

Remote sensing colour image semantic segmentation of large herbivorous mammal trails

José Francisco Díez-Pastor, Francisco Javier González-Moya, Pedro Latorre-Carmona, Francisco Javier Pérez-Barbería, Ludmila Kuncheva, Antonio Canepa-Oneto, Álar Arnaiz-González & César García-Osorio

To cite this article: José Francisco Díez-Pastor, Francisco Javier González-Moya, Pedro Latorre-Carmona, Francisco Javier Pérez-Barbería, Ludmila Kuncheva, Antonio Canepa-Oneto, Álar Arnaiz-González & César García-Osorio (05 Feb 2026): Remote sensing colour image semantic segmentation of large herbivorous mammal trails, International Journal of Remote Sensing, DOI: [10.1080/01431161.2026.2618658](https://doi.org/10.1080/01431161.2026.2618658)

To link to this article: <https://doi.org/10.1080/01431161.2026.2618658>

 [View supplementary material](#) 

 Published online: 05 Feb 2026.

 [Submit your article to this journal](#) 

 Article views: 59

 [View related articles](#) 

 [View Crossmark data](#) 



Remote sensing colour image semantic segmentation of large herbivorous mammal trails

José Francisco Díez-Pastor ^a, Francisco Javier González-Moya ^a, Pedro Latorre-Carmona ^a, Francisco Javier Pérez-Barbería ^b, Ludmila Kuncheva ^c, Antonio Canepa-Oneto ^b, Álvor Arnaiz-González ^a and César García-Osorio ^a

^aDepartment of Computer Engineering, Universidad de Burgos, Burgos, Spain; ^bBiodiversity Research Institute, Spanish Research Council, University of Oviedo, Principado de Asturias, Spain; ^cSchool of Computer Science and Electronic Engineering, Bangor University, Bangor, UK

ABSTRACT

Detection of spatial areas where biodiversity is at risk is of paramount importance for the conservation and monitoring of ecosystems. Large terrestrial mammalian herbivores are keystone species as their activity not only has deep effects on soils, plants, and animals but also shapes landscapes, as large herbivores act as allogenic ecosystem engineers. One key landscape feature that indicates intense herbivore activity and potentially impacts biodiversity is the formation of grazing trails. Grazing trails are formed by the continuous trampling activity of large herbivores that can produce complex networks of tracks of bare soil. Here, we evaluated different algorithms based on machine learning techniques to identify grazing trails. Our goal is to automatically detect potential areas with intense herbivory activity, which might be beneficial for conservation and management plans. We have applied five *semantic segmentation* methods combined with fourteen encoders aimed at mapping grazing trails on aerial images. Our results indicate that in most cases the chosen methodology successfully mapped the trails, although there were a few instances where the actual trail structure was underestimated. The UNet architecture with the MambaOut encoder was the best architecture for mapping trails. The proposed approach could be applied to develop tools for mapping and monitoring temporal changes in these landscape structures to support habitat conservation and land management programmes. This is the first time, to the best of our knowledge, that competitive image segmentation results are obtained for the detection and delineation of trails of large herbivorous mammals.¹

ARTICLE HISTORY

Received 2 June 2025
Accepted 11 January 2026

KEYWORDS

Semantic segmentation; deep Learning; grazing trails; herbivory; biodiversity; monitoring

1. Introduction

Large terrestrial herbivores are keystone species in ecosystem functions, as their type of biological interactions, known as herbivory (plant defoliation, trampling, defaecation, and urination), have profound trophic effects. They also act as allogenic ecosystem engineers,

CONTACT Pedro Latorre-Carmona  plcarmona@ubu.es  Department of Computer Engineering, Universidad de Burgos, Avenida Cantabria s/n, Burgos, Spain

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/01431161.2026.2618658>

© 2026 Informa UK Limited, trading as Taylor & Francis Group

by creating physical changes in biotic or abiotic materials (Jones, Lawton, and Shachak 1994). Herbivory maintains, modifies, and creates new habitats and niches at various spatial scales, promoting habitat heterogeneity and, consequently, biodiversity (Filazzola et al. 2020; Gordon and Prins 2019; Mills, Soulé, and Doak 1993). Low and medium intensities of herbivory positively affect biodiversity at both taxonomic and trait levels. However, high-intensity herbivory reduces habitat heterogeneity, compromising the number of niches and leading to a decrease in biodiversity (Wieren and Bakker 2008).

One of the landscape features was created by the activity of large mammals, mainly ungulates, known as 'grazing trails' (Chemekova and Chemekov 1975; Jin et al. 2016, 2022). These are narrow trails shaped by continuous animal trampling in a dominant direction towards a resource or to reduce energy expenditure during motion on steep terrain (May 1981; Robbins 1993). Trails can constitute networks of very different shapes, from resembling the contour lines of a topographic map, when trails develop on a steep slope (Jin et al. 2016, 2022), to radial networks due to the attraction to a concentrated resource such as a savanna waterhole (Washington-Allen et al. 2004).

A trail consists of a depressed longitudinal surface, often bare soil, bordered by raised shoulders (Goudie 2013; Stavi et al. 2021). These microtopography changes significantly influence some ecological processes, including water distribution, drainage, soil moisture (Stavi et al. 2008), soil erosion (both positive and negative) (Higgins 1982; Watanabe 1994), as well as nutrients, litter, root biomass, and vegetation distribution (Hiltbrunner et al. 2012; Stavi et al. 2008, 2021).

Since trails result from intense and concentrated activity of livestock and wild ungulates, monitoring their location over time and space is crucial to identifying herbivory hotspots that may compromise biodiversity in grazed ecosystems. Developing automatic tools can facilitate the mapping and analysis of these complex landscape features in aerial images, supporting habitat conservation and land management programmes.

The automatic characterization of trails has received limited attention, with a few examples relying on landscape features such as relief, abrupt elevation changes, topographic position, among others (Bailly and Levavasseur 2012; Danilo Godone 2018; Diaz-Varela et al. 2014; Pijl et al. 2020; Sas et al. 2012; Sofia, Marinello, and Tarolli 2014).

To the best of our knowledge, the only study that aimed to automatically detect livestock trails from high-resolution satellite imagery is the work by Hellman et al. (2020). However, their method is limited to predicting trail presence at the patch level (37×37 m). We improve upon this approach by using semantic segmentation, enabling pixel-level detection that ultimately provides an effective means of reconstructing the full extent of grazing trails.

Semantic segmentation (Hao, Zhou, and Guo 2020) is currently an active area of research that started within the computer vision community but extended later to a wide and diverse group of application fields (Guo et al. 2018; Hao, Zhou, and Guo 2020; Minaee et al. 2022). It is somewhat related to image classification since it produces per-pixel category prediction rather than of an image-level prediction.

Semantic segmentation using deep learning architectures has been applied to artificial structures, like roads and buildings, or even parts of cities, as shown by Jamali et al. (2024), Fibaek et al. (2024), Hong et al. (2023) and Wurm et al. (2019). This approach takes into account both local information provided by convolutional neural network methodologies

and global contextual information provided by architectures like the so-called vision transformers (Khan et al. 2022).

Other semantic segmentation strategies are applied to identify natural objects/structures, such as rivers (Chen et al. 2023; Wieland et al. 2023), ice in rivers and open seas (C. Zhang, Chen, and Ji 2022; X. Zhang et al. 2020), forests (Bragagnolo, da Silva, and Grzybowski 2021; da Costa et al. 2022), or agricultural structures (Jadhav and Singh 2018; Luo et al. 2023).

In some cases, the application areas may involve considering truly remote sensing images, while in others, the images are captured much closer to the objects of interest, which may not strictly qualify as remote sensing images (Dong et al. 2021; Hosseini et al. 2022).

Another approach has been to consider the identification of natural or artificial structures as a classification-related problem (Pal, Akshay Rohilla, and Teja 2020; Parajuli et al. 2022). For instance, water bodies are identified in remote sensing images acquired by the European Space Agency (ESA) Sentinel-2 satellite (ESA 2025) by extracting patches and classifying their central pixels as either water or non-water pixels.

Note that any local-based approach would not consider non-local information that an expert might use to determine a grazing trail such as its shape, extent, width, etc. This is the rationale for using a semantic segmentation approach, which not only identifies the pixels constituting a trail but also incorporates non-local information during the segmentation process (*i.e.*, features from the image that are further away but still influence the correct identification or classification of the pixels forming a trail).

The aim of this research is to assess the effectiveness of different semantic segmentation methodologies for mapping grazing trails, ensuring that the methods enable the detection of these landscape features without relying on additional topographic information. This presents a challenge due to the wide variety of shapes, continuity, and complexity within these trail networks.

2. Materials and methods

This section details the image processing techniques, from image acquisition to the algorithms used for its processing. The main features regarding the specific geographical area and morphology considered in our analysis, as well as the type of processed images are presented in subsection 2.1. The creation of *groundtruth* images is then explained in Subsection 2.2. Semantic segmentation and the architectures used in the study are detailed in Subsection 2.3, and finally, in Subsection 2.4 the details of the experimental setup are presented.

