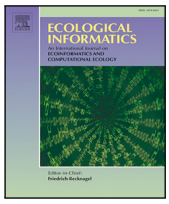
Ecological Informatics 93 (2026) 103523

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Deep tracks: Using deep learning and procedurally simulated data for automated vertebrate footprints classification

Carolina S. Marques [a,*], Afonso Mota [b,c], Matteo Belvedere [d], Diego Castanera [e], Ignacio Díaz-Martínez [f], Elisabete Malafaia [g,h], Soraia Pereira [a,i,j], Luís Miguel Rosalino [k,l], Vanda F. Santos [g,m,n], Lara Sciscio [o,p], Emmanuel Dufourq [q]

[a] Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[b] Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), Universidade do Porto, Porto, Portugal
[c] Instituto de Astrofísica e Ciências do Espaço, CAUP, Universidade do Porto, Porto, Portugal
[d] Dipartimento di Scienze della Terra, Università degli Studi di Firenze, Via G. La Pira, Firenze, Italy
[e] Grupo Aragosaurus-IUCA. Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain
[f] Departamento de Ciencias de la Tierra y Física de la Materia Condensada, Universidad de Cantabria, Santander, Spain
[g] Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[h] Grupo de Biología Evolutiva, Universidad Nacional de Educación a Distancia, Madrid, Spain
[i] Centro de Matemática (CMAT), Universidade do Minho, Guimarães, Portugal
[j] Departamento de Matemática, Universidade do Minho, Guimarães, Portugal
[k] Center for Ecology, Evolution and Environmental Change (cE3c), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[l] Global Change and Sustainability Institute (Change), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[m] Departamento de Geologia, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[n] PaleoIbérica Research Group, University of Alcalá, Alcalá de Henares, Spain
[o] Jurassica Museum, Porrentruy, Switzerland
[p] Department of Geosciences, University Fribourg, Fribourg, Switzerland
[q] Computer Science, University of Northern Iowa, Cedar Falls, United States of America

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.15196257, Vertebrate-Tracks-Classifier, https://zenodo.org/records/15092442

Keywords:
Simulated data
Vertebrate footprints
Transfer learning
Deep learning
Automatic classification

ABSTRACT

The study of vertebrate footprints provides useful information on animal behavior, locomotion, and ecology. However, automatically classifying these records using photographs is difficult due to the significant morphological variation in footprints and the lack of readily available labeled datasets. To address this issue, this study developed Deep Tracks, a novel Unity application to procedurally create a dataset of simulated footprint images. Two datasets were used to evaluate the influence and impact of the simulated dataset on real footprint classification: (1) a dataset comprising 40,000 simulated footprints, (2) approximately 1,500 real vertebrate footprints from 10 different vertebrate groups. Both simulated and real footprints belong to the following clades: Mammalia (coyotes, foxes, bears, otters, squirrels, raccoons, deer), avian Dinosauria (turkeys) and non-avian Dinosauria (theropods, sauropods). Convolutional Neural Networks (CNNs) were used to classify the different datasets either from the simulated or real footprints. An initial comparison of five different architectures (DenseNet-121, ResNet-18, ResNet-50, EfficientNet-b0, and InceptionNet-v3) was done using the simulated dataset, with EfficientNet-b0 presenting better metrics results. Seven experimental configurations were designed to evaluate different strategies for incorporating the real data into the model development. The first configuration involved training and testing exclusively on real footprints, without any simulated data. The second configuration trained the model on real data, but tested it on simulated footprints. The third configuration used transfer learning to fine-tune a CNN, initially trained on simulated data, for classifying real footprint images. The remaining four configurations incorporated simulated data into the training process alongside a fixed percentage of real data — 20%, 50%, 80%, or 100%. The application of fine-tuning led to an accuracy improvement of over 30% in classifying real footprints, compared to a CNN trained solely on real data. These results highlight the significance of advanced data augmentation techniques in improving both accuracy and reliability in vertebrate footprint classification, particularly in scenarios with limited real data availability.

* Corresponding author.
E-mail address: csmarques@fc.ul.pt (C.S. Marques).

https://doi.org/10.1016/j.ecoinf.2025.103523
Received 17 April 2025; Received in revised form 10 October 2025; Accepted 10 November 2025
Available online 19 November 2025

1. Introduction

The study of vertebrate footprints – whether of living animals or fossils – offers insights into the behavior, distribution and displacement of their producers, as footprints capture traces of animals in motion. They enable paleobiological interpretations regarding taxonomic identity, limb anatomy, locomotion (including posture, gait, and trackmaker speed), and social structure (Lockley, 1986, 1998; Buatois and Mángano, 2011; Romano and Whyte, 2012; Buatois and Mángano, 2016; Milàn et al., 2016; Toledo and Arregui, 2023; Falkingham, 2025; Díaz-Martínez et al., 2023). Therefore, the study of footprints helps understanding both modern and ancient ecosystems (Lockley, 1986, 1998; Buatois and Mángano, 2011; Romano and Whyte, 2012; Buatois and Mángano, 2016; Milàn et al., 2016; Toledo and Arregui, 2023). For extant species, footprints serve as non-invasive biometric markers that can be used to monitor population dynamics, landscape use, and species distribution without disturbing the animals (e.g., Jewell and Alibhai, 2013; Riordan, 1998). Fossil footprints, on the other hand, provide complementary data to skeletal remains by revealing information on behavior, posture, and substrate interaction— details often not available in the body fossil record (e.g., Lockley, 1998; Lallensack et al., 2022).

Despite their potential, the broader application of vertebrate footprint analysis has been limited by the scarcity of extensive, high-quality, and well-annotated datasets. Traditional manual measurements are labor-intensive, subjective, and struggle to capture the inherent variability introduced by substrate differences, preservation conditions, and environmental factors (Russell et al., 2009; Geng et al., 2012; Jewell and Alibhai, 2013; Milàn et al., 2016; Frey, 1975; Belvedere et al., 2018; Bennett and Budka, 2018; Marchetti et al., 2019). Moreover, interpreting footprint morphology is often challenging due to weathering, erosion, and, in the case of extant vertebrates, due to the difficulty of precisely correlating prints with the trackmaker's anatomical structures. For example, Russell et al. (2009) noted that automated track recognition methods often struggle due to the scarcity of large-scale, accurately labeled footprint datasets, which are essential to capture the full variability inherent in natural prints. Similarly, Geng et al. (2012) highlighted the challenges in collecting and annotating footprint images – especially for small species – due to issues such as degradation and variability in substrate conditions, which complicate reliable classification. Lallensack et al. (2022) also emphasized that the scarcity of comprehensive, high-fidelity datasets limits the performance of machine learning techniques in discriminating between morphologically similar track types. Additionally, Lockley (1998) emphasized that while the fossil footprint record is abundant, its utility is undermined by inconsistent preservation and annotation practices, thereby restricting its broader application in both ecological and paleobiological research. Collectively, these studies illustrate the need for standardized, large-scale, and correctly annotated datasets to fully reach the potential of vertebrate footprint analysis.

Recent advances in simulation, deep learning and synthetic data generation offer promising approaches to overcome these challenges. For instance, Bonetto et al. (2023) introduced Generating Realistic And Dynamic Environments (GRADE) – a highly customizable framework based on NVIDIA IsaacSim – that generates realistic and dynamic virtual environments. By allowing precise control over simulation parameters such as lighting, physics, and asset placement, GRADE has been shown to significantly reduce the gap between synthetic and real-world data. It does so by simulating animated humans to improve human detection and segmentation, thereby improving the robustness of models on a subset of the COCO-Humans dataset (Lin et al., 2014). Building on this concept, Bonetto and Ahmad (2023) showed the effectiveness of synthetic data in the detection of zebras in aerial drone imagery. Their approach used the GRADE framework to produce richly annotated datasets that capture the inherent variability of natural scenes, leading to object detectors that outperform those trained on limited real-world data alone. Models trained exclusively on synthetic data outperformed COCO-pretrained models on all but one dataset, achieving performance improvements ranging from approximately 20% to 45% in average precision at an Intersection over Union threshold of 0.50. This study highlights the potential of simulation pipelines to overcome the challenges of data scarcity and annotation errors that commonly affect wildlife monitoring efforts.

Furthermore, Vierbergen et al. (2023) showed that a carefully engineered synthetic pipeline can effectively bridge the sim-to-real gap in challenging detection tasks. In their study on automated Calendula flower harvesting, the authors generated synthetic data through photogrammetry, and a dedicated flower field simulator, achieving an F1-score (the balance between recall and precision) of up to 86% on real datasets despite training exclusively on synthetic images. This result highlights how environmental variability – such as changes in illumination and substrate – can be modeled in simulation to build robust detectors.

In recent work on animal pose estimation, researchers have addressed the domain adaptation challenge using complementary approaches. For example, Cao et al. (2019) dealt with the domain adaptation problem by transferring pose knowledge from extensive human pose datasets to unlabeled animal images. Their cross-domain adaptation technique mitigated the severe domain shift between human and animal images, thereby allowing more accurate estimation of animal poses with minimal labeled animal data. This approach highlights the importance of using shared anatomical features across species to improve model performance in ecological applications. Similarly, Mu et al. (2020) proposed a method that learns from synthetic animals generated from Computer-Aided Design (CAD) models. Their consistency-constrained semi-supervised learning framework utilizes both spatial and temporal consistencies to bridge the domain gap between synthetic and real images. By incorporating pseudo-labels derived from high-confidence predictions, their method achieves competitive performance in keypoint detection and animal parsing, even when real image labels are scarce.

Building on similar principles, replicAnt Plum et al. (2023) provided a comprehensive simulation pipeline implemented in Unreal Engine 5 (Epic Games, Inc., North Carolina, USA) that generates large-scale, annotated image datasets of animals in complex environments. By automating the annotation process and introducing extensive variability in appearance, pose, and lighting, replicAnt significantly reduced the need for hand-labeled data. The synthetic datasets produced using replicAnt have been successfully applied to train deep neural networks for animal detection, tracking, pose estimation, and semantic segmentation, thereby improving model robustness and domain invariance. Notably, in object detection tasks, YOLOv4 (Bochkovskiy et al., 2020) models trained solely on synthetic data achieved up to 91% mean Average Precision (mAP), outperforming real-only models (which reached up to 88%). Mixed models combining synthetic with limited real data achieved the highest performance with up to 93% mAP. This highlights the effectiveness of using information from both synthetic and real-world samples.

The current literature illustrates the potential of simulation-based approaches and cross-domain adaptation in different ecological problems. We hypothesized that similar performance gains could be achieved in vertebrate footprint classification. This study is the first to train multiple convolutional neural networks (CNNs) on a large Unity-generated (Unity Technologies, 2024) synthetic dataset, capturing variations in footprint shape, substrate texture, and lighting. These CNNs integrate real data through different learning strategies, including pre-training on synthetic images followed by fine-tuning on a smaller real-world dataset. This approach enhances classification accuracy while improving the scalability and reliability of ecological monitoring.

