Appendix B. Transfer-Learning Results

This appendix presents the results of our transfer-learning experiments conducted using satellite images from 2022. Two tables are included to illustrate these results.

Table A2: Satellite Image and Ground Truth Mask

The first table provides the original satellite image used in our study and the associated semantic segmentation ground truth mask, which was manually annotated by the researchers. This served as a reference for the expected outcomes of our segmentation models.

Table A3: Semantic and Instance Segmentation Results

The second table showcases the semantic segmentation and instance segmentation results obtained from the different transfer-learning experiments conducted during the second phase of our study. The experiments included the following strategies:

- Freezing Initial Encoder Layers
- Freezing Half of the Encoder Layers
- Freezing the Entire Encoder
- Unfreezing the Entire Encoder

These images provide a qualitative visual comparison of how each approach performed, highlighting the effectiveness and limitations of each transfer-learning strategy in segmenting sinkholes from satellite imagery.

Table A2. Satellite Image and Ground Truth Mask.
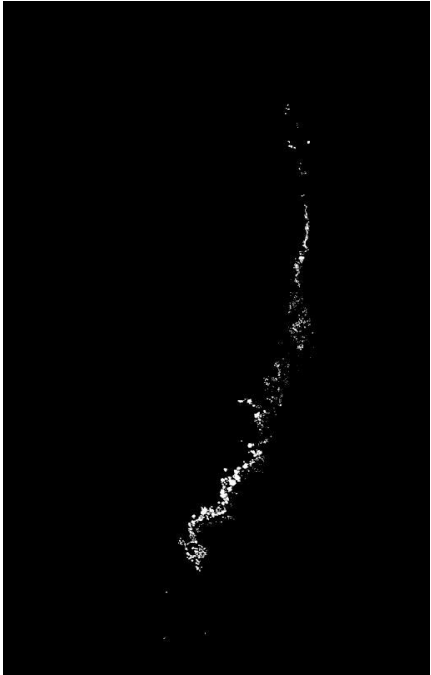
The Satellite Images from The Year 2022	Associated Ground Truth Semantic Segmentation Mask Image

Table A3. Semantic and Instance Segmentation Results (Phase II). Note: As the images are very large, we will provide a sample from a selected area (reflecting the same area covered by the drone dataset) to show the different results obtained from the different experiments.
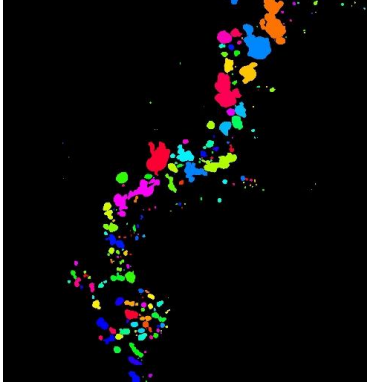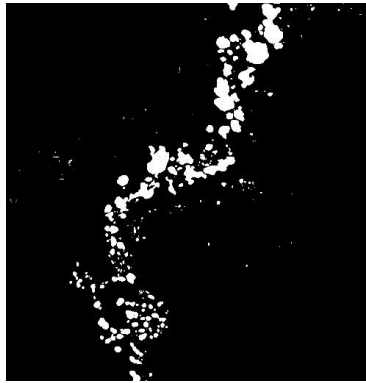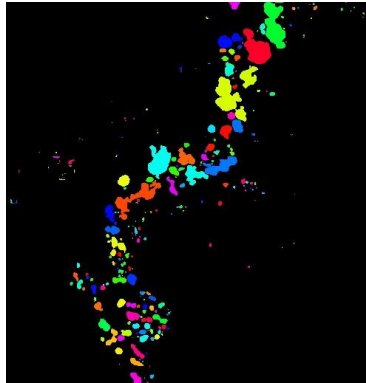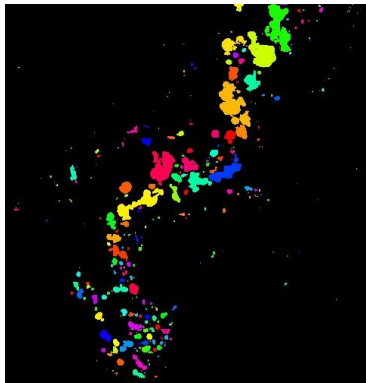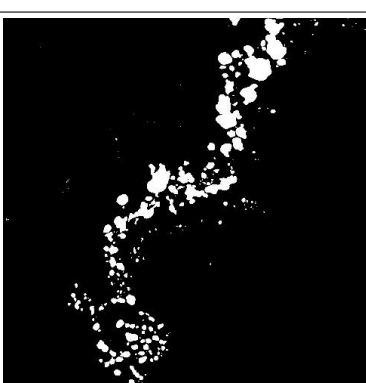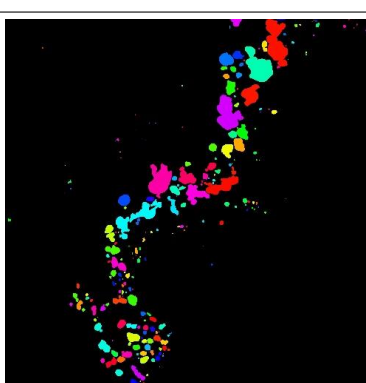
	Semantic Segmentation	Instance Segmentation
Freezing Initial Encoder Layers		