2.1. Aerial images

One hundred aerial images were selected from grazed ecosystems in five mountain ranges in Spain (see Figure 1): (1) the 'Cantabrian' mountain range; (2) the 'Pyrenees'; (3) 'Sierra de la Demanda' (4); 'Sierra de Béjar', and (5) 'Montaña Palentina'. The selection followed a sampling strategy designed to (i) ensure the presence of well-developed grazing path networks and (ii) capture a broad range of path morphologies. One of the co-authors (F.J. Pérez-Barbería), who has extensive experience in grazing ecology, used orthoimages from Google Earth Pro to conduct a systematic search focusing on areas

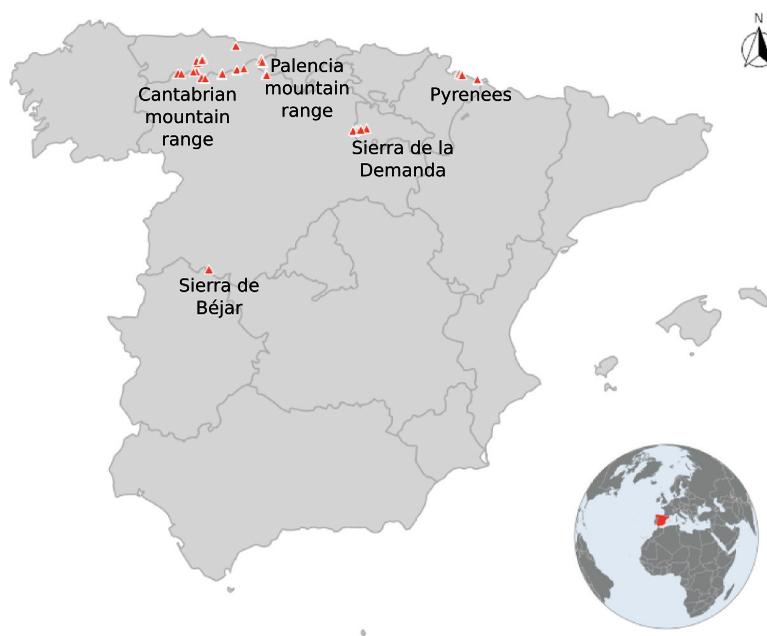


Figure 1. Study area. Map of Spain showing the areas (red triangles) where images were obtained from. The selected mountain systems correspond to: (i) “Cantabrian” mountain range, (ii) “Palencia” mountain range, (iii) “Pyrenees”, (iv) “Sierra de la Demanda”, and (v) “Sierra de Béjar”.

with a documented history of extensive grazing, particularly high-altitude summer trans-humance pastures where livestock typically move freely over large areas.

Within each of these regions, sectors in which grazing paths were clearly discernible in the imagery were visually identified. For each sector, the morphology of the grazing paths was then evaluated, distinguishing, for example, between predominantly parallel networks, bifurcating networks, mesh-like networks, and more radial configurations. From each morphological type and sector, images in which the grazing paths were most clearly visible were selected, prioritizing high image resolution and avoiding areas with strong shadows or other artefacts that could compromise interpretability.

Subsequently, images were projected at a nominal height of 100 m above ground level (approximate ground coverage: $70m \times 115m$). From these, the observer selected 100 images that represented the diversity in grazing-path morphology, density, and continuity of the trail networks in the area (e.g., network shape, branching complexity, and continuity; [Figure 2](#)). The centre of each image was recorded as a ‘.kmz’ file for georeferencing purposes, and the corresponding image was saved as a ‘.jpg’ file at maximum resolution (8192×4968 pixels). The selected trails were located at altitudes between 1091 and 1880m above sea level (mean = 1487m, $sd = 234.7m$).

In our study, all images were obtained using Google Earth Pro. It aggregates very high-resolution satellite and aerial orthophotography from multiple commercial and institutional providers. However, unlike remotely sensed products accessed directly from the original providers, Google Earth Pro exposes only a limited amount of metadata at the image level (typically, an ‘imagery date’, an approximate scale, and a generic provider credit). Details

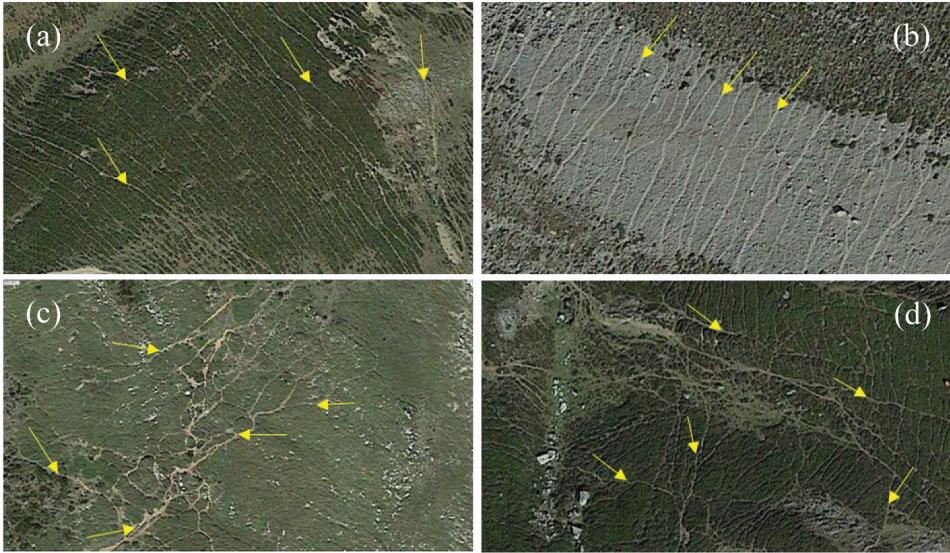


Figure 2. Examples of different grazing trails networks in the “Cantabrian” mountain range. (A) parallel trails on heather, their layout dominant direction is perpendicular to the slope, $43^{\circ} 1'37.14''\text{N}$, $5^{\circ} 30'4.25''\text{O}$ WGS84; (B) as (A) but on a rocky slope, coordinates $43^{\circ} 11'17.56''\text{N}$, $4^{\circ} 45'24.56''\text{O}$; (C) trails on mountain grassland, no dominant layout direction, coordinates $43^{\circ} 2'25.66''\text{N}$, $6^{\circ} 13'20.43''\text{O}$; (D) trails on heather, two dominant directions, coordinates $43^{\circ} 1'35.50''\text{N}$, $5^{\circ} 29'40.07''\text{O}$. Some trails are marked with arrows.

given by Google Earth Pro related to the acquisition time and sensor for the images can be found as a separate document, as stated in the supplementary material (Appendix A).

2.2. Groundtruth image generation

To train the models and assess the segmentation quality, *groundtruth* images were created by the same expert. A copy of each image was imported to ImageJ (Schneider, Rasband, and Eliceiri 2012) and all visible trails were drawn along the centreline using the yellow pencil tool (10 pixels in width). This RGB-labelled image I_{rgb} was converted to a binary-labelled image C , the *groundtruth* M was obtained following the methodology described by Mnih and Hinton (2010).

Let I_{rgb} be an image in RGB format, manually labelled by the expert with yellow lines marking the trails. Let I_{hsi} be the image resulting from converting I_{rgb} to the HSI colour space (being the three coordinates, normalized to the $[0, 1]$ interval). Let P_{ref} be the reference pixel (colour) of the labels in the HSI space, given by $P_{\text{ref}} = [H_{\text{ref}} = 0.6, S_{\text{ref}} = 1, I_{\text{ref}} = 1]$.

Since the pencil tool creates traces with some degree of smoothing, further processing is required to generate the corresponding binary masks. For each pixel $P_{ij} = [H_{ij}, S_{ij}, I_{ij}]$ in I_{hsi} , the Euclidean distance D_{ij} between P_{ij} and P_{ref} is obtained as:

$$D_{ij} = \sqrt{(H_{ij} - H_{\text{ref}})^2 + (S_{ij} - S_{\text{ref}})^2 + (I_{ij} - I_{\text{ref}})^2} \quad (1)$$

These distances are then scaled by dividing them by the maximum distance D_{\max} in the entire image:

$$D_{ij\text{-norm}} = \frac{D_{ij}}{D_{\max}} \quad (2)$$

Finally, a binary-labelled mask image C is created where pixels with $D_{\text{norm}} < Th$ are labelled as true, *i.e.*:

$$C_{ij} = \begin{cases} 1, & \text{if } D_{ij\text{-norm}} < Th \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

In our case, we set $Th = 0.3$. The threshold 0.3 was selected as the 2.5th percentile of the D_{norm} distribution (normalized HSI Euclidean distances), maximizing specificity given the class imbalance. This value is necessary to capture the full width of the annotations, accounting for the expert's tool smoothing/feathering effect. C_{ij} is then 1 if the (i, j) pixel location in the remote sensing image S belongs to a trail centreline and 0 otherwise. The mask C is later used to define the *groundtruth* map M as follows:

$$M_{ij} = \exp \left[-\frac{d_{ij}^2}{\sigma^2} \right] \quad (4)$$

where d_{ij} is the Euclidean distance between the location with coordinates (i, j) in the image, and the nearest nonzero pixel in C . σ is a smoothing parameter that depends on the scale of the aerial images.

Mnih and Hinton (2010) set the parameter σ such that the distance equivalent to $2\sigma + 1$ pixels roughly corresponds to the width of a typical road lane. In our case, the value of σ was determined using an empirical sampling process. The value corresponds to half the average width of a representative sample of path segments, selected from the original training set, using the measurement tool embedded in traditional drawing tools. M_{ij} can be interpreted as the probability that location (i, j) belongs to a trail, given that it is d_{ij} pixels away from the nearest centreline pixel.

The aerial images were scaled down by a factor of 8 to reduce the computational cost. Scale factors of 2 and 4 were also tested without detecting any difference in the results. This reduction was applied consistently to both the input images (RGB) and the generated Ground Truth probability map (M). By generating the Gaussian map M using the high-resolution σ and subsequently applying the same interpolation and downsampling factor, we ensured the geometric fidelity and effective σ of the labels were preserved, maintaining consistency with the reduced computational scale.