By combining insights from both extant and extinct footprint studies, our work highlights the dual utility of footprint analysis. Both fossil and modern footprints contain key information for the study of animal

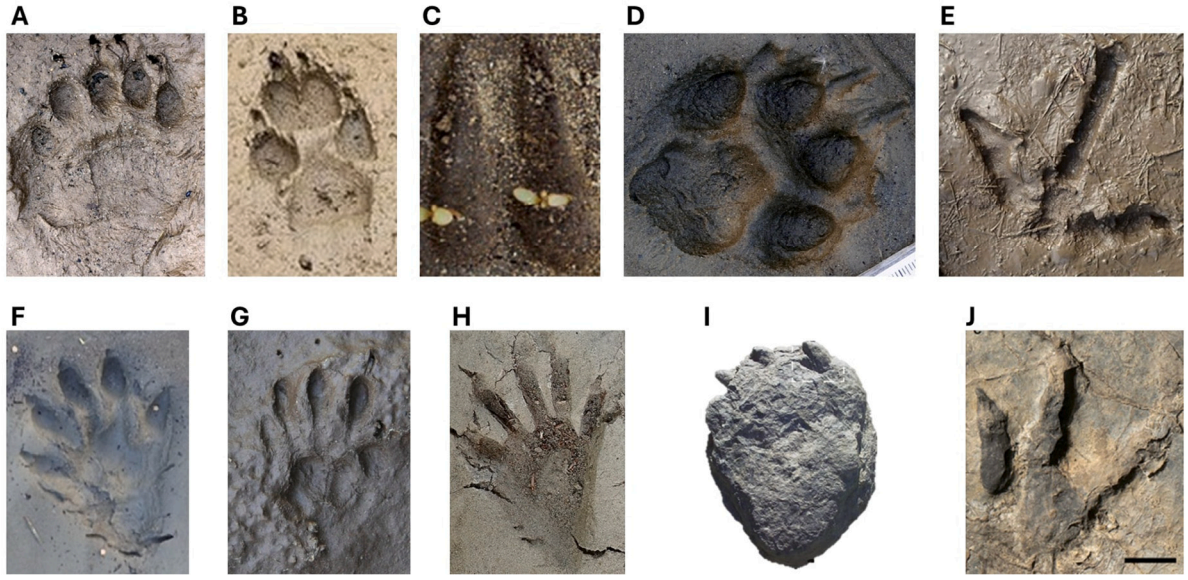


Fig. 1. Examples of real vertebrate footprints that were used to train the models, illustrating variation in morphology across the taxa. A- bear, B- coyote, C- deer, D- fox, E- turkey, F- otter, G- squirrel, H- raccoon, I- sauropod natural cast, J-theropod.
Source: A-H were retrieved from Shinoda and Shiohara (2024), I was retrieved from Díaz-Martínez et al. (2023), and J was retrieved from Castanera et al. (2015).

behavior. With fossil footprints mainly contributing to our understanding of locomotion, behavioral evolution, and environmental conditions in deep time (Lockley, 1998). Whereas modern footprints provide real-time data on species presence, species identity, and population dynamics (Riordan, 1998; Jewell and Alibhai, 2013). To our knowledge, this is the first study to develop a synthetic footprint creation tool, named Deep Tracks, for the purpose of vertebrate classification. Using Deep Tracks, we generated a large-scale simulated dataset that contributes 40,000 images to assist in model pre-training for vertebrate classification. This study is also the first study that created a general automatic classifier of vertebrate footprints images, including both extant and extinct vertebrates and using fine-tuning.

2. Materials and methods

2.1. Data collection and simulation

The approach adopted in this study integrates simulated and real vertebrate footprints to train and evaluate a CNN across 10 broader animal classes — bears, coyotes, deer, foxes, otters, raccoons, sauropods, squirrels, theropods, and turkeys. Two main sources of images were compiled, firstly, a simulated dataset generated under systematically varied conditions, secondly, a real-footprint dataset gathered from previously published sources (Shinoda and Shiohara, 2024; Marty et al., 2010; Faria dos Santos et al., 2024; Sciscio et al., 2023, 2022; Razzolini et al., 2017; Sciscio et al., 2016, 2017; Abrahams et al., 2017, 2020; Razzolini et al., 2016, 2014; Lockley and Meyer, 1998; Martill et al., 2024; Díaz-Martínez et al., 2015, 2023; Castanera et al., 2020, 2023, 2011, 2012, 2014, 2015; Paratte et al., 2017, 2018a,b, 2017; Castanera et al., 2016a,b- Fig. 1). Each class in the simulated dataset included 3000 training images and 1000 validation images. The real dataset contained 100 to 213 training images per class. Both datasets assigned 75% of the images to training, and the remaining 25% to validation (Table 1).

2.2. Deep tracks: procedural image generation

To overcome limitations related to the availability of real footprint picture datasets, Deep Tracks was developed for the procedural generation of realistic animal footprints. For this task, we used the Unity Engine, due to its flexible 3D rendering capabilities, accessibility, and integration with C# for scripting. The final software leverages Unity's 3D built-in tools, like its flexible terrain system, to simulate diverse environmental conditions, based on template footprint textures provided by the user.

Users have extensive control over image generation parameters via a user interface that includes adjustable footprint dimensions (size and depth), substrate texture, image resolution, the maximum number of footprints per image, camera orientation (including randomized angles), lighting conditions (Sun's position), and mold options (regular mold or counter-mold) (Fig. S1). This controlled diversity allowed the synthetic images to capture some of the morphological and contextual variations often seen in field-collected footprints (Fig. S2).

During footprint simulation, Unity's terrain height map is manipulated using the footprint textures. The textures consisted of grayscale images with a maximum of 50 × 50 pixels — larger images added negligible detail to the imprint while increasing the computational cost. The textures are randomly rotated, scaled, and positioned onto terrains, with additional variability introduced through random offsets and tiling adjustments. Environmental realism is further enhanced by simulated lighting and weather effects that alter the terrain appearance. The application supports creating both realistic and simplified visual outputs including grayscale and silhouette (grayscale) representations.

For each generated footprint image, a metadata file is output along with the image, recording the simulation parameters. This includes the animal footprint type and dimensions (size, depth), terrain substrate, and other relevant data. This facilitates detailed analysis and reproducibility of the simulation, as well as serving as useful input for model training. Deep Tracks is openly accessible for the research community in Mota et al. (2025).

The simulated dataset created by Deep Tracks aimed to address the limited availability and inherent variability of real vertebrate footprint samples. By systematically controlling substrate properties, and trackmaker morphology in the Unity-based simulation, we generated diverse footprints capturing morpho-dynamic variations that are difficult to obtain empirically. This dataset was used both to augment real data during model training and to assess how simulated variations influence classification performance, demonstrating its utility as a controlled proxy for real footprints in scenarios of data scarcity. Fig. 2 compares examples of simulated and real footprints.

Table 1
Overview of the simulated and real datasets used for each footprint class, including training and validation splits.

Class	Simulated data (training)	Simulated data (validation)	Real data (training)	Real data (validation)
Bear	3,000	1,000	200	66
Coyote	3,000	1,000	129	43
Deer	3,000	1,000	204	68
Fox	3,000	1,000	135	45
Turkey	3,000	1,000	116	39
Otter	3,000	1,000	105	35
Squirrel	3,000	1,000	213	71
Raccoon	3,000	1,000	210	70
Sauropod	3,000	1,000	195	65
Theropod	3,000	1,000	158	54

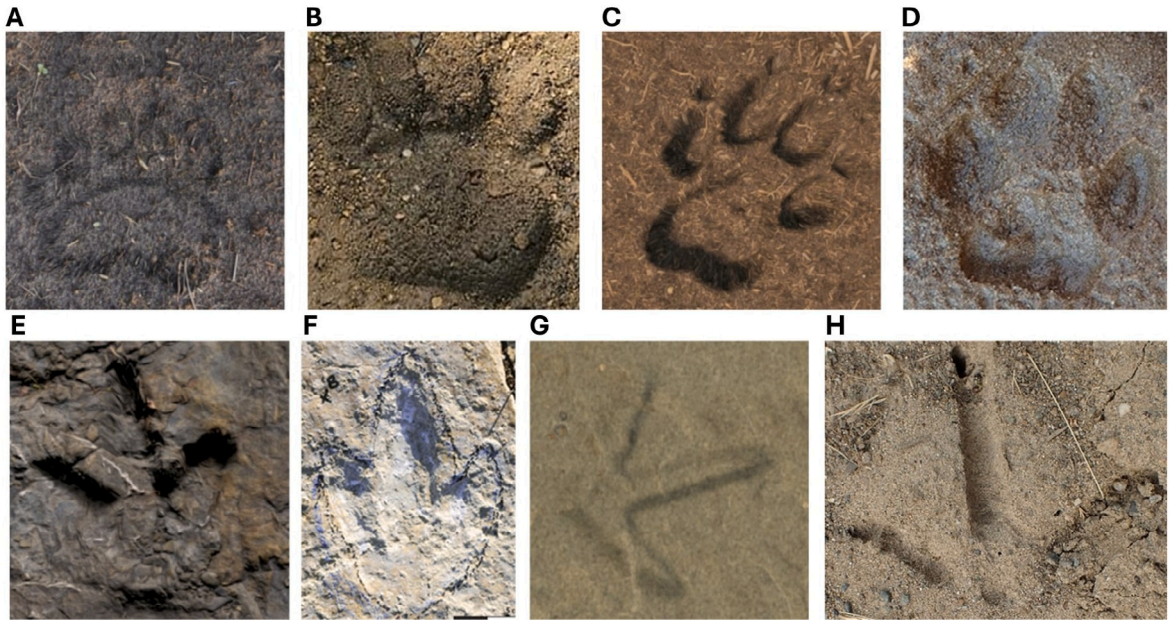


Fig. 2. Examples of real and simulated vertebrate footprints that were used to train the models, illustrating different scenarios and similarities between simulated and real data.
Source: A- simulated bear, B- real bear, C- simulated fox, D- real fox, E- simulated theropod, F- real theropod, G- simulated turkey, H- real turkey. B, D, and H were retrieved from Shinoda and Shiohara (2024), F was retrieved from Paratte et al. (2018b).

2.3. Simulating footprint textures

The simulation process begins by inputting black-and-white footprint silhouettes, each with a maximum width or height of 50 pixels (Fig. 3, top left). The silhouettes were obtained through previously published animal guides and studies (e.g., Olsen, 2013; Walker, 1996; Hanson and Hanson, 2023; Sanz et al., 2004; Santos et al., 2009; Torcida Fernández-Baldor et al., 2021; Castanera et al., 2018). To increase variability within each animal class, random grayscale points were added with their intensity values being drawn from a normal distribution, $U(0, 255)$, and their position from $U(min(width, height), max(width, height))$ (Fig. 3, all except top left).

