Adaptive Deep Learning for Automated Detection, Density Estimation and Species Identification of Bats

Abstract

Bats play vital ecological roles as pollinators, seed dispersers, and insect regulators, yet their nocturnal habits and high sensitivity to disturbance make them difficult to monitor using traditional field methods. Mist-netting, acoustic surveys, and visual emergence counts each have significant limitations in terms of accuracy, cost, and scalability. In this study, we present a novel deep learning-based pipeline for the automatic detection, counting, and classification of bats in static images, with applications in biodiversity monitoring and ecological impact assessments.

The presented approach integrates three complementary models: YOLOv11s-seg for instance-level detection and segmentation in low-density scenes, CSRNet for density map estimation in crowded conditions, and Swin Transformer V2 for fine-grained species classification. We introduce both a rule-based thresholding mechanism and a learned meta-decision module (Random Forest) to dynamically switch between YOLO and CSRNet based on image characteristics.

Trained on a curated dataset of over 2,000 annotated images sourced from iNaturalist, the system demonstrates high detection precision (93.2%) and recall (84.3%) in sparse images, while CSRNet maintains competitive performance (MAE = 8.15) under high-density conditions. The classification model achieves an overall F1-score of 0.78, with >0.96 scores for

the most represented species. Background removal and Grad-CAM analysis confirm that model predictions rely on biologically relevant features.

This modular, interpretable pipeline offers a robust, scalable alternative to manual monitoring methods and aligns with legal requirements under the EU Habitats Directive and French environmental regulations. It paves the way for efficient, AI-assisted monitoring of bat populations in support of global conservation efforts.

I) Introduction

Global biodiversity loss is intensifying, undermining ecosystem functions. The world is facing an unprecedented biodiversity crisis: a comprehensive review found that 67 % of monitored animal populations have declined by an average of 45 % since 1970, reflecting pervasive "defaunation" and marking the onset of the planet's sixth mass extinction (Dirzo et al., 2014). Human activities have accelerated species losses to rates up to 100 times above natural background levels, indicating that a mass extinction is already underway (Ceballos et al., 2015). At the same time, Earth's biosphere is approaching a planetary-scale tipping point where rapid, irreversible shifts could occur under continued pressure (Barnosky et al., 2012). The 2019 IPBES assessment further warns that up to one million species are threatened with extinction in the coming decades, putting essential ecosystem services, such as food provision, water purification, and climate regulation, at grave risk (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 2019).

Given these challenges, systematic biodiversity inventories are crucial for cataloging all species within an ecosystem and providing the baseline data needed to develop effective conservation

strategies and preserve ecosystem integrity (Bungartz et al., 2012). Among the various taxa that benefit from such inventories, bats are especially illustrative. Indeed, as nocturnal flying mammals, bats offer vital ecosystem services by pollinating a wide range of plants, dispersing seeds, and controlling insect populations, including numerous agricultural pests. They also act as sensitive bioindicators of habitat quality. However, their small size, cryptic nocturnal behavior, and high disturbance sensitivity make them particularly challenging to survey (Kunz et al., 2011).

Therefore, robust inventory and monitoring programs are essential to protect bat populations and ensure ecosystem stability. At the international level, legal instruments such as the Bern Convention (1979) and the EUROBATS Agreement (1991/1994) commit signatory states to the protection and monitoring of bat populations across Europe and neighboring regions (Council of Europe, 1979; United Nations Environment Programme (UNEP), Convention on Migratory Species Secretariat, 1991). These frameworks illustrate a broader global consensus on the need for standardized biodiversity inventories to support conservation. Even at the national scale, many countries have implemented strict regulations for bat protection and survey requirements. For example, in France, the EU Habitats Directive (92/43/EEC) (Council of the European Communities, 1992) has been transposed into the Environmental Code (République française, 2021, 2016), together with the “Avoid, Reduce, Compensate” (ERC – Éviter, Réduire, Compenser) mitigation hierarchy, making fauna–flora inventories mandatory for any project likely to affect protected species (Ministère de la Transition écologique et solidaire, 2019). Similar national regulations exist in other countries, such as Germany, where the Federal Species Protection Regulations (Bundesartenschutzverordnung) ensure strict protection of all bat species (Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2005). This multi-level legal framework underscores the need for robust and scalable monitoring methods.

Despite these imperatives, traditional bat monitoring methods reveal significant limitations. Mist-netting, which involves deploying nets to capture individuals, suffers from highly variable capture rates depending on net type and placement, leading to systematic underestimation of population size and strong sampling bias, especially in cluttered or elevated habitats (Ferreira et al., 2024; Larsen, 2007). Emergence counts are constrained by a limited temporal window, the dispersal of individuals in flight, and overlapping entries and exits, which require complex manual corrections to estimate total colony size accurately (Jung and Kukka, 2013). Acoustic detectors can identify species by their ultrasonic calls but only provide activity indices rather than quantitative abundance estimates; they have a limited detection range and high equipment costs that hinder large-scale deployment (Whiting et al., 2022). Finally, direct counts within roosts require at least three to four visits to achieve statistically reliable results, increasing both field effort and budget requirements significantly (Froidevaux et al., 2020). These operational and methodological constraints underscore the need for more robust, accurate, and non-invasive automated solutions for bat population monitoring.