Table A3. *Cont.*

	Semantic Segmentation	Instance Segmentation
Freezing Half of the Encoder		
Freezing the Entire Encoder		
Unfreezing the Entire Encoder		

References

1. De Waele, J.; Gutiérrez, F. *Karst Hydrogeology, Geomorphology and Caves*; Wiley Blackwell: Hoboken, NJ, USA, 2022; ISBN 9781119605348.
2. Orhan, O.; Haghshenas Haghighi, M.; Demir, V.; Gökkaya, E.; Gutiérrez, F.; Al-Halbouni, D. Spatial and Temporal Patterns of Land Subsidence and Sinkhole Occurrence in the Konya Endorheic Basin, Turkey. *Geosciences* **2024**, *14*, 5. [CrossRef]
3. Gutiérrez, F.; Parise, M.; De Waele, J.; Jourde, H. A review on natural and human-induced geohazards and impacts in karst. *Earth-Sci. Rev.* **2014**, *138*, 61–88. [CrossRef]
4. Gutiérrez, F.; Cooper, A.H.; Johnson, K.S. Identification, prediction, and mitigation of sinkhole hazards in evaporite karst areas. *Environ. Geol.* **2008**, *53*, 1007–1022. [CrossRef]
5. De Waele, J.; Gutiérrez, F.; Parise, M.; Plan, L. Geomorphology and natural hazards in karst areas: A review. *Geomorphology* **2011**, *134*, 1–8. [CrossRef]
6. Galve, J.P.; Bonachea, J.; Remondo, J.; Gutiérrez, F.; Guerrero, J.; Lucha, P.; Cendrero, A.; Gutiérrez, M.; Sánchez, J.A. Development and validation of sinkhole susceptibility models in mantled karst settings. A case study from the Ebro valley evaporite karst (NE Spain). *Eng. Geol.* **2008**, *99*, 185–197. [CrossRef]
7. Galve, J.P.; Gutiérrez, F.; Lucha, P.; Guerrero, J.; Bonachea, J.; Remondo, J.; Cendrero, A. Probabilistic sinkhole modelling for hazard assessment. *Earth Surf. Process. Landf.* **2009**, *34*, 437–452. [CrossRef]