The dataset consisting of RGB images and their corresponding *groundtruth* images was generated using HuggingFace,² accessible on HuggingFace Hub.³

2.3. Semantic segmentation

In this paper, five semantic segmentation architectures to identify grazing trails in aerial images were used:

- U-Net (UNet) is the *seminal* work that proposed the so-called *fully convolutional network* architecture/methodology for computer vision (Ronneberger and Brox 2015). One of its main advantages is that it works with very few training images. It also yields precise segmentations. *Fully convolutional networks* supplement a usual *contracting network* by successive layers, where pooling operators are replaced by upsampling operators (see Figure 3). In a *contracting network*, the spatial information is reduced, while the information about the features is increased. Therefore, upsampling operators make these layers to increase the resolution of the output. High-resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.
- Feature Pyramid Networks (FPNet) are generic feature extractors that exploit multi-level feature representations in an inherent and pyramidal-type hierarchy (Lin et al. 2017). They use a top-down architecture with *lateral* connections to fuse high-level semantic information into middle and low levels, with little extra cost. In particular, the aim in this method is to leverage a convolutional neural network's pyramidal feature hierarchy, which has semantics from low to high levels, and build a feature pyramid with high-level semantics throughout it. The resulting feature pyramid network is general, and the focus is on sliding window strategies, and region-based detectors. FPNet is generalized to instance segmentation proposals. They take a single-scale image of an arbitrary size as input data and give as an output a series of proportionally sized feature maps at multiple levels. This process is independent of the convolutional architectures for feature extraction.

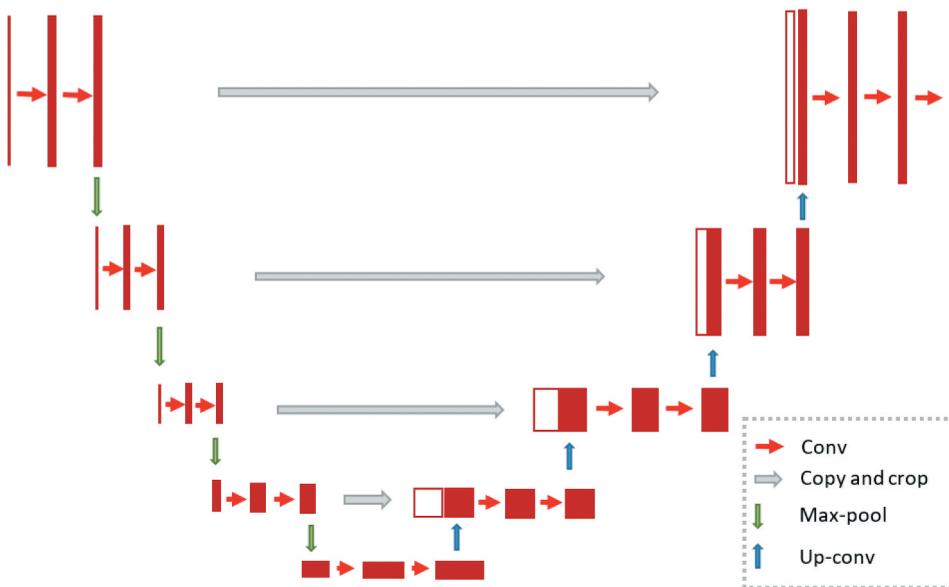


Figure 3. Schematic representation of UNet architecture. Figure inspired in (Ronneberger and Brox 2015). Conv means convolution operation. Up-conv is an upsampling operation followed by a 2×2 convolution. Max-pool refers to a downsampling operation that takes only the maximum value out of a window.

Table 1. List of encoders used: name, number of parameters (in millions), and reference. All of them were pretrained with the so-called *ImageNet* dataset.

Encoder	# of parameters (M)	Reference
ConvNeXt (small)	50	Liu et al. (2022)
EfficientViT B3	12	Cai et al. (2023)
EfficientNet B7	63	Tan and Le (2020)
MambaOut (Base)	85	Gu and Dao (2024)
Mobilenet v3 (Large)	15	Howard et al. (2019)
SAM 2 Hiera (Large)	224	Ryali et al. (2023)
Densenet-161	26	Huang et al. (2018)
Inception v4	41	Szegedy et al. (2016)
MIT B5	81	Xie et al. (2021)
MobileOne s4	12	Vasu et al. (2022)
ResNet-34	21	He et al. (2016)
Xception-71	20	Chollet (2017)
Vgg16	14	Simonyan and Zisserman (2015)
Vgg19	20	Simonyan and Zisserman (2015)

- The Pyramid Scene Parsing Network (PSPNet) method tries to take advantage of the capability to use global context information by an aggregation methodology of different region-based features (Zhao et al. 2017). In a deep neural network, the size of the so-called *receptive field* (i.e. the size of the area in the input image that creates the feature) may roughly indicate how much context information we use/consider. In order to further reduce context information loss between different sub-regions, researchers proposed a hierarchical global prior, containing information with different scales and varying among different sub-regions (Zhao et al. 2017).
- The transformer framework for semantic segmentation (Semantic Segmentation with Transformers or Segformer) can be understood as a semantic segmentation method which unifies the so-called *transformers* with lightweight multilayer perceptron (MLP) decoders (Xie et al. 2021). A vision transformer (ViT) decomposes an input image into a series of patches, serializes each patch into a vector, and maps it to a smaller dimension. These vector embeddings are then processed by a transformer encoder as if they were token embeddings. ViTs were designed as alternatives to convolutional neural networks. In particular, they present a new positional-encoding-free hierarchical transformer and a lightweight All-Multilayer Perceptron⁴ (MLP) decoder design.
- The Unified Perceptual Parsing Network (UperNet) is a network based on FPNets (Xiao et al. 2018), previously described in (Lin et al. 2017). Its main drawback is that, even when the theoretical receptive field of deep CNNs might be large enough, the empirical receptive field of deep CNNs may also be relatively smaller. To address this issue, a Pyramid Pooling Module (PPM) from PSPNet was proposed (Zhao et al. 2017). This PPM is applied on the last layer of the backbone network, just before feeding it into the top-down branch, in FPNets.

These methods were combined with 14 encoders, listed in Table 1 along with their number of parameters. Encoders accomplished two main purposes: (a) to generate a high-dimensional feature vector from the input image, and (b) to aggregate features across multiple resolution levels. Similarly, decoders also serve two key functions: (a) to produce a semantic segmentation mask from the high-dimensional feature vector and (b) to decode the multi-level features aggregated by the encoder.

2.4. Experimental setup

To assess the best combination of architecture and encoder for grazing trail identification, the Pytorch Segmentation Model library was used (Iakubovskii 2019). This framework allows different architectures to be paired with various encoders, facilitating a thorough evaluation of multiple configurations.

For the assessment of the performance of the models, a 10-fold cross-validation strategy was followed on every architecture-encoder pair, *i.e.*, $5 \times 14 = 70$ combinations. In every single fold, from the set of 100 aerial images (with their corresponding 100 *groundtruth* images), 80 images were used for model training, 10 images for internal model validation and, finally, 10 images were used for testing. The set of training, validation, and test images alternates on every fold to ensure the use of each instance in the test part only once.

To compare the performance of the different models, the following measures were used: the Intersection Over Union (IoU, also called the *Jaccard coefficient*), *Precision*, *Recall*, and the *F1-score*. IoU can be used to measure the similarity between finite sample sets (**A** and **B**), and it is defined as follows:

$$\text{IoU}(\mathbf{A}, \mathbf{B}) \equiv J(\mathbf{A}, \mathbf{B}) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A} \cup \mathbf{B}|} \quad (5)$$

with: $0 \leq J(\mathbf{A}, \mathbf{B}) \leq 1$. In our case, **A** and **B** would be the *groundtruth* image, and the image segmented by the corresponding method, respectively. In addition, a pixel associated with a trail can be regarded as belonging to the positive class. Taking into account the concept of *true positives TP* (case where a pixel belongs to a trail and the segmentation method classifies it as trail), *true negatives TN* (a pixel does not belong to a trail and the segmentation method classifies it as not belonging to a trail), *false positives FP* (a pixel does not belong to a trail and the segmentation method classifies it as belonging to a trail), and *false negatives FN* (a pixel belongs to a trail but the segmentation method classifies it as not belonging to a trail), the Precision, Recall, and F0 measures would be defined as follows: Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$, F1 = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

To compare the performance of the different architectures and encoders, average rankings were used (Wilcoxon 1992). For each encoder, architectures were ranked based on their mean IoU scores, with tied architectures assigned the average of their competing ranks (*e.g.*, two architectures tied for first place receive a ranking value of 1.5). The final performance metric for each architecture was derived by averaging these per-encoder rankings across all encoders, where lower average rankings indicate superior overall performance.

Bayesian comparisons (Benavoli et al. 2017) were conducted using the Baycomp library,⁵ which employs non-informative priors to avoid introducing subjective bias. Specifically, for the Bayesian correlated *t*-test, the library utilized the Normal-Gamma prior $NG(\mu, \nu; \mu_0, k_0, a, b)$, which is conjugate to the likelihood. The parameters are fixed to the specific values known as the ‘matching prior’: $\{\mu_0 = 0, k_0 \rightarrow \infty, a = -1/2, b = 0\}$. This particular prior choice ensures that the resulting posterior distribution of the mean difference (μ) coincides numerically with the Student distribution used in the frequentist correlated *t*-test. No domain-specific priors or subjective information is introduced. The primary utility of this Bayesian approach is therefore not derived from the prior, but from

querying the resulting posterior distribution (which describes the magnitude and uncertainty of the effect) against the Region of Practical Equivalence (ROPE), allowing for practical conclusions about equivalence. This test takes as a parameter the width of the ROPE, or in other words, the maximum difference that can exist between two results in order to be considered equivalent. In our analysis, we adopted a ROPE width of 0.01, which is the default parameter of the library used and consistent with previous literature (Costa, Costa, and Rocha 2025; Jiménez-Navarro et al. 2023). This particular value reflects a conservative threshold below which differences are considered practically negligible.

The Bayesian test gives three probabilities as its outcome: (a) p_{left} , the probability that the first method outperforms the second, (b) p_{rope} , the probability that both methods can be considered *equivalent*, and (c) p_{right} , the probability that the second method outperforms the first. Whether one method is *practically equivalent* to the other is determined by the ROPE parameter.

