Automatic Creation of High Fidelity Open Terrain Digital Twins for Off-Road Autonomous Vehicles Training and Validation

Ido Ariav, David Zaphir
Elbit Systems LTD.
Netanya, Israel
{ido.ariav,david.zaphir}@elbitsystems.com

Asaf Avinoam, Alon Faraj, Yisachar Shapira
Elbit Systems LTD.
Netanya, Israel
{asaf.avinoam,alon.faraj,yisachar.shapira}@elbitsystems.com

ABSTRACT

High-fidelity open-terrain digital twins are widely used in various applications such as training and validating artificial intelligence, mission rehearsal, and traversability prediction. However, creating large-scale, high-fidelity digital replicas of open terrain scenarios requires expert intervention and tedious and time-consuming work. We present a novel algorithm and implementation for the automatic and efficient generation of large-scale, high-fidelity open-terrain digital twins. The algorithm relies on standard, publicly available geospatial data, such as a 2D raster map obtained from aerial or satellite images and possibly a digital surface model (DSM). This algorithm comprises three consecutive processes: First, an AI method converts the raster map into a material map, where each pixel represents a predefined material (road, soil type, rock type, vegetation type, etc.). Then, an enhancement process combines the material map with the raster map and DSM to create a Multi-Layered Geospatial Representation (MLGR). Finally, an automatic pipeline converts the MLGR into a simulated environment that is perceptually realistic and geo-specific. No specialized graphic design or programming skills are required for these processes. Our automated pipeline is demonstrated with Unreal Engine 5, although other simulation platforms can also be used. Our results show that high-fidelity open-terrain digital twins can be generated automatically with minimal human intervention. In rigorous field tests, we demonstrate the applicability of these digital twins for AI training and validation, AV traversability analysis, and more. Finally, we discuss limitations and future work.

ABOUT THE AUTHORS

Ido Ariav is leading the AI team at Elbit Systems C4I, with almost 10 years of experience in research and development of various AI and Computer Vision solutions. Ido's main interest areas include semantic segmentation, depth prediction, and autonomous platforms. Ido received his B.Sc, M.Sc, and Ph.D. (summa Cum Laude) degrees in electrical engineering from the Technion – Israel Institute of Technology, Haifa, Israel.

David Zaphir is a product leader at Elbit Systems C4I, possessing over 20 years of experience as a development and product leader in the fields of computer graphics, visual simulation, GIS, robotics, and artificial intelligence. For the past six years, he has been leading the development of Elbit Systems C4I's robotic kit for autonomous vehicles, with a primary focus on the challenging task of using simulation for training and validating AI-based autonomous solutions.

Asaf Avinoam is a researcher in the field of Deep Neural Networks at Elbit Systems. He has 5 years of experience in developing various AI solutions in Computer Vision for Robotics & GIS applications. Asaf holds a B.Sc degree in Electrical & Electronics Engineering from Bar-Ilan University and is currently in the final stages of his M.Sc.

Alon Faraj is a leading software development specialist at Elbit Systems C4I for the last 8 years. He has extensive experience in simulation, ROS, cloud, autonomous vehicles stack, hardware, and integration. Alon received his B.Sc in Electrical and Electronics Engineering from Tel Aviv University.

Yisachar Shapira is leading the Operation Research & Simulation activities in Elbit Systems C4I. He holds an M.Ba in Business Management from Ben Gurion University and a B.Sc in Electronic & Electronics Engineering from Ariel University.

Automatic Creation of High Fidelity Open Terrain Digital Twins for Off-Road Autonomous Vehicles Training and Validation

Ido Ariav, David Zaphir
Elbit Systems LTD.
Netanya, Israel
{ido.ariav,david.zaphir}@elbitsystems.com

Asaf Avinoam, Alon Faraj, Yisachar Shapira
Elbit Systems LTD.
Netanya, Israel
{asaf.avinoam,alon.faraj,yisachar.shapira}@elbitsystems.com

INTRODUCTION

Demand for geo-specific virtual environments and simulations has grown substantially in recent years. Such simulated "digital twins" can be used for various use cases, e.g., mission rehearsal, training, and even validation of developed artificial intelligence (AI) capabilities. However, generating a high-fidelity, visually compelling geo-specific simulated environment is challenging, requiring much expert manual work.

The challenge of automatically generating high-fidelity virtual environments has been studied for many years. Some approaches suggested using laser scanners for terrain modeling (Elmqvist, Jungert, Lantz, Persson, & Soderman, 2001). Other approaches reconstruct the virtual environment from aerial oblique images (Yang et al., 2022; Spicer, McAlinden, Conover, & Adelphi, 2016). While both approaches excel at providing precise measurements of ground and solid objects' geometry, their effectiveness is restricted when it comes to reconstructing highly intricate objects such as trees and bushes. Moreover, the generated virtual environment lacks semantic significance, which is crucial for fundamental simulation operations such as physical interactions. To address these limitations, certain studies have proposed the utilization of deep learning techniques to classify point clouds into predefined categories that may be extracted into segmented objects (M. Chen et al., 2020, 2021). These approaches are highly dependent on the availability of the necessary photogrammetric data, posing scalability challenges.

Several works propose to partially automate the generation of virtual environments from globally available satellite imagery (Hollosi, Menzel-Berger, Walter, & Lahm, 2022; Bodhiswatta Chatterjee & Patel, 2020). However, these approaches are focused primarily on urban environments, prioritizing tasks such as the accurate reconstruction of buildings. In addition, the majority of these studies do not adequately address the representation of open-terrain environments, which are typically very complex, crowded, and unorganized.

This paper presents a fully automated pipeline for creating open terrain geo-specific simulated environments from readily available aerial imagery and a digital surface model (DSM). First, an AI-based segmentation module classifies each pixel of the aerial imagery into a set of predefined categories. Next, we use the segmented material map, DSM, and aerial imagery to generate a Multi-Layered Geospatial Representation (MLGR) that incorporates both ground and land-cover elements. Finally, we use the MLGR to automatically generate a digital twin in any simulation engine. Figure 1 shows the different components in our pipeline.

Our main contributions are as follows: (1) We present a semantic segmentation module capable of classifying readily available aerial maps into semantic meaning and material. Moreover, our segmentation module outputs soft rather than hard classification, i.e., each pixel is assigned with the probability of belonging to each predefined category, improving the visual quality of the generated environment. (2) We introduce a workflow that generates MLGR automatically from the enhanced mapping materials. (3) The generated MLGR is automatically translated into high-fidelity 3D terrain in any simulation engine.