The modified silhouettes were then input into Deep Tracks, where the simulation's goal was to add noise and variability when creating the simulated images. The application used a uniform distribution to randomly select variables, including the silhouette to imprint on the soil, mold or counter-mold type, soil type and humidity, footprint depth, size, position, rotation, Sun position, and both camera position (physical location of the camera relative to the soil) and angle (orientation of the camera relative to the surface).

In our synthetic data generation process, each footprint is "imprinted" in a background texture that simulates soil or ground surfaces. These textures, referred to as "soil images", are not derived from real photographs but instead originate from procedural and texture-based materials used within the Unity game engine. The soil dataset includes

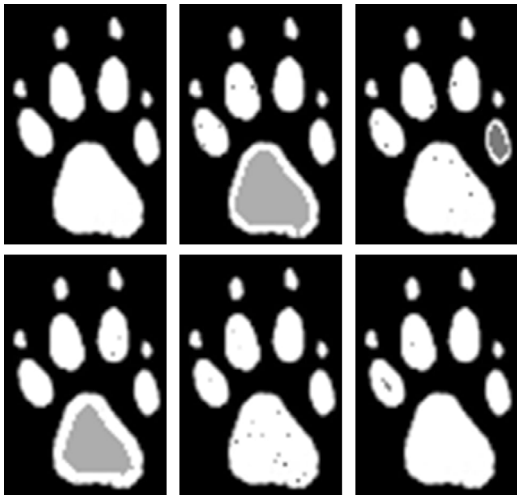


Fig. 3. An example of a wolf footprint silhouette after the addition of grayscale points to the original silhouette (top left). The grayscale intensity values, position and size was selected randomly following Uniform distribution.

a variety of textures that represent different surface types (e.g., sand, mud, compact dirt, rock), each with distinct visual properties. Each texture is represented as a 2D image indexed by i, denoted as $s_i(x, y)$, where x and y correspond to pixel positions in the image.

Before imprinting the silhouette, the soil was normalized using the equation:

$$\tilde{s}i(x, y) = \frac{s_i(x, y) - s_{\min}(x, y)}{s_{\max}(x, y) - s_{\min}(x, y)}, \quad (1)$$

where $\tilde{s}i(x, y)$ is the normalized value of the soil data at position (x, y) for sample i, $s_i(x, y)$ is the original value at that position, and $s_{\min}(x, y)$ and $s_{\max}(x, y)$ are the minimum and maximum values across all soil data at that position, respectively.

In addition to generating simulated images, corresponding black-and-white silhouette images were created for each simulated image. These silhouettes were produced using a white layer (pixel values of 1) that turned black (pixel values of 0) where soil distortion occurred (e.g., where a footprint was imprinted). The silhouette creation process is defined as:

$$S(x, y) = \begin{cases} 0 & \text{if } D(x, y) \geq T \\ 1 & \text{if } D(x, y) < T, \end{cases} \quad (2)$$

where $S(x, y)$ is the new silhouette value at position (x, y), $D(x, y)$ is the soil deformity measure (e.g., depth or pressure) at position (x, y), and T is a threshold value that determines the boundary between deformed and non-deformed areas.

Through the use of the newly simulated silhouettes S, the bounding box for the simulated footprints was obtained and the simulated images were cropped to contain only the relevant information of the image (Fig. 4). The simulated silhouettes were used to extract bounding boxes around the footprints, ensuring that simulated images contained only relevant footprint information. This step helped align the simulated dataset with the real footprint dataset (as the real images were cropped around the footprints). Additionally, to further enhance realism – particularly for fossil footprints (sauropod and theropod) – the synthetic images were randomly processed into three variations. Firstly, the original footprint with a white background, secondly, the original footprint with a black background, and lastly, the unmodified synthetic image. These variations simulated different fossil preservation contexts, such as curated specimens in collections (variations 1 and 2) or footprints observed in the field (variation 3).

2.4. Experimental configurations

Seven experimental configurations were designed to evaluate the CNN performance when transitioning from simulation-only training, mixed real and simulated training, and real-only training scenarios (Fig. 5). In real-only training (Fig. 5 - config. A), a model was trained from randomly initialized weights, solely on the real dataset, without using simulated data.

In real and simulated mixed training (Fig. 5 - config. B-E), models were trained using all of the simulated data and a certain percentage of real training data (20%, 50%, 80%, 100%). Lastly, in simulation-only training (Fig. 5 - config. F and G) the model was trained on the simulated dataset (3000 images per class). In the first configuration (Fig. 5 - config. F), the trained model was tested on the real validation dataset to assess how well synthetic training footprints generalized to real examples. In the second configuration (Fig. 5 - config. G), the pre-trained model was fine-tuned on the real training set and validated on the real validation set. Additionally, the validation dataset was kept fixed across all configurations to maintain comparability of accuracy results. The impact of simulated training dataset size on the fine-tuning approach was assessed by training models using 500, 1000, 2000, and 3000 images per class and by evaluating their performance.

2.5. Data pre-processing and neural networks

Initial model training compared five CNN architectures - DenseNet-121 (Huang et al., 2017), ResNet-18 (He et al., 2016), ResNet-50 (He et al., 2016), EfficientNet-b0 (Tan and Le, 2019), and InceptionNet-v3 (Szegedy et al., 2016). The dataset used for the initial model selection was the simulated dataset, which included 3000 training images and 1000 validation images for each of the 10 classes. The architectures were configured for multi-class classification by replacing the final layer with a linear layer outputting predictions for 10 classes. All of the off-the-shelf architectures that were used for comparisons were pretrained on ImageNet (Deng et al., 2009) to accelerate convergence. We did not include a Softmax layer since the loss function used was the Pytorch (Paszke et al., 2019) cross-entropy loss function, which internally applies the log-softmax transformation. The cross-entropy loss function used in during the classification task is given by

$$\mathcal{L}(\mathbf{x}, y) = -\log\left(\frac{\exp(x_y)}{\sum_{j=1}^{C}\exp(x_j)}\right), \quad (3)$$

which can be rewritten as

$$\mathcal{L}(\mathbf{x}, y) = -x_y + \log\left(\sum_{j=1}^{C}\exp(x_j)\right), \quad (4)$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_C]$ is the vector of raw logits produced by the model, x_y is the logit corresponding to the true class y, and C is the total number of classes.

All images were normalized and resized to 224×224 pixels (to align with the model's input size requirements), except for InceptionNet-v3, which required 299×299 pixels.

Different data augmentation techniques were evaluated, and the final models applied random rotation with a probability of 0.2. All images were normalized prior to modeling. Preliminary experiments revealed that the optimal batch size was 32. Following preliminary experimentation, the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.001 was selected to train the models. All model configurations were set to run for 30 epochs, where early stopping was performed if two consecutive epochs resulted in a decrease of the validation accuracy (Hussein and Shareef, 2024). Additionally, none of the models' layers were frozen during training.

Model performance was evaluated using accuracy (the ratio of correct predictions to total samples), precision (the proportion of true positives among all positive predictions), recall (the proportion of true positives among all actual positive samples), F1-score (the harmonic mean of precision and recall), and the normalized confusion matrix. Systematic comparisons across the seven configurations assessed the generalizability of synthetic footprints to real images, and the impact of adding real data to the training set on final model performance. Energy consumption and memory usage were evaluated across the different architectures and across the different configurations, where the validation dataset was always the real data validation dataset.

All image pre-processing, modeling, and evaluation were performed using Python (Python Software Foundation, 2023) (version 3.12.7). The PyTorch library was used for modeling, including fine-tuning, training, validation, and testing. Image pre-processing was conducted with OpenCV (Itseez, 2015). Energy consumption was obtained by the Codecarbon library (Benoit Courty et al., 2024) and memory usage was measured by the psutil library (Loden and Rodolà, 2024). The CNNs were trained on an NVIDIA GeForce RTX 4060 Ti GPU and the image pre-processing was done using Intel(R) Core(TM) i5-13400 CPU. Experiments were conducted on a computer running (Windows-11-10.0.26100-SP0) with 32 GB RAM.

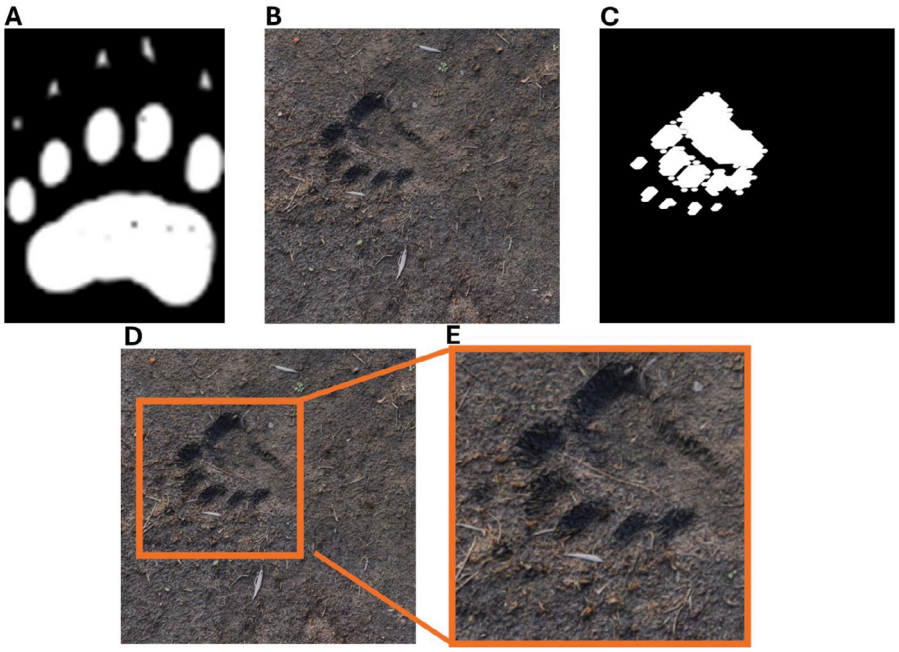


Fig. 4. An example of the pre-processing steps done for the simulated images. A- silhouette used to simulate footprints, fed to the application and "printed" on a randomly selected soil image, B- resulting simulated image, C- resulting black and white version of the footprint simulated, D- simulated image with the obtained bounding box, E- footprint image obtained after the removal of the information outside the bounding box.

