



RESEARCH ARTICLE

BatNet: a deep learning-based tool for automated bat species identification from camera trap images

Gabriella Krivek¹ , Alexander Gillert², Martin Harder³, Marcus Fritze¹, Karina Frankowski¹, Luisa Timm¹, Liska Meyer-Olbersleben¹, Uwe Freiherr von Lukas^{2,4}, Gerald Kerth¹ & Jaap van Schaik¹ 

¹Zoological Institute and Museum, Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany

²Fraunhofer Institute for Computer Graphics Research IGD, Rostock, Germany

³Forschungsgruppe Höhle und Karst Franken e.V. (FHKF), Almoshofer Hauptstraße 51, Nürnberg, Germany

⁴Institute for Visual and Analytic Computing, University of Rostock, Rostock, Germany

Keywords

Automated monitoring, bat conservation, camera trap, Chiroptera, deep learning, infrared light barrier

Correspondence

Gabriella Krivek, Zoological Institute and Museum, Applied Zoology and Nature Conservation, University of Greifswald, Loitzer Strasse 26, 17489 Greifswald, Germany. Tel.: +49 (0)3834 420-4358; Fax: +49 (0)3834 420-4252; E-mail: krivek.g@gmail.com

Funding Information

This work was funded by a joint research project DIG-IT! That is supported by the European Social Fund (ESF), reference: ESF/14-BM-A55-0014/19, and the Ministry of Education, Science and Culture of Mecklenburg-Vorpommern, Germany.

Editor: Marcus Rowcliffe

Associate Editor: Rahel Sollmann

Received: 12 January 2023; Revised: 21 April 2023; Accepted: 25 April 2023

doi: 10.1002/rse2.339

Remote Sensing in Ecology and Conservation 2023; **9** (6):759–774

Introduction

Effective conservation depends on the ability to quantify biodiversity and monitor species-level population dynamics in threatened ecosystems (Primack, 1995). Bats are an integral part of nearly all terrestrial ecosystems, where they provide essential ecosystem services and act as ecological indicators of general ecosystem health (Kunz et al., 2011). Despite their essential ecological role, bat

Abstract

Automated monitoring technologies can increase the efficiency of ecological data collection and support data-driven conservation. Camera traps coupled with infrared light barriers can be used to monitor temperate-zone bat assemblages at underground hibernacula, where thousands of individuals of multiple species can aggregate in winter. However, the broad-scale adoption of such photo-monitoring techniques is limited by the time-consuming bottleneck of manual image processing. Here, we present *BatNet*, an open-source, deep learning-based tool for automated identification of 13 European bat species from camera trap images. *BatNet* includes a user-friendly graphical interface, where it can be retrained to identify new bat species or to create site-specific models to improve detection accuracy at new sites. Model accuracy was evaluated on images from both trained and untrained sites, and in an ecological context, where community- and species-level metrics (species diversity, relative abundance, and species-level activity patterns) were compared between human experts and *BatNet*. At trained sites, model performance was high across all species (F1-score: 0.98–1). At untrained sites, overall classification accuracy remained high (96.7–98.2%), when camera placement was comparable to the training images (<3 m from the entrance; <45° angle relative to the opening). For atypical camera placements (>3 m or >45° angle), retraining the detector model with 500 site-specific annotations achieved an accuracy of over 95% at all sites. In the ecological case study, all investigated metrics were nearly identical between human experts and *BatNet*. Finally, we exemplify the ability to retrain *BatNet* to identify a new bat species, achieving an F1-score of 0.99 while maintaining high classification accuracy for all original species. *BatNet* can be implemented directly to scale up the deployment of camera traps in Europe and enhance bat population monitoring. Moreover, the pretrained model can serve as a baseline for transfer learning to automatize the image-based identification of bat species worldwide.

populations across the globe face multiple threats, such as the loss and degradation of suitable roosting and foraging sites, the introduction of new infectious diseases, and global warming coupled with increasingly unpredictable climatic conditions (Frick et al., 2020). These effects are especially problematic for bats, which exhibit slow life strategies, and thus, their populations may require decades to recover from individual mortality events (Fleischer et al., 2017). Therefore, the need for accurate

estimates of population trends and a fundamental understanding of how these effects are changing bat behavior and life history has never been more pressing.

One of the primary techniques used for monitoring temperate-zone bat populations is visually counting bats at their winter hibernacula. Hibernation sites are attractive for monitoring as they are used by individuals of multiple species and by individuals from multiple summer maternity colonies (Dekeukeleire et al., 2016). However, as bats are small and many prefer to hibernate in deep crevices, there can be large discrepancies between winter hibernation counts and the actual population sizes (Battersby, 2008), and some species may be entirely missed by these visual surveys (e.g., Toffoli & Calvini, 2021).

More accurately monitoring bat activity and population dynamics at hibernacula is possible with the combination of infrared light barriers and custom-made camera traps (Krivek et al., 2023). These camera traps consist of a mirrorless digital camera and a white flash, which provides high image quality (Fig. 1A) and thus allows reliable species-level identification (Fig. 1B–D). Moreover, bats do not change their behavior in response to the fast, white flash of such camera traps (1/5500 s, 1/16 power), making these photo-monitoring systems suitable as a minimally invasive method for bat monitoring (Krivek et al., 2022). These camera traps can either be installed on the inner side of the light barrier and be triggered by each bat entering the hibernaculum (i.e., ‘entry’ camera), and/or on the outer side and be triggered by each bat leaving the hibernaculum (i.e., ‘exit’ camera). In this study, we focused on the use of entry cameras to describe ecological metrics, such as species diversity, relative abundance of species and species-level activity patterns at hibernation sites. To obtain these metrics, thus far, the species of the bat that triggered the camera trap had to be manually identified, which is a time-consuming task that requires extensive experience with the subtle morphological differences between species. Given that a site with around 600 hibernating bats may yield up to 30 000 camera trap images every year (Krivek et al., 2023), manual image analysis represents a substantial hurdle for large-scale monitoring projects. Although deep learning-based species identification from camera trap images is now commonplace for many terrestrial mammals (e.g., Norouzzadeh et al., 2018; Tabak et al., 2019) and several automated species identification tools have been developed for bats from acoustic recordings (see examples in Rydell et al., 2017), such resources do not exist for identifying bat species from camera trap images. While manual validation of some identifications (e.g., with low confidence or of rare species) should be performed prior to ecological inference, such automated solutions can nevertheless

considerably speed up the identification process. Here, we present *BatNet*, an open-source, deep learning-based tool for automated bat species identification from camera trap images. This tool was developed to identify 13 bat species or species-complexes (i.e., similar species within a genus that cannot be reliably distinguished based on the morphological characteristics visible on the camera trap images), encompassing all species commonly observed at hibernacula in Northwestern Europe. *BatNet* consists of three main stages: a detector that localizes all bats in an image, a segmentation network that removes the background around the detected bats and a classifier that uses the image crops for species identification. To train the baseline model, we used an imbalanced training dataset of 16 333 camera trap images of 13 bat species from 32 hibernation sites (range: 375–3576 images; see Table S1 for sample sizes per species). For new locations and species, both the detector and the classifier stages can be retrained from within the user-friendly, coding-free graphical interface of *BatNet*. Here, the detector of the baseline model was retrained to create site-specific models for six new hibernation sites, and the classifier was retrained to identify one additional bat species. Model performance was evaluated in four ways: (1) accuracy on test images of all 13 species from trained sites ($N = 2163$) using the baseline model; (2) accuracy on images from six new, untrained sites ($N = 49\,873$) using the baseline and the site-specific models; (3) in an ecological case study, where community- and species-level ecological metrics (species diversity, relative abundance, and species-level activity patterns) were compared between human and *BatNet* using 5-month datasets from three sites ($N = 54\,748$), encompassing the entire hibernation-entry phase; and (4) accuracy on test images of the original 13 species ($N = 2163$) supplemented by images of a newly added species ($N = 1143$) using the retrained classifier model. *BatNet* is freely available under a CC BY-NC-SA 4.0 license (<https://github.com/GabiK-bat/BatNet>).