8. Sevil, J.; Gutiérrez, F. Morphometry and evolution of sinkholes on the western shore of the Dead Sea. *Implic. Susceptibility Assess. Geomorphol.* **2023**, *434*, 108732. [CrossRef]
9. Gutiérrez, F. Sinkhole Hazards. In *Oxford Research Encyclopedia of Natural Hazard Science*; Gutiérrez, F., Ed.; Oxford University Press: Oxford, UK, 2016; ISBN 9780199389407.
10. Bondesan, A.; Meneghel, M.; Sauro, U. Morphometric analysis of dolines. *IJS* **1992**, *21*, 1–55. [CrossRef]
11. Williams, P.W. The analysis of spatial characteristics of karst terrains. In *Spatial Analysis in Geomorphology*, 1st ed.; Chorley, R.J., Ed.; Routledge: London, UK, 1972; pp. 135–164. ISBN 9780429273346.
12. Huang, T.S.; Kohonen, T.; Schroeder, M.R.; Lotsch, H.K.V.; Maybank, S. *Theory of Reconstruction from Image Motion*; Springer: Berlin/Heidelberg, Germany, 1993; ISBN 978-3-642-77559-8.
13. Westoby, M.J.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **2012**, *179*, 300–314. [CrossRef]
14. Liu, X. Airborne LiDAR for DEM generation: Some critical issues. *Prog. Phys. Geogr. Earth Environ.* **2008**, *32*, 31–49. [CrossRef]
15. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, 51. [CrossRef]
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
17. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
20. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
21. O'Callaghan, J.F.; Mark, D.M. The extraction of drainage networks from digital elevation data. *Comput. Vis. Graph. Image Process.* **1984**, *28*, 323–344. [CrossRef]
22. Jenson, S.K.; Domingue, J.O. Extracting topographic structure from digital elevation data for geographic information-system analysis. *Photogramm. Eng. Remote Sens.* **1988**, *54*, 1593–1600.
23. Wang, L.; Liu, H. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 193–213. [CrossRef]
24. Wu, Q.; Liu, H.; Wang, S.; Yu, B.; Beck, R.; Hinkel, K. A localized contour tree method for deriving geometric and topological properties of complex surface depressions based on high-resolution topographical data. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 2041–2060. [CrossRef]
25. Wu, Q.; Deng, C.; Chen, Z. Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology* **2016**, *266*, 1–10. [CrossRef]
26. Wu, Q.; Lane, C.R.; Wang, L.; Vanderhoof, M.K.; Christensen, J.R.; Liu, H. Efficient Delineation of Nested Depression Hierarchy in Digital Elevation Models for Hydrological Analysis Using Level-Set Methods. *J. Am. Water Resour. Assoc.* **2019**, *55*, 354–368. [CrossRef]
27. Wu, Q. lidar: A Python package for delineating nested surface depressions from digital elevation data. *JOSS* **2021**, *6*, 2965. [CrossRef]
28. Lee, E.J.; Shin, S.Y.; Ko, B.C.; Chang, C. Early sinkhole detection using a drone-based thermal camera and image processing. *Infrared Phys. Technol.* **2016**, *78*, 223–232. [CrossRef]
29. Zhu, J.; Pierskalla, W.P. Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *J. Hydrol.* **2016**, *533*, 343–352. [CrossRef]
30. Kang, M.-S.; Kim, N.; Im, S.B.; Lee, J.-J.; An, Y.-K. 3D GPR Image-based UcNet for Enhancing Underground Cavity Detectability. *Remote Sens.* **2019**, *11*, 2545. [CrossRef]
31. Mihevc, A.; Mihevc, R. Morphological characteristics and distribution of dolines in Slovenia, a study of a lidar-based doline map of Slovenia. *AC* **2021**, *50*, 11–36. [CrossRef]
32. Nefeslioglu, H.A.; Tavus, B.; Er, M.; Ertugrul, G.; Ozdemir, A.; Kaya, A.; Kocaman, S. Integration of an InSAR and ANN for Sinkhole Susceptibility Mapping: A Case Study from Kirikkale-Delice (Turkey). *IJGI* **2021**, *10*, 119. [CrossRef]
33. Rafique, M.U.; Zhu, J.; Jacobs, N. Automatic Segmentation of Sinkholes Using a Convolutional Neural Network. *Earth Space Sci.* **2022**, *9*, 448. [CrossRef]
34. Abelson, M.; Yechieli, Y.; Baer, G.; Lapid, G.; Behar, N.; Calvo, R.; Rosensaft, M. Natural versus human control on subsurface salt dissolution and development of thousands of sinkholes along the Dead Sea coast. *J. Geophys. Res. Earth Surf.* **2017**, *122*, 1262–1277. [CrossRef]
35. Al-Halbouni, D.; Holohan, E.P.; Saberi, L.; Alrshdan, H.; Sawarieh, A.; Closson, D.; Walter, T.R.; Dahm, T. Sinkholes, subsidence and subrosion on the eastern shore of the Dead Sea as revealed by a close-range photogrammetric survey. *Geomorphology* **2017**, *285*, 305–324. [CrossRef]
36. Watson, R.A.; Holohan, E.P.; Al-Halbouni, D.; Saberi, L.; Sawarieh, A.; Closson, D.; Alrshdan, H.; Abou Karaki, N.; Siebert, C.; Walter, T.R.; et al. Sinkholes and uvalas in evaporite karst: Spatio-temporal development with links to base-level fall on the eastern shore of the Dead Sea. *Solid Earth* **2019**, *10*, 1451–1468. [CrossRef]