The revolution in artificial intelligence opens new perspectives for automated wildlife monitoring. Deep learning models now classify large-scale camera-trap datasets automatically with high accuracy, and protocols have been proposed to integrate these predictions robustly into ecological analyses (Tabak et al., 2019). Moreover, specialized algorithms have emerged for in-field detection and recognition of wildlife: WildARe-YOLO, a lightweight model dedicated to wild animal recognition; YOLO-SAG, an enhanced YOLOv8 variant optimized to balance high precision (mAP ≈ 96–97 %) with rapid inference in natural contexts (Bakana et al., 2024; Chen et al., 2024). Additionally, convolutional neural networks have proven effective at automatically identifying a wide range of animal species in images, whether spanning multiple taxonomic groups or targeting specific species (e.g., endemic bird recognition, African mammals, forest wildlife) with dedicated models achieving state-of-the-art performance

(Bothmann et al., 2023; Huang and Basanta, 2021; Norouzzadeh et al., 2021, 2018; Whytock et al., 2021). In France, the DeepFaune initiative trained a model to classify 26 faunal taxa using camera-trap images, achieving approximately 97 % accuracy on validation datasets (Noa Rigoudy et al., 2022).

In comparison, few AI-based methods have been developed specifically for bats. A recent study introduced an EfficientNet-based model for rapid identification of seven *Rhinolophus* species, achieving approximately 94 % accuracy by extracting fine-grained facial features (Cao et al., 2023). BatNet, a convolutional neural network trained on over 5,000 camera-trap images, automatically classifies 13 European bat species with near-expert performance (F1 $\approx$ 1 on familiar sites) (Gabriella Krivek et al., 2023). The “Counting in the Dark” projects use infrared light barriers combined with cameras to estimate hibernacula populations, demonstrating that traditional visual winter counts can miss over 90 % of individuals in large roosts, while the automated approach yields highly precise estimates even in complex or hard-to-access sites (G. Krivek et al., 2023).

However, these solutions exhibit key limitations when applied to roosting bat monitoring. None were trained on images captured inside confined boxes, where bats often cluster in dense groups and fold their wings in ways that obscure essential morphological markers. Such conditions impede both accurate counting and reliable species identification. Moreover, the training datasets remain restricted in variety, causing these models to struggle when generalizing beyond their original acquisition sites and lighting conditions. Consequently, it may be necessary to fine-tune the model to adapt it to new roosts or novel nest-box configurations.

The presented approach implements an adaptive pipeline combining two complementary counting modules and a fine-grained classifier to meet the challenges of nest-box monitoring. When images contain densely clustered bats, a CSRNet-based network produces high-resolution density maps to estimate total colony size, even under conditions of overlapping

individuals and folded wings (Li et al., 2018). On the contrary, in lower-density scenes, YOLO11-Seg performs instance-level detection and precise counting, isolating each bat for further analysis (Khanam and Hussain, 2024; Redmon et al., 2016). These detected regions of interest then feed into a Swin Transformer V2 classifier, which assigns a species or genus label to each individual by leveraging detailed visual features (Liu et al., 2022). Trained on a highly diverse image set covering multiple nest-box designs and lighting conditions, this modular system demonstrates strong potential for deployment across varied sites, ensuring robust adaptability to local monitoring contexts.

This work stands out from existing studies by proposing an integrated pipeline that combines complementary models for bat detection, counting, and species identification. Rather than relying on a single architecture, this approach integrates multiple models and applies them selectively to match the ecological context. A key innovation is the use of a Random Forest meta-classifier, which adaptively selects the most suitable model and consistently outperforms fixed threshold strategies. The study also distinguishes itself through the robustness of its validation: the models were trained and tested across heterogeneous environments, variable lighting conditions, and a wide range of colony densities, ensuring ecological relevance. Although the dataset remains limited in size, it still offers enough diversity to demonstrate the potential and applicability of the proposed pipeline.

II) Materials and methods

2.1) Dataset

All images were collected from the iNaturalist website via its API (www.inaturalist.org). Only those belonging to species located in France or included in the predefined list of species of interest (Pipistrellus, Myotis daubentonii, Myotis mystacinus, Myotis bechsteinii, Eptesicus,

Myotis brandtii, Nyctalus) were kept. In total, 2860 images were extracted, showing a wide variety of environments (caves, walls, trees, human hands, etc.), bat positions (in the roost, on the ground, in flight, etc.), density (solitary individuals or dense clusters), and scales (close-ups and distant views) (Figure 1). Each image was then evaluated for quality (correct framing, visible individual, absence of blur, etc.), and 2109 images were kept. All segmentation masks were generated in Roboflow using the Enhanced Smart Polygon tool, which is based on the Segment Anything Model (SAM) to generate initial contours. Each mask was then manually refined to ensure accurate annotations. These annotations were subsequently used to train the detection and classification models.