Materials and Methods

Training data and model architecture

In total 18 496 images of bats were collected at the entrance of 32 hibernacula across Germany using custom-built camera traps (Fig. 1) that were triggered by infrared light barriers (ChiroTEC, Lohra). For each image, two human experts classified the bat to species level (*Barbastella barbastellus*, *Eptesicus serotinus*, *Myotis bechsteinii*, *M. dasycneme*, *M. daubentonii*, *M. emarginatus*, *M. nattereri* and *Nyctalus noctula*) or to species complex (the whiskered bats: *Myotis alcathoe*, *M. brandtii*, *M. mystacinus*; the mouse-eared bats: *Myotis blythii*, *M. myotis*; the long-eared bats: *Plecotus*

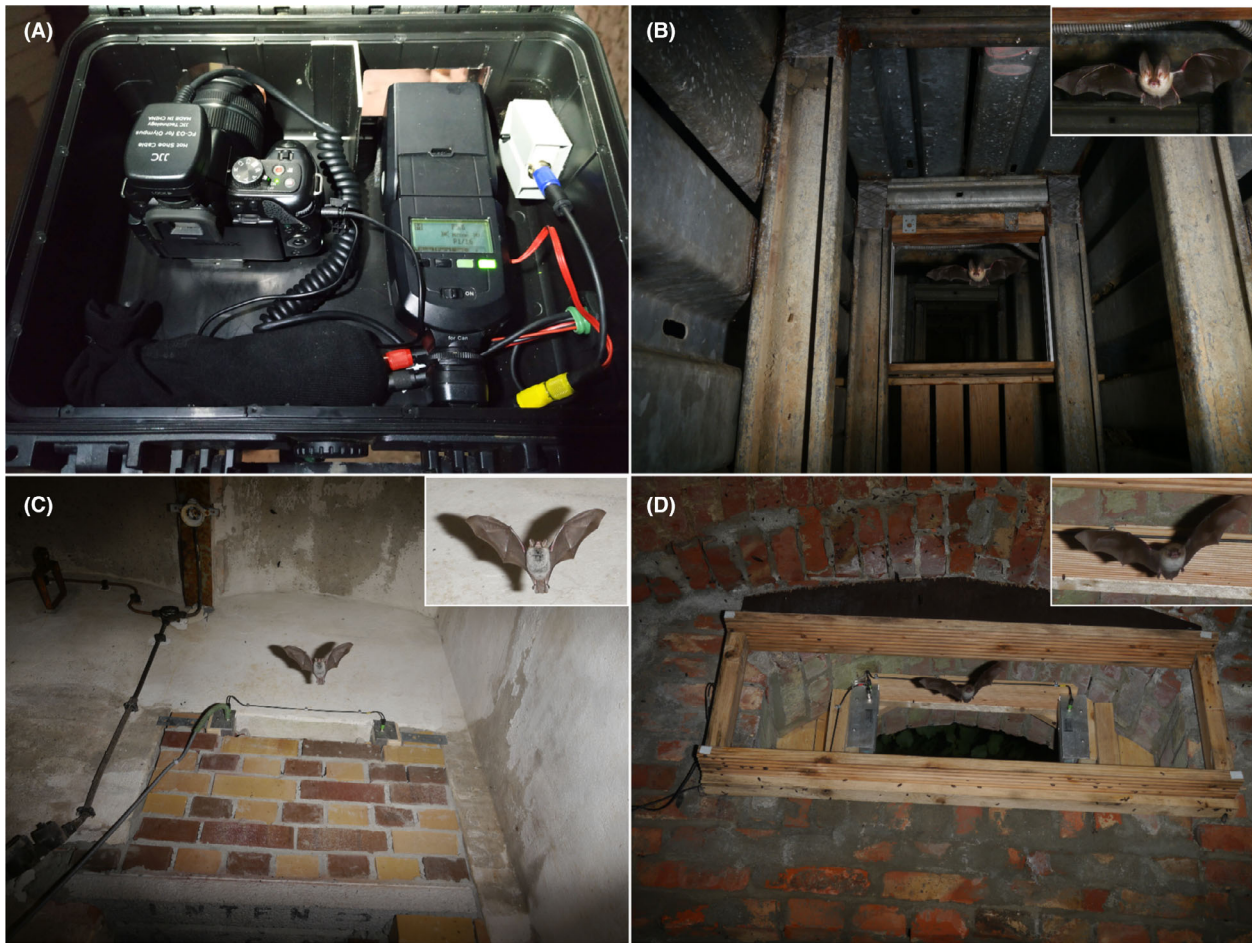


Figure 1. (A) The camera trap setup used in this study for bat monitoring at hibernation sites, composed of a mirrorless digital camera and an external white flash. (B–D) Camera trap images of bats (insets show enhanced image crops of the captured bats) entering three of the investigated hibernacula in Germany: (B) Batzbach (*Plecotus auritus*), (C) Comthurey (*Myotis nattereri*) and (D) Eldena (*Myotis daubentonii*). The entrances of these sites were monitored with infrared light barriers, which automatically triggered the camera trap when the bat entered the hibernaculum.

auritus, *P. austriacus*; the pipistrelles: *Pipistrellus pipistrellus*, *Pi. pygmaeus*; and the horseshoe bats: *Rhinolophus ferrumequinum*, *R. hipposideros*). The location and species identity of each bat in all images were annotated with bounding boxes using the *LabelMe* software (Torralba et al., 2010), and these annotations were used to train an object detector and classifier networks. Using the same annotation tool, a random subset of 3685 images were subsequently manually traced to crop the bat out from the background, which were used to train a segmentation network. From the total dataset, 90% ($N = 16\,333$) was used to train the detector and the classifier, and 10% ($N = 2163$) was used for testing final model performance (see Table S1 for sample sizes per species). All networks were trained for 30 epochs with a learning rate of 0.05 and a stochastic gradient descent optimizer.

BatNet is composed of three distinct stages: detection, segmentation, and species classification (Figure S1). First, a Faster-R-CNN object detector (Ren et al., 2015) with a ResNet50 (He et al., 2016) Feature Pyramid Network (Lin et al., 2017) places a bounding box around each bat detected in an image. Localizing classifications within an image rather than classifying the image as a whole was preferred, as this approach decreases the noise resulting from the image background and provides the ability to count and identify all animals in an image (Schneider et al., 2020). Second, the image is cropped to the bounding box and a U-Net segmentation network (Ronneberger et al., 2015) with a MobileNet V3 backbone (Howard et al., 2019) removes the background. Because deep learning models have the tendency to learn static background features (Miao et al., 2019), this segmentation step

ensures that the actual bat characteristics are used for classification in the next step and not the background features that remain within the bounding box. Finally, the segmented crop of the image is classified by an ensemble of three MobileNet V3 networks (Howard et al., 2019). This configuration was selected, because ensemble networks are less prone to make highly confident yet incorrect predictions than a single neural network (Li & Hoiem, 2020). Each of the three networks classifies the original and the flipped version of the image crop, and the six resulting predictions are then averaged. This technique, called test-time augmentation, is known to improve the performance of image classification models (e.g., Kim et al., 2020). The final output of the classifier is composed of the predicted identification for each detected bat and a confidence value between 0 and 1 for each prediction, which indicates the level of certainty in the species identification.

Since transfer learning is an established technique to improve neural network performance and reduce training time (Yosinski et al., 2014), the object detector was pre-trained on the COCO (Common Objects in Context) dataset (Lin et al., 2014), and all other networks were pre-trained on ImageNet (Russakovsky et al., 2015). In addition, the training dataset was augmented with random horizontal flips of the original camera trap images, with image crops of bats from these modified camera trap images, and with empty images (i.e., only background without any bats). Since outlier exposure (i.e., training with natural images that do not contain the target objects) is commonly used to improve detection performance at untrained background locations (Hendrycks et al., 2018), random images from the ImageNet dataset (Russakovsky et al., 2015) were also included in the training dataset of the baseline model as negative examples (i.e., images of anything else than a bat).

Using *BatNet*

Camera trap images can either be processed in a fully automated way using the command-line interface ('batch mode', no limit to the number of images within the processed folder), or in a semi-automated workflow using a browser-based graphical user interface, which supports manual validation of the output ('manual mode', optimized to process around 1000 locally stored images at a time). Both approaches result in an output that includes species labels with confidence levels for each detected bat, the coordinates of the corresponding bounding boxes, metadata from the images (e.g., file name, timestamp) and flags to denote images where the confidence level of any identifications is below a user-defined threshold, or where multiple bats were detected in an image or where

no bats were detected (i.e., empty images). Using these flags, users can quickly sort and filter images that require manual review, which is further supported by the possibility to zoom in and change the brightness of the images. Although empty images as a result of false triggers are uncommon in this photo-monitoring system due to high light barrier accuracy (Krivek et al., 2023), the camera traps can be set to trigger at regular intervals (i.e., time lapse mode) when the light barrier is blocked for an extended period (e.g., by a spiderweb or a predator sitting in the hibernaculum entrance), which can result in large numbers of empty images. Flagging these empty images drastically reduces the time required for filtering and allows users to focus on images containing the species of interest (Beery et al., 2019).

Within the *BatNet* graphical user interface, both the object detector and the species classifier can be retrained on new images in a coding-free environment. In both cases, new images can either be manually annotated with bounding boxes and species identifications in the graphical user interface, or the baseline model output can be corrected within the user interface (i.e., possibility for adding, removing, and modifying both bounding boxes and species labels) and used directly. All training parameters (i.e., species of interest, number of epochs, learning rate) are adjustable, and the resulting retrained model can be selected from a drop-down menu within the user interface. A step-by-step guide for *BatNet* image processing and retraining is provided on GitHub (<https://github.com/GabiK-bat/BatNet>).

Evaluation on test data

As an initial evaluation, we quantified *BatNet* performance on the 2163 test images that were withheld from the training dataset but were taken at trained background locations. To evaluate the performance of the object detector, we compared the intersection between the predicted and the manually created bounding boxes around each labeled bat. We considered predictions as true positive above 0.4 Intersection over Union (IoU; 0 – no overlap, 1 – perfect overlap) and false negative if the overlap was below the threshold. Predicted bounding boxes without any bats were considered false positive errors. To evaluate classifier performance, identifications were considered true positive when the human and predicted classifications were the same, false negative when the species of interest was incorrectly classified as a different species, and false positive when a different species was incorrectly classified as the species of interest.