37. Al-Halbouni, D.; Watson, R.A.; Holohan, E.P.; Meyer, R.; Polom, U.; Dos Santos, F.M.; Comas, X.; Alrshdan, H.; Krawczyk, C.M.; Dahm, T. Dynamics of hydrological and geomorphological processes in evaporite karst at the eastern Dead Sea—A multidisciplinary study. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 3351–3395. [CrossRef]
38. Closson, D.; Abou Karaki, N. Salt karst and tectonics: Sinkholes development along tension cracks between parallel strike-slip faults, Dead Sea, Jordan. *Earth Surf. Process. Landf.* **2009**, *34*, 1408–1421. [CrossRef]
39. Shviro, M.; Haviv, I.; Baer, G. High-resolution InSAR constraints on flood-related subsidence and evaporite dissolution along the Dead Sea shores: Interplay between hydrology and rheology. *Geomorphology* **2017**, *293*, 53–68. [CrossRef]
40. Yechieli, Y.; Abelson, M.; Bein, A.; Crouvi, O.; Shtivelman, V. Sinkhole "swarms" along the Dead Sea coast: Reflection of disturbance of lake and adjacent groundwater systems. *Geol. Soc. Am. Bull.* **2006**, *118*, 1075–1087. [CrossRef]
41. Yechieli, Y.; Abelson, M.; Baer, G. Sinkhole formation and subsidence along the Dead Sea coast, Israel. *Hydrogeol. J.* **2016**, *24*, 601. [CrossRef]
42. Avni, Y.; Lensky, N.; Dente, E.; Shviro, M.; Arav, R.; Gavrieli, I.; Yechieli, Y.; Abelson, M.; Lutzky, H.; Filin, S.; et al. Self-accelerated development of salt karst during flash floods along the Dead Sea Coast, Israel. *JGR Earth Surf.* **2016**, *121*, 17–38. [CrossRef]
43. Arav, R.; Filin, S.; Avni, Y. Sinkhole swarms from initiation to stabilisation based on in situ high-resolution 3-D observations. *Geomorphology* **2020**, *351*, 106916. [CrossRef]
44. Ezersky, M.G.; Frumkin, A. Identification of sinkhole origin using surface geophysical methods, Dead Sea, Israel. *Geomorphology* **2020**, *364*, 107225. [CrossRef]
45. Al-Halbouni, D.; Holohan, E.P.; Taheri, A.; Schöpfer, M.P.J.; Emam, S.; Dahm, T. Geomechanical modelling of sinkhole development using distinct elements: Model verification for a single void space and application to the Dead Sea area. *Solid Earth* **2018**, *9*, 1341–1373. [CrossRef]
46. Al-Halbouni, D.; Holohan, E.P.; Taheri, A.; Watson, R.A.; Polom, U.; Schöpfer, M.P.J.; Emam, S.; Dahm, T. Distinct element geomechanical modelling of the formation of sinkhole clusters within large-scale karstic depressions. *Solid Earth* **2019**, *10*, 1219–1241. [CrossRef]
47. Schulten, H.Z.; Watson, R.A.; Al-Halbouni, D.; Al-Rabayah, O.A.-R.; Abdulla, F.; Holohan, E.P. Dynamics of sinkhole and uvala development on the eastern shore of the Dead Sea, 1980–2022. In Proceedings of the EGU23, the 25th EGU General Assembly, Vienna, Austria, 23–28 April 2023.
48. El-Isa, Z.; Rimawi, O.; Jarrar, G.; Abou Karaki, N.; Taqieddin, S.; Atallah, M.; Seif El-Din, N.; Al Saed, A. *Assessment of the Hazard of Subsidence and Sinkholes in Ghor Al-Haditha Area*; University of Jordan: Amman, Jordan, 1995.
49. Al-Halbouni, D.; AlRabayah, O.; Rüpke, L. A Vision on a UNESCO Global Geopark at the Southeastern Dead Sea in Jordan—Geosites and Conceptual Approach. *Land* **2022**, *11*, 549. [CrossRef]
50. Al-Halbouni, D.; AlRabayah, O.; Nakath, D.; Rüpke, L. A Vision on a UNESCO Global Geopark at the Southeastern Dead Sea in Jordan—How Natural Hazards May Offer Geotourism Opportunities. *Land* **2022**, *11*, 553. [CrossRef]
51. Polom, U.; Alrshdan, H.; Al-Halbouni, D.; Holohan, E.P.; Dahm, T.; Sawarieh, A.; Atallah, M.Y.; Krawczyk, C.M. Shear wave reflection seismic yields subsurface dissolution and subrosion patterns: Application to the Ghor Al-Haditha sinkhole site, Dead Sea, Jordan. *Solid Earth* **2018**, *9*, 1079–1098. [CrossRef]
52. Sevil, J.; Gutiérrez, F. Temporal variability of sinkhole hazard illustrated in the western shore of the Dead Sea. *Nat. Hazards* **2024**, *114*, 2395–2414. [CrossRef]
53. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
54. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. Available online: http://arxiv.org/pdf/1505.04597 (accessed on 17 June 2024).
55. He, L.; Ren, X.; Gao, Q.; Zhao, X.; Yao, B.; Chao, Y. The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognit.* **2017**, *70*, 25–43. [CrossRef]
56. Kurian, N.C.; Lohan, A.; Verghese, G.; Dharamshi, N.; Meena, S.; Li, M.; Liu, F.; Gillet, C.; Rane, S.; Grigoriadis, A.; et al. Deep Multi-Scale U-Net Architecture and Label-Noise Robust Training Strategies for Histopathological Image Segmentation. 2022. Available online: http://arxiv.org/pdf/2205.01777 (accessed on 17 June 2024).
57. Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D.J.; Norouzi, M. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4713–4726. [CrossRef]
58. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
59. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1483–1498. [CrossRef]
60. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026.
61. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