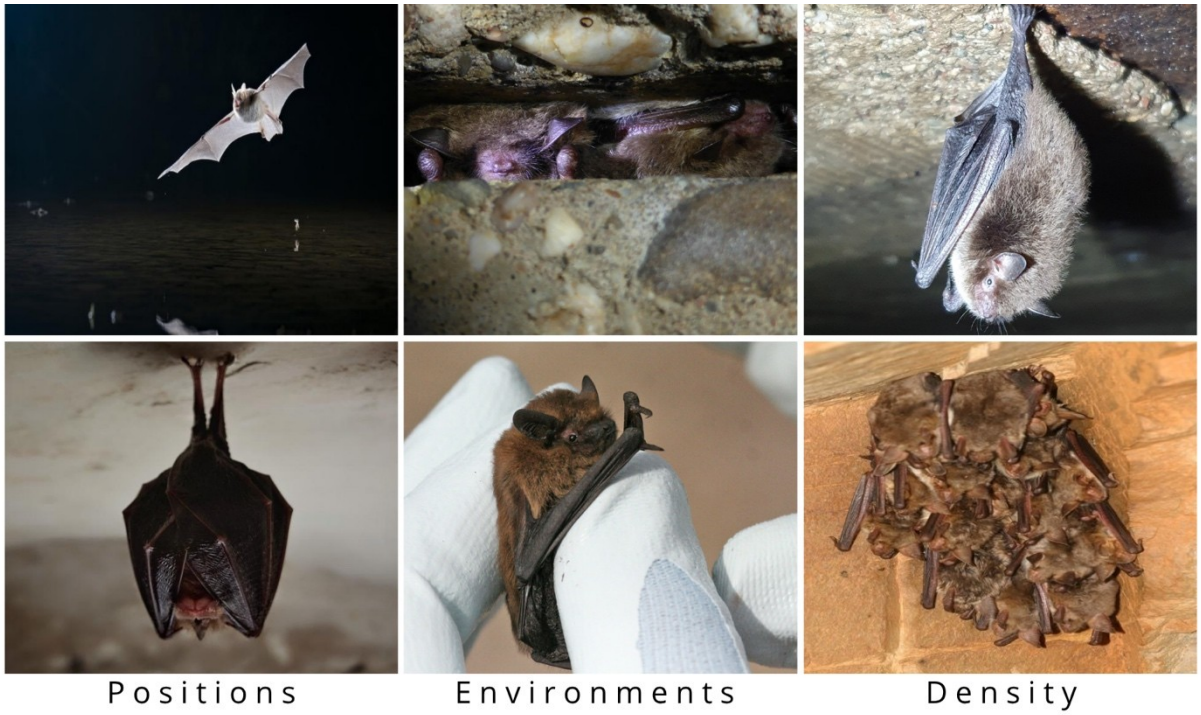


Figure 1. Illustration of dataset diversity: bat positions, environments, and density

Of the 2109 images, 179 were retained as test datasets; supplementary details are provided (Fig. S1, table S1). Then, to train the three models, three distinct datasets were created:

Detection (YOLO): This dataset includes all annotated images extracted from iNaturalist. In total, it contains 1930 images of bats and 121 of background (negative examples).

Density map (CSRNet): This set contains only images showing more than five individuals. To increase the training volume, “sub-images” were created by cropping images with high concentrations of bats (more than 15 individuals). Negative examples were also added, containing only background (caves, trees, or areas without bats). In total, 193 bat images and 19 background images were used for this dataset.

Classification (Swin): This set includes all images containing one of the seven genera or species of interest and 10% of unidentified species. It comprises 1312 images.

2.2. Models

2.2.1. Segmentation model

The detection model is based on the YOLOv11s-seg architecture (Khanam and Hussain, 2024), implemented via transfer learning from weights pre-trained on COCO. To improve robustness and generalization, several data-augmentation strategies were applied: standard geometric transformations (vertical and horizontal flips, translations, saturation, and luminosity variations) combined with composite augmentations (mosaic and copy-paste), and negative examples were added.

Training was carried out for 100 epochs with a batch size of 8 and an initial learning rate of 0.01, optimized using the Adam algorithm. A five-fold cross-validation scheme (80 % training / 20 % validation) ensures reproducibility and allows evaluation of model stability (Table 1).

To monitor performance over time, mAP@[0.5: 0.9], precision, and recall were measured for both bounding boxes and segmentation masks, while tracking loss curves on the training and validation sets.