The object detection and classifier performance were quantified by three common accuracy metrics: *precision* (i.e., ratio of correctly predicted positive observations to

the total predicted positive observations; high precision minimizes false positive errors), *recall* (i.e., ratio of correctly predicted positive observations to all observations in the actual class; high recall minimizes false negative errors) and *F1-score* (i.e., weighted average of precision and recall; used for evaluation when both false negative and false positive errors are equally undesirable).

Untrained sites and model retraining

Next, we evaluated the baseline model performance on 49 873 images from six untrained sites that were spatially and temporally independent from the training data (for example images see Figure S2). Untrained sites were categorized based on their similarity to the training dataset and included three typical sites, where the camera angle and distance from the entrance were similar to the training images (i.e., camera installed <3 m from the entrance and at a <45° angle relative to the opening), one with atypical camera distance (i.e., >3 m from the entrance), and two sites with atypical camera angle (i.e., >45° angle relative to the opening). Images from the untrained sites were classified by one human expert and annotated with bounding boxes and species labels.

Besides using the baseline object detector model, a total of 24 site-specific detector models were trained for the six sites (10 epochs, learning rate 0.001) using 25, 50, 100 or 500 site-specific annotations (i.e., bounding boxes without species labels). As for the baseline detector model, F1-scores were calculated for each of these detector models to evaluate their performance.

Ecological case study

We explored the utility of *BatNet* for describing community- and species-level ecological metrics using a continuous 5-month camera trap dataset comprising the complete hibernation-entry phase (01 August–01 January) from one trained (Eldena) and two untrained locations (Batzbach, Comthurey). In these datasets ($N = 54\,748$ images), the human expert only identified the species of the bat that triggered the camera trap without considering the bats flying in the background or annotating them with bounding boxes. This represents the typical manual identification procedure, where the primary goal is to quantify the number of bats per species that entered the hibernaculum. In terms of the automated identifications, *BatNet* predictions were based on the baseline model for images from the trained site (Eldena) and from the untrained site with typical camera angle (Batzbach). For the untrained site with atypical camera angle (Comthurey), images were identified using a site-specific model that was trained with 500 site-specific bounding boxes of bats.

In addition to the overall accuracy as described above, we focused on three ecological metrics: species diversity (i.e., the list of species detected at a site), relative abundance (i.e., the percentage of identifications attributed to each species at a site) and species-level activity patterns of bats throughout their hibernation-entry phase (i.e., the dates at which the total number of identifications per species within a site had reached the 5th, 25th, 50th, 75th and 95th percentiles). For these applications, different confidence thresholds can be applied to the output of *BatNet* to optimize the balance between high accuracy (i.e., F1-score) and the proportion of identifications that are retained in the final output (i.e., above confidence threshold). Instead of using the test data (i.e., images withheld from the baseline training), we used the data from the ecological case study to generate an optimal confidence threshold for each ecological application, because these were considered more informative for real-world applications. To define the optimal thresholds for each application, we evaluated the proportion of false positive errors (i.e., errors retained in the final output) versus the false negative errors and the identifications below the selected threshold (i.e., identifications not retained in the final output) across all confidence thresholds (Fig. 2). Based on these results, species diversity at a hibernaculum was determined using a 95% confidence threshold, which minimizes the proportion of false positive errors while still retaining each species, including the rare ones. To eliminate the small number of remaining false positives, we manually reviewed all identifications of species that constitute less than 1% of the total dataset based on the *BatNet* output. To estimate the relative abundance of each species and describe species-specific activity patterns, we selected a 70% confidence threshold and discarded all identifications below this threshold. At this threshold, the proportion of false-positive errors is strongly reduced, but the proportion of false-negative errors and identifications that are discarded as below the threshold has not started to exponentially increase yet (Fig. 2).

To describe overall accuracy in the ecological case study, we generated confusion matrices using a 70% confidence threshold for the *BatNet* output. As *BatNet* provides predictions for all bats detected in an image, including the ones in the background, some images yielded multiple bat identifications. Since true species labels were missing for the bats that were not considered to have triggered the camera trap by the human evaluator, *BatNet* predictions for these images were manually corrected so that only the bat that triggered the camera trap was retained for the accuracy assessment (if the associated confidence value exceeded the 70% confidence threshold). To correct for human error, if there was a mismatch between the human

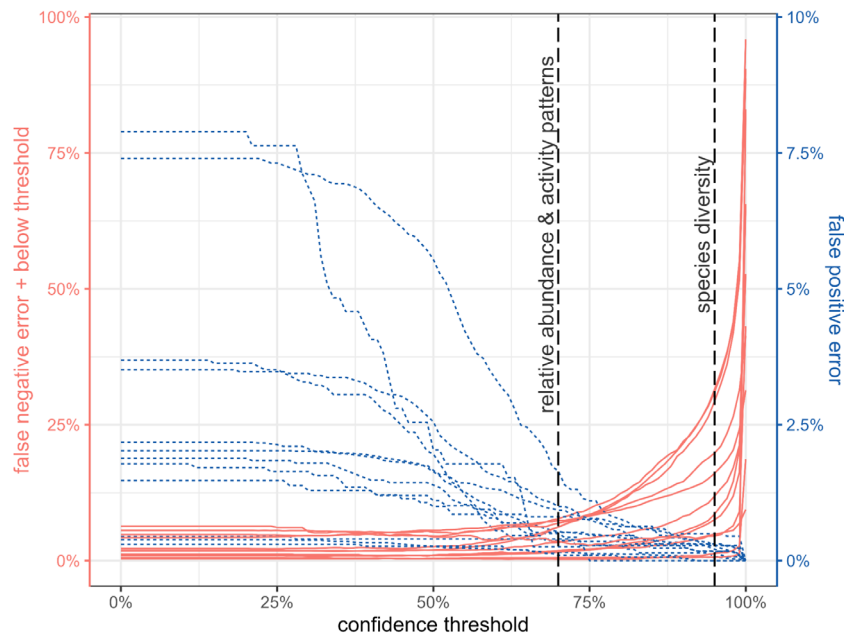


Figure 2. Proportion of false positive errors (i.e., errors retained in the final output; blue dashed lines) versus the proportion of false negative errors and the identifications below threshold (i.e., identifications not retained in the final output; red solid lines) across all confidence thresholds (0–100%), when using the baseline model of *BatNet* to process camera trap images of bats collected at three hibernation sites in Germany. For the visualization, only those species were considered that had at least 100 identifications within a site in the *BatNet* output, represented by the blue and red lines. Vertical dashed black lines indicate the confidence thresholds used for describing relative abundance and activity patterns of species (70%) and species diversity (95%), when using 5-month camera trap datasets that encompass the entire hibernation-entry phase of bats at these sites.

label and the prediction above 70% confidence threshold ($N = 243$ out of 54 748 images), two additional human experts manually reviewed the identifications. Based on this consensus scoring, the original human identification was either considered correct (i.e., *BatNet* prediction was incorrect, 76.5%), or incorrect (23.5%), and thus, the original human label was corrected.

For the investigated ecological metrics, all *BatNet* predictions above the selected confidence threshold were considered, including cases where multiple bats per image met these criteria. To investigate the ability of *BatNet* to accurately describe species diversity from a camera trap dataset, we compared the list of species identified by *BatNet* using a 95% confidence threshold with the species that were truly present at the site based on human identifications. For relative abundance, we compared the percentage of the dataset assigned to each species by human identification (i.e., the bat that triggered the image) and by *BatNet* using a 70% confidence threshold (i.e., including multiple identifications per image when they were above the threshold). Finally, we compared the activity patterns of the four most common bat species at the investigated sites (*Myotis nattereri*, *M. daubentonii*, *M. myotis* and *P. auritus*) between a human expert and *BatNet*. Specifically, we quantified the differences in the

species-level activity patterns between the two datasets by calculating the dates at which certain percentiles (5, 25, 50, 75 and 95%) of the total number of identifications had been reached per species and per site. For each percentile, differences were quantified as the number of days between the date of the percentile obtained by the human expert and *BatNet*. Additionally, we used Lin's concordance correlation coefficients (CCC) to quantify the agreement between the human expert and *BatNet* regarding the number of identifications per species per night throughout the hibernation-entry phase.