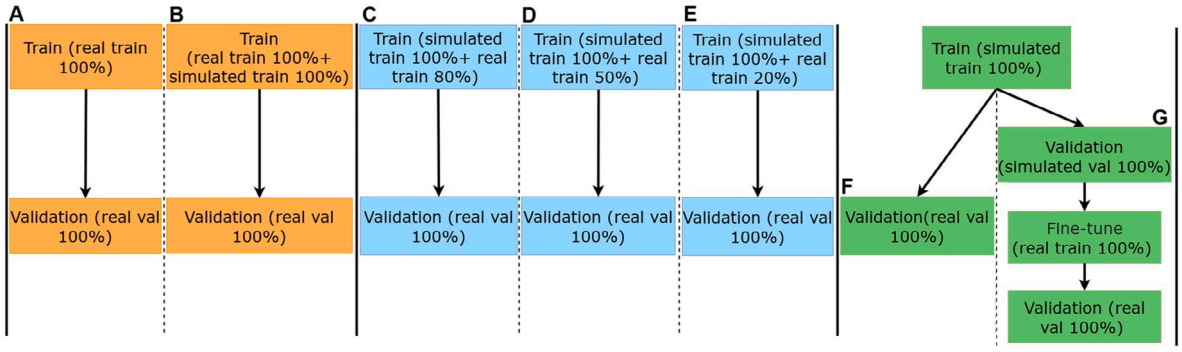


Fig. 5. Schematic representation of the seven different experimental configurations. A and B both use the full training real image dataset, with B also including the simulated dataset in the training. C, D and E correspond to training combining the simulated dataset with a fixed fraction of the real training data — 20%, 50%, 80%, respectively. F (no fine-tuning) and G (fine-tuning) are both initially trained using only simulated images. G uses the model pre-trained on the simulated data, and fine-tunes it to the real images. Both F and G are evaluated on the real validation dataset.

3. Results

Using Deep Tracks, 40,000 synthetic images and their corresponding black and white representations were generated. This took approximately 10 h, averaging 0.9 s per footprint. The processed images exhibited substantial variability in backgrounds and styles (Fig. 6). For example, Fig. 6 – sub-figure I features a black background – a deliberate artifact of the post-simulation process, where fossil footprints were randomly assigned a black background to simulate curated specimens from collections.

The findings revealed that EfficientNet-b0, the model with the least network parameters, performed best, achieving 97.0% accuracy on the simulated validation dataset, while ResNet-50 had the lowest accuracy at 91.7% (Fig. 7). EfficientNet-b0 converged quickly and had the highest performance after the first epoch. The energy consumption and RAM usage results indicate that ResNet-18 had the lowest energy consumption, emitting 0.0053 kilograms of carbon dioxide equivalent (kgCO2eq) per five epochs of training, while EfficientNet-b0 had the lowest RAM usage, requiring 957.8 megabytes (MB) to train for five epochs (Table 2). In contrast, Inception exhibited the highest values for both metrics, consuming 0.011 kgCO2eq per five epochs of training and using 1,068.66 MB of RAM over the same training duration. After five epochs of training, EfficientNet-b0 achieved the highest accuracy on both the training and validation datasets, exceeding 95% in both cases. In contrast, ResNet-50 performed the worst, with accuracy remaining below 55% (Table 2). Additional performance metrics are presented in Fig. S3.

The performance of EfficientNet-b0 across the different experimental configurations revealed that the highest validation accuracy was achieved in config. G (74.0%), and the next best result was obtained by config. B (72.4%) (Fig. 8). In contrast, the worst performance was observed in config. A (36.0% accuracy) or in config. F (24.0% accuracy) (Fig. 8).

The energy consumption and RAM usage obtained showed that the experimental configuration with less energy consumption was config. A (0.00031 kgCO2eq per five epochs of training), and that the model config. F used less RAM (936.7 MB of RAM) to run five epochs of training (Table 3). The configuration that presented the highest energy

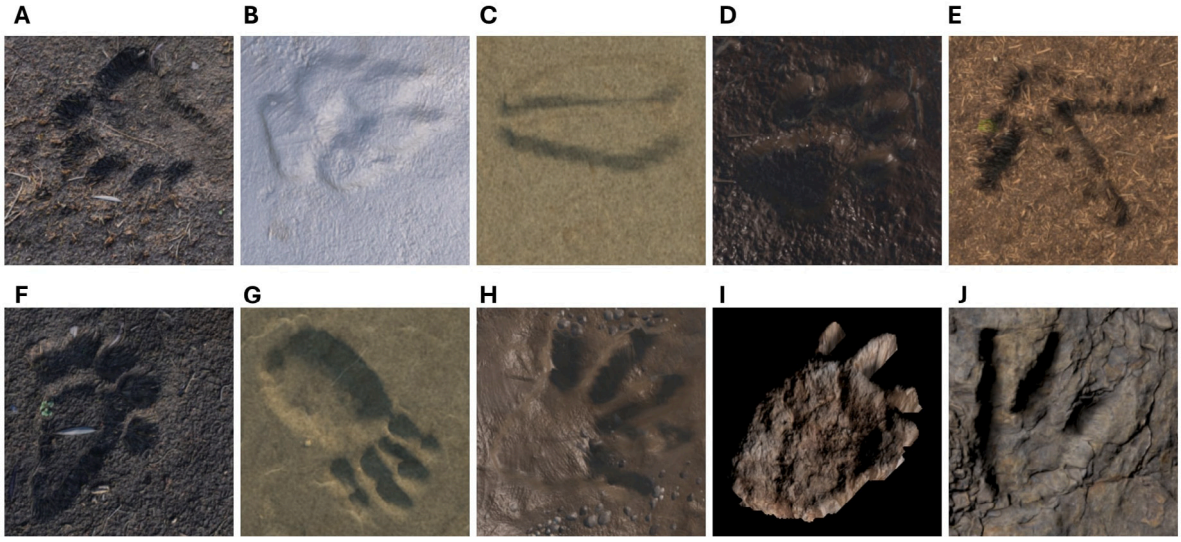


Fig. 6. Examples of simulated vertebrate footprints from Deep Tracks. A- bear, B- coyote, C- deer, D- fox, E- turkey, F- otter, G- squirrel, H- raccoon, I- sauropod, J-theropod.

Table 2
Energy consumption in kgCO2eq, and RAM usage, across different network architectures after five training epochs.

Configuration	Energy consumed (kgCO2eq)	Final RAM (MB)	Train accuracy (%)	Val accuracy (%)
EfficientNet-b0	0.0072	957.81	96.27	95.14
ResNet50	0.0095	1,042.30	54.78	51.95
ResNet18	0.0053	1,039.42	93.52	91.95
DenseNet121	0.0099	1,052.74	93.36	92.00
Inception	0.0110	1,068.66	95.18	94.28

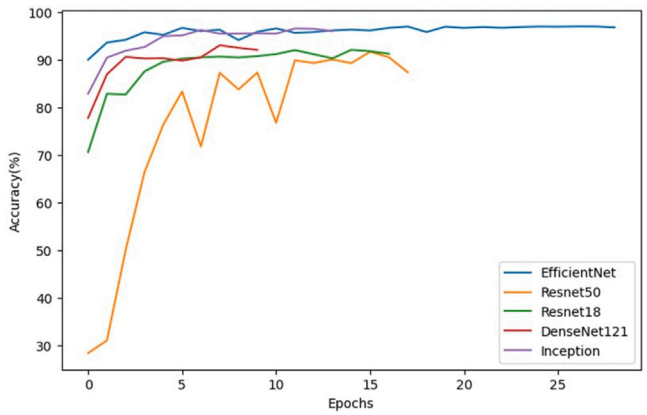


Fig. 7. Accuracy percentages achieved by the different CNN architectures trained and validated on the simulated dataset. The number of training epochs varied due to the application of early stopping, as described in the methodology section.

consumption values (0.0065 kgCO2eq per five epochs of training) was the model trained on config. F; the model with higher usage of RAM was config. B, using 977.9 Mb of RAM to run the five epochs of training (Table 3). The accuracy after the five epochs of training showed that the best configuration for both validation and training dataset was config. G, with an accuracy of 96.8% on the training and 70.8% for the validation. The worst configuration was the model using config. F, with the validation accuracy not reaching values higher than 21% (Table 3). Additional performance metrics are presented in Fig. S4 and the best metric results are presented in Table S1.

Analyzing the performance of different EfficientNet-b0 models across varying sizes of simulated training datasets, the best results were achieved using 2000 simulated images during the initial training phase. This configuration yielded an accuracy of 97.0% on the simulated validation dataset and 76.0% on the real validation dataset (Fig. 9). Additional performance metrics are presented in Fig. S5 for the simulated footprints dataset, and in Fig. S6 for the real footprints dataset.

In the best-performing configuration (config. G, using only 2000 simulated images per class), the model correctly classified over 70% of images for all classes except coyote, fox, otter, and theropod (Fig. 10). The highest accuracies were observed for bear (85%), deer (82%), and sauropod (82%). The most common misclassifications were: (1) theropod as sauropod, (2) otter as raccoon, (3) squirrel as raccoon, (4) coyote as fox, (5) otter as fox, (6) otter as squirrel, (7) sauropod as theropod (Fig. 10).

Considering the top-2 predicted classes – i.e., the two most probable classes output by the model – an accuracy of 86.6% was achieved on the validation dataset. The only class with a proportion below 75% was the otter, which continued to be frequently misclassified as a raccoon or as a squirrel (Fig. 11).

4. Discussion

Among the architectures evaluated, EfficientNet-b0 outperformed InceptionV3, ResNet50, ResNet18, and DenseNet121 in footprint classification. This superior performance can be related to its compound scaling strategy, which simultaneously optimizes depth, width, and resolution, allowing a more effective capture of the morphological features inherent to both simulated and real footprint data. In contrast, more complex models tended to overfit or struggled with the high variability and limited quantity of real samples, leading to suboptimal generalization. EfficientNet-b0's relatively lower parameter count (5.3 million compared to 11.7–25.6 million in the other models) and efficient architecture allowed it to extract discriminative features with greater precision, making it particularly suited for vertebrate footprint classification.

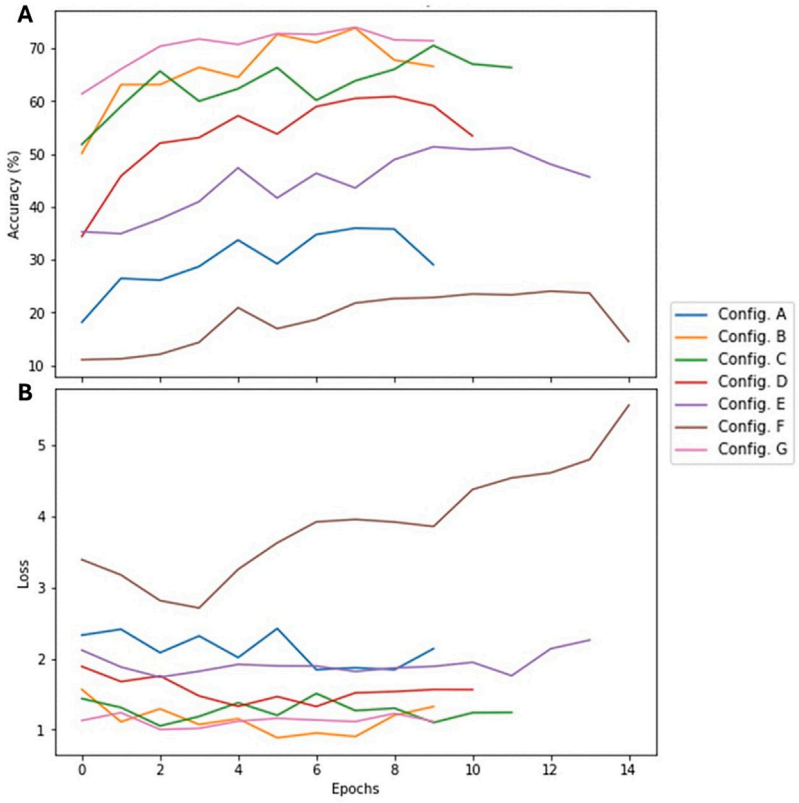


Fig. 8. Validation accuracy percentages achieved for the different experimental configurations evaluated on the real dataset. Config. G obtained the highest performance. The epochs of training varied due to the application of early stopping, which was described in the methodology section. See Fig. 5 for details on the different configurations.