The remainder of this paper is organized as follows: In the Proposed Method section, we introduce our automatic pipeline for simulated terrain creation. In the Results section, we demonstrate and discuss our results in several different environments and use cases. Finally, in the last section, we conclude and discuss future research directions.

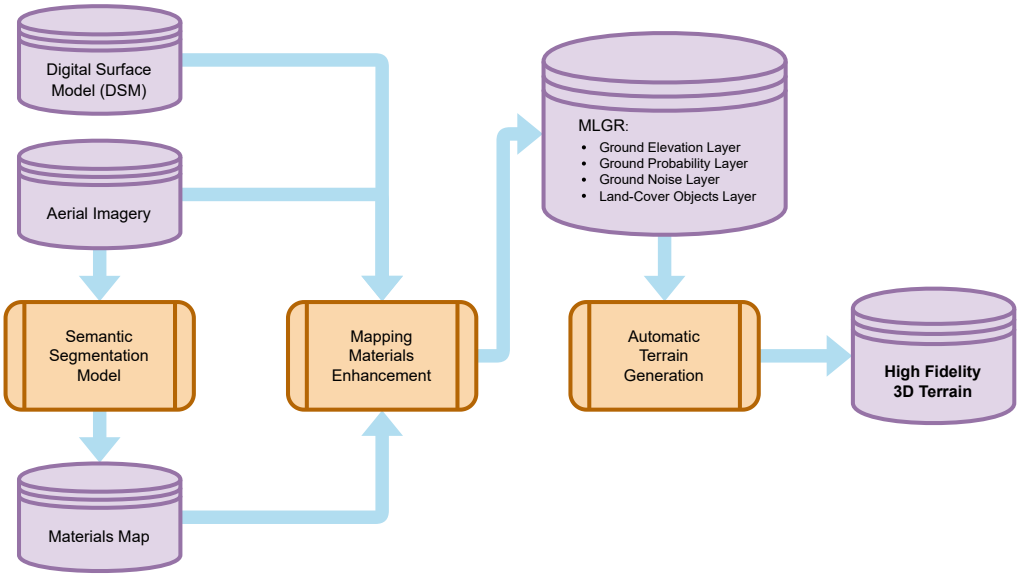


Figure 1: The proposed pipeline for the creation of open terrain geo-specific simulated environments.

PROPOSED METHOD

Material Map Creation

As part of our proposed Automated Simulated Terrain Creation Pipeline (ASTCP), we introduce a framework for segmenting readily available aerial mapping data into predefined categories (such as soils, rocks, man-made objects, and vegetation types). We refer to the output of this segmentation module as a material map. A material map identifies the category that each pixel in a raster image belongs to so that it can be used, e.g., in creating a geo-specific virtual environment, as will be discussed in the following sections. An example of aerial maps used in our system and the corresponding material map generated by our segmentation module can be found in Figure 2.

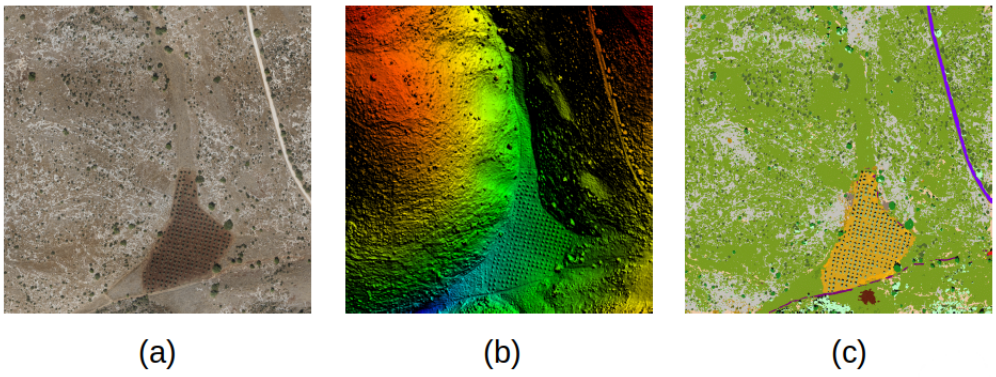


Figure 2: An example of mapping materials used in our system. (a) Aerial imagery. (b) Digital surface model (DSM) of the same area. (c) The output of our semantic segmentation module. (best viewed in color)

The semantic segmentation of 2D images is a fundamental problem in computer vision that has attracted much attention recently. It was previously modeled as a texture classification problem. The main focus was on determining and developing statistical approaches to quantify handcrafted texture features using different texture filters (Sobel filters, color histograms, filter banks, etc.). Several methods for segmenting images have been proposed based on deep neural networks (DNN), specifically convolutional neural networks (CNN), with recent advances in deep learning.

These recent advances, however, still face several challenges when applied to aerial imagery segmentation: (1) In comparison to perspective real-world images, satellite or airborne imagery (i.e., orthophotos) tend to have poor quality and

may look distorted or blurry. Additionally, orthophotos typically contain numerous small-scale objects, which could seriously challenge the segmentation module. (2) Due to the limited number of publicly available datasets for open terrain segmentation, training a CNN that requires a large amount of annotated data is challenging. Hence, we construct an annotated dataset to be used throughout our experiments, as explained in the Quantitative Results section. (3) Our ASTCP is required to work across a variety of geographical regions, each with its own land characteristics (such as deserts, forests, mountains, etc.). However, using a segmentation model pre-trained on existing datasets cannot be easily applied to a newly collected dataset. Different environments may have completely different types of land cover (e.g., different soils, different vegetation types, etc.). As a result, training a model that generalizes well presents additional difficulties.

To overcome these limitations, we propose a hierarchical multi-task deep neural network for the segmentation of aerial imagery. The proposed architecture is based on DeeplabV3+ (L.-C. Chen, Zhu, Papandreou, Schroff, & Adam, 2018), a fully convolutional network that has been successfully applied to many 2D image segmentation tasks. Our implementation utilizes the Xception backbone (Chollet, 2017). We design our hierarchical network to segment each pixel in a given orthophoto into a top-level and fine-grained category. It differs from a "flat" architecture, which assigns each pixel only a specific fine-grained class. Categories at the top level are referred to as "semantics", while categories at the finer level are called "materials". Table 1 shows our hierarchical categories structure. For the network to classify semantics and materials, we added two additional classification blocks, one for classifying materials and one for classifying rocks' morphology. Figure 3 shows a scheme of the proposed architecture. A classification block is constructed by concatenating two convolution layers of $K = 3$ kernel size. After each convolutional layer, batch normalization (Ioffe & Szegedy, 2015) and rectified linear unit (ReLU) activation are performed. As a final step, a convolution layer with a kernel size of $\hat{K} = 1$ is followed by a softmax layer to output the final class probabilities.