Classifier retraining: adding new species

We explored the feasibility of adding a new species to the classifier, while maintaining the classification accuracy for the original 13 species. The baseline classifier was retrained with 58 annotations of a new species (*Miniopterus schreibersii*) and 40–50 annotations per species originally included in the baseline training. The classifier was retrained for 10 epochs at a learning rate of 0.001. This comparatively small number of epochs and low learning rate were selected to lead to smaller weight updates, which is needed to minimize forgetting of the original species classes (i.e., catastrophic forgetting). The

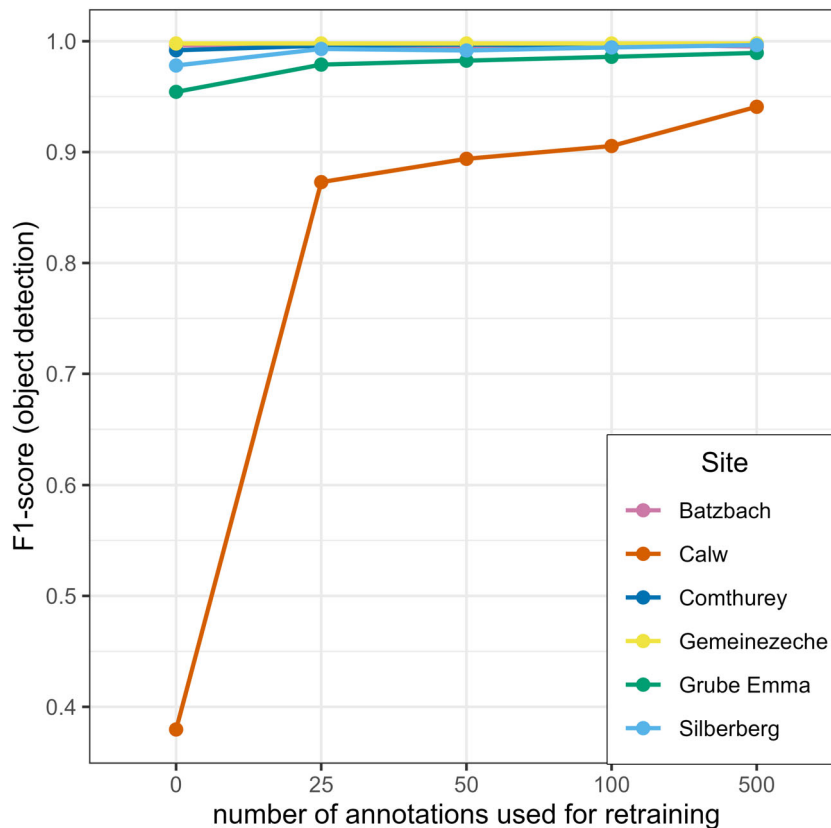


Figure 3. The object detection performance of *BatNet* on camera trap images of bats from six untrained hibernacula from Germany using the baseline model (i.e., no retraining) and using site-specific models after retraining the baseline detection model with a varying number of site-specific annotations (25, 50, 100 or 500 bounding boxes without species labels). The performance was quantified by the F1-score without using a confidence threshold. At three of the sites (Batzbach, Gemeinezeche, Silberberg), the camera angle and distance from the entrance were similar to the training images (i.e., camera installed <3 m from the entrance and at a <45° angle relative to the opening). In Calw, the camera distance was atypical (i.e., >3 m from the entrance), and in Comthurey and Grube Emma, the camera angle was atypical (i.e., >45° angle relative to the opening).

performance of the retrained model was evaluated on 1143 test images of *Mi. schreibersii*, in addition to the original 2163 test images of the other 13 bat species.

Results

Test dataset evaluation

Out of the 2163 *BatNet* identifications on test images from trained background locations, 15 were incorrect (12 misidentifications and 3 missed detections), yielding an overall classification accuracy of 99.3% (CI 98.9–99.6%). Precision, recall and F1-score ranged from 0.97 to 1.00 for all 13 bat species (for confusion matrix see Figure S3).

Untrained sites

Object detection performance of the baseline model, quantified using the F1-score, ranged from 0.95 to 1.00 at

five of six untrained locations. It was noticeably lower at one site (0.38 in Calw; Fig. 3), where the camera trap was installed further from the entrance than usual (>3 m). After retraining the baseline detector using 500 site-specific annotations (i.e., bounding boxes without species labels) for each of the six previously untrained sites, the F1-score of the site-specific object detection model increased to 0.94 in Calw and to over 0.98 at the other five previously untrained locations.

Classification accuracy of the baseline model varied depending on the camera angle and the distance between the camera and the entrance (Table 1; example camera trap images: Figure S2, confusion matrices: Figure S4). Classification accuracy was high (96.7–98.2%) at untrained locations with typical backgrounds (i.e., similar camera angle and distance to the training dataset). It was markedly lower and more variable at sites with atypical camera placement (17.8, 86.3 and 90.8%; Table 1), presumably because many bats were not detected or

Table 1. *BatNet* classification accuracy of bat species from camera trap images with 95% confidence interval at six untrained background locations from Germany, using the baseline model and the site-specific detector models retrained with 500 site-specific annotations (r500).

Site category	Site	N_{images}	Accuracy (95% CI) baseline	Accuracy (95% CI) r500
Typical	Batzbach	39 430	98.2 (98.1–98.3)	98.1 (97.9–98.2)
	Gemeinezeche	997	97.6 (96.4–98.5)	99.9 (99.4–100)
	Silberberg	1000	96.7 (95.4–97.7)	99.8 (99.3–100)
Atypical angle	Comthurey	6472	90.8 (90.1–91.5)	97.3 (96.9–97.7)
	Grube Emma	979	86.3 (84–88.4)	97.5 (96.3–98.3)
Atypical distance	Calw	995	17.8 (15.5–20.3)	95.5 (94–96.7)

N indicates the number of images used for evaluation. Hibernation sites were categorized based on their similarity to the training dataset in terms of the camera angle and distance from the entrance. In a typical monitoring setup, the camera was installed <3 m from the entrance and at a <45° angle relative to the opening. The setup was considered atypical when the camera was installed more than 3 m away from the entrance or it was positioned at a >45° angle relative to the opening.

incorrectly segmented. Notably, after retraining the detector with 500 site-specific annotations for each of the six previously untrained sites, classification accuracy improved to over 95% at all sites (Table 1; 95.5–99.9%).

Ecological case study

Species diversity

BatNet detected all species that were identified by human experts at all three sites (Batzbach, Comthurey, Eldena; Table 2). Across the three evaluated datasets

Table 2. Bat species diversity (i.e., species present at the site) based on *BatNet* predictions of species identity with 95% confidence threshold and human expert species identifications at three hibernation sites in Germany.

Site	Species	N_{BatNet}	N_{human}
Batzbach	<i>Myotis nattereri</i>	13 304 (43.2%)	19 416
	<i>Myotis bechsteinii</i>	11 077 (36%)	11 901
	<i>Plecotus</i> sp.	2202 (7.16%)	2191
	<i>Myotis daubentonii</i>	1827 (5.94%)	2666
	<i>Myotis myotis</i>	1470 (4.78%)	1653
	<i>Myotis brandtii</i>	879 (2.86%)	1363
	<i>Myotis dasycneme</i>	2 (0.01%)	0
Comthurey	<i>Myotis nattereri</i>	2836 (45.7%)	3019
	<i>Myotis myotis</i>	2239 (36.1%)	2263
	<i>Myotis daubentonii</i>	1024 (16.5%)	1071
	<i>Barbastella barbastellus</i>	73 (1.18%)	76
	<i>Plecotus</i> sp.	36 (0.58%)	37
Eldena	<i>Myotis nattereri</i>	5542 (72.4%)	6403
	<i>Myotis daubentonii</i>	1743 (22.8%)	2192
	<i>Plecotus</i> sp.	345 (4.51%)	375
	<i>Myotis myotis</i>	19 (0.25%)	71
	<i>Myotis brandtii</i>	5 (0.07%)	51

The number of identifications (N) and the proportion of all identifications within the site that it represents (%) are provided for each species identified by *BatNet*. Bold text indicates that the total proportion of predicted *BatNet* identifications for that species was below the 1% threshold, which was used to recommend manual review of these identifications.

($N = 54\,748$), manual review for the species that constituted less than 1% of the total dataset was required for 62 images (0.1% of total), resulting in the confirmation of three true positive species ($N = 60$) and the detection of one false positive species ($N = 2$).

Relative species abundance

To describe relative species abundance, we used a 70% confidence threshold that maintained high precision and recall for all species (Fig. 4) and retained over 90% of the dataset at all sites (Eldena 90.1%, Batzbach 93.7%, Comthurey 92.1% of images above the confidence threshold). The difference in the relative abundance of all species was within 1.1% at all three sites when comparing *BatNet* predictions with 70% confidence threshold to human identifications (Table 3).

Species-specific activity patterns

Species-specific activity patterns of the four investigated species (*M. daubentonii*, *M. myotis*, *M. nattereri* and *P. auritus*) across a 5-month period were nearly identical between the human and *BatNet* identifications (see Fig. 5A for one example per species, all other combinations in Figure S5A). When the activity patterns were compared based on the percentiles (5, 25, 50, 75 and 95%) obtained from the human and *BatNet* outputs, the difference between the two methods was always less than 3 days across all percentiles per species and per site. Only one exception occurred, when a 6-day discrepancy was observed between the 95th percentile obtained by human vs. *BatNet* (*Myotis daubentonii* in Comthurey; Figure S5A).