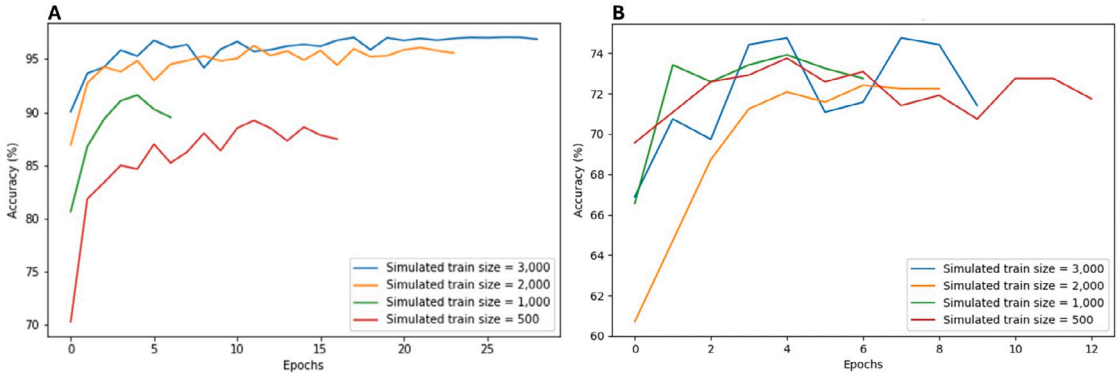


Fig. 9. Accuracy percentages for the different simulated training sizes. A- validation accuracy for the simulated validation dataset, B- validation accuracy for the real validation dataset, during fine-tuning. The epochs of training varied due to the application of early stopping, which was described in the methodology section.

Table 3
Energy consumption in kgCO2eq and RAM usage for EfficientNet-b0 across different experimental configurations evaluated after five training epochs.

Configuration	Energy consumed (kgCO2eq)	Final RAM (MB)	Train accuracy (%)	Val accuracy (%)
Config. A	0.00031	959.64	30.40	33.74
Config. B	0.00610	977.89	95.17	64.63
Config. C	0.00600	953.43	95.05	62.37
Config. D	0.00600	968.54	95.44	57.27
Config. E	0.00600	964.93	95.82	47.40
Config. F	0.00650	936.68	28.75	20.93
Config. G	0.00050	970.96	96.82	70.76

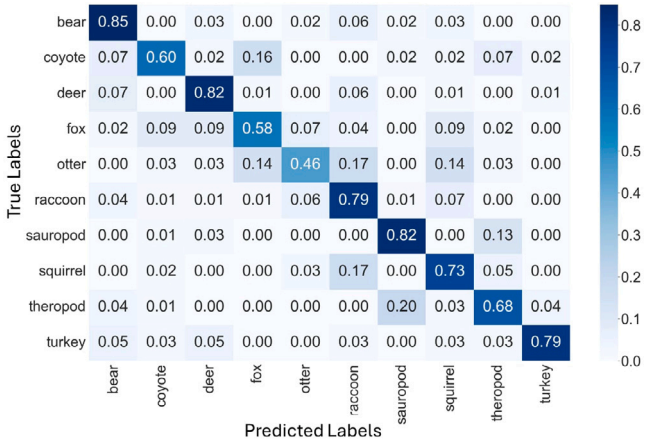


Fig. 10. Normalized confusion matrix showing the top-1 predictions for config. G.

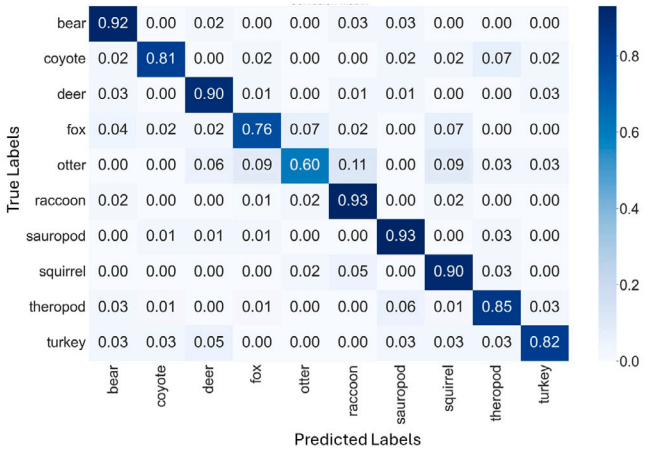


Fig. 11. Normalized confusion matrix showing the top-2 predictions for config. G.

Our results show the effectiveness of including simulated data in various configurations to improve model performance in footprint classification tasks. It is important to note that while most configurations show an overall increase in validation accuracy across epochs, some curves present fluctuating or declining trends after the initial increase, particularly those combining limited real data with simulated data. This may reflect overfitting to specific features in the simulated dataset, discrepancies between simulated and real data distributions, or the small size of the real dataset affecting generalization. The use of early stopping was intended to mitigate these effects by preventing overtraining when validation accuracy plateaued or decreased. When comparing models trained on different combinations of real and simulated data (Fig. 5 - configs. B–E), we observed a consistent increase in validation accuracy as the proportion of real data in the training set increased. This suggests that, while simulated data provides a useful baseline, the inclusion of more real data helps the model better generalize to real-footprints variability. Notably, the best performance was achieved when 100% of the available real data was combined with simulated data from the outset (Fig. 5 - config. B), highlighting the value of hybrid datasets for training robust models. However, our findings revealed that pre-training on simulated footprints followed by fine-tuning on real data (Fig. 5 - config. G) obtained better metrics, compared to models trained on a mixture of simulated and real images. This improvement stems from several factors. Firstly, training solely on simulated data allowed the model to learn a broad and robust representation of the underlying morphological features in a controlled environment, free from the noise and inconsistencies inherent in real-world images. Secondly, the simulated dataset, being abundant and balanced, allowed the model to capture essential characteristics across all classes without being influenced by domain-specific irregularities. Thirdly, fine-tuning with real data provides a focused adaptation step, where the model adjusts its learned representations to accommodate the unique textures, lighting conditions, and variabilities found in actual footprints. Moreover, a comparison between real data only (Fig. 5 - config. A) and a combination of real and simulated data (Fig. 5 - config. B) showed that including simulated samples in the training dataset can significantly improve performance. The unique contribution of the simulated dataset lies in its ability to systematically expand footprint variability under controlled conditions, complementing scarce real data and enabling more robust classifier training. The limited size of the dataset, relative to the model's extensive parameter count, makes training from scratch impractical. Training only on the real dataset, without any fine-tuning (Fig. 5 - config. A), yielded some of the lowest metrics values.

In addition to classification accuracy, the computational efficiency of the different architectures and training configurations were evaluated in terms of energy consumption and memory usage. Among all tested models, the fine-tuned EfficientNet-b0 architecture presented the most favorable trade-off between performance and efficiency. It achieved high validation accuracy while maintaining relatively low memory requirements and energy consumption across different training setups. EfficientNet-b0 consumed approximately 0.0072 kgCO2eq and required 957 MB of memory at peak usage (i.e., the maximum amount of memory used during training), outperforming heavier models like ResNet-50, which consumed more memory (approx. 1042 MB) and that delivered lower accuracy. When evaluating training configurations using different proportions of real data, it was possible to conclude that memory usage increased with dataset size and complexity, but not always linearly with performance improvements. In particular, fine-tuning configurations maintained a high accuracy (70.8%) with minimal memory increase (approx. 45%) and very low energy consumption (approx. 0.00050 kgCO2eq).

The results revealed several recurrent misclassifications, most notably coyote footprints labeled as fox, theropod footprints labeled as sauropod, otter footprints labeled as fox, raccoon or squirrel, and squirrel as raccoon. The imbalance in the real dataset likely played a significant role in the misclassifications observed. As shown in Table 1, while the simulated data included 3000 training images and 1000 validation images per class, the real dataset presented considerable variations across classes. For example, coyote footprints were represented by only 129 real training images and 43 validation images, while otter footprints had just 105 real training images and 30 validation images. Similarly, the validation set contained only 35 otter footprints – a group known for its high variability (Kistner et al., 2022, 2023) – and approximately 40 images for both coyotes and foxes. Additionally, squirrels had the highest number of testing images (71), followed by the raccoons (70), which may have biased the model toward predicting footprints as belonging to these classes. Given that track morphology is closely related to substrate properties and locomotion dynamics – for example, wetter, more pliable surfaces can lead to increased distortion – the limited size of the real dataset likely constrains the model's ability to robustly distinguish between naturally occurring variations in footprint shape and true inter-class differences. This scarcity of real-world samples may prevent the model from fully learning the range of substrate-induced deformations, thereby contributing to the observed misclassifications.

Another reason that might have led to the misclassifications were morphological similarities between groups (Fig. 12). Douglas de Carvalho et al. (2015), discusses the existence of overlap in shape and size of some carnivores, in particular in small fields, and De Angelo et al. (2010) discusses similarities between large felids and large canids. Since coyotes and foxes are both canids, their footprints share four toe

Fig. 12. Variability that can be noted in the real footprint examples of the most commonly misidentified extant vertebrate footprint groups. These differences can arise from multiple factors, including sediment composition and saturation, animal behavior (i.e., speed/gait), ontogenetic stage, and post-registration degradation (i.e., weathering), or a combination of these influences.

pads arranged in a similar pattern, making partial or low-quality prints hard to distinguish. Fox footprints were also occasionally labeled as squirrel footprints, possibly because size cues were lost in some images, and because partial impressions did not clearly show the canid pad structure. Additionally, confusion between squirrel and fox prints may also appear since squirrel footprints typically display four toe pads on the front feet and five on the hind feet; thus, if only the front footprint is analyzed, a squirrel print can easily be mistaken for fox footprint. Future work should focus on improving the quality of photographs and standardizing the photographic methodology – using a consistent camera position, zoom level, and including a scale of uniform size – to better differentiate similar footprints, especially among groups with morphological similarities. Collecting more real data is also essential to support these improvements.

In some cases, during the analysis it was possible to see possible misclassifications within the original dataset in the extant vertebrate groups (e.g., first photograph of the fox footprints in Fig. 12, which appears to exhibit five toe pads with an unusual arrangement). These possible misclassifications in the source database could propagate through the training process and compromise the accuracy of subsequent automated identification.