As the Results section explains, some images in our dataset are labeled purely for semantics, whereas others are also labeled for materials and morphology. To train our network, we use all annotations available for each image.

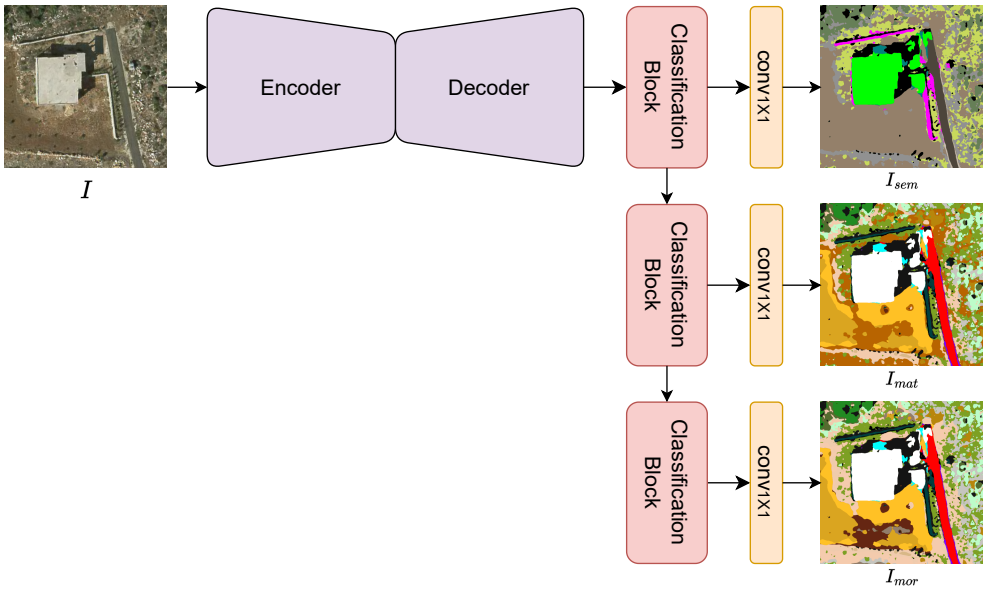


Figure 3: The proposed hierarchical multi-task segmentation model.

Formally, given an orthophoto $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$ as input, features are extracted from $\mathbf{I}$ via the backbone network. To control the trade-off between feature map size and computational complexity, atrous convolution is used in the last few blocks of the backbone. A global average pooling method is also applied to the last feature map of the backbone to incorporate global context information. Using such a network architecture, long-range feature maps can be extracted. Furthermore, atrous convolutions were found to improve segmentation by exponentially increasing the receptive field without reducing the spatial dimension. To improve training stability and efficiency, we include a batch normalization layer between each convolution layer and incorporate dropout (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) with a 50 percent dropout rate throughout the network's backbone.

An Atrous Spatial Pyramid Pooling (ASPP) network is applied to the extracted features to capture contextual information at multiple scales. We denote the ASPP network output as $\mathbf{F}_{ASPP} \in \mathbb{R}^{H/2 \times W/2 \times \hat{C}}$. Our experiments used $\hat{C} = 256$

Table 1: Semantics, Materials, and Morphologies used in our material map creation module.

Semantics	Materials	Morphologies
Clutter		
Shadow		
Buildings		
Vehicles		
Water		
Impervious Surface	Paved Road	
	Dirt Road	
	Pavement	
Low Vegetation	Dry Grassland	
	Green Grassland	
	Batha	
High Vegetation	Unirrigated Orchard	
	Irrigated Orchard	
	Garigue	
	Maquis	
Soil	TerraRosa	
	Clayey Soil	
	Rendzina	
	Clayey Deep Soil	
Rocks	Limestone	Rocky Terrain
		Boulder
		Stoney Terrain
		Bedded Rock
		Smooth Rock Slopes
		Rock Dip Slope
		Terrace
	Dolomite	Rocky Terrain
		Boulder
		Stoney Terrain
	Nari	Rocky Terrain
		Stoney Terrain
		Smooth Rock Slopes
		Rock Dip Slope
	Basalt	Rocky Terrain
		Boulder
		Stoney Terrain
	Chalk	Smooth Rock Slopes

channels. $\mathbf{F}_{ASPP}$ is fed to the three classification blocks as described above to generate class probabilities for semantics, materials, and morphology, denoted as $\mathbf{P}_{sem}$, $\mathbf{P}_{mat}$ and $\mathbf{P}_{mor}$, respectively. A combination of cross-entropy loss and focal loss (Lin, Goyal, Girshick, He, & Dollár, 2017) is used to train our network. We found that adding the focal loss term improved performance on the classification of small objects and classes with few annotations. We note that when an image only had semantic annotations, we trained only the semantic classification block in this training iteration without updating the materials or morphology blocks.

Using such a hierarchical multi-loss structure has several benefits. First, our experiments showed that training a network to predict both semantics and materials improved the prediction's accuracy as compared to training the network to predict only materials. Second, because our architecture learns semantics and materials, fine-tuning is easier for a new scenario since fewer data is required. Specifically, it involves adding materials annotations to a small amount of data, according to the soil types, vegetation types, etc., that exist in the new scenario. In addition, since we train the network only using RGB colors, the segmentation labels are determined solely by scene textures and spatial layout. Further details are provided in the Results section.

Another notable implication of the generated material map is that it allows for the implementation of physically based rendering (PBR) techniques (Pharr, Jakob, & Humphreys, 2016). PBR heavily relies on the radiometric properties of materials and introduces variations across different types of materials. Additionally, as we discuss in the Results section, the physical properties of materials can be used to predict vehicle maneuverability on diverse terrain surfaces.