The overall sample sizes between the human and *BatNet* datasets differed due to classifications being discarded below threshold (reduces the *BatNet* sample size), and the classification of multiple bats per image where humans only scored a single bat per image (increases the *BatNet*

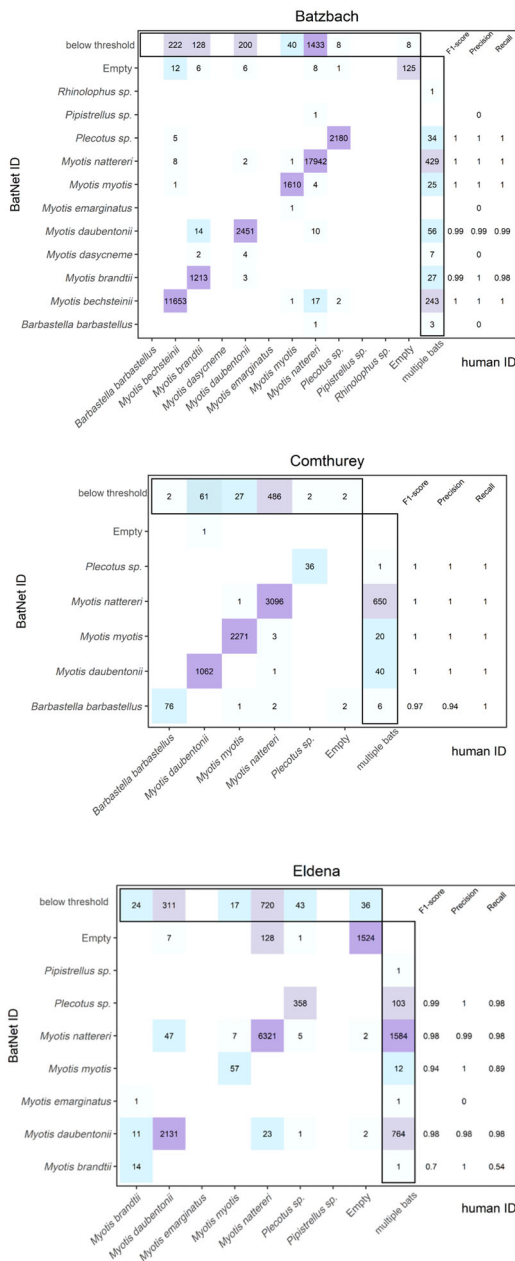


Figure 4. Confusion matrix of human species identifications and *BatNet* predictions of species identity with 70% confidence threshold for camera trap images of bats collected at three hibernation sites in Germany. The confusion matrix shows the distribution of classification error within a species, where the diagonal represents the number of accurate classifications and all other cells in the matrix describe the number of errors (i.e., missed detections or misclassifications). The color of the cells reflects the number of classifications within each category, with dark purple cells indicating high numbers and light blue cells indicating low numbers. Identifications below the confidence threshold (70%) were summarized according to their true species label ('below threshold'). For images when multiple bats were detected by *BatNet*, but humans only identified the bat that triggered the camera trap, additional *BatNet* identifications were summarized according to their predicted species label ('multiple bats'). Precision refers to the ratio of correctly predicted positive observations to the total predicted positive observations. Recall indicates the ratio of correctly predicted positive observations to all observations in the actual class. F1-score is the weighted average of precision and recall.

Table 3. Relative bat species abundance per site based on *BatNet* predictions of species identity with 70% confidence threshold and human species identifications at three hibernation sites in Germany.

Site	Species	<i>BatNet</i> %	Human %
Batzbach <i>N</i> _{images} = 39 190	<i>Myotis nattereri</i>	48.40	49.50
	<i>Myotis bechsteini</i>	31.40	30.40
	<i>Myotis daubentonii</i>	6.67	6.80
	<i>Plecotus</i> sp.	5.86	5.59
	<i>Myotis myotis</i>	4.32	4.22
	<i>Myotis brandtii</i>	3.28	3.48
	<i>Myotis dasycneme</i>	0.03	0.00
	<i>Barbastella barbastellus</i>	<0.01	0.00
	<i>Myotis emarginatus</i>	<0.01	0.00
	<i>Pipistrellus</i> sp.	<0.01	0.00
Comthurey <i>N</i> _{images} = 6466	<i>Myotis nattereri</i>	47.30	46.70
	<i>Myotis myotis</i>	34.70	35.00
	<i>Myotis daubentonii</i>	16.20	16.60
	<i>Barbastella barbastellus</i>	1.24	1.18
	<i>Plecotus</i> sp.	0.55	0.57
Eldena <i>N</i> _{images} = 9092	<i>Myotis nattereri</i>	70.90	70.40
	<i>Myotis daubentonii</i>	24.30	24.10
	<i>Plecotus</i> sp.	3.99	4.12
	<i>Myotis myotis</i>	0.64	0.78
	<i>Myotis brandtii</i>	0.16	0.56
	<i>Myotis emarginatus</i>	0.01	0.00
	<i>Pipistrellus</i> sp.	0.01	0.00

The number of images evaluated (*N*) is indicated for each site.

sample size). Despite these differences, we observed high concordance between human and *BatNet* classifications per species, per day (range: 0.989–0.999; Fig. 5B and Figure S5B).

New species

After retraining the baseline model with 58 annotations of *Mi. schreibersii*, *BatNet* achieved an F1-score of 0.99 for the new species (Fig. 6). The performance for the

original 13 species remained high (F1-score range: 0.94–0.99). Overall classification accuracy of the model was 98% (CI 97.3–98.3%). When applying a 70% confidence threshold, out of the 1413 *Mi. schreibersii* identifications 196 were below the threshold and only 1 identification was incorrect (F1-score 1.00; Fig. S6).

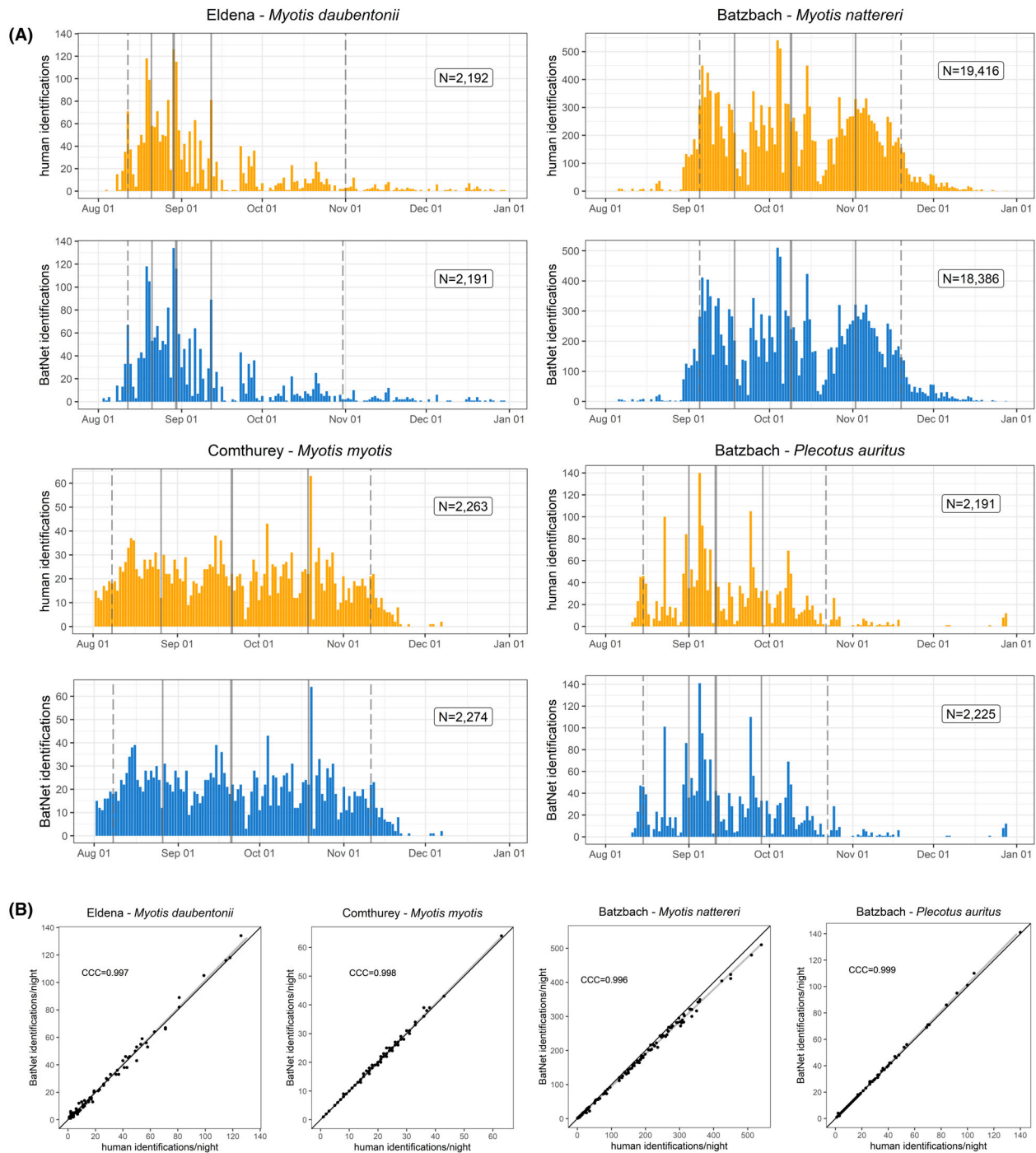


Figure 5. (A) Activity patterns of four bat species (*Myotis daubentonii*, *M. myotis*, *M. nattereri* and *Plecotus auritus*) throughout the hibernation-entry phase (01 August–01 January), based on species identifications from camera trap images by human experts (orange) and *BatNet* predictions with 70% confidence threshold (blue). Camera trap images of bats were collected at three hibernation sites in Germany (Batzbach, Comthurey, Eldena). To quantify the differences between the activity patterns obtained by human experts vs. *BatNet*, percentiles were used across the 5-month datasets (5% and 95% indicated with vertical dashed gray lines, 25% and 75% indicated with vertical solid black lines, and 50% indicated with vertical solid gray lines). The sample size (N) indicates the total number of identifications across the season. (B) Concordance plots indicate the agreement between the number of human and *BatNet* identifications per bat species per night, quantified by the Lin's CCC (range: 0–1). These coefficients indicate how far the observed data deviate from the line of perfect concordance (black solid line).