Otter footprints were sometimes misclassified as raccoon or squirrel. These misclassifications could be due to the fact that all three animal groups can show five toe impressions, and in muddy or worn conditions the presence of webbing on otter prints may not be readily apparent. Otter footprints have been previously confused with rabbit, raccoon, opossum, fox, dog, cat, nutria, bobcat, turtle, rat, armadillo, and bullfrog as mentioned by Evans et al. (2009), where the most common mistakes were between otters and (1) raccoons, (2) opossums, (3) dogs, and (4) cats. Although otters and foxes are taxonomically distinct (Family Mustelidae and Canidae, respectively), their footprints occasionally share overlapping morphological features, particularly when image quality is compromised by factors such as partial impressions, poor lighting, or substrate irregularities, which can result in otter footprint images displaying only four toe pads. Otter prints are typically characterized by a broader pad and subtle indications of webbing (Sanz et al., 2004), features that may be less pronounced or obscured in degraded samples. In contrast, fox footprints tend to be more elongated (Sanz et al., 2004); however, when the prints are incomplete or blurred, the distinctiveness of these features can diminish, leading to misclassifications. Moreover, the relatively small number of real otter and fox footprints in the training set limits the model's ability to learn the full range of variability inherent to each class.

Similarly, confusion between squirrel and raccoon footprints appears to be based on subtle morphological similarities and the inherent variability of footprint impressions. Both species produce prints with five digits, and when the prints are partial or degraded, the distinct characteristics that typically differentiate a squirrel's compact, delicate paw from a raccoon's more robust and defined pad can become blurred. This ambiguity is worsened by the limited number of high-quality real-footprint examples available for training and by the loss of footprint size information, which constrains the model's ability to fully capture and learn the differences between these classes.

Analysis of the predicted classes showed that sauropod footprints yielded reliable classification results, with occasional misclassifications as theropod footprints, likely due to their distinctive shape and the unique characteristics of the terrain in which they were found. In contrast, theropod footprints were associated with lower classification accuracy, often being misidentified as sauropod footprints due to overlapping terrain features, variability associated with the fossil record,

and the difference in the number of images available (theropod - 159 for training and 54 for validation; sauropod - 195 for training and 65 for validation).

Taking into account the top-2 predicted classes, rather than only the top-1 prediction, resulted in a further improvement in accuracy (Fig. 11). In this analysis, only otter footprints exhibited accuracy values below 75%, with most misclassifications involving confusion between otters and raccoons (Fig. 11). These findings indicate that the model often faced uncertainty in its predictions; when the top-ranked class was incorrect, the second-ranked class was frequently the correct one. This demonstrates the importance of human oversight in evaluating model outputs, as the correct label may not always be the highest-ranked prediction but can still be among the top candidates.

Beyond the classification results, this study presents several important contributions with implications that extend beyond the scope of vertebrate footprint analysis. First, we introduce Deep Tracks, a novel Unity simulation tool designed specifically for the generation of vertebrate footprints. The use of Unity, a widely adopted and accessible game engine, opens the door for developing additional interactive tools aimed at paleontology, conservation and ecological research. While in this study this approach was used to generate synthetic data, it also offers potential for expansion into object detection, segmentation, or even auto-encoding workflows in future applications. Second, we release a high-quality, balanced dataset containing approximately 40,000 simulated images, which represents a significant contribution for the community. This dataset can be used for fine-tuning, pre-training, or benchmarking in similar classification tasks. While this study focuses on model performance and its limitations, we emphasize that the availability of this large, reliable dataset, along with associated model weights and experimental code, is a valuable resource for future research in paleoichnology, biodiversity monitoring, and conservation. Deep Tracks, a publicly available tool, provides a flexible and extensible framework that could easily be adapted to other vertebrate footprint groups, animal groups, or even entirely different domains of natural history data. An important step for future work should involve extending Deep Tracks to simulate additional real-world conditions, such as partial footprints resulting from erosion or incomplete registration. Incorporating different silhouettes to simulate varying degrees of erosion would improve the realism of the synthetic dataset and improve its applicability to real data. Furthermore, although Deep Tracks has the capability to generate overlapping footprints, this functionality was not used in the present study, as the real dataset comprised only isolated footprints. Future research should benefit from this feature to simulate and analyze more complex trackway assemblages involving overlapping or sequential footprints.

This study presents a stepping stone in vertebrate footprint automatic classification. The use of synthetic images will help overcome challenges in ecological monitoring and wildlife management, where data scarcity and variability frequently hold up accurate species identification. This study also carries important implications for paleontology, as refining the classification of fossil footprints can lead to valuable insights into the behavior, ecology, diversity, and evolutionary history of extinct species.

5. Conclusion

This study is the first to develop a tool for synthetic image generation of vertebrate footprints for both paleontological and ecological research. We developed Deep Tracks, an open-source Unity-based application that simulates vertebrate's footprints in different environments and conditions. Our findings show that the combination of using both simulated and real footprint images improves classification performance. By initially training on a large, balanced simulated dataset, and subsequently fine-tuning on real data, our approach achieved improved generalization and superior performance compared to the other models analyzed in this study. The analysis revealed that the EfficientNet-b0 architecture was particularly effective in capturing subtle morphological features across both simulated and real domains, thereby reducing misclassifications even under conditions of data scarcity and variability. The fine-tuned EfficientNet-b0 architecture achieved high validation accuracy with relatively low energy consumption and memory usage, showing its strong performance–efficiency tradeoff. The results indicated that as the size of the simulated pre-training dataset increased, validation accuracy improved, and the highest performance on the real dataset was achieved with the largest synthetic set. However, challenges remain, as misclassifications were observed among classes with similar footprint characteristics, such as between coyotes and foxes, and between squirrels and raccoons. These findings highlight the importance of developing tools, such as our Deep Tracks generator, as a means to mitigate data scarcity issues. Moreover, the study highlights the importance of creating an open-source real footprint database to improve the accuracy of machine learning models, as they could also be used to complement training sets. While showing promising results, this study presents certain limitations. First, while the simulated data improved classification performance, differences exist when we look at the domain between simulated and real footprints that can affect generalization. Second, the real dataset used was limited in size, potentially constraining the model's ability to capture the full variability of natural footprints. Third, this study focused on isolated footprints, and the Deep Track's ability to model overlapping footprints or partial preservation was not explored. Finally, the simulation parameters, although diverse, may not cover the entire spectrum of substrate and trackmaker interactions observed in nature, including cases of overlapping footprints and erosion. Future work should focus on improving the realism of simulated data by improving the silhouettes that are used, and by improving Deep Tracks (e.g. adding more soil types, improving the printing process of the silhouette on the soil), exploring advanced augmentation techniques (such as generative adversarial networks or diffusion models), and integrating additional contextual information (such as including a scale in the image, including different morphometric data extracted from the footprints – e.g., interpad distance, footprint width, and length) to improve discrimination among closely related classes; and creating more focused classification models (e.g., canid footprints classifier). Future research should also explore more advanced methodologies to simulate datasets to improve real data classification – especially for fossil footprints, which present additional challenges such as poor preservation and class similarities (i.e., tridactyl tracks that could be produced by ornithopod or theropod dinosaurs). Given the flexibility of the methodology, future work could include exploring other neural network architectures (e.g., transformers).

CRediT authorship contribution statement

Carolina S. Marques: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Afonso Mota:** Writing – review & editing, Methodology, Conceptualization. **Matteo Belvedere:** Writing – review & editing, Conceptualization. **Diego Castanera:** Writing – review & editing, Conceptualization. **Ignacio Díaz-Martínez:** Writing – review & editing, Conceptualization. **Elisabete Malafaia:** Writing – review & editing, Conceptualization. **Soraia Pereira:** Writing – review & editing, Conceptualization. **Luís Miguel Rosalino:** Writing – review & editing, Conceptualization. **Vanda F. Santos:** Writing – review & editing, Conceptualization. **Lara Sciscio:** Writing – review & editing, Conceptualization. **Emmanuel Dufourq:** Writing – review & editing, Supervision, Conceptualization.

Funding

This work was supported by Portuguese government funds through FCT- Fundação para a Ciência e Tecnologia under the doctoral scholarship FCT/CEAUL (UI/BD/154258/2022), the FCT doctoral scholarship

2024.02676.BD, the individual contract CEECIND/01770/2018 (https://doi.org/10.54499/CEECIND/01770/2018/CP1534/CT0004), CEAUL's strategic projects: UID/00006/2025 and UIDB/00006/2020 (https://doi.org/10.54499/UIDB/00006/2020), and FCT, I.P./MCTES through national funds (PIDDAC): UID/50019/2025, UIDB/50019/2020 (https://doi.org/10.54499/UIDB/50019/2020) and LA/P/0068/2020 (https://doi.org/10.54499/LA/P/0068/2020). Diego Castanera was supported by a postdoctoral fellowship of the Manuel López Programme, funded by the Programa Propio de Política Científica of the Universidad de Zaragoza (Spain). Ignacio Díaz-Martínez was subsidized by Project PID2022-140205NA-I00 funded by Ministerio de Ciencia e Innovación (Spanish Government). Ignacio Díaz-Martínez is supported by a Ramón *y* Cajal fellowship (RYC-2022) granted by the Ministerio de Ciencia e Innovación (Spanish Government). There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of competing interest

ED is part of the editorial board. The authors declare no other competing interests.

Acknowledgments

This work was supported by Portuguese government funds through FCT- Fundação para a Ciência e Tecnologia under the doctoral scholarship FCT/CEAUL (UI/BD/154258/2022 - https://doi.org/10.54499/UI/BD/154258/2022), the FCT doctoral scholarship 2024.02676.BD, the individual contract CEECIND/01770/2018 (https://doi.org/10.54499/CEECIND/01770/2018/CP1534/CT0004), CEAUL's strategic project UID/00006/2023, FCT, I.P./MCTES through national funds (PIDDAC): UID/50019/2023, LA/P/0068/2020 and UID/50019/2025 (https://doi.org/10.54499/LA/P/0068/2020 and https://doi.org/10.54499/UID/PRR/50019/2025), cE3c (UID/00329/2025) and CHANGE (LA/P/0121/2020). Diego Castanera was supported by a postdoctoral fellowship of the Manuel López Programme, funded by the Programa Propiode Política Científica of the Universidad de Zaragoza (Spain). Ignacio Díaz-Martínez was subsidized by Project PID2022-140205NA-I00 funded by Ministerio de Ciencia e Innovación (Spanish Government). Ignacio Díaz-Martínez is supported by a Ramón y Cajal fellowship (RYC-2022) granted by the Ministerio de Ciencia e Innovación (Spanish Government). This study was also funded in part by the University of Northern Iowa's support for Open Access publishing. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish,or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103523.