2.2.2. Density map model

The density map model is based on the CSRNet architecture (Li et al., 2018). Its encoder uses transfer learning from a VGG16 network pre-trained on ImageNet to extract low-level features. Before training, each image underwent data augmentation, including rotations, flips, and zooms, to enhance robustness to varying acquisition conditions.

Training was performed over 100 epochs with a batch size of 8 and an initial learning rate of 1×10^{-5}, optimized using the Adam algorithm. The loss function combined mean absolute error (MAE) for counting accuracy and mean squared error (MSE) for pixel-wise fidelity, ensuring both global and spatial precision of the generated density maps (Table 1). During map generation, the Gaussian kernel's σ parameter was dynamically adjusted for each image scale.

To evaluate stability and reproducibility, a five-fold cross-validation (70 % training, 20 % validation, 10 % test) was used, measuring performance via MAE, root mean squared error (RMSE), and mean absolute percentage error (MAPE). An early-stopping and checkpointing mechanism automatically saved the best weights on the validation set according to the MAE metric.

2.2.3. Classification Model – Swin V2

The classification model is based on the SwinV2_base_window16_256 architecture, implemented through transfer learning from weights pre-trained on ImageNet. Before training, each image had its background removed to isolate the bat silhouette using the segmentation masks generated by YOLO. This step ensures that the model learns features only related to the bat itself and not the background, as confirmed by Grad-CAM visualizations. After background removal, images went through a data-augmentation pipeline (rotations, flips, zooms, and variations in brightness and contrast) to improve robustness against different acquisition

conditions. Training was conducted for 100 epochs with a batch size of 8 and an initial learning rate of 3×10^{-4}, which was dynamically managed by a OneCycleLR scheduler coupled with the AdamW optimizer. To address class imbalance between rare and abundant species, Focal Loss was used as a loss function, allowing us to weight classes according to their frequency (Table 1). A stratified five-fold cross-validation scheme (80 % training / 20 % validation) was employed to ensure reproducibility and to assess the model's stability.

Performance was evaluated using multiple metrics: overall accuracy, weighted F1-score, precision, recall, and a confusion matrix to analyze inter-species misclassifications. Additionally, Grad-CAM visualizations were generated to confirm that the network's decisions were based on relevant regions (i.e., bat body parts). This configuration produces a model that is both accurate and interpretable for fine-grained recognition of the seven target species.

Table 1: Summary of hyperparameters of the three models

Model	Batch	Epoch	LR	Loss	Optimizer	Method	Scheduler
Yolo v11	8	100	1e-2	Yolo loss function	Adam	5-fold	OneCycleLR
CSRNet	8	100	1e-5	MAE on number + MSE on mask	Adam	5-fold	ReduceLROnPlateau
Swin V2	8	100	3e-4	Focal Loss	AdamW	5-fold	OneCycleLR

2.2) Pipeline

To determine the most appropriate strategy for estimating the number of bats and identifying their species, two alternative decision mechanisms were implemented and evaluated within the pipeline.

In the first version, a simple rule-based approach was applied: the pipeline starts by running YOLO on the input image to detect individual bats. If the number of detections is below a fixed threshold (e.g., eleven), the pipeline performs counting directly from YOLO detections, followed by species identification on each individual through the SwinV2 classifier. When the number of detections exceeds the threshold, indicating a high colony density, the CSRNet model is used to estimate bat abundance through a density map. In this case, only the patch corresponding to the most confident YOLO detection is selected, its background is removed, and the patch is passed to SwinV2 for one-shot species classification (Figure 2).

In the second version, the decision process was modeled using a Random Forest classifier. The pipeline initially computes the detection count and mean confidence score obtained from YOLO, together with the density estimation derived from CSRNet. These quantities are concatenated into a feature vector, which is subsequently provided as input to the classifier.

The classifier determines which method (YOLO or CSRNet) is likely to yield the most accurate count for the given image. Based on the selected method, the subsequent species classification step is adjusted accordingly, as described above. This strategy allows the pipeline to adapt to varying image characteristics and colony densities dynamically (Figure 2).