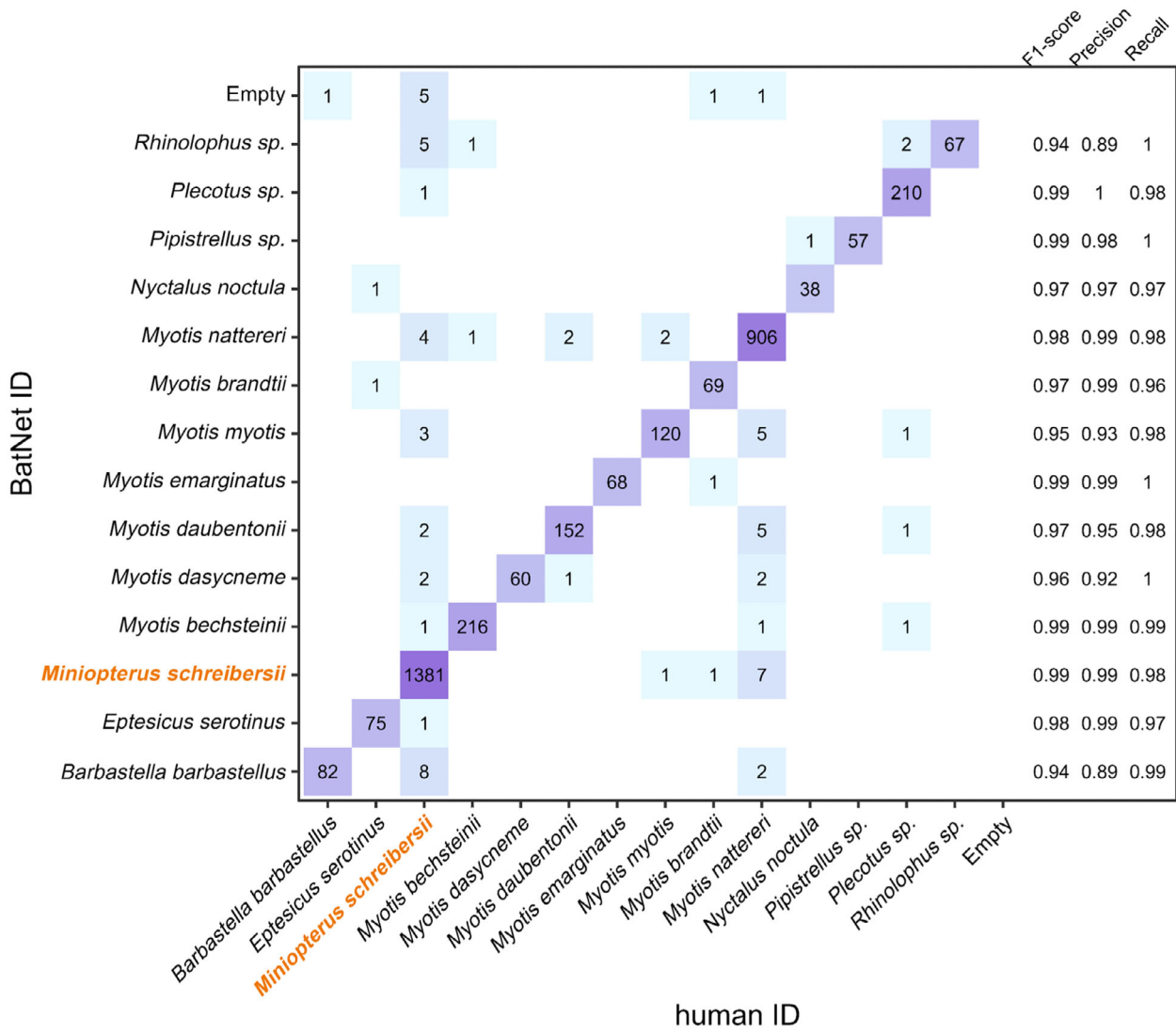


Figure 6. Confusion matrix of human species identifications and *BatNet* predictions of species identity after retraining the baseline model to be able to identify a new European bat species, *Miniopterus schreibersii*, in addition to the 13 bat species included in the original training data. The confusion matrix shows the distribution of classification error within a species, where the diagonal represents the number of accurate classifications and all other cells in the matrix describe the number of errors (i.e., missed detections or misclassifications). The color of the cells reflects the number of classifications within each category, with dark purple cells indicating high numbers and light blue cells indicating low numbers. Precision refers to the ratio of correctly predicted positive observations to the total predicted positive observations. Recall indicates the ratio of correctly predicted positive observations to all observations in the actual class. F1-score is the weighted average of precision and recall.

Discussion

BatNet is a deep learning-based tool for automated identification of 13 Northwestern European bat species, that can be retrained to adjust to new sites and to include new species within a coding-free environment. On test images from trained locations, the baseline model achieved high species-level classification accuracy across all 13 bat species (F1-score range: 0.98–1.00). This is likely a result of the localization and segmentation steps implemented before species classification, which were not

used in other image-based identification studies that found highly variable model performance for different species (e.g., Vélez et al., 2023; Whytock et al., 2021). Overall classification accuracy of the baseline model remained remarkably high at untrained sites (96.7–98.2%), where the camera angle and distance from the entrance were comparable to the training images. At untrained sites with an atypical camera setup, site-specific models reached an overall classification accuracy above 95% after retraining with 500 annotations. These results are particularly important, as classification accuracy

measures derived from trained locations are known to decrease significantly when applied on images from new locations (Schneider et al., 2020), despite models being trained with broad and diverse image datasets. The possibility to retrain the object detector and create site-specific models with minimal manual annotation effort allows *BatNet* to overcome detection difficulties related to new backgrounds and camera setups. Beyond overall accuracy, we showed that *BatNet* yields nearly identical results to manual identification when used to quantify ecologically relevant community- and species-level metrics, such as species diversity, relative abundance, and species-specific activity patterns. Finally, retraining the baseline model with an additional, morphologically similar, new bat species resulted in high classification accuracy, both for the new species (F1-score: 0.99), and for all other 13 species (F1-score: 0.94–0.99). Consequently, *BatNet* represents an accurate and highly adaptable platform for automation of camera trap-based bat monitoring.

Improving the speed and scalability of camera trap-based monitoring of bats has large implications for bat conservation, given the improvement this method constitutes over winter hibernation counts for monitoring bat population dynamics (Krivek et al., 2023). Importantly, camera traps attached to infrared light barriers can be used to accurately describe species diversity at a hibernaculum, since they are able to detect all species entering the site, including those that are often vastly undercounted or not detected at all during visual surveys (e.g., crevice-roosting species; Toffoli & Calvini, 2021). Furthermore, the continuous nature of camera trap-based monitoring allows us to describe the activity patterns of different species. Here this was exemplified using percentiles, where the 5th and 95th percentiles can serve as a reliable measure of the start and end of the species-specific activity during the hibernation-entry phase, and the combination of the 25th, 50th and 75th percentiles can indicate the peak activity of different species. These measures can be used then to compare activity patterns between species, sites and years in a standardized way. Exploring these fine-scale changes in bat activity can help describe how species differ in their hibernation phenology and in terms of their response to changing weather conditions (cf. Meier et al., 2022) and contribute to data-driven conservation actions. Finally, the installation of camera traps with infrared light barriers could be a promising new survey method to minimize direct contact with bats and thus, prevent human disturbance and possible introduction of pathogens to new sites (e.g., WNS, Covid-19; Blehert et al., 2009; Kingston et al., 2021).

Although not investigated here, dual camera trap setups (i.e., both entry and exit camera) have the potential to also quantify the absolute abundances of bat species at

hibernacula, which remains difficult for many species based on traditional monitoring methods (Van der Meij et al., 2015). By adding up the net number of entries (i.e., identifications in the entry camera) and exits (i.e., identifications in the exit camera) per species throughout the hibernation entry or emergence phases, species-level population sizes could be estimated – an approach similar to estimating population sizes of mixed species assemblages using light barrier data (Krivek et al., 2023). For such applications, *BatNet* should be implemented in a semi-automated workflow, where identifications below the confidence threshold of all species are manually reviewed in the graphical user interface. Additionally, in images with multiple bats, users must select the identification of the bat that triggered the camera trap and discard the identifications of bats in the background.