Enhancement of Mapping Materials

The original mapping materials and the generated material map can be used directly to create a simulated environment. However, this will result in a simulation with rather poor visual quality and will hinder the usability of such an environment for the following reasons: (1) The mapping materials represent an orthographic view of the world, where the land cover is merged with its surroundings, unlike in the real world, where each land-cover object is distinct and separated from its surroundings. Furthermore, this approach overlooks the inclusion of 3D properties, such as tree branches and leaves. As a consequence, this approach introduces a limitation that results in compromised visual quality, particularly when observed at close distances. (2) The resolution of the mapping materials is usually insufficient for an accurate representation of the world, especially from close range. (3) Mapping materials represent the season, weather conditions, and time of day of their capture and might not be sufficient to simulate different environmental conditions.

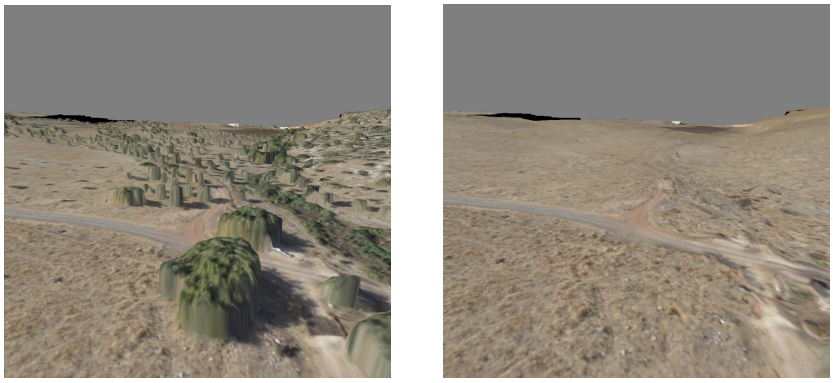
In order to address the specified challenges, we fuse vital information from each mapping material, resulting in much more detailed and refined representations. The enhanced geospatial materials are represented as MLGR and will be later used for the terrain composition. The process of generating the MLGR is described below.

Ground Elevation Layer

The original DSM represents world geometry in a 2.5-dimensional format, merging ground and land cover geometries together, unlike the physical reality where each land-cover object maintains separation from its surroundings. Our proposal introduces an automated solution that utilizes material classification to generate ground geometry from the DSM, producing a digital ground model (DGM). First, according to their classification, we remove land cover objects' geometry from the DSM. Then, we fill the geometry gaps by interpolating neighboring ground heights, using four directions conic search for each pixel in the eliminated areas. Figure 4 shows an example of land cover cleanup.

Ground Noise Layer

As will be described in the following section, we use the ground representation from the original aerial imagery as a noise texture to enhance the terrain visualization. The ground representation is generated from the original aerial imagery by replacing land cover objects and shadows with a mixture of their surroundings, using patch matching and in-painting techniques. An example of the cleaning process of the aerial imagery can be found in Figure 4.



(a) **(b)**

Figure 4: An example of the proposed land cover and shadow removal process. (a) Original terrain composed of aerial imagery layered on DSM (b) The enhanced terrain composed from processed aerial imagery and DGM.

Ground and Land Cover Layers

As stated in the Material Map Creation section, our segmentation module provides material classification predictions for individual pixels. Although the predictions derived from our segmentation module showcase a rich array of materials, assigning a solitary class to each pixel is insufficient for a comprehensive depiction for the following reasons: (a) Open

terrain is compound from ground and land cover objects, however from an orthographic point of view the land cover occludes the ground material. (b) Material classification accuracy is strongly dependent on aerial imagery resolution, which is usually insufficient for differentiating between open terrain materials such as soil and dry grass. (c) Open terrain ground is usually composed of a mixture of soils. Solitary representation of ground material leads to poor visual quality, especially in the boundaries between different materials.

In order to overcome the specified limitations we are taking the following actions: (a) using soft predictions rather than hard predictions, thus enabling multi-class probabilistic distribution for each classified pixel. (b) separating between ground and land cover objects according to their classification. Occluded ground areas are interpolated from their surroundings by applying four directions conic search. (c) Rocks in nature are integrated within terrain ground (like bedded rocks) or placed above terrain ground (like boulders), according to their morphology. Hence classified morphology is used to assign rocks to their relevant ground or land cover layer. The outcome is soil, ground-integrated rocks, and low vegetation layers represented by their geospatial probabilistic distribution and land cover layers. The dimension of each land cover object is estimated according to its classification and height difference between the original DSM and generated DGM layer.

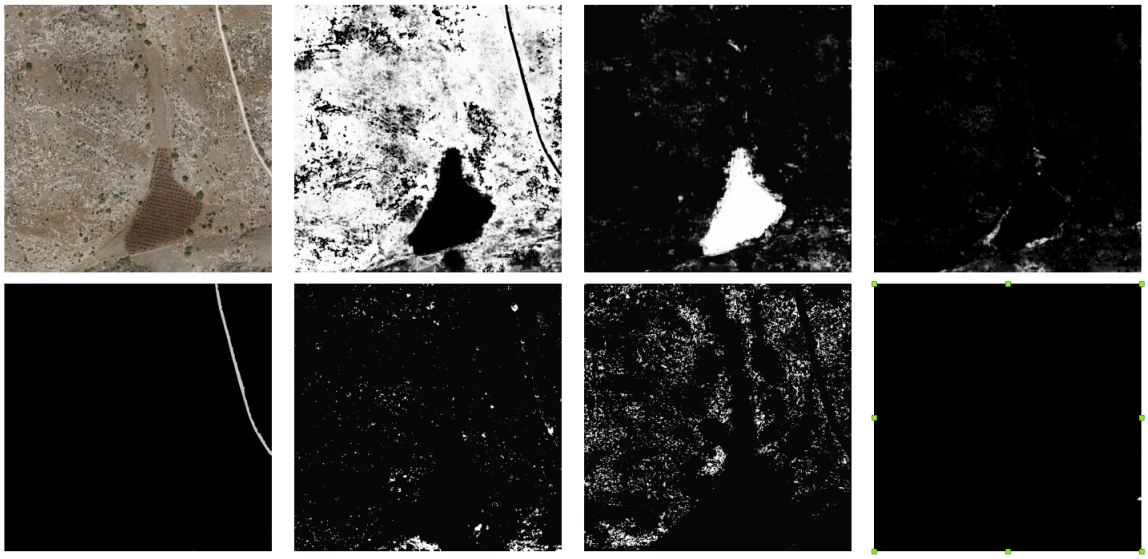


Figure 5: Probability layers of different materials. White areas indicate a higher probability.