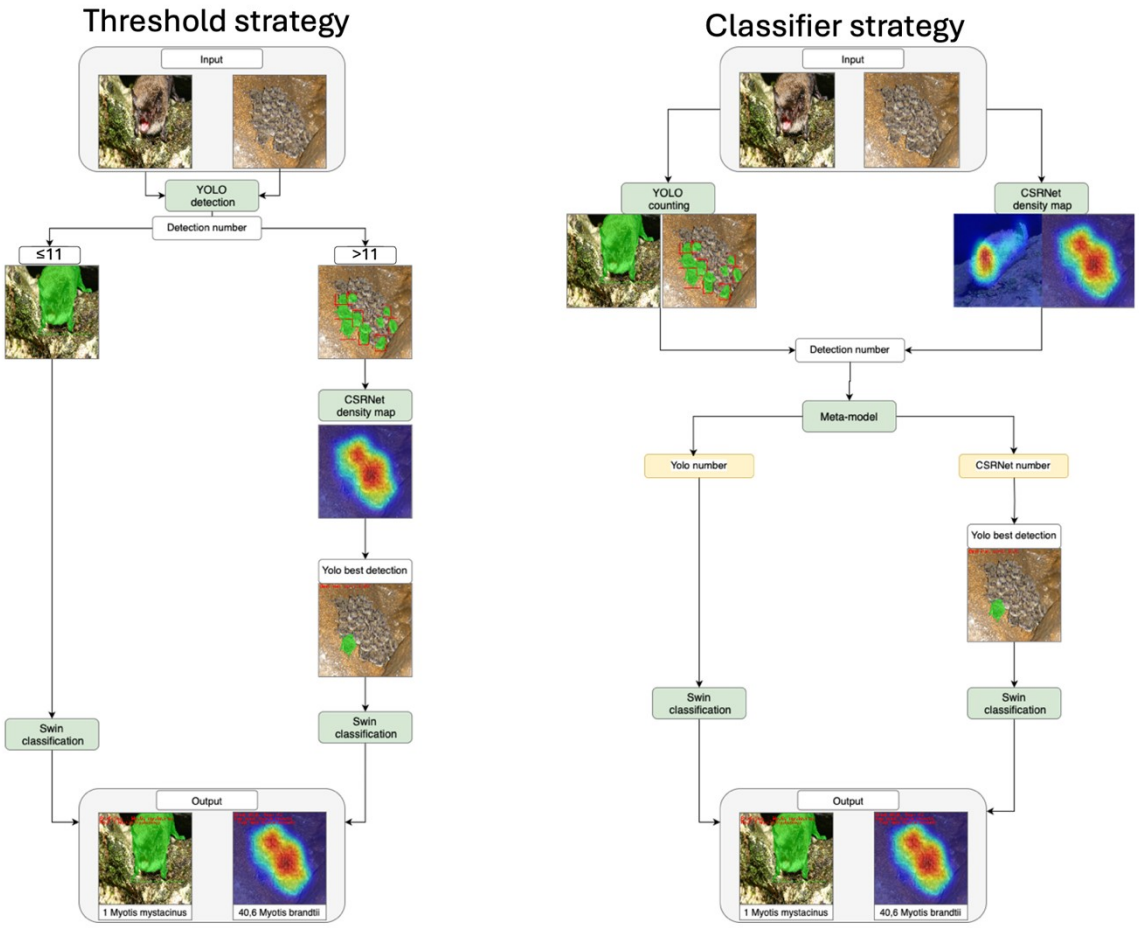


Figure 2: Pipeline to count bats and identify species in an image, using two different strategies: the threshold strategy and the classifier strategy.

2.3) Metrics

Detection performance is assessed via precision, recall and their harmonic mean (F1 score). These metrics are defined on the basis of true positives (TP), false positives (FP), and false negatives (FN), as follows:

-

-

-

Counting accuracy is measured with three error metrics. For a dataset of images, with true counts and predicted counts :

-

-

-

These metrics together provide a balanced view of detection quality and counting accuracy.

III) Results

3.1. Counting step

3.1.1. Yolo

The YOLO model was evaluated on the validation set. With a precision of 93.2% and a recall of 84.3%, YOLO reliably localizes the vast majority of individuals. The mean Average Precision (mAP) over IoU thresholds from 0.50 to 0.95 is 72.4 %. This indicates that the model remains robust even when stricter overlap is required. At the single IoU threshold of 0.50, mAP rises to 92.6 %, showing that most detections meet at least a moderate overlap requirement.

To assess YOLO's counting ability across different crowding levels, the 179 test images were split into three density classes: "Low" (1 individual, 152 images), "Medium" (2–10 individuals,

22 images) and “High” (> 10 individuals, 5 images). This split reflects the data distribution: 85 % of the images contain only one individual, while fewer than 3 % contain more than ten. It thus supports both evaluation under common, low-density conditions and assessment in very crowded scenes.

In low-density scenes, YOLO achieves a negligible MAE (0.03 individual) and a MAPE of only 2.6 %, confirming its high reliability when a single individual is present. As density increases, the average error grows: in medium-density images, MAE remains moderate (0.64) but MAPE rises to 23.4 %, reflecting variability caused by slight occlusions and scale heterogeneity. Under high-density conditions, performance degrades sharply (MAE = 23.2; MAPE = 67.8 %) due to the challenge of separating closely spaced individuals and reducing false negatives (table 2). These results confirm that, while YOLO is highly robust at low densities, its counting accuracy declines markedly once more than ten individuals are present.