The test uses the results from a 10-fold cross-validation and applies Monte Carlo simulation to generate 5.0×10^4 samples. Convergence was experimentally verified by increasing the sample size to 150,000, yielding negligible differences ($< 10^{-4}$) in posterior probabilities. Based on these samples, a distribution is modelled, and the three probabilities are estimated.

More specifically, about hyper-parameters setup, every *encoder-architecture* pair used the same parametrization: *i.e.*, a 9×10^{-5} learning rate value and a total amount of 60 *epochs*, per execution. For training, a 10-fold cross-validation scheme was used with 80% of images for training, 10% for validation, and 10% for testing every model. This means that for every fold, it is selected rotatively 80 images from training, the next 10 for internal model validation and, finally, the lasting 10 for testing purposes. The following fold shifts 10 images to select the next 80 training ones, the next 10 for validation and the last 10 for testing, and so on.

All the tests were made using a single NVIDIA RTX-A6000 with a 49 GB RAM GPU memory. Using GPU got experimental results in a reasonable time, spending only a few hours per experiment, although it varies significantly on every *encoder-architecture* pair. The 1×8 set of scaled 100 images was employed on every test, with a 1024×640 pixels per image, which offers the collected results that will be presented. This ratio for whole images provides a good relationship between execution time and experiment result metrics in comparison with other tested scales.

3. Results

This section presents the results of the semantic segmentation experiments described above. First, the performance of the different architectures and encoders is reported. Second, a subset of six new images from various locations worldwide is used to assess performance across different landscape types.

3.1. Results on Spanish mountain grasslands

Our analysis indicates that the proposed semantic segmentation approach effectively classified herbivore trails by accurately predicting the probability that a pixel is part of

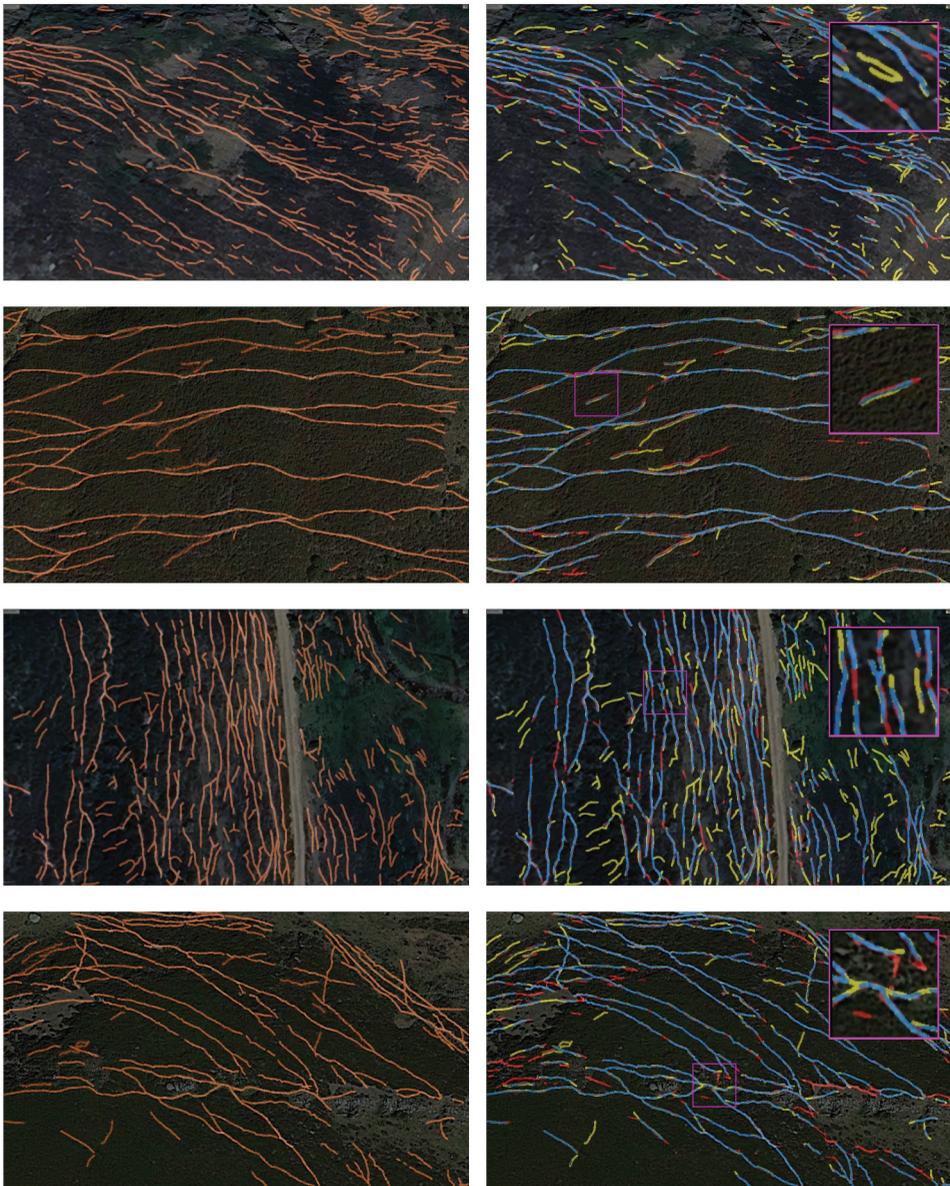


Figure 4. Example of the original RGB *groundtruth* image (left columns) and the result of the automatic trail segmentation. The *groundtruth* trails are highlighted in orange, the prediction trails are highlighted in blue (TP), yellow (FN), and red (FP). A zoomed-in part is also created, for illustration purposes. Best viewed in colour.

a grazing trail with high precision. **Figure 4** presents examples of segmentation results produced by the optimal encoder–architecture combination (right), alongside their corresponding *groundtruth* images (left). In **Figure 4** (right) the true positives (TP) are coloured in blue, false positives (FP) in red, and false negatives (FN) in yellow. A zoomed-in part has been created as well, in each one of these figures, for illustration purposes.

It should be noted that semantic segmentation predictions effectively differentiate between grazing trails and anthropogenic structures, such as motorized dirt tracks. For instance, in [Figure 4](#), the dirt track and its ditches were not misclassified as grazing trails, indicating the efficacy of semantic segmentation and its ability to (i) use non-local contextual information effectively and (ii) maintain scale invariance in trail characterization.

This discriminative capability extends consistently to other landscape features in our dataset, including woodland canopy edges, rock formations, and artificial structures, among others. These results show significant promise for the implementation of such models in automated detection tools within geographic information systems, particularly for ecological monitoring and landscape management applications.

A closer look at the errors ([Figure 4](#), right column) reveals that many of the (FP and FN) errors concentrate in what could be considered as the edges of the trail. This result is expected, since the *groundtruth* aims at creating the best potential segmentation to compare against, but not even the *groundtruth* is free of assumptions or decisions made to create them. Therefore, we can consider this as a somehow unavoidable by-product of the image segmentation quality assessment process.

Given that the five architectures were tested in combination with 14 encoders, the results exhibit considerable variability. Thus, it is essential to consider both the architecture and the encoder, as some encoders outperform others.

The UNet architecture with the MambaOut encoder achieved the highest overall performance across combinations of architectures and encoders, with UNet and MIT B5 encoder yielding a similar performance. Overall, the findings are consistent across the two metrics (IoU and F_1 score), used to compare the performance of the methods, with UNet achieving the best performance, and PSPNet the least. Regarding encoders, MIT B5, MambaOut, and ConvNeXt exhibit a superior performance than the rest.

[Figure 5](#) presents the Intersection over Union (IoU) performance metrics across all encoder-architecture combinations. The visualization consists of:

(1) Main heatmap matrix:

- Each cell displays the mean IoU score for an encoder-architecture pair.
- Colour intensity (green gradient) represents performance level (the darker the higher IoU).

(2) Performance Rankings:

- *Row above matrix*: Average architecture ranking across encoders.
- *Left column*: Average encoder ranking across architectures.
- Ranking convention: Lower values indicate better performance (with corresponding darker green shading).

(3) Top Performers Identification:

- Underlined values: Best encoder for each architecture.
- Bold value: Overall best-performing encoder-architecture combination.

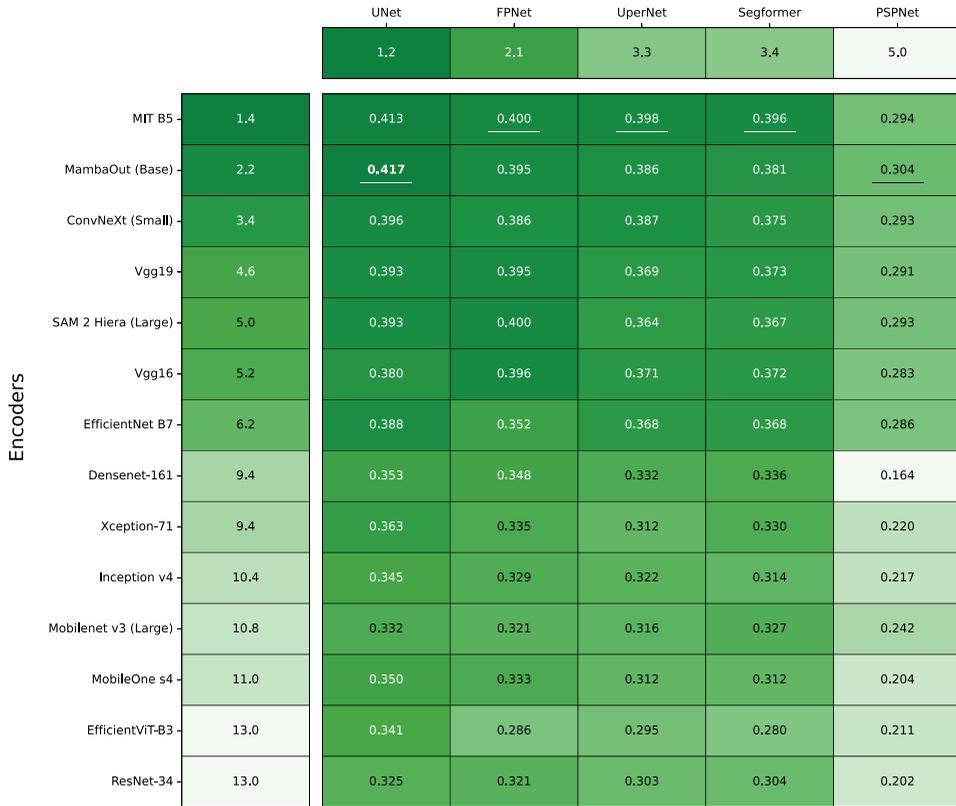


Figure 5. Heatmap of Intersection over Union (IoU) performance across all encoder-architecture combinations. Each cell displays the mean IoU for a specific pair, with underlined values indicating the best encoder for each architecture and bold values showing the overall top combination. The row above the heatmap matrix presents architecture rankings (averaged across encoders), while the column at the left shows encoder rankings (averaged across architectures), where lower values indicate better performance. Throughout the figure, darker green shading consistently represents better performance: in the main matrix this corresponds to higher (better) IoU values, while in the rankings it indicates lower (better) average rank values.