Automatic Generation of Simulated Environment

The third phase of the ASTCP uses MLGR, which was introduced in the previous section, to generate a photo-realistic terrain. MLGR uses standard image and vector formats and, therefore, may be used as source data for the generation of terrain in any given format. For the scope of this work, we use Unreal Engine as the rendering engine. Our proposed terrain generation process employs the multi-layer geospatial representation as follows: the DGM layer is utilized to create the ground geometry, the ground and noise layers are employed for ground painting, the low vegetation layers are utilized for procedurally generating corresponding 3D objects, and the land cover layers are used to automatically generate instances of corresponding objects.

Whenever feasible, the generation process employs conventional methods such as constructing tiled world geometry using the DGM layer and positioning land-cover objects based on their corresponding representations. In certain scenarios, we employ advanced techniques, which are outlined as follows: The soil is depicted by a composite of ground materials based on their assigned probabilities. Each ground material is associated with a characteristic texture. The ground representation derived from aerial imagery is combined with soil textures to augment realism and reduce the repetition of the ground's typical texture. Bedded Rocks are generated according to their masks by applying advanced rendering techniques such as parallax occlusion mapping (Tatarchuk, 2006), and height field meshes (Lindstrom et al., 1996). Low vegetation is automatically generated according to the relevant probability layers using procedural foliage spawner. An example of terrain composed of ground, bedded rocks, and low vegetation can be found in Figure: 6.

Figure 6: Sample terrain generated via our automated pipeline.

RESULTS

Quantitative Results

For training the hierarchical classification module, we utilized 2 databases: (1) composed of 150 orthophoto tiles in the Arbel area in the north district of Israel. Each tile T_i consists of $512 \times 512 \times 3$ pixels with a 12.5cm/pixel resolution. Tiles were manually labeled according to semantics, material, and morphology. See Table 1 for a complete list of classes used. We used geologists to label materials and morphologies due to their specialized knowledge. (2) composed of 500 orthophoto tiles from several open terrain locations worldwide. Each tile $\hat{T}_i$ consists of $512 \times 512 \times 3$ pixels with a 25cm/pixel resolution. Tiles were manually labeled for semantics alone. We used 70% of the labeled tiles for training, 15% for testing, and 15% for validation. The proposed hierarchical network achieved a weighted average F1-score of 0.78 for the classification of semantics, materials, and morphologies of the Arbel test tiles.

A publicly available dataset (Rottensteiner et al., 2012), emphasizing urban areas, was used to compare the proposed hierarchical approaches to a "flat" model as a baseline. This dataset was used for testing only, and no data was used in training either network. The hierarchical network outperformed the baseline network with an F1-score of 0.81 versus 0.62, respectively. This cross-dataset evaluation confirmed that the hierarchical approach enables consistent training from diverse and balanced data sources, resulting in a more robust network with improved generalization capabilities.

Visual Results

This section provides an overview of ASTCP procedures and performance. The results are demonstrated in the Arbel by focusing on 2 points of view as displayed in Figure: 7. In Figures: 8 and 9, we describe the major steps of generating high fidelity simulated environment from MLGR. As described in the Enhancement of Mapping Materials section, we use the material map to remove land cover elements from both DSM and aerial imagery. Then we use ground probability masks to compose terrain ground texture. The enhanced aerial imagery and ground probability masks are integrated as a noise texture to decrease repetition and increase realism. The original DSM resolution is usually insufficient for representing complex geometries such as bedded rocks. By representing each rock material and morphology as a separate mask, we can apply advanced rendering techniques such as parallax offset mapping and height fields according to the complexity of the geometry of each rock morphology. Low vegetation is generated according to its probability mask. Land cover elements such as trees and bushes are generated according to their metadata. Since the rendering techniques are aligned with the material physical properties, the simulated environment remains realistic in different lighting conditions.

Use Cases

AI Training & Validation

Deep neural networks have provided a tremendous improvement in modern machine perception systems. However, it

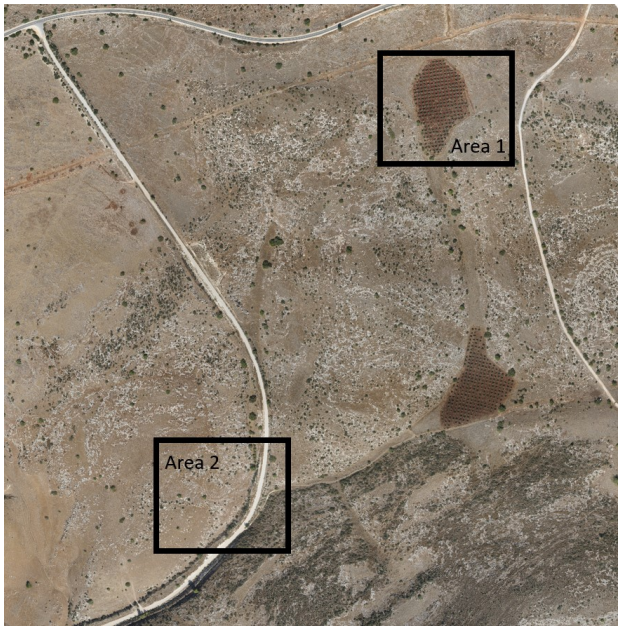


Figure 7: Aerial imagery of the Arbel area. Area 1 and 2 were selected to visualize the ASTCP performance