3.1.2. CSRnet

To complement the YOLO results, CSRNet was next evaluated on the same validation set. CSRNet achieves a mean absolute error (MAE) of 3.68 individuals, a root-mean-square error (RMSE) of 7.63, and a mean absolute percentage error (MAPE) of 12.9 % (table 2).

To investigate density-dependent behavior, the same three classes as for YOLO were used. In low-density images, CSRNet yields a MAE of 4.02 and a MAPE of 402.3 %. Such a high relative error reflects the ratio of a four-unit error to a single real individual. In medium-density scenes, MAE falls to 2.79, but MAPE remains high at 116.5 %, indicating persistent overestimation when only a few individuals are present. Under high-density conditions, CSRNet produces larger absolute errors (MAE = 8.15; RMSE = 9.61) yet achieves its lowest MAPE of 29.1 %. This happens because true counts in these images are much higher, making relative errors smaller. It also reflects that only five images fall in this class, so a single large

error heavily influences the MAE. Overall, these findings show that CSRNet struggles in near-empty scenes but becomes increasingly reliable as crowding grows. CSRNet was also trained on image sets containing more than five individuals, which explains its poor performance in low-density scenarios. The goal was to specialize the model for high-density cases, where YOLO's performance tends to decline.

Table 2: Counting errors by density class for YOLO and CSRNet models.

	YOLO			CSRNet		
Density	**MAE**	**RSME**	**MAPE (%)**	**MAE**	**RSME**	**MAPE (%)**
Low (1)	0.03	0.16	2.63	4.02	9.06	402.30
Medium (2-10)	0.64	1.04	23.42	2.79	4.29	116.49
High (>10)	23.20	25.22	67.82	8.15	9.61	29.06

3.1.3. Comparison of YOLO and CSRNet

Figure 3a presents boxplots of absolute error for both models in each density class. In low-density images, the YOLO median error is near zero, whereas CSRNet exhibits high variability. In medium density, YOLO error remains low and stable; CSRNet error decreases but still exceeds that of YOLO. In high-density scenes, YOLO error rises sharply, whereas CSRNet error median falls below YOLO, despite a wider spread due to the small sample size.

Figure 3b,c shows the predicted counts from YOLO and CSRNet plotted against the true counts. YOLO points lie close to the diagonal line for up to ten individuals but scatter heavily at higher counts, indicating under-detection in crowded scenes. CSRNet predictions fall below the diagonal for low counts with some predictions reaching 75 for a single bat. However, it aligns more linearly with true values as density grows.

Some examples can be seen in Figure 4, which provides qualitative examples for each density class. In the low-density image, YOLO correctly localizes the single bat, whereas CSRNet's density map underestimates the count and fails to pinpoint the individual. In medium density, YOLO detects most bats but misses one occluded individual, while CSRNet captures the group size more closely. In high density, YOLO detects only a fraction of individuals, whereas CSRNet produces a smooth density map whose integral closely matches the true count.

Together, these results confirm that YOLO excels at counting in sparse scenes but fails when many individuals overlap. CSRNet, by contrast, overestimates at very low densities yet becomes more accurate in crowded conditions. This complementarity suggests a hybrid approach could leverage YOLO's precision in sparse images and CSRNet's robustness in dense ones.

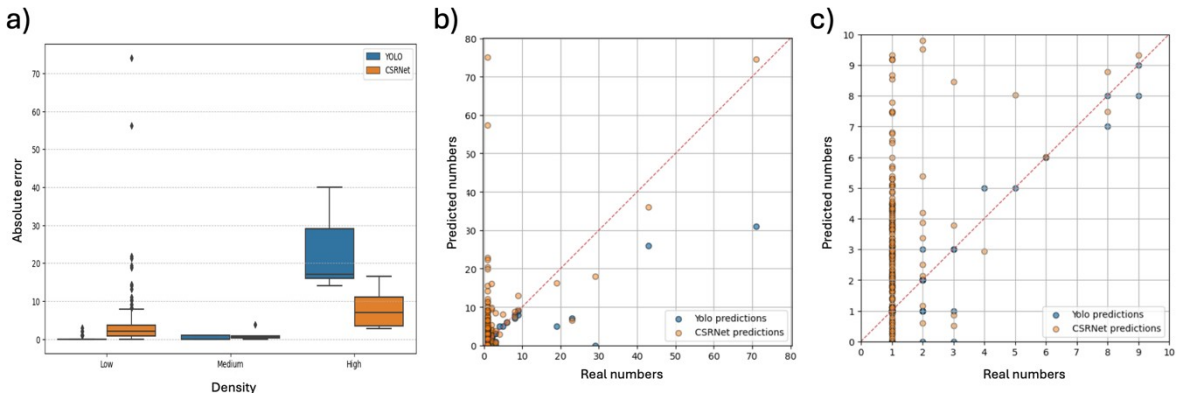


Figure 3: Counting performance of YOLO and CSRNet across density conditions: (a) distribution of absolute counting errors by density class, (b) predicted versus true counts across the full range of colony sizes, (c) predicted versus true counts restricted to the 0–10 range.