Data availability

All simulated images, and code used for the training and evaluation of the fine-tuned model, with the respective trained model weights are available onlineon Zenodo (https://doi.org/10.5281/zenodo.15196257). The trained model with better results is also publicly available through a user-friendly application created using Hugging Face (Hugging Face, 2024) and Gradio (Abid et al., 2019). The application (Vertebrate-Tracks-Classifier), takes photographs of footprints as inputs and returns a table where each row corresponds to a photograph, and the columns contain information of the top three predicted classes with their probabilities (values closer to 1 indicate higher confidence on the prediction). The Deep Tracks application is available on Zenodo (https://zenodo.org/records/15092442).

References

Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., Zou, J., 2019. Gradio: hassle-free sharing and testing of ml models in the wild. arXiv preprint arXiv:1906.02569.

Abrahams, M., Bordy, E.M., Sciscio, L., Knoll, F., 2017. Scampering, trotting, walking tridactyl bipedal dinosaurs in southern Africa: ichnological account of a Lower Jurassic palaeosurface (upper Elliot Formation, Roma Valley) in Lesotho. Hist. Biol. 29 (7), 958–975. http://dx.doi.org/10.1080/08912963.2016.1267164.

Abrahams, M., Sciscio, L., Reid, M., Haupt, T., Bordy, E., 2020. Large tridactyl dinosaur tracks from the Early Jurassic of southern Gondwana – uppermost Elliot Formation, Upper Moyeni, Lesotho. Ann. Soc. Geol. Pol. http://dx.doi.org/10.14241/asgp.2020.07.

Belvedere, M., Bennett, M.R., Marty, D., Budka, M., Reynolds, S.C., Bakirov, R., 2018. Stat-tracks and mediotypes: powerful tools for modern ichnology based on 3D models. PeerJ 6, e4247. http://dx.doi.org/10.7717/peerj.4247.

Bennett, M.R., Budka, M., 2018. Vertebrate ichnology: Issues and case studies. In: Digital Technology for Forensic Footwear Analysis and Vertebrate Ichnology. Springer International Publishing, pp. 189–219. http://dx.doi.org/10.1007/978-3-319-93689-5_6.

Benoit Courty, Schmidt, V., Luccioni, S., Goyal-Kamal, MarionCoutarel, Feld, B., Lecourt, J., LiamConnell, Saboni, A., Inimaz, supatomic, Léval, M., Blanche, L., Cruveiller, A., ouminasara, Zhao, F., Joshi, A., Bogroff, A., de Lavoreille, H., Laskaris, N., Abati, E., Blank, D., Wang, Z., Catovic, A., Alencon, M., echły, M.S., Bauer, C., de Araújo, L.O.N., JPW, MinervaBooks, 2024. mlco2/codecarbon: v2.4.1. http://dx.doi.org/10.5281/zenodo.11171501.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.

Bonetto, E., Ahmad, A., 2023. Synthetic data-based detection of zebras in drone imagery. In: 2023 European Conference on Mobile Robots. ECMR, IEEE, pp. 1–8. http://dx.doi.org/10.1109/ecmr59166.2023.10256293.

Bonetto, E., Xu, C., Ahmad, A., 2023. GRADE: Generating realistic and dynamic environments for robotics research with Isaac Sim. http://dx.doi.org/10.48550/ARXIV.2303.04466, arXiv.

Buatois, L.A., Mángano, M.G., 2011. Ichnology: Organism-Substrate Interactions in Space and Time. Cambridge University Press, http://dx.doi.org/10.1017/S0016756811001038.

Buatois, L.A., Mángano, M.G., 2016. Recurrent patterns and processes: The significance of ichnology in evolutionary paleoecology. In: The Trace-Fossil Record of Major Evolutionary Events: Volume 2: Mesozoic and Cenozoic. Springer, pp. 449–473. http://dx.doi.org/10.1007/978-94-017-9597-5_9.

Cao, J., Tang, H., Fang, H.-S., Shen, X., Tai, Y.-W., Lu, C., 2019. Cross-domain adaptation for animal pose estimation. In: 2019 IEEE/CVF International Conference on Computer Vision. ICCV, IEEE.

Castanera, D., Aurell, M., Canudo, J.I., Cuenca-Bescós, G., Gasca, J.M., Bádenas, B., 2023. Paleoecology and paleoenvironment of the Early Cretaceous theropod-dominated ichnoassemblage of the Los Corrales del Pelejón tracksite, Teruel Province, Spain. Palaeogeogr. Palaeoclimatol. Palaeoecol. 630, 111761. http://dx.doi.org/10.1016/j.palaeo.2023.111761.

Castanera, D., Barco, J.L., Díaz-Martínez, I., Gascón, J.H., Pérez-Lorente, F., Canudo, J.I., 2011. New evidence of a herd of titanosauriform sauropods from the lower Berriasian of the Iberian range (Spain). Palaeogeogr. Palaeoclimatol. Palaeoecol. 310 (3–4), 227–237. http://dx.doi.org/10.1016/j.palaeo.2011.07.015.

Castanera, D., Belvedere, M., Marty, D., Paratte, G., Lapaire-Cattin, M., Lovis, C., Meyer, C.A., 2018. A walk in the maze: variation in Late Jurassic tridactyl dinosaur tracks from the Swiss Jura Mountains (NW Switzerland). PeerJ 6, e4579. http://dx.doi.org/10.7717/peerj.4579.

Castanera, D., Colmenar, J., Sauqué, V., Canudo, J.I., 2015. Geometric morphometric analysis applied to theropod tracks from the Lower Cretaceous (Berriasian) of Spain. Palaeontology 58 (1), 183–200. http://dx.doi.org/10.1111/pala.12132.

Castanera, D., Díaz-Martínez, I., Moreno-Azanza, M., Canudo, J., Gasca, J., 2016a. Mirambel formation in the Iberian range (NE Spain). Cretac. Period: Biot. Divers. Biogeogr.: Bull. 71 (ISSN: 1524-4156) 71, 65.

Castanera, D., Pascual, C., Canudo, J.I., Hernández, N., Barco, J.L., 2012. Ethological variations in gauge in sauropod trackways from the Berriasian of Spain. Lethaia 45 (4), 476–489. http://dx.doi.org/10.1111/j.1502-3931.2012.00304.x.

Castanera, D., Piñuela, L., García-Ramos, J.C., 2016b. Grallator theropod tracks from the Late Jurassic of Asturias (Spain): ichnotaxonomic implications. Span. J. Palaeontol. 31 (2), 283–296. http://dx.doi.org/10.7203/sjp.31.2.17156.

Castanera, D., Piñuela, L., García-Ramos, J.C., 2020. Grallator theropod tracks from the Late Jurassic of Asturias (Spain): ichnotaxonomic implications. Span. J. Palaeontol. (ISSN: 2255-0550) 31 (2), 283–296. http://dx.doi.org/10.7203/sjp.31.2.17156.

Castanera, D., Vila, B., Razzolini, N., Santos, V., Pascual, C., Canudo, J., 2014. Sauropod trackways of the Iberian Peninsula: palaeoetological and palaeoenvironmental implications. J. Iber. Geol. (ISSN: 1698-6180) 40 (1), http://dx.doi.org/10.5209/rev_jige.2014.v40.n1.44087.

De Angelo, C., Paviolo, A., Di Bitetti, M.S., 2010. Traditional versus multivariate methods for identifying Jaguar, Puma, and large canid tracks. J. Wildl. Manag. 74 (5), 1141–1151. http://dx.doi.org/10.2193/2009-293.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 248–255, Retrieved from https://ieeexplore.ieee.org/document/5206848.

Díaz-Martínez, I., Castanera, D., Gasca, J.M., Canudo, J.I., 2015. A reappraisal of the Middle Triassic chirotheriid *Chirotherium ibericus* Navás, 1906 (Iberian range NE Spain), with comments on the Triassic tetrapod track biochronology of the Iberian Peninsula. PeerJ (ISSN: 2167-8359) 3, e1044. http://dx.doi.org/10.7717/peerj.1044.

Díaz-Martínez, I., Citton, P., Castanera, D., 2023. What do their footprints tell us? Many questions and some answers about the life of non-avian dinosaurs. J. Iber. Geol. (ISSN: 1886-7995) 50 (1), 5–26. http://dx.doi.org/10.1007/s41513-023-00226-6.

Douglas de Carvalho, W., Rosalino, L.M., Dalponte, J.C., Santos, B., Adania, C.H., Esbérard, C.E.L., 2015. Can footprints of small and medium sized felids be distinguished in the field? Evidences from Brazil's Atlantic forest. Trop. Conserv. Sci. 8 (3), 760–777. http://dx.doi.org/10.1177/194008291500800313.

Evans, J.W., Evans, C.A., Packard, J.M., Calkins, G., Elbroch, M., 2009. Determining observer reliability in counts of river otter tracks. J. Wildl. Manag. 73 (3), 426–432. http://dx.doi.org/10.2193/2007-514.

Falkingham, P.L., 2025. Reconstructing dinosaur locomotion. Biology Lett. (ISSN: 1744-957X) 21 (1), http://dx.doi.org/10.1098/rsbl.2024.0441.

Frey, R.W., 1975. The realm of ichnology, its strengths and limitations. In: The Study of Trace Fossils. Springer Berlin Heidelberg, pp. 13–38. http://dx.doi.org/10.1007/978-3-642-65923-2_2.

Geng, H., Russell, J., Shin, B.-S., Nicolescu, R., Klette, R., 2012. A flexible method for localisation and classification of footprints of small species. In: Advances in Image and Video Technology: 5th Pacific Rim Symposium, PSIVT 2011, Gwangju, South Korea, November 20-23, 2011, Proceedings, Part II 5. Springer, pp. 274–286. http://dx.doi.org/10.1007/978-3-642-25346-1_25.

Hanson, J., Hanson, R., 2023. Basic Illustrated Animal Tracks. Rowman & Littlefield Publishers, Incorporated, p. 112.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4700–4708. http://dx.doi.org/10.1109/CVPR.2017.243.

Hugging Face, 2024. Hugging face hub. https://huggingface.co.

Hussein, B.M., Shareef, S.M., 2024. An empirical study on the correlation between early stopping patience and epochs in deep learning. In: Kattan, A. (Ed.), In: ITM Web of Conferences, vol. 64, EDP Sciences, (ISSN: 2271-2097) p. 01003. http://dx.doi.org/10.1051/itmconf/20246401003.

Itseez, 2015. Open source computer vision library. https://github.com/itseez/opencv.

Jewell, Z., Alibhai, S., 2013. Identifying endangered species from footprints. In: International Society for Optics and Photonics (SPIE) Newsroom, vol. 2013, pp. 1–3. http://dx.doi.org/10.1117/2.1201212.004636.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kistner, F., Slaney, L., Jewell, Z., Ben-David, A., Alibhai, S., 2022. It's otterly confusing-distinguishing between footprints of three of the four sympatric Asian otter species using morphometrics and machine learning. OTTER J. Int. Otter Surviv. Fund 2022, 108–125.