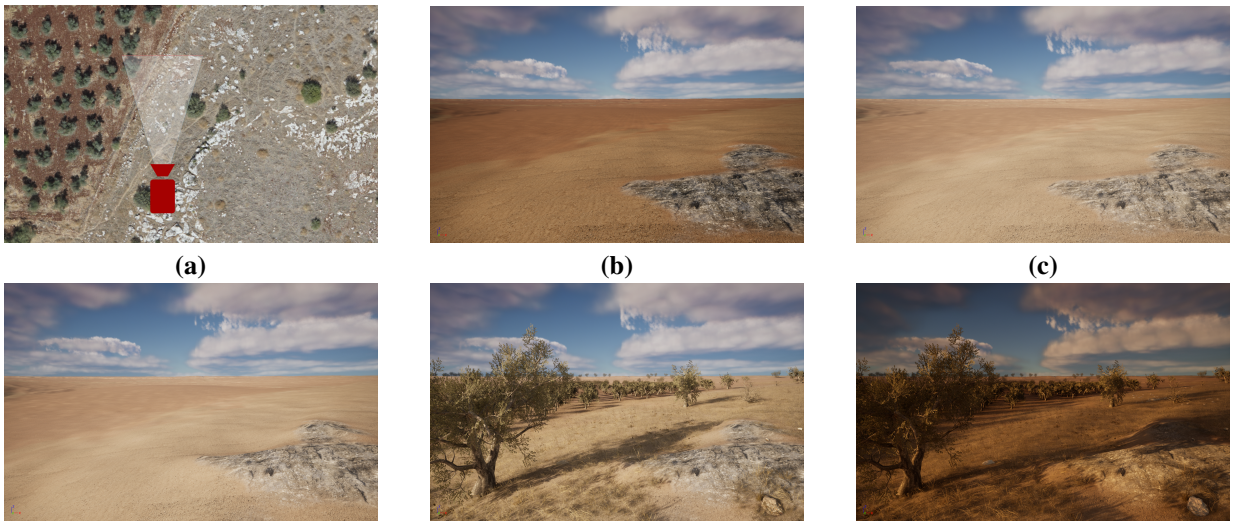


Figure 8: Visual results in Arbel area rendered from Cam1 point of view. (a) Cam1 field of view. (b) Rendering terrain ground by assigning a typical texture to each ground element. (c) Integrating noise texture generated from the aerial imagery to overcome ground repetition and enhance realism. (d) Enhancing bedded rocks geometrical representation. (e) Integrating land cover geospatial layers. (f) Rendering the scene from the same location using different lighting conditions.

is imperative to note that these neural models require enormous amounts of high-quality data to train on to take advantage of their exceptional learning capacity and improved generalization ability. This is even more prominent in recent architectures based on Transformers (Dosovitskiy et al., 2020).

Due to this, synthetic data is increasingly being used to train DNNs. Despite its vast potential, synthetic data presents several challenges, including the domain gap, sometimes called the reality gap. Training models on data collected in one domain generally results in poor accuracy in other domains. The domain gap between simulated and real data refers to how a DNN perceives synthetic data versus what it perceives from real data. This limits the performance of machine learning models trained only in simulation when deployed in the real world.

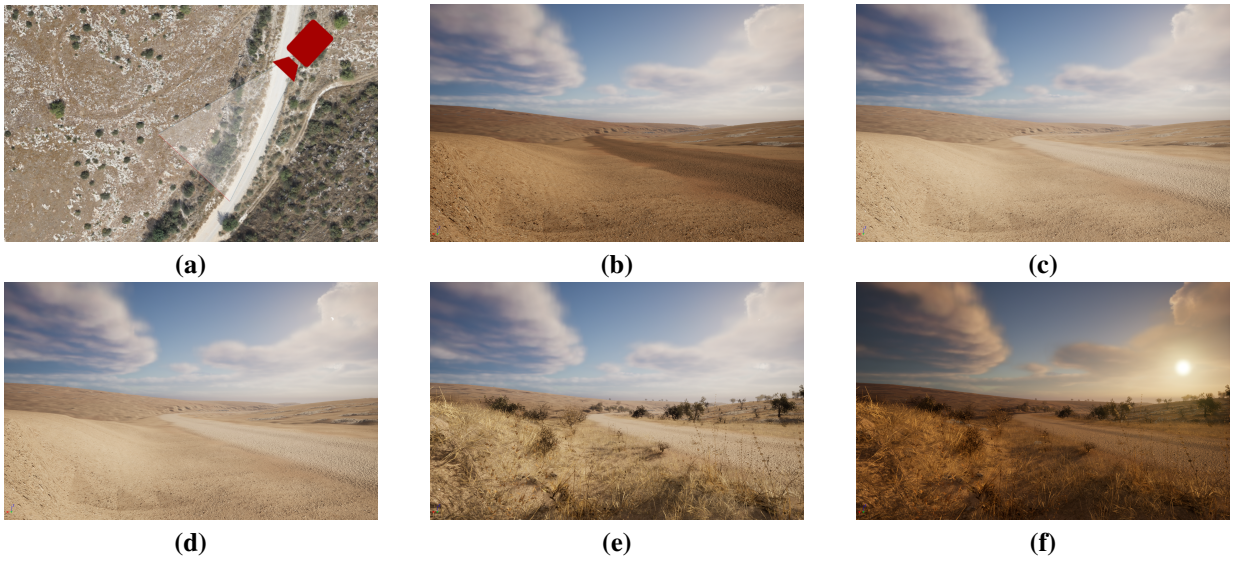


Figure 9: Visual results in Arbel area rendered from Cam2 point of view. (a) Cam2 field of view. (b) Rendering terrain ground by assigning a typical texture to each ground element. (c) Integrating noise texture generated from the aerial imagery to overcome ground repetition and enhance realism. (d) Enhancing bedded rocks geometrical representation. (e) Integrating land cover geospatial layers. (f) Rendering the scene from the same location using different lighting conditions.

For synthetic data to be effectively used, closing or reducing the domain gap is imperative. To reduce the domain gap, most methods require re-training models, acquiring annotated target domain data, or making changes to the models' architecture. The most feasible approach, however, would be to generate synthetic data with precise ground truths and make these data as similar to real-world data as possible.

In real-world data, there is often a great deal of variability and a structure that is hard to define by predefined rules. Our simulated environment is highly photo-realistic and geo-specific, so we believe the domain gap between synthetic and real data can be reduced. As described in previous sections, our improved generated simulated environments improve the fidelity of the data provided to subsequent AI entities.

Using only data from our simulated environment to train, for example, a semantic segmentation model, we achieved good results on real sensor data. Additionally, when we complemented manually labeled real-world data with additional simulated data to train our model, we saw improved performance (on real data) compared to training without the simulated data.

Moreover, the validation of AI-based solutions for open terrain is faced with significant challenges due to the vast diversity of terrains, weather conditions, and scenarios encountered. The number of available testing scenarios is expanded by employing high-fidelity digital replicas of the physical environment, thereby enhancing the evaluation process. An example of a complex scenario generated within our simulated environment is illustrated in Figure: 10.