The primary limitation to the implementation of this photo-monitoring method is that the light barriers that are used to trigger the camera trap can only monitor entrance sizes of up to 35×300 cm. However, in Germany and many other European countries, the entrances of many large complex mines and caves, where gains in monitoring resolution are expected to be greatest, have already been reduced in size to limit human disturbance and access (Krivek et al., 2023). Thus, although modifications to the entrance should always be performed with caution (e.g., Pugh & Altringham, 2005), the method may be nevertheless widely applicable to monitor temperate zone bats that predominantly make use of underground sites as hibernacula. Finally, it should be noted that if absolute population estimates are not needed, the system could also be installed to only cover a portion of the total entrance, in which case ecological metrics could still be estimated under the assumption that the bats flying through the monitored area constitute a random sample of the total assemblage.

Comparison with other automated species identification approaches

The accuracy of *BatNet*, both at trained and untrained sites, is remarkably high in comparison to other deep learning solutions for automated, image-based mammal species identification (e.g., Norouzzadeh et al., 2018; Tabak et al., 2019). In large part, this may be explained by several key differences between classic wildlife camera trap setups and the camera traps triggered by infrared light barriers here used for bat monitoring. First, these custom-made camera traps are installed at the entrance of hibernation sites that are nearly exclusively used by bats. Therefore, only a relatively narrow species range had to be considered for training the networks. Second, since these camera traps are triggered by bats flying through an

infrared light barrier, their distance from the camera when the image is taken remains highly consistent. Thus, the camera can be manually focused at a fixed depth to ensure that most bats appear sharp on the images. Third, the environment is often comparatively simple and artificial, and the bats are only rarely partially occluded, which contrasts sharply with the complex, vegetation-rich backdrop of most camera trap studies. This allows for relatively simple segmentation and isolation of the target from the background. Finally, the use of white flash with standardized settings provides a fixed amount of white light in an otherwise completely dark environment. This results in a better and more standardized image quality than afforded by infrared flashes and variable lighting conditions in most traditional wildlife camera setups. The resulting high image quality allows identification of different bat species with high certainty, even though the morphological differences between bat species are far more subtle than between many other mammals. However, the image quality also depends on the camera trap angle and distance from the entrance, as demonstrated here for several sites with atypical camera placements (Calw, Comthurey, Grube Emma). Therefore, for optimal performance of *BatNet*, camera traps should be placed 2–3 m away from the entrance and ideally in a 45° angle or less to ensure the best possible image conditions for reliable species identification.

The performance of *BatNet* was further improved by implementing techniques that have not been commonly used in other automated, image-based species identification pipelines (e.g., Norouzzadeh et al., 2018; Tabak et al., 2019). First, deep learning models can learn the background features of specific camera trap stations instead of the focal animals (Miao et al., 2019), which introduces bias. To ensure that the classifier focuses on the characteristics of bats instead of the common background features, we trained a U-Net segmentation network to automatize background removal. While such approaches may be more difficult to implement for datasets with more complex backgrounds, they may nevertheless be worthwhile. Second, single neural networks are more prone to make highly confident yet incorrect predictions (Li & Hoiem, 2020). Here, we used an ensemble of three neural networks for classification, where each network classified the original and the flipped version of the image (i.e., test-time augmentation). This resulted in more informative confidence levels that could be used for discarding low-confidence identifications or filtering them out for manual review. Exploring the adoption of these techniques in other deep learning-based species identification approaches may similarly improve their performance.

For bats, deep learning-based species identification of passive acoustic recordings have become increasingly

popular, with several automated classifiers of echolocation calls being developed (Mac Aodha et al., 2018; Rydell et al., 2017; Tabak et al., 2022). Such approaches can similarly be applied to characterize bat assemblages at underground sites. However, several taxa, most notably the genus *Myotis*, remain difficult to identify automatically (Rydell et al., 2017) due to the high variability in call features within species. This is exacerbated when multiple individuals of several species are calling simultaneously (Bergmann et al., 2022). Moreover, these issues similarly affect the manual validation of acoustic recordings, whereas validation from camera trap images is readily feasible. Despite these shortcomings, acoustic surveys represent an important method for wide-spread surveillance and scouting and remain one of the only methods for monitoring hibernation sites where light barriers cannot be readily installed and that cannot be visually counted (e.g., complex sites with many large entrances, rock crevices and piles; Blomberg et al., 2021).

Application in bat monitoring and conservation

Automated monitoring of hibernacula combined with the implementation of *BatNet* has the potential to improve bat population monitoring worldwide. In Northwestern Europe, the ability to retrain *BatNet* for new locations allows it to be directly applied to vastly scale up camera trap-based bat monitoring while maintaining high accuracy. In other regions, the pretrained model of *BatNet* can be used as a baseline for transfer learning to automatize identification of a broad range of bat species, beyond our target species list. In adjacent regions this may only require minor modification of the species list, to add species such as illustrated here for *Mi. schreibersii*. In other areas, using the pretrained model as a baseline is expected to produce more accurate and stable results with less computational expense than pretraining on conventional image datasets, because of the general features the baseline model learned from a diverse, yet bat-specific camera trap dataset.

Prior to ecological inference for new datasets, model performance should always be carefully evaluated by a human using a subsample of manually identified images to detect any new or hidden biases (Norouzzadeh et al., 2021; Schneider et al., 2020). In this context, *BatNet* and its graphical user interface improve the efficiency of camera trap analysis in several ways. First, it allows sorting and filtering images based on their flags (i.e., empty, with multiple bats, below confidence threshold). Second, it is possible to manually review the final output within a user-friendly graphical interface (i.e., add, remove, or modify bounding boxes around bats and their species labels). Finally, the user interface also supports the

coding-free retraining of the baseline detector model for new sites and of the classifier model for new bat species. Overall, these aspects can help ecologists establish more efficient workflows for processing large camera trap datasets (Vélez et al., 2023). Given the numerous stressors affecting global bat populations (Frick et al., 2020) and the legal obligation to monitor bat populations worldwide, a greater flow of monitoring data is essential to support data-driven wildlife management and conservation decisions. *BatNet* drastically improves our ability to achieve these objectives.

Acknowledgements

We would like to thank the entire ChiroTEC team for providing an identified set of camera trap images and Karl Kugelschafter for his valuable insights, to Alexander Seliger and Marvin Marzenberger for helping with the training data preparation, to Jonas Denck for advice regarding the development of *BatNet*, and to Thomas Lilley and two anonymous reviewers for their helpful comments on a previous version of this manuscript. This work was funded by a joint research project DIG-IT! That is supported by the European Social Fund (ESF), reference: ESF/14-BM-A55-0014/19, and the Ministry of Education, Science and Culture of Mecklenburg-Vorpommern, Germany. G. Kr. is an associate member of the DFG Research training Group 'Biological Responses to Novel and Changing Environments'; RTG 2010. Open Access funding enabled and organized by Projekt DEAL.

Author Contributions

Conception: GKe, JvS, GKr; Training data collection: GKr, MH, MF, JvS; Training data preparation: GKr, KF, LT, LM; Human image identification: KF, MH, GKr, JvS; Software development: AG, UFvL; Evaluation: GKr, JvS; Writing: GKr, JvS. All authors commented on the manuscript and gave final approval for publication.

Data Availability Statement

BatNet is freely available under a CC BY-NC-SA 4.0 license at <https://github.com/GabiK-bat/BatNet>, along with data and scripts used for evaluation, under a CC BY-NC-ND 4.0 license.

References

Battersby, J. (2008) *Surveillance and monitoring methods for European bats*. Guidelines Produced by the Agreement on the Conservation of Populations of European Bats (EUROBATS). p. 85.