62. Görtler, J.; Hohman, F.; Moritz, D.; Wongsuphasawat, K.; Ren, D.; Nair, R.; Kirchner, M.; Patel, K. Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels. In *CHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2022; pp. 1–13.

63. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

64. Ma, Z.; Mei, G. Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth-Sci. Rev.* **2021**, *223*, 103858. [CrossRef]

65. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1544–1554. [CrossRef]

66. Palmer, A.N. *Cave Geology*; Cave Books: Dayton, OH, USA, 2007; ISBN 978-0939748662.

67. Ford, D.; Williams, P.W. *Karst Hydrogeology and Geomorphology*; Wiley: Chichester, UK, 2007; ISBN 978-0-470-84996-5.

68. Kobal, M.; Bertoncelj, I.; Pirotti, F.; Dakskobler, I.; Kutnar, L. Using lidar data to analyse sinkhole characteristics relevant for understory vegetation under forest cover-case study of a high karst area in the dinaric mountains. *PLoS ONE* **2015**, *10*, e0122070. [CrossRef]

69. Vennari, C.; Parise, M. A Chronological Database about Natural and Anthropogenic Sinkholes in Italy. *Geosciences* **2022**, *12*, 200. [CrossRef]

70. Pellicani, R.; Spilotro, G.; Gutiérrez, F. Susceptibility mapping of instability related to shallow mining cavities in a built-up environment. *Eng. Geol.* **2017**, *217*, 81–88. [CrossRef]

71. Gong, R.; Li, W.; Chen, Y.; van Gool, L. Dlow: Domain flow for adaptation and generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2477–2486.