Traversability Analysis

The evaluation of traversability in a specific geographic location has been studied for numerous years. Several methodologies investigate traversability by analyzing terrain slopes using the DSM (Toscano-Moreno, Mandow, Martínez, & García-Cerezo, 2023). Although terrain slopes are pivotal in assessing traversability, challenges arise when encountering flat regions like lakes, rivers, and puddles that may impede passage. To address these challenges, our proposed material-based virtual environment substantially enhances the fidelity of traversability analysis. This improvement is achieved by integrating material mechanical properties with terrain slopes, resulting in a more accurate evaluation of traversability in complex terrains.

Vehicle traversability can be simulated using discrete body dynamics (Franco, Shani, Gat, & Shmulevich, 2020). The tire–soil interaction may be modeled using Brixius prediction, and the specific soil properties can be obtained from the classification system for each tire–soil interaction, size, and geometric area. The tire–ground contact can be determined by topographic surface and adjustment of the forces and direction acting on the tires. The classification system considers

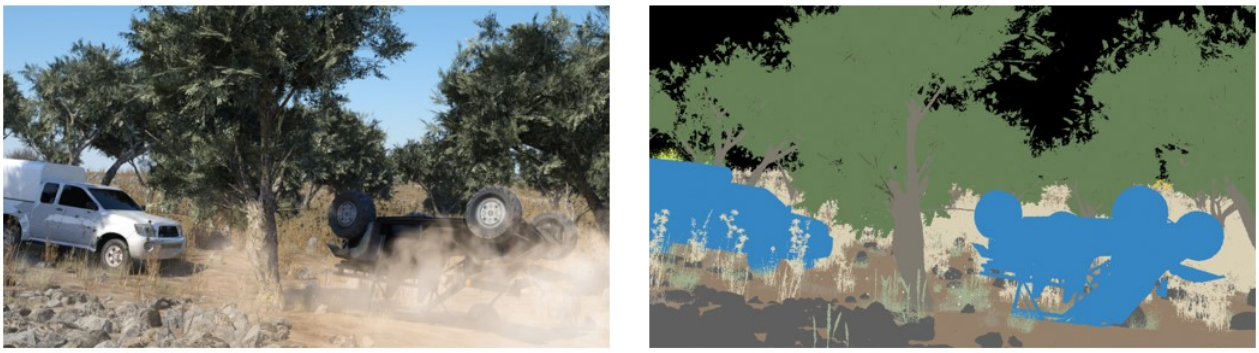
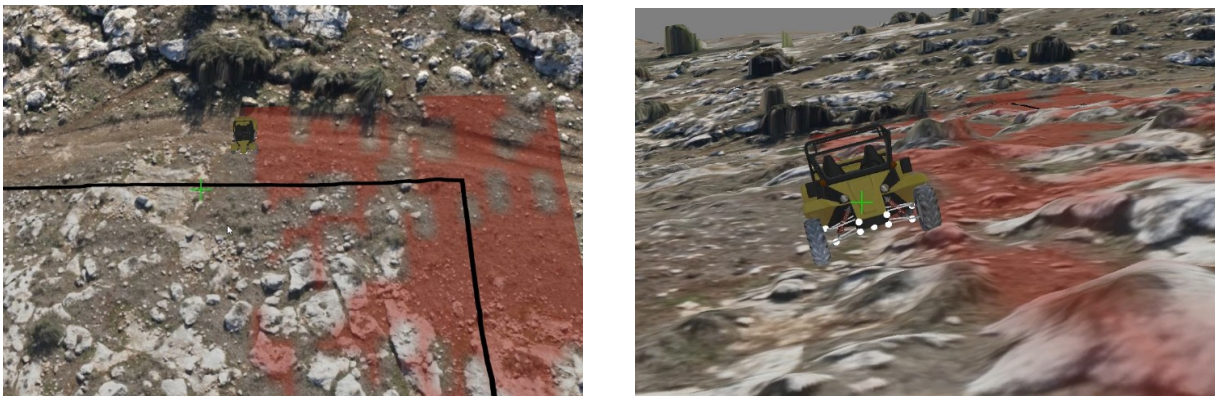


Figure 10: Simulated scenario in the Arbel area with its associated ground-truth label.

the ground type and the prevailing moisture conditions.

Traversability assessment for a geo-specific world location exhibits variability across different vehicle types and covers various parameters. By conducting simulations that capture the dynamics of each vehicle within material-based, geo-specific virtual environments, traversability analysis can comprehensively evaluate performance under various conditions, including different velocities, directions, and moisture conditions. Figure: 11 displays the generation process of a traversability map in the Arbel area. Vehicle dynamics is simulated using the discrete body dynamics method (Franco et al., 2020). The traversability map is generated by simulating vehicle interactions with the ground according to DGM and ground mechanical properties in different velocities, directions, and moisture conditions. Traversability limitations may arise during specific scenarios, such as when the vehicle chassis makes contact with the ground, when the wheels slip due to their interaction with the ground, or when the vehicle engine cannot generate sufficient power to propel the wheels, e.g., on steep slopes.



(a) **(b)**

Figure 11: An example of the traversability analysis process. (a) Vehicle dynamics are simulated in different directions, velocities, and moisture conditions. (b) The interaction between the simulated vehicle model determines its traversability.

Open Terrain Robotics Verification

Verification plays a significant role in ensuring open-terrain robotic systems' reliability, safety, and effectiveness. As autonomous machines venture into unstructured and challenging environments, it becomes imperative to establish robust verification processes to validate their capabilities and mitigate potential risks. The verification processes enable comprehensive testing and validation of the system's algorithms, perception capabilities, decision-making processes, and physical interactions. This ensures the system can reliably and accurately navigate, perceive, and respond to its dynamic environment.

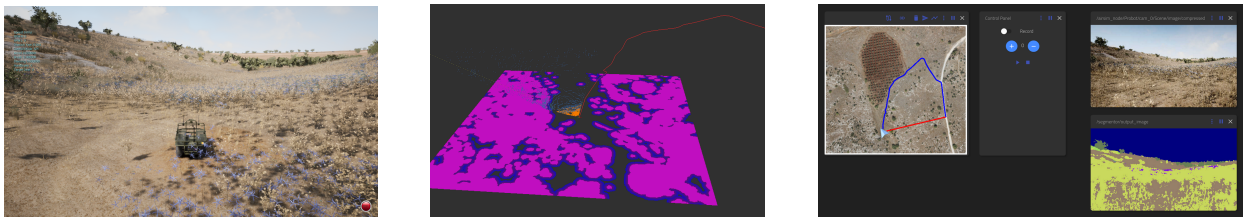
An illustrative instance of such a robotic system is an autonomous off-road vehicle. The architecture for such vehicles comprises a command and control system that generates routes to the desired destination based on the vehicle's traversabil-

ity capabilities. Additionally, a robotic kit is employed on the vehicle to analyze sensor data, interpret the surrounding environment, and transmit control commands to the vehicle. To verify the performance of such autonomous robots in diverse locations, seasons, times of day, environmental conditions, and scenarios, extensive physical testing is traditionally required, which is both costly and time-consuming.