The heatmap provides immediate visual comparison of segmentation performance, with colour intensity consistently representing superior performance (higher IoU or lower ranking values) through darker green shading.

Figure 6 follows the same structure but presents the results for the F_1 score. The findings are consistent across both metrics, with UNet being the top-performing architecture, followed by FPNet, while PSPNet consistently yields the worst performance. Among encoders, MIT B5, MambaOut, and ConvNeXt give the best results. Notably, the UNet-MambaOut combination achieves the highest overall performance, though UNet paired with MIT B5 produces nearly comparable results, suggesting that both encoders are highly effective choices for the UNet architecture.

With the aim of assessing whether the differences among the architectures are significant or not, Bayesian test was used. Since Bayesian test can only perform pair comparisons, the best architecture-encoder pairs were selected. Overall, UNet performed

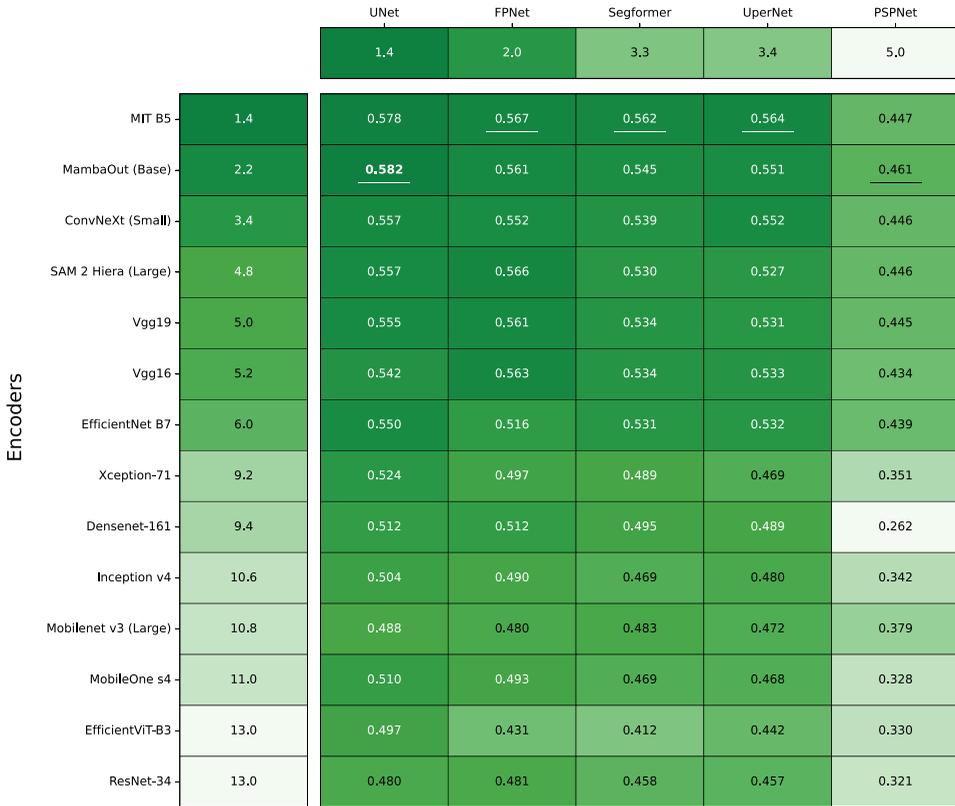


Figure 6. Heatmap of the F_1 performance measure across all encoder-architecture combinations. Each cell displays the mean IoU for a specific pair, with underlined values indicating the best encoder for each architecture and bold values showing the overall top combination. The row above the heatmap matrix presents architecture rankings (averaged across encoders), while the column at the left shows encoder rankings (averaged across architectures), where lower values indicate better performance. Throughout the figure, darker green shading consistently represents better performance: in the main matrix this corresponds to higher (better) F_1 values, while in the rankings it indicates lower (better) average rank values.

significantly better than the rest of the architectures. The differences between FPNet, UperNet, and Segformer were not statistically significant, while PSPNet performed significantly worse than the rest.

Figure 7 presents the summarized results of the Bayesian tests. Each cell shows the p_{left} (top) and p_{right} (bottom) values. The colour associated with each cell represents/codifies the value: $(p_{\text{left}} - p_{\text{right}})$, i.e., the probability that the method in the row is better than the method in the column minus the probability that the opposite occurs. The main diagonal has not been represented since it is always 0, i.e., a method is compared against itself. It can be observed that UNet significantly outperforms all the other architectures. Conversely, PSPNet is significantly worse than the other approaches. Finally, the performance differences between the other three models are not statistically significant, indicating that their results are very similar to one another.

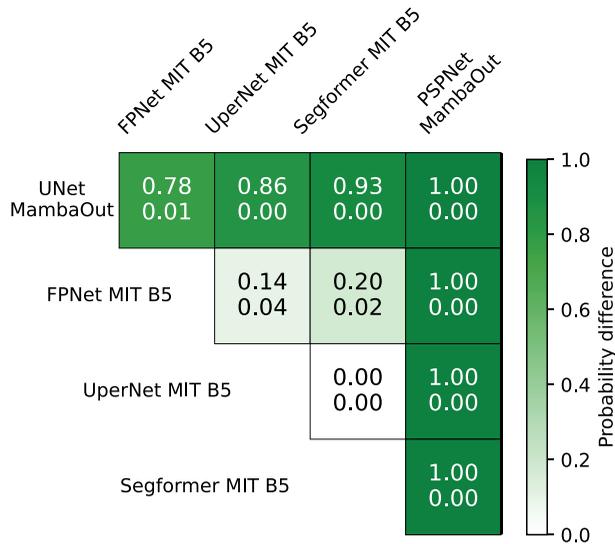


Figure 7. Bayesian tests heatmap, comparing the performance of the five architectures with their best encoder. Each cell contains the probability that the method in the row is better than the method in the column (p_{left}) on top and the probability that the opposite occurs (p_{right}) on the bottom. The green colour *intensity* associated with each cell represents the ($p_{left} - p_{right}$) value.

3.2. Results for other grassland habitats

To assess the model performance on unseen grassland types and landscapes (*i.e.*, images not used in model calibration and sourced from different locations across continents), a set of six images was selected (Figure 8). These images correspond to grassland, woodland, and agriculture-dominated landscapes in Africa, the Americas, Australia, and Europe. Predictions were generated using the best-performing architecture: UNet with MambaOut (Base).

In Figure 8, the first row shows human constructions and paths. The algorithm is generally not misled by building infrastructures, which it does not confuse with grazing trails. However, it incorrectly classifies parts of these structures, as well as parallel agricultural furrows, as grazing trails. Furrows were deliberately included in the image selection process, to create challenging identification conditions. In the second row, the pattern of a canopy of an Atlantic mixed forest (left) does not substantially mislead the algorithm, only some very short segments of false grazing trails are detected in the canopy; these could plausibly be interpreted as grazing trails, although they are not. On the right, the Mount Kenya is shown. Its heath and moorland belt (the ericaceous zone) constitutes a subalpine transitional band between the upper forest/treeline and the afro-alpine zone, characterized by Ericaceae-dominated shrublands (*e.g.*, *Erica* spp.) interspersed with tussock grasses, sedges, and herbs. In this image, we show how the existing trails are correctly identified indeed. In the third row, the methodology also yields satisfactory results, detecting paths that are actually present across different landscape types.