- Beery, S., Morris, D., Yang, S., Simon, M., Norouzzadeh, A. & Joshi, N. (2019) Efficient pipeline for automating species ID in new camera trap projects. *Biodiversity Information Science and Standards*, **3**, e37222.
- Bergmann, A., Burchardt, L.S., Wimmer, B., Kugelschafter, K., Gloza-Rausch, F. & Knörnschild, M. (2022) The soundscape of swarming: proof of concept for a noninvasive acoustic species identification of swarming *Myotis* bats. *Ecology and Evolution*, **12**(11), e9439.
- Blehert, D.S., Hicks, A.C., Behr, M., Meteyer, C.U., Berlowski-Zier, B.M., Buckles, E.L. et al. (2009) Bat white-nose syndrome: an emerging fungal pathogen? *Science*, **323**(5911), 227.
- Blomberg, A.S., Vasko, V., Meierhofer, M.B., Johnson, J.S., Eeva, T. & Lilley, T.M. (2021) Winter activity of boreal bats. *Mammalian Biology*, **101**, 609–618.
- Deukeleire, D., Janssen, R., Haarsma, A.-J., Bosch, T. & Van Schaik, J. (2016) Swarming behaviour, catchment area and seasonal movement patterns of the Bechstein's bats: implications for conservation. *Acta Chiropterologica*, **18**(2), 349–358.
- Fleischer, T., Gampe, J., Scheuerlein, A. & Kerth, G. (2017) Rare catastrophic events drive population dynamics in a bat species with negligible senescence. *Scientific Reports*, **7**(1), 1–9.
- Frick, W.F., Kingston, T. & Flanders, J. (2020) A review of the major threats and challenges to global bat conservation. *Annals of the New York Academy of Sciences*, **1469**(1), 5–25.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Hendrycks, D., Mazeika, M. & Dietterich, T. (2018) Deep anomaly detection with outlier exposure. *arXiv [Preprint] Arxiv:1812.04606*.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M. et al. (2019) Searching for MobileNetV3. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1314–1324.
- Kim, I., Kim, Y. & Kim, S. (2020) Learning loss for test-time augmentation. *Advances in Neural Information Processing Systems*, **33**, 4163–4174.
- Kingston, T., Frick, W., Kading, R., Leopardi, S., Medellin, R., Mendenhall, I.H. et al. (2021) IUCN SSC Bat Specialist Group (BSG) recommended strategy for researchers to reduce the risk of transmission of SARS-CoV-2 from humans to bats. Version 2.0, AMP: Assess, Modify, Protect.
- Krivek, G., Mahecha, E.P.N., Meier, F., Kerth, G. & van Schaik, J. (2023) Counting in the dark: estimating population size and trends of bat assemblages at hibernacula using infrared light barriers. *Animal Conservation*. Available from: <https://doi.org/10.1111/acv.12856>
- Krivek, G., Schulze, B., Poloskei, P.Z., Frankowski, K., Mathgen, X., Douwes, A. et al. (2022) Camera traps with white flash are a minimally invasive method for long-term

- bat monitoring. *Remote Sensing in Ecology and Conservation*, **8**(3), 284–296.
- Kunz, T.H., Braun de Torrez, E., Bauer, D., Lobova, T. & Fleming, T.H. (2011) Ecosystem services provided by bats. *Annals of the New York Academy of Sciences*, **1223**(1), 1–38.
- Li, Z. & Hoiem, D. (2020) Improving confidence estimates for unfamiliar examples. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2686–2695.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. & Belongie, S. (2017) Feature pyramid networks for object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. et al. (2014) Microsoft COCO: common objects in context. *European Conference on Computer Vision*, **8693**, 740–755.
- Mac Aodha, O., Gibb, R., Barlow, K.E., Browning, E., Firman, M., Freeman, R. et al. (2018) Bat detective—deep learning tools for bat acoustic signal detection. *PLoS Computational Biology*, **14**(3), e1005995.
- Meier, F., Grosche, L., Reusch, C., Runkel, V., van Schaik, J. & Kerth, G. (2022) Long-term individualized monitoring of sympatric bat species reveals distinct species- and demographic differences in hibernation phenology. *BMC Ecology and Evolution*, **22**(1), 1–12.
- Miao, Z., Gaynor, K.M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M.S. et al. (2019) Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, **9**(1), 1–9.
- Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jovic, N. & Clune, J. (2021) A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, **12**(1), 150–161.
- Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(25), E5716–E5725.
- Primack, R.B. (1995) *Essentials of conservation biology*, Vol. 23. Sunderland: Sinauer Associates.
- Pugh, M. & Altringham, J.D. (2005) The effect of gates on cave entry by swarming bats. *Acta Chiropterologica*, **7**(2), 293–299.
- Ren, S., He, K., Girshick, R. & Sun, J. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, **28**, 1–9.
- Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, **9351**, 234–241.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. et al. (2015) ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, **115**(3), 211–252.
- Rydell, J., Nyman, S., Eklöf, J., Jones, G. & Russo, D. (2017) Testing the performances of automated identification of bat echolocation calls: a request for prudence. *Ecological Indicators*, **78**, 416–420.
- Schneider, S., Greenberg, S., Taylor, G.W. & Kremer, S.C. (2020) Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, **10**(7), 3503–3517.
- Tabak, M.A., Murray, K.L., Reed, A.M., Lombardi, J.A. & Bay, K.J. (2022) Automated classification of bat echolocation call recordings with artificial intelligence. *Ecological Informatics*, **68**, 101526.
- Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., VerCauteren, K.C., Snow, N.P. et al. (2019) Machine learning to classify animal species in camera trap images: applications in ecology. *Methods in Ecology and Evolution*, **10**(4), 585–590.
- Toffoli, R. & Calvini, M. (2021) Long term trends of hibernating bats in North-Western Italy. *Biologia*, **76**(2), 633–643.
- Torralla, A., Russell, B.C. & Yuen, J. (2010) LabelMe: online image annotation and applications. *Proceedings of the IEEE*, **98**(8), 1467–1484.
- Van der Meij, T., Van Strien, A., Haysom, K., Dekker, J., Russ, J., Biala, K. et al. (2015) Return of the bats? A prototype indicator of trends in European bat populations in underground hibernacula. *Mammalian Biology*, **80**(3), 170–177.
- Vélez, J., McShea, W., Shamon, H., Castiblanco-Camacho, P.J., Tabak, M.A., Chalmers, C. et al. (2023) An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods in Ecology and Evolution*, **14**(2), 459–477.
- Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Ślupski, T., Koumba Pambo, A.F., Rogala, M. et al. (2021) Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods in Ecology and Evolution*, **12**(6), 1080–1092.
- Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014) How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, **27**, 1–9.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Number of camera trap images per bat species used for training *BatNet* and testing the baseline model performance.

Figure S1. Schematic overview of *BatNet*, a deep learning-based tool that automatically identifies bat species from camera trap images in three steps: bat detection (object detector), background removal (segmentation network) and species classification (ensemble of classifiers). The final

output includes a species prediction with a confidence level. Optionally, low-confidence predictions can be manually reviewed in the graphical user interface by human experts.

Figure S2. Example camera trap images from six untrained locations that were categorized based on their similarity to the training dataset, including three typical hibernation sites (camera installed <3 m from the entrance and at a <45° angle relative to the opening; A – Batzbach, B – Gemeinezeche, C – Silberberg), two sites with atypical camera angle (>45° angle relative to the opening; D – Comthurey, E – Grube Emma) and one with atypical camera distance (>3 m from the entrance; F – Calw).

Figure S3. Confusion matrix of human identifications and *BatNet* predictions (without confidence threshold) for test images from trained background locations. The confusion matrix shows the distribution of classification error within a species, where the diagonal represents the number of accurate classifications and all other cells in the matrix describe the number of errors (i.e., missed detections or misclassifications). The color of the cells reflects the number of classifications within each category, with dark purple cells indicating high numbers and light blue cells indicating low numbers. Precision refers to the ratio of correctly predicted positive observations to the total predicted positive observations. Recall indicates the ratio of correctly predicted positive observations to all observations in the actual class. F1-score is the weighted average of precision and recall.

Figure S4. Confusion matrix of human identifications and *BatNet* predictions (without confidence threshold) for camera trap images from six untrained background locations using the baseline model and the site-specific models retrained with 500 local annotations (r500). The confusion matrix shows the distribution of classification error within a species, where the diagonal represents the number of accurate classifications and all other cells in the matrix describe the number of errors (i.e., missed detections or misclassifications). The color of the cells reflects the number of classifications within each category, with dark purple cells indicating high numbers and light blue cells indicating low numbers. Precision refers to the ratio of correctly predicted positive observations to the total predicted positive observations. Recall indicates the

ratio of correctly predicted positive observations to all observations in the actual class. F1-score is the weighted average of precision and recall.

Figure S5. (A) Activity patterns of *Myotis daubentonii*, *Myotis myotis*, *Myotis nattereri* and *Plecotus auritus* throughout the hibernation-entry phase (01 August–01 January), based on species identifications from camera trap images by human experts (orange) and *BatNet* predictions with 70% confidence threshold (blue). Camera trap images of bats were collected at three hibernation sites in Germany (Batzbach, Comthurey, Eldena). To quantify the differences between the activity patterns obtained by humans versus *BatNet*, percentiles were used across the 5-month datasets (5 and 95% indicated with vertical dashed gray lines, 25 and 75% indicated with vertical solid gray lines, and 50% indicated with vertical solid black lines). The sample size (N) indicates the total number of identifications across the season. (B) Concordance plots indicate the agreement between the number of human and *BatNet* identifications per bat species per night, quantified by the Lin's concordance correlation coefficient (CCC, range: 0–1). These coefficients indicate how far the observed data deviate from the line of perfect concordance (black solid line).

Figure S6. Confusion matrix of human identifications and *BatNet* predictions with 70% confidence threshold after retraining the baseline model to be able to identify a new European bat species, *Miniopterus schreibersii*, in addition to the 13 bat species included in the original training data. The confusion matrix shows the distribution of classification error within a species, where the diagonal represents the number of accurate classifications and all other cells in the matrix describe the number of errors (i.e., missed detections or misclassifications). The color of the cells reflects the number of classifications within each category, with dark purple cells indicating high numbers and light blue cells indicating low numbers. Precision refers to the ratio of correctly predicted positive observations to the total predicted positive observations. Recall indicates the ratio of correctly predicted positive observations to all observations in the actual class. F1-score is the weighted average of precision and recall.