Our proposed approach is to utilize the "Digital Twin" open terrain concept, which has been previously discussed, to generate virtual environments for the verification process. The method demonstrated in Figure 12 offers a compelling solution. In this methodology, the command and control system will employ automatically generated traversability maps. Furthermore, the vehicle physical simulation, as outlined in (Franco et al., 2020), will replicate the behavior of the physical vehicle while interacting with the virtual environment. Simulated sensors, including cameras, laser scanners, and GPS, will emulate the functionality of actual sensors by actively engaging with the virtual environment.

The Robotic Kit will interpret the sensor data received from the simulation as if it were actual real-world sensor input. Subsequently, it will send control commands to the simulated vehicle, simulating real-world operations and interactions.

The fidelity attained within the virtual environment, and the accuracy of traversability prediction, vehicle physical interaction, and sensor simulation are strongly tied to the proposed open terrain generation process. By leveraging this methodology, we can achieve a comprehensive and reliable verification process, ultimately enhancing the performance evaluation of open-terrain robotics systems.



(a) **(b)** **(c)**

Figure 12: Open terrain robotics development and verification. (a) Experimental Arbel area virtual environment rendered in Unreal Engine. (b) Simulated sensor data is being processed by the robotic kit, which is implemented in ROS2. (c) The calculation of the global path involves utilizing the traversability layer. The path is then transmitted to the robotic kit via the command and control center application.

CONCLUSIONS

This paper presents our automated pipeline for generating geo-specific and high-fidelity simulations of open terrain. The proposed pipeline generates photorealistic and geo-specific virtual environments using readily available aerial maps, making it highly applicable to real-world use cases worldwide.

In our pipeline, aerial maps are first classified according to their semantics and materials. The classification is achieved using a novel hierarchical deep neural network with a multi-task objective. Our network has proven accurate and readily adaptable to new terrains with minimal labeling.

Based on these classifications, MLGR are automatically generated, which are then converted into simulated environments. We have demonstrated the fidelity of our generated environments and highlighted several potential applications that can greatly benefit from our pipeline and classification module, i.e., mission rehearsal, data generation for AI algorithms, and traversability analysis.

In future work, we plan to enhance our capabilities to support the automatic production of larger high-fidelity open terrains on a variety of locations worldwide and introduce new categories such as snow, trench, and gate to the generation pipe.

DISCLAIMER

This article is published by Elbit Systems Ltd. ("Elbit") for information purposes only. It should not be relied upon to make investment decisions, nor does it purport to summarize Elbit's business or financial information. Those wishing to learn more about Elbit are referred to its public filings with the U.S. Securities and Exchange Commission and/or the

Tel Aviv Stock Exchange. In addition, this article may contain forward-looking statements. Such statements are based on management's current estimates and assumptions about future events. Actual results may differ materially from these forward-looking statements due to a variety of factors. All forward-looking statements speak only as of the date of this article, and Elbit does not undertake any obligation to update its forward-looking statements.

References

Bodhiswatta Chatterjee, H. B., & Patel, B. (2020). Creating geo-specific synthetic environments using deep learning and process automation.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Eccv*.

Chen, M., Feng, A., Hou, Y., McCullough, K., Prasad, P. B., & Soibelman, L. (2021). Ground material classification for uav-based photogrammetric 3d data a 2d-3d hybrid approach. *arXiv preprint arXiv:2109.12221*.

Chen, M., Feng, A., McCullough, K., Prasad, P. B., McAlinden, R., Soibelman, L., & Enloe, M. (2020). Fully automated photogrammetric data segmentation and object information extraction approach for creating simulation terrain. *arXiv preprint arXiv:2008.03697*.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1251–1258).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... others (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Elmqvist, M., Jungert, E., Lantz, F., Persson, A., & Soderman, U. (2001). Terrain modelling and analysis using laser scanner data. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, *34*(3/W4), 219–226.

Franco, Y., Shani, M., Gat, G., & Shmulevich, I. (2020). Three-dimensional dynamic model for off-road vehicles using discrete body dynamics. *Journal of Terramechanics*, *91*, 297–307.

Hollosi, A., Menzel-Berger, T., Walter, H., & Lahm, D. (2022). Automated 3d building generation at global scale based on satellite imagery. *2022 – Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456).

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the ieee international conference on computer vision* (pp. 2980–2988).

Lindstrom, P., Koller, D., Ribarsky, W., Hodges, L. F., Faust, N., & Turner, G. A. (1996). Real-time, continuous level of detail rendering of height fields. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques* (pp. 109–118).

Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically based rendering: From theory to implementation*. Morgan Kaufmann.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., & Breitkopf, U. (2012). The isprs benchmark on urban object classification and 3d building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 (2012), Nr. 1*, *1*(1), 293–298.

Spicer, R., McAlinden, R., Conover, D., & Adelphi, M. (2016). Producing usable simulation terrain data from uas-collected imagery. In *Interservice/industry training, simulation, and education conference (i/itsec)* (pp. 1–13).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Tatarchuk, N. (2006). Dynamic parallax occlusion mapping with approximate soft shadows. In *Proceedings of the 2006 symposium on interactive 3d graphics and games* (pp. 63–69).

Toscano-Moreno, M., Mandow, A., Martínez, M. A., & García-Cerezo, A. (2023). Dem-aia: Asymmetric inclination-aware trajectory planner for off-road vehicles with digital elevation models. *Engineering Applications of Artificial Intelligence*, *121*, 105976.

Yang, B., Ali, F., Zhou, B., Li, S., Yu, Y., Yang, T., ... Zhang, K. (2022). A novel approach of efficient 3d reconstruction for real scene using unmanned aerial vehicle oblique photogrammetry with five cameras. *Computers and Electrical Engineering*, *99*, 107804.