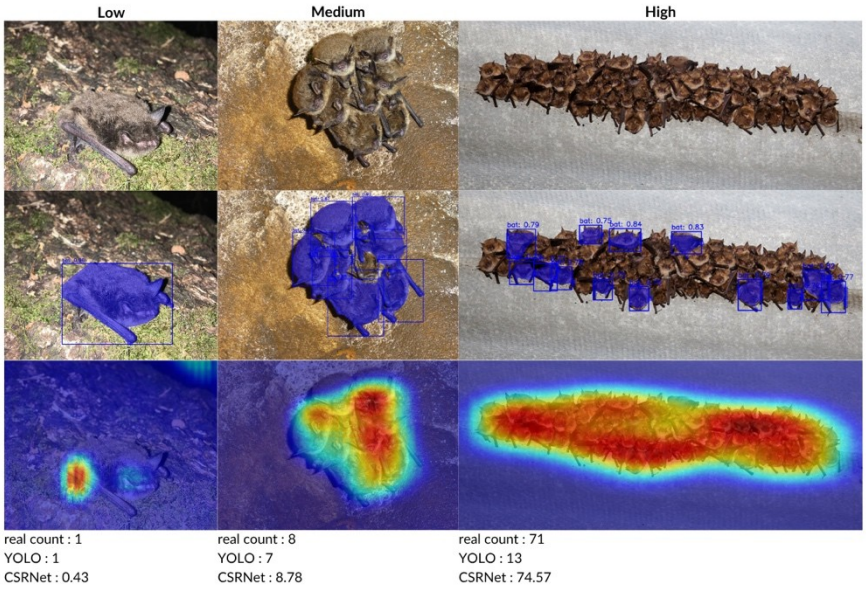


Figure 4: Qualitative examples of YOLO detections and CSRNet density maps across different density classes. The first row shows the original images, the second row shows YOLO detections, and the third row shows CSRNet density maps.

3.1.3. Swin transformer– classification step

The classification model achieves an overall precision of 0.79, recall of 0.79 and F1 score of 0.78. Performance generally scales with the size of the training set. Pipistrellus and Myotis daubentonii, which account for 401 and 389 images, respectively, reach precision and recall above 0.96. Nyctalus, despite appearing in only 38 images, also achieves strong scores

(precision 1.00, recall 0.96), demonstrating that the network can learn robust features even from limited samples (figure 5).

The confusion matrix shows that most errors happen between morphologically similar Myotis species. Myotis daubentonii is occasionally misclassified as M. mystacinus and vice versa. Myotis brandtii, which has fewer examples, has lower recall (0.42) and experiences the highest confusion rates. Eptesicus and the “unidentified bat” class also show moderate misclassification (Figure 6). Representative Grad-CAM visualizations for different species are provided in Supplementary Figure S2. Overall, these results confirm that abundant classes are detected most accurately, while rarer species can still achieve excellent performance when their visual features are distinctive.

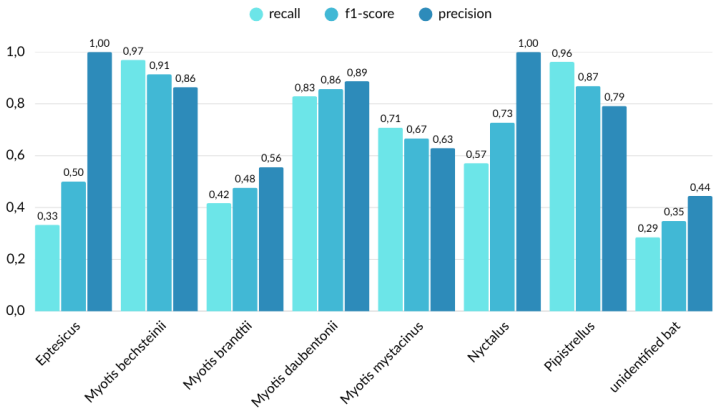


Figure 5: Per-Species Precision, Recall, and F1 Score

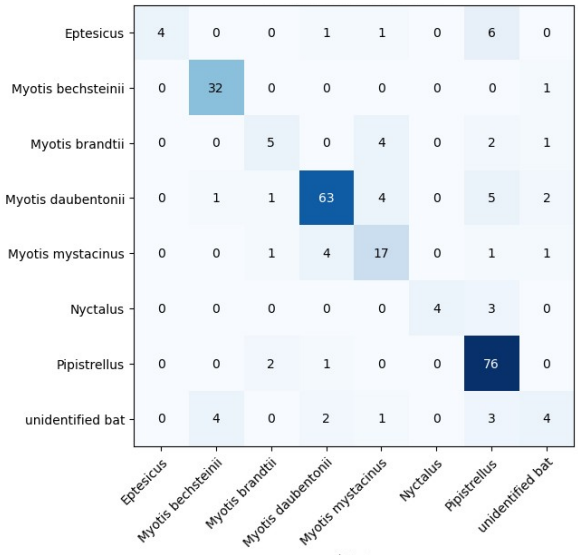


Figure 6: Confusion Matrix for Bat Species Classification

3.2. Pipeline

The optimal selection threshold was initially estimated on the training set (n = 1930). Figure 7a shows MAE as a function of the true-count threshold, ranging from 1 to 50. The error decreases rapidly at low thresholds, then levels off, reaching a minimum MAE of 0.10 at a threshold of 21 individuals (relative MAE = 0.0196).