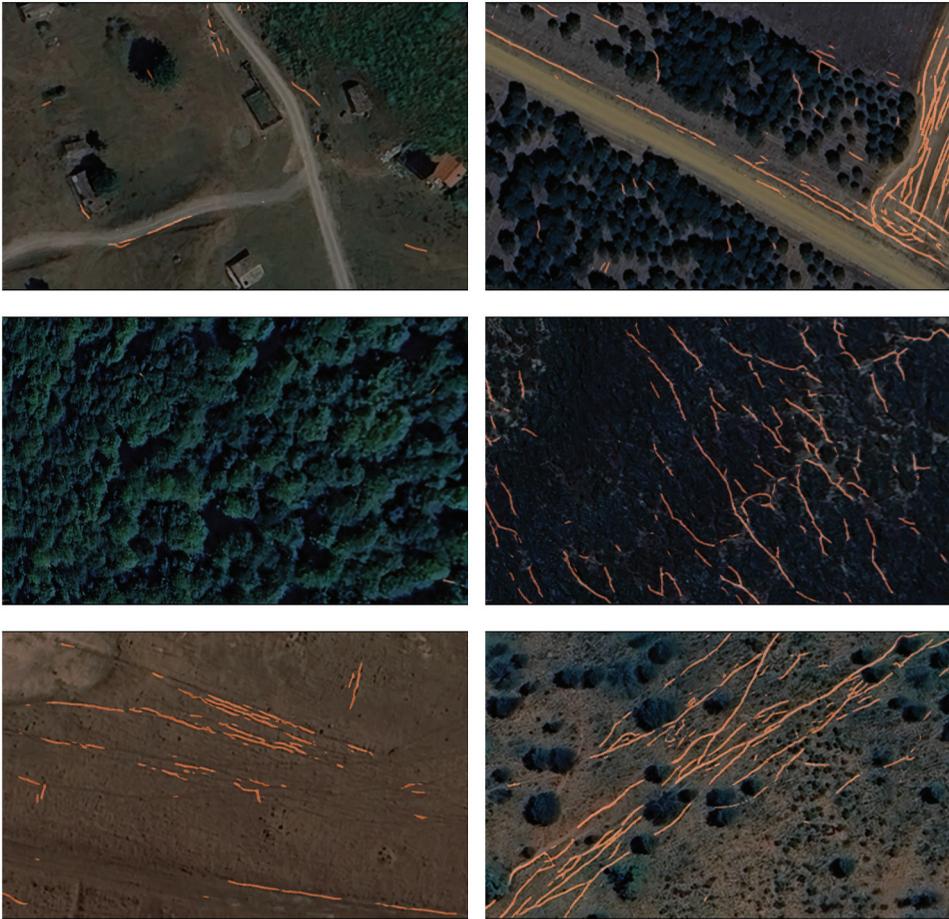


Figure 8. Prediction of herbivory trails on images from other habitats of the world. The predicted trails are highlighted in orange. Best viewed in colour.

4. Discussion

Our study shows that semantic segmentation is a powerful tool that can accurately identify grazing trails produced by large herbivorous mammals, and this was based on the capability of semantic segmentation for pixel classification, an advantage against, other approaches used in previous studies (Hellman et al. 2020).

Some approaches in automatic characterization have included the detection of continuous trampling activity of large herbivores based, among other factors, on the slope angle and relief (Danilo Godone 2018), abrupt elevation changes (Diaz-Varela et al. 2014), topographic position index (Sas et al. 2012), differentiation of terraces from natural plains using correlation (Sofia, Marinello, and Tarolli 2014), line segment detection methods applied to terrace walls (Bailly and Levvasseur 2012), the use of digital elevation models (Pijl et al. 2020), and the analysis of high-resolution images and classification based on the Fourier transformation (Hellman et al. 2020).

The morphology of the grazing trails network in our images mainly featured parallel paths over branched or radial ones. However, our models were able to detect these other morphologies, making them useful for applications in other *movement* scenarios, such as spatial displacement towards focal resources like water puddles in the savannah, natural salt licks, and grazing-rich spots, and to consider a few examples. It should be particularly stressed that (as previously stated) semantic segmentation predictions effectively differentiate between grazing trails and anthropogenic structures, such as motorized dirt tracks. This is an important feature since it may often appear in remote sensing imagery.

An important advantage of the approach we considered, when compared to that shown by Hellman et al. (2020), is that our method classifies grazing trails at pixel level, whereas Hellman et al. (2020) classify a complete image window as a window where a grazing trail appears. In particular, Hellman et al. (2020) detect areas with the presence of grazing trails (using public accessible images from Google or Bing) and then created windows of 250×250 pixels in size (37×37 m). For each one of them, they inferred the so-called Fourier domain periodogram. Periodograms display the squared magnitude of the discrete Fourier transform (DFT) (the so-called power spectrum) at all spatial frequencies of the image. From the selected image, and for each window, a 95th percentile histogram thresholding technique is applied, and a Fourier band-pass post-processing operation, afterwards. The inverse DFT completes the processing step. This resulting enhanced image is partitioned into patches, and each patch is then classified as grazing trail or not. However, this discretization on a window level fails to characterize grazing trails as a continuous and quantified feature and depending on the environment they can be more or less detectable.

Our approach supersedes Hellman et al. (2020) methodology as we are able to identify pixels that composed the trails (i.e., classification at a pixel level). This facilitates comparing different images of the same area over time and applying topology and spatial analysis to understand the morphological changes, reduction or increase in trails density, and their relation to changes in the activity of large herbivores. All these features are considerably useful in landscape monitoring programmes.

Grazing trails develop over heterogeneous land cover and topography that produces a gradient in the visibility of the trails, ranging from the highly distinguishable to the barely discernible Hellman et al. (2020) This leads to unavoidable errors, inherent to observer subjectivity when labelling these features (Powell et al. 2004) and our work is not free of this issue. In order to minimize subjectivity error in the labelling of trail images (i.e., when a fading trails begins or ends) it is advisable to use high-resolution images and to define a protocol describing the threshold to decide when a trail exists or not, which could be based on a trail continuity distance parameter Hellman et al. (2020). Our proposal is to use these methodologies for the automatic monitoring of grazing trails, allowing for their identification, quantification, and tracking over time, for example within Geographic Information System (GIS) tools.

For the semantic segmentation algorithms to be effectively used in GIS tools, it is essential that these algorithms (i) minimize type 1 errors, meaning they should not mistakenly identify other landscape features or infrastructures as grazing trails (e.g., structures with edges such as tree patches, dirt roads, fences, walls, agricultural furrows, etc.), and (ii) ensure they do not underestimate the amount of grazing trails

present. The first point is particularly crucial, as the second would still allow for the quantification and monitoring of changes in grazing trails. Our results suggest exactly that: while some grazing trail segments are not identified, other non-grazing trail structures have not been incorrectly classified as grazing trails. A clear example of this can be seen in [Figure 4](#), where the edges of a dirt road have not been classified as grazing trails. Despite these caveats, our methodology has proven to be useful to monitor grazing trails in a variety of habitats and shapes, which can be used to monitor trail changes over time and its potential impact on habitats as a proxy of herbivory intensity.

5. Conclusions and future work

This paper presents the first pixel-level semantic segmentation approach for mapping grazing trails created by large herbivores, advancing beyond the limitations of current patch-based detection methods. Whereas existing approaches can only determine the presence/absence of trails within coarse $37 \times 37m$ patches (Hellman et al. 2020), our method achieves precise trail detection at the original pixel resolution of the image. This could be used to reconstruct the full extent of the grazing trails, enabling precise measurement of trail morphology, connectivity, and spatial distribution patterns.

Through the systematic evaluation of several modern semantic segmentation architectures and encoding methods, it was determined that the best model architecture is UNet, closely followed by FPNNet, UperNet, and Segformer. In contrast, PSPNet showed the worst performance. Based on these findings, we recommend UNet for future studies in this domain and advise against the use of PSPNet. Additionally, another interesting result of this study is the important influence of the encoding method on the model performance, with MIT B5 and MambaOut (Base) performing the best across all the architectures tested.

The work presented here demonstrates that semantic segmentation can be effectively used to identify herbivory trails through transfer learning, even when relying on a limited set of manually labelled images. Future work could explore the use of data augmentation to expand the training dataset and assess whether increasing the amount of training data further improves performance without incurring additional labelling costs.

Importantly, this study shows that accurate detection of grazing trails can be achieved using only aerial RGB imagery, without the need to incorporate ancillary information such as Digital Elevation Models or other terrain descriptors. This independence from additional geospatial layers makes the approach logistically simpler and more pragmatic. Nonetheless, future work could explicitly test the incorporation of Digital Elevation Models into the modelling framework to evaluate whether terrain slope helps to characterize distinct grazing-trail geometries that may be conditioned by topography.

One of the most interesting potential uses of the proposed semantic segmentation algorithms is the mapping and analysis of the temporal variation of grazing trails to assess changes in herbivory intensity. On the other hand, it would be interesting to test the efficiency of the methods used in our study in other areas and grazed ecosystems. Future work would include the integration of these methods into geographic information system tools that would facilitate their use by land-managers and stakeholders to assist in landscape management and conservation programmes.

Another promising research direction is the integration of curvilinear-feature enhancement techniques to improve the segmentation of herbivory trails. The use of these techniques (*i.e.*, Carlotto (2007)) could help to identify complete trails that sometimes are unconnected, amplifying weak linear structures that are easily missed by conventional pixel-wise segmentation networks. Exploring hybrid pipelines that couple deep semantic segmentation with curvilinear-feature enhancement therefore represents a valuable direction for future work.

Notes

1. An Author's Original Manuscript (AOM) version (in arXiv) of this paper can be found at <https://www.arxiv.org/abs/2504.12121>.
2. HuggingFace is a repository of ready-to-use datasets for machine learning applications.
3. https://huggingface.co/datasets/jfdpastor/Comp_Geo_Herb_Trails_8x.
4. Multilayer Perceptron is a classical artificial neural network that consists of fully connected artificial neurons which are organized in layers. It has at least three layers, input layer (the first one), output layer (the last one and gives the prediction), and hidden layer (placed in between the other two and can be more than one).
5. <https://baycomp.readthedocs.io/>.

Author contributions

CRedit: **José Francisco Díez-Pastor:** Conceptualization, Formal Analysis, Funding acquisition, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing; **Francisco Javier González-Moya:** Software, Validation, Visualization, Writing – original draft, Writing – review and editing; **Pedro Latorre-Carmona:** Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing; **Francisco Javier Pérez-Barbería:** Ecological conceptualization, Data collection, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing; **Ludmila I. Kuncheva:** Formal analysis, Writing – original draft, Writing – review and editing; **Antonio Canepa-Oneto:** Conceptualization, Writing – original draft, Writing – review and editing; **Álvar Arnaiz-González:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review and editing; **César García-Osorio:** Formal analysis, Funding acquisition, Writing – review and editing.