Kistner, F., Slaney, L., Morant, N., 2023. Can you tell the species by a footprint?-identifying three of the four sympatric southeast Asian otter species using computer vision and deep learning. IUCN Otter Spec. Group Bull 40 (4), 197–210.

Lallensack, J.N., Romilio, A., Falkingham, P.L., 2022. A machine learning approach for the discrimination of theropod and ornithischian dinosaur tracks. J. R. Soc. Interface 19 (196), 20220588. http://dx.doi.org/10.1098/rsif.2022.0588.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context. In: European Conference on Computer Vision. ECCV, Springer, pp. 740–755.

Lockley, M.G., 1986. The paleobiological and paleoenvironmental importance of dinosaur footprints. PALAIOS (ISSN: 0883-1351) 1 (1), 37. http://dx.doi.org/10.2307/3514457.

Lockley, M.G., 1998. The vertebrate track record. Nature (ISSN: 1476-4687) 396 (6710), 429–432. http://dx.doi.org/10.1038/24783.

Lockley, M.G., Meyer, C., 1998. Megalosauripus and the problematic concept of megalosaur footprints. GAIA: Rev. GeociÊncias (ISSN: 0871-5424) 15, 313.

Loden, J., Rodolà, G., 2024. Psutil: Cross-platform lib for process and system monitoring in python. https://github.com/giampaolo/psutil. (Accessed 25 March 2025).

Marchetti, L., Belvedere, M., Voigt, S., Klein, H., Castanera, D., Díaz-Martínez, I., Marty, D., Xing, L., Feola, S., Melchor, R.N., Farlow, J.O., 2019. Defining the morphological quality of fossil footprints. Problems and principles of preservation in tetrapod ichnology with examples from the Palaeozoic to the present. Earth-Sci. Rev. (ISSN: 0012-8252) 193, 109–145. http://dx.doi.org/10.1016/j.earscirev.2019.04.008.

Martill, D.M., Smith, R.E., Romano, M., 2024. Sauropod manus and pes prints with impressions of integument from the Ravenscar Group (Middle Jurassic) of Whitby, Yorkshire, England. Proc. Geologists' Assoc. (ISSN: 0016-7878) 135 (2), 196–207. http://dx.doi.org/10.1016/j.pgeola.2024.02.003.

Marty, D., Belvedere, M., Meyer, C.A., Mietto, P., Paratte, G., Lovis, C., Thüring, B., 2010. Comparative analysis of late Jurassic sauropod trackways from the Jura Mountains (NW Switzerland) and the central High Atlas Mountains (Morocco): implications for sauropod ichnotaxonomy. Hist. Biol. (ISSN: 1029-2381) 22 (1–3), 109–133. http://dx.doi.org/10.1080/08912960903503345.

Milàn, J., Falkingham, P.L., Marty, D., Richter, A., 2016. Experimental and comparative ichnology. In: Dinosaur Tracks: The Next Steps. Indiana University Press Bloomington and Indianapolis, pp. 14–27.

Mota, A., Segurado Marques, C., Belvedere, M., Castanera, D., Díaz-Martínez, I., Malafaia, E., Pereira, S., Rosalino, L.M., Faria dos Santos, V., Sciscio, L., Dufourq, E., 2025. Deep tracks: Procedural footprint generator. http://dx.doi.org/10.5281/zenodo.15092442.

Mu, J., Qiu, W., Hager, G.D., Yuille, A.L., 2020. Learning from synthetic animals. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR), IEEE, pp. 12383–12392. http://dx.doi.org/10.1109/cvpr42600.2020.01240.

Olsen, L.-H., 2013. Tracks and Signs of the Animals and Birds of Britain and Europe. Princeton University Press.

Paratte, G., Lapaire, M., Lovis, C., Marty, D., 2017. Traces de dinosaures jurassiques – Courtedoux - Tchâfouè. In: Catalogues Du Patrimoine Paléontologique Jurassien – A16, Office de la culture – Paléontologie A16, Porrentruy, p. 490.

Paratte, G., Lapaire, M., Lovis, C., Marty, D., 2018a. Traces de dinosaures jurassiques – Courtedoux - Bois de Sylleux. In: Catalogues Du Patrimoine Paléontologique Jurassien – A16, Office de la culture – Paléontologie A16, Porrentruy, p. 500.

Paratte, G., Lapaire, M., Lovis, C., Marty, D., 2018b. Traces de dinosaures jurassiques – Courtedoux - Sur Combe Ronde. In: Catalogues du patrimoine paléontologique jurassien – A16, Office de la culture – Paléontologie A16, Porrentruy, p. 192.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. NeurIPS, In: Advances in Neural Information Processing Systems, vol. 32, Retrieved from https://arxiv.org/abs/1912.01703.

Plum, F., Bulla, R., Beck, H.K., Imirzian, N., Labonte, D., 2023. Replicant: a pipeline for generating annotated images of animals in complex environments using Unreal Engine. Nat. Commun. 14 (1), 7195. http://dx.doi.org/10.1038/s41467-023-42898-9.

Python Software Foundation, 2023. Python language reference, version 3.12.7. Python is a registered trademark of the Python Software Foundation. https://www.python.org.

Razzolini, N.L., Belvedere, M., Marty, D., Paratte, G., Lovis, C., Cattin, M., Meyer, C.A., 2017. Megalosauripus transjuranicus ichnosp. nov. A new Late Jurassic theropod ichnotaxon from NW Switzerland and implications for tridactyl dinosaur ichnology and ichnotaxomy. PLoS ONE (ISSN: 1932-6203) 12 (7), e0180289. http://dx.doi.org/10.1371/journal.pone.0180289.

Razzolini, N.L., Oms, O., Castanera, D., Vila, B., Santos, V.F.d., Galobart, À., 2016. Ichnological evidence of Megalosaurid dinosaurs crossing middle Jurassic tidal flats. Sci. Rep. (ISSN: 2045-2322) 6 (1), http://dx.doi.org/10.1038/srep31494.

Razzolini, N.L., Vila, B., Castanera, D., Falkingham, P.L., Barco, J.L., Canudo, J.I., Manning, P.L., Galobart, À., 2014. Intra-trackway morphological variations due to substrate consistency: The el frontal dinosaur tracksite (Lower Cretaceous, Spain). PLoS ONE (ISSN: 1932-6203) 9 (4), e93708. http://dx.doi.org/10.1371/journal.pone.0093708.

Riordan, P., 1998. Unsupervised recognition of individual tigers and snow leopards from their footprints. Anim. Conserv. (ISSN: 1469-1795) 1 (4), 253–262. http://dx.doi.org/10.1111/j.1469-1795.1998.tb00036.x.

Romano, M., Whyte, M.A., 2012. Information on the foot morphology, pedal skin texture and limb dynamics of sauropods: evidence from the ichnological record of the Middle Jurassic of the Cleveland Basin, Yorkshire, UK. Zubia (ISSN: 0213-4306) (30), 45.

Russell, J.C., Hasler, N., Klette, R., Rosenhahn, B., 2009. Automatic track recognition of footprints for identifying cryptic species. Ecology (ISSN: 1939-9170) 90 (7), 2007–2013. http://dx.doi.org/10.1890/08-1069.1.

Santos, V.F., Moratalla, J.J., Royo-Torres, R., 2009. New sauropod trackways from the Middle Jurassic of Portugal. Acta Palaeontol. Pol. (ISSN: 0567-7920) 54 (3), 409–422. http://dx.doi.org/10.4202/app.2008.0049.

Faria dos Santos, V., Moratalla, J., Royo-Torres, R., Belvedere, M., Castanera, D., Cerífaco, L., Mocho, P., Merino, M., Meyer, C., Sciscio, L., 2024. A revised name and new insights into the Middle Jurassic sauropod trackways from Portugal. A correction of Santos et al. 2009. Acta Palaeontol. Pol. (ISSN: 0567-7920) 69, http://dx.doi.org/10.4202/app.01167.2024.

Sanz, B., Artigas, J.V.T., Balmorí, A., 2004. Huellas y rastros de los mamíferos de la Península Ibérica. Librería Félix de Azara, Zaragoza, Spain.

Sciscio, L., Belvedere, M., Meyer, C.A., Marty, D., 2022. Sauropod trackway morphometrics: An exploratory study using highway A16 excavation at the courtedoux-tchâfouè track site (Late Jurassic, NW Switzerland). Front. Earth Sci. (ISSN: 2296-6463) 10, http://dx.doi.org/10.3389/feart.2022.805442.

Sciscio, L., Bordy, E.M., Abrahams, M., Knoll, F., McPhee, B.W., 2017. The first megatheropod tracks from the Lower Jurassic upper Elliot Formation, Karoo Basin, Lesotho. In: Wong, W.O. (Ed.), PLOS ONE (ISSN: 1932-6203) 12 (10), e0185941. http://dx.doi.org/10.1371/journal.pone.0185941.

Sciscio, L., Bordy, E.M., Lockley, M.G., Abrahams, M., 2023. Basal sauropodomorph locomotion: ichnological lessons from the Late Triassic trackways of bipeds and quadrupeds (Elliot Formation, main Karoo Basin). PeerJ (ISSN: 2167-8359) 11, e15970. http://dx.doi.org/10.7717/peerj.15970.

Sciscio, L., Bordy, E.M., Reid, M., Abrahams, M., 2016. Sedimentology and ichnology of the Mafube dinosaur track site (Lower Jurassic, eastern Free State, South Africa): a report on footprint preservation and palaeoenvironment. PeerJ (ISSN: 2167-8359) 4, e2285. http://dx.doi.org/10.7717/peerj.2285.

Shinoda, R., Shiohara, K., 2024. OpenAnimalTracks: A dataset for animal track recognition. arXiv preprint arXiv:2406.09647.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2818–2826. http://dx.doi.org/10.1109/CVPR.2016.308.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. ICML, In: Proceedings of the 36th International Conference on Machine Learning, pp. 6105–6114. http://dx.doi.org/10.48550/arXiv.1905.11946.

Toledo, N., Arregui, M., 2023. Concurrent evidence from ichnology and anatomy: the scelidotheriine ground sloths (Xenarthra, Folivora) from the Pleistocene of Argentina. Hist. Biol. 35 (2), 284–292. http://dx.doi.org/10.1080/08912963.2022.2035379.

Torcida Fernández-Baldor, F., Díaz-Martínez, I., Huerta, P., Montero Huerta, D., Castanera, D., 2021. Enigmatic tracks of solitary sauropods roaming an extensive lacustrine megatracksite in Iberia. Sci. Rep. (ISSN: 2045-2322) 11 (1), http://dx.doi.org/10.1038/s41598-021-95675-3.

Unity Technologies, 2024. Unity. Retrieved from https://unity.com/.

Vierbergen, W., Willekens, A., Dekeyser, D., Cool, S., et al., 2023. Sim2real flower detection towards automated Calendula harvesting. Biosyst. Eng. 234, 125–139. http://dx.doi.org/10.1016/j.biosystemseng.2023.08.016.

Walker, C., 1996. Signs of the Wild. New Holland Publishers, p. 216.