Acknowledgments

This work is partially supported by the Asturias Biodiversity Complementary Program BIO06 (Next Generation EU/PRTR), Strategic Projects Oriented Towards Ecological and Digital Transition (TED2021-131388B-100), Spanish Knowledge Generation Projects (PID2023-146074OB-I00) funded by EU Next Generation and Spanish Research Agency, and Spanish Research Council Tenured Scientist Incorporation Grants 2022 (202230I041). J.F. Díez-Pastor was supported by a mobility grant PRX22/00634 from the Spanish Ministry of Universities.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Asturias Biodiversity Complementary Program BIO06 (Next Generation EU/PRTR), Strategic Projects Oriented Towards Ecological and Digital Transition (TED2021-131388B-100), Spanish Knowledge Generation Projects (PID2023-146074OB-I00) funded by EU Next Generation and Spanish Research Agency, Spanish Research Council Tenured Scientist Incorporation Grants 2022 (202230I041). J.F. Díez-Pastor was supported by a mobility grant PRX22/00634 from the Spanish Ministry of Universities and Ministerio de Ciencia e Innovación (ID2023-146074OB-I00).

ORCID

José Francisco Díez-Pastor  <http://orcid.org/0000-0001-5013-7505>
Francisco Javier González-Moya  <http://orcid.org/0000-0002-9473-9572>
Pedro Latorre-Carmona  <http://orcid.org/0000-0001-6984-5173>
Francisco Javier Pérez-Barbería  <http://orcid.org/0000-0001-7513-5418>
Ludmila Kuncheva  <http://orcid.org/0000-0002-0415-6964>
Antonio Canepa-Oneto  <http://orcid.org/0000-0002-0608-2743>
Álvar Arnaiz-González  <http://orcid.org/0000-0001-6965-0237>
César García-Osorio  <http://orcid.org/0000-0002-1206-1084>

Computer code availability

Datasets will be accessible via the HuggingFace Hub (https://huggingface.co/datasets/jfdpastor/Comp_Geo_Herb_Trails_8x). Raw Images and code will be accessible in GitHub. (<https://github.com/joseFranciscoDiez/HerbiSeg>).

References

- Bailly, J. S., and F. Levvasseur. 2012. "Potential of Linear Features Detection in a Mediterranean Landscape from 3D VHR Optical Data: Application to Terrace Walls." In *2012 IEEE International Geoscience and Remote Sensing Symposium*, 7110–7113. <https://doi.org/10.1109/IGARSS.2012.6352024>.
- Benavoli, A., G. Corani, J. Demšar, and M. Zaffalon. 2017. "Time for a Change: A Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis." *Journal of Machine Learning Research* 18:1–36. URL: <http://jmlr.org/papers/v18/16-305.html>.
- Bragagnolo, L., R. da Silva, and J. Grzybowski. 2021. "Amazon Forest Cover Change Mapping Based on Semantic Segmentation by U-Nets." *Ecological Informatics* 62:101279. <https://doi.org/10.1016/j.ecoinf.2021.101279>.
- Cai, H., J. Li, M. Hu, C. Gan, and S. Han. 2023. "EfficientViT: Lightweight Multi-Scale Attention for High-Resolution Dense Prediction." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17302–17313.
- Carlotto, M. J. 2007. "Enhancement of Low-Contrast Curvilinear Features in Imagery." *IEEE Transactions on Image Processing* 16 (1): 221–228. <https://doi.org/10.1109/TIP.2006.884949>.
- Chemekova, T. Y., and Y. F. Chemekov. 1975. "Hillslope Terracettes." *Soviet Geography* 16 (9): 609–615. <https://doi.org/10.1080/00385417.1975.10640098>.
- Chen, J., M. Xia, D. Wang, and H. Lin. 2023. "Double Branch Parallel Network for Segmentation of Buildings and Waters in Remote Sensing Images." *Remote Sensing* 15 (6): 1536. <https://doi.org/10.3390/rs15061536>.
- Chollet, F. 2017. "Xception: Deep Learning With Depthwise Separable Convolutions." arXiv:1610.02357.

- Costa, D., V. V. Costa, and E. Rocha. 2025. "A New Parametric Information-Gain Criterion for Tree-Based Machine Learning Algorithms." *PeerJ Computer Science* 11:e3319. <https://doi.org/10.7717/peerj-cs.3319>.
- da Costa, L. B., O. L. F. de Carvalho, A. O. de Albuquerque, R. A. T. Gomes, R. F. Guimarães, and O. A. de Carvalho Júnior. 2022. "Deep Semantic Segmentation for Detecting Eucalyptus Planted Forests in the Brazilian Territory Using Sentinel-2 Imagery." *Geocarto International* 37 (22): 6538–6550. <https://doi.org/10.1080/10106049.2021.1943009>.
- Danilo Godone, D. G. 2018. "Rapid Mapping Application of Vegetated Terraces Based on High Resolution Airborne LiDAR." *Geomatics, Natural Hazards and Risk* 9 (1): 970–985. <https://doi.org/10.1080/19475705.2018.1478893>.
- Diaz-Varela, R., P. Zarco-Tejada, V. Angileri, and P. Loudjani. 2014. "Automatic Identification of Agricultural Terraces Through Object-Oriented Analysis of Very High Resolution DSMs and Multispectral Imagery Obtained from an Unmanned Aerial Vehicle." *Journal of Environmental Management* 134:117–126. <https://doi.org/10.1016/j.jenvman.2014.01.006>.
- Dong, G., Y. Yan, C. Shen, and H. Wang. 2021. "Real-Time High-Performance Semantic Image Segmentation of Urban Street Scenes." *IEEE Transactions on Intelligent Transportation Systems* 22 (6): 3258–3274. <https://doi.org/10.1109/TITS.2020.2980426>.
- ESA. 2025. "S2 Mission." Accessed April 9, 2025. <https://sentiwiki.copernicus.eu/web/s2-mission>.
- Fibaek, C., L. Camilleri, A. Luyts, N. Dionelis, and B. Le Saux. 2024. "Phileo Bench: Evaluating Geo-Spatial Foundation Models." ArXiv:2401.04464.
- Filazzola, A., C. Brown, M. A. Dettlaff, A. Batbaatar, J. Grenke, T. Bao, I. Peetoom Heida, and J. F. Cahill Jr. 2020. "The Effects of Livestock Grazing on Biodiversity Are Multi-Trophic: A Meta-Analysis." *Ecology Letters* 23 (8): 1298–1309. <https://doi.org/10.1111/ele.13527>.
- Gordon, I. J., and H. H. T. Prins. 2019. *The Ecology of Browsing and Grazing II*: 239. Springer Cham.
- Goudie, A. 2013. *Encyclopedia of Geomorphology*. Taylor & Francis. <https://books.google.com/books?id=E01saQ-xwK0C>.
- Gu, A., and T. Dao. 2024. "Mamba: Linear-Time Sequence Modeling with Selective State Spaces." *arXiv preprint arXiv:2312.00752*.
- Guo, Y., Y. Liu, T. Georgiou, and M. Lew. 2018. "A Review of Semantic Segmentation Using Deep Neural Networks." *International Journal of Multimedia Information Retrieval* 7 (2): 87–93. <https://doi.org/10.1007/s13735-017-0141-z>.
- Hao, S., Y. Zhou, and Y. Guo. 2020. "A Brief Survey on Semantic Segmentation with Deep Learning." *Neurocomputing* 406:302–321. <https://doi.org/10.1016/j.neucom.2019.11.118>.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Hellman, I., R. Heinse, J. W. Karl, and M. Corrao. 2020. "Detection of Terracettes in Semi-Arid Rangelands Using Fourier-Based Image Analysis of Very-High-Resolution Satellite Imagery." *Earth Surface Processes and Landforms* 45 (13): 3368–3380. <https://doi.org/10.1002/esp.4971>.
- Higgins, C. G. 1982. "Grazing-Step Terracettes and Their significance." *Zeitschrift für Geomorphologie* 26 (4): 459–472. <https://doi.org/10.1127/zfg/26/1982/459>.
- Hiltbrunner, D., S. Schulze, F. Hagedorn, M. W. Schmidt, and S. Zimmermann. 2012. "Cattle Trampling Alters Soil Properties and Changes Soil Microbial Communities in a Swiss Sub-Alpine Pasture." *Geoderma* 170:369–377. <https://doi.org/10.1016/j.geoderma.2011.11.026>.
- Hong, D., B. Zhang, H. Li, Y. Li, J. Yao, C. Li, M. Werner, J. Chanussot, A. Zipf, and X. X. Zhu. 2023. "Cross-City Matters: A Multimodal Remote Sensing Benchmark Dataset for Cross-City Semantic Segmentation Using High-Resolution Domain Adaptation Networks." *Remote Sensing of Environment* 299:113856. <https://doi.org/10.1016/j.rse.2023.113856>.
- Hosseini, M., F. Miranda, J. Lin, and C. T. Silva. 2022. "Citysurfaces: City-Scale Semantic Segmentation of Sidewalk Materials." *Sustainable Cities and Society* 79:103630. <https://doi.org/10.1016/j.scs.2021.103630>.
- Howard, A., M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan, W. Wang, et al. 2019. "Searching for Mobilenetv3." <https://arxiv.org/abs/1905.02244>, arXiv:1905.02244.

- Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger. 2018. "Densely Connected Convolutional Networks." <https://arxiv.org/abs/1608.06993>, arXiv:1608.06993.
- Iakubovskii, P. 2019. "Segmentation Models PyTorch." https://github.com/qubvel/segmentation_models.pytorch.
- Jadhav, J. K., and R. P. Singh. 2018. "Automatic Semantic Segmentation and Classification of Remote Sensing Data for Agriculture." *Mathematical Models in Engineering* 4 (2): 112–137. <https://doi.org/10.21595/mme.2018.19840>.
- Jamali, A., S. K. Roy, J. Li, and P. Ghamisi. 2024. "Neighborhood Attention Makes the Encoder of ResUNet Stronger for Accurate Road Extraction." *IEEE Geoscience and Remote Sensing Letters* 21:1–5. <https://doi.org/10.1109/LGRS.2024.3354560>.
- Jiménez-Navarro, M., M. Martínez-Ballesteros, F. Martínez-Álvarez, and G. Asencio-Cortés. 2023. "Philnet: A Novel Efficient Approach for Time Series Forecasting Using Deep Learning." *Information Sciences* 632:815–832. <https://doi.org/10.1016/j.ins.2023.03.021>.
- Jin, B., H. Cheng, G. Sun, F. Li, and X. B. Wu. 2022. "Multi-Parallel Structure and a Generalized Conceptual Model of Livestock Track Network." *Catena* 216:106380. <https://doi.org/10.1016/j.catena.2022.106380>.
- Jin, B., G. Sun, Y. Zhang, M. Zou, X. Ni, K. Luo, X. Zhang, H. Cheng, F. Li, and X. B. Wu. 2016. "Livestock Tracks Transform Resource Distribution on Terracette Landscapes of the Loess Plateau." *Ecosphere* 7 (4): e01337. <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecs2.1337>.
- Jones, C. G., J. H. Lawton, and M. Shachak. 1994. "Organisms as Ecosystem Engineers." In *Oikos*, 373–386. Vol. 69. Wiley: Nordic Society Oikos. <https://doi.org/10.2307/3545850>.
- Khan, S., M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah. 2022. "Transformers in Vision: A Survey." *ACM Computing Surveys (CSUR)* 54 (10s): 1–41. <https://doi.org/10.1145/3505244>.
- Lin, T. Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. 2017. "Feature Pyramid Networks for Object Detection." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944. <https://doi.org/10.1109/CVPR.2017.106>.
- Liu, Z., H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. 2022. "A ConvNet for the 2020s." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Luo, Z., W. Yang, Y. Yuan, R. Gou, and X. Li. 2023. "Semantic Segmentation of Agricultural Images: A Survey." *Information Processing in Agriculture* 11 (2): 172–186. <https://doi.org/10.1016/j.inpa.2023.02.001>.
- May, R. M. 1981. "Mammal Tracks on Mountain Slopes." *Nature* 292 (5825): 672–673. <https://doi.org/10.1038/292672a0>.
- Mills, L. S., M. E. Soulé, and D. F. Doak. 1993. "The Keystone-Species Concept in Ecology and Conservation." *Bioscience* 43 (4): 219–224. <https://doi.org/10.2307/1312122>.
- Minaee, S., Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. 2022. "Image Segmentation Using Deep Learning: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44:1–1. <https://doi.org/10.1109/TPAMI.2021.3059968>.
- Mnih, V., and G. E. Hinton. 2010. "Learning to Detect Roads in High-Resolution Aerial Images." In *European Conference on Computer Vision (ECCV)*, 210–223. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-15567-3_16.
- Pal, M., H. Akshay Rohilla, and B. C. Teja. 2020. "Patch Based Land Cover Classification: A Comparison of Deep Learning, SVM and NN Classifiers." In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 1933–1936. <https://doi.org/10.1109/IGARSS39084.2020.9323755>.
- Parajuli, J., R. Fernandez-Beltran, J. Kang, and F. Pla. 2022. "Attentional Dense Convolutional Neural Network for Water Body Extraction from Sentinel-2 Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:6804–6816. <https://doi.org/10.1109/JSTARS.2022.3198497>.
- Pijl, A., J. S. Bailly, D. Feurer, M. A. El Maoui, M. R. Boussema, and P. Tarolli. 2020. "Terra: Terrain Extraction from Elevation Rasters through Repetitive Anisotropic Filtering." *International Journal of Applied Earth Observation and Geoinformation* 84:101977. <https://doi.org/10.1016/j.jag.2019.101977>.

- Powell, R., N. Matzke, C. de Souza, M. Clark, I. Numata, L. Hess, and D. Roberts. 2004. "Sources of Error in Accuracy Assessment of Thematic Land-Cover Maps in the Brazilian Amazon." *Remote Sensing of Environment* 90 (2): 221–234. <https://doi.org/10.1016/j.rse.2003.12.007>.
- Robbins, C. 1993. *Wildlife Feeding and Nutrition*. Vol. 2nd. Academic Press.
- Ronneberger, O. F., P. T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 234–241. Berlin Heidelberg, Cham: Springer. https://doi.org/10.1007/978-3-319-24574-4_28.
- Ryali, C., Y. T. Hu, D. Bolya, C. Wei, H. Fan, P. Y. Huang, V. Aggarwal, et al. 2023. "Hiera: A Hierarchical Vision Transformer without the Bells-and-Whistles." In *Proceedings of the 40th International Conference on Machine Learning, PMLR*, edited by A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, 29441–29454.
- Sas, R. J., J. P. Yu, C. Y. Y. Fugro Pau, and K. A. Styles. 2012. "Detection of Old Agricultural Terraces in Steep, Vegetated Terrain Using Airborne Lidar: Case Studies from Hong Kong." In *33rd Asian Conference on Remote Sensing*, 408–417. URL: <https://api.semanticscholar.org/CorpusID:15246203>.
- Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. "NIH Image to ImageJ: 25 Years of Image Analysis." *Nature Methods* 9 (7): 671–675. <https://doi.org/10.1038/nmeth.2089>.
- Simonyan, K., and A. Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." URL: <https://arxiv.org/abs/1409.1556>, arXiv:1409.1556.
- Sofia, G., F. Marinello, and P. Tarolli. 2014. "A New Landscape Metric for the Identification of Terraced Sites: The Slope Local Length of Auto-Correlation (SLLAC)." *ISPRS Journal of Photogrammetry and Remote Sensing* 96:123–133. <https://doi.org/10.1016/j.isprsjprs.2014.06.018>.
- Stavi, I., E. D. Ungar, H. Lavee, and P. Sarah. 2008. "Grazing-Induced Spatial Variability of Soil Bulk Density and Content of Moisture, Organic Carbon and Calcium Carbonate in a Semi-Arid Rangeland." *Catena* 75 (3): 288–296. <https://doi.org/10.1016/j.catena.2008.07.007>.
- Stavi, I., H. Yizhaq, Y. Osem, and E. Argaman. 2021. "Positive Impacts of Livestock and Wild Ungulate Routes on Functioning of Dryland Ecosystems." *Ecology and Evolution* 11 (20): 13684–13691. <https://doi.org/10.1002/ece3.8147>.
- Szegedy, C., S. Ioffe, V. Vanhoucke, and A. Alemi. 2016. "Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning." <https://arxiv.org/abs/1602.07261>, arXiv:1602.07261.
- Tan, M., and Q. V. Le. 2020. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." <https://arxiv.org/abs/1905.11946>, arXiv:1905.11946.
- Vasu, P. K. A., J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan. 2022. "An Improved One Millisecond Mobile Backbone." ArXiv Preprint arXiv:2206.04040.
- Washington-Allen, R. A., T. G. V. Niel, R. D. Ramsey, and N. E. West. 2004. "Remote Sensing-Based Piosphere Analysis." *GIScience and Remote Sensing* 41 (2): 136–154. <https://doi.org/10.2747/1548-1603.41.2.136>.
- Watanabe, T. 1994. "Soil Erosion on Yak-Grazing Steps in the Langtang Himal, Nepal." *Mountain Research and Development* 14 (2): 171–179. <https://doi.org/10.2307/3673799>.
- Wieland, M., S. Martinis, R. Kiefl, and V. Gstaiger. 2023. "Semantic Segmentation of Water Bodies in Very High-Resolution Satellite and Aerial Images." *Remote Sensing of Environment* 287:113452. <https://doi.org/10.1016/j.rse.2023.113452>.
- Wieren, S. E., and J. P. Bakker. 2008. "The Impact of Browsing and Grazing Herbivores on Biodiversity." In *The Ecology of Browsing and Grazing*, edited by I. J. Gordon and H. H. Prins, 263–292. Vol. 195. Berlin Heidelberg: Springer.
- Wilcoxon, F. 1992. "Individual Comparisons by Ranking Methods." In *Breakthroughs in Statistics: Methodology and Distribution*, 196–202. Springer. https://doi.org/10.1007/978-1-4612-4380-9_16.
- Wurm, M., T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck. 2019. "Semantic Segmentation of Slums in Satellite Images Using Transfer Learning on Fully Convolutional Neural Networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 150:59–9. <https://doi.org/10.1016/j.isprsjprs.2019.02.006>.
- Xiao, T., Y. Liu, B. Zhou, Y. Jiang, and J. Sun. 2018. "Unified Perceptual Parsing for Scene Understanding." In *Proceedings of the European Conference on Computer Vision (ECCV)*, 432–448. Cham: Springer International Publishing.

- Xie, E., W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. 2021. "Segformer: Simple and Efficient Design for Semantic Segmentation with Transformers." In *Advances in Neural Information Processing Systems, Curran Associates, Inc*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, 12077–12090.
- Zhang, C., X. Chen, and S. Ji. 2022. "Semantic Image Segmentation for Sea Ice Parameters Recognition Using Deep Convolutional Neural Networks." *International Journal of Applied Earth Observation and Geoinformation* 112:102885. <https://doi.org/10.1016/j.jag.2022.102885>.
- Zhang, X., J. Jin, Z. Lan, C. Li, M. Fan, Y. Wang, X. Yu, and Y. Zhang. 2020. "ICENET: A Semantic Segmentation Deep Network for River Ice by Fusing Positional and Channel-Wise Attentive Features." *Remote Sensing* 12 (2): 221. <https://doi.org/10.3390/rs12020221>.
- Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. 2017. "Pyramid Scene Parsing Network." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6230–6239. Los Alamitos, CA, USA: IEEE Computer Society. <https://doi.org/10.1109/CVPR.2017.660>.