To simulate a real deployment, the true count was replaced with the number of YOLO detections. The MAE curve (Figure 7b) shows a similar shape but reaches its lowest MAE of 0.11 at a threshold of 11 detections (relative MAE = 0.019). The slight rise in MAE reflects YOLO's tendency to under-detect in crowded scenes. Underestimation by YOLO at higher true densities causes some dense images to be classified as low-density, preventing CSRNet from being applied and thus increasing the overall error. These findings confirm that a threshold-based rule can effectively guide model selection, but careful consideration of YOLO's bias in dense conditions is necessary.

Given these limitations of the fixed-threshold approach, a classifier model was therefore developed to improve the selection between YOLO and CSRNet.

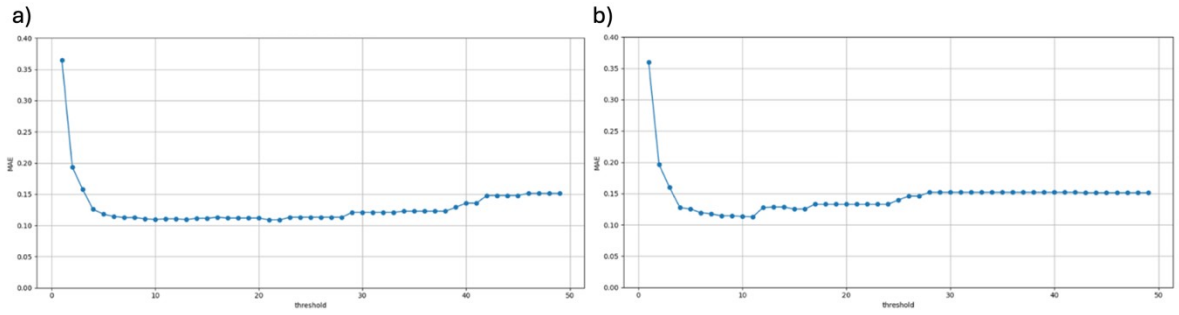


Figure 7: MAE evolution by threshold, using true count (a) or YOLO predictions (b)

Under the fixed-threshold strategy, YOLO is selected whenever it detects more than eleven objects and CSRNet otherwise. On the test set (n = 179), this rule yields very high accuracy for YOLO (F1-score 0.97) but leaves CSRNet with poor recall (0.18, F1-score 0.31). Switching to the learned classifier markedly improves CSRNet's recall (0.55, F1-score 0.71) while preserving YOLO's near-perfect performance (F1-score 0.99, Table 3). Because more than 85% of images contained only a few individuals, where YOLO already performs strongly, the overall reduction in counting error was relatively modest. Nevertheless, the classifier-based strategy consistently outperformed the simple threshold rule, reducing the mean absolute error from 0.49 to 0.41, the root means square error from 2.78 to 2.57, and the mean absolute percentage error from 6.6% to 5.3%. These results confirm that the classifier provides a more reliable selection mechanism for optimizing counting performance across the dataset.

Table 3. Classification performance under threshold and classifier strategies

	Threshold strategy			Classifier strategy		
Model	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
CSRNet	1.00	0.18	0.31	1.00	0.55	0.71
Yolo	0.95	1.00	0.97	0.97	1.00	0.99

Figure 8 compares predicted versus true counts under the fixed-threshold and classifier strategies, each shown at full scale (top) and zoomed in on 0–10 bats (bottom). On the left, the threshold rule selects CSRNet only when YOLO detects eleven or fewer objects. In crowded scenes, many points fall well below the diagonal, indicating that YOLO under-detects and prevents CSRNet from correcting the count. The zoomed-in view confirms that nearly all low-density images are counted correctly by YOLO, while CSRNet's few applications still produce large over- and under-estimates. On the right, the learned classifier switches more effectively between YOLO and CSRNet. Its full-range plot shows far fewer under-estimates at high counts, with predictions clustering closer to the ideal line. In the 0–10 zoom, the classifier reduces CSRNet's extreme errors in low-density cases and preserves YOLO's accuracy, yielding a tighter alignment around the diagonal. Together, these panels demonstrate that a dynamic, learned selection improves count accuracy in dense scenes while retaining YOLO's precision when bats are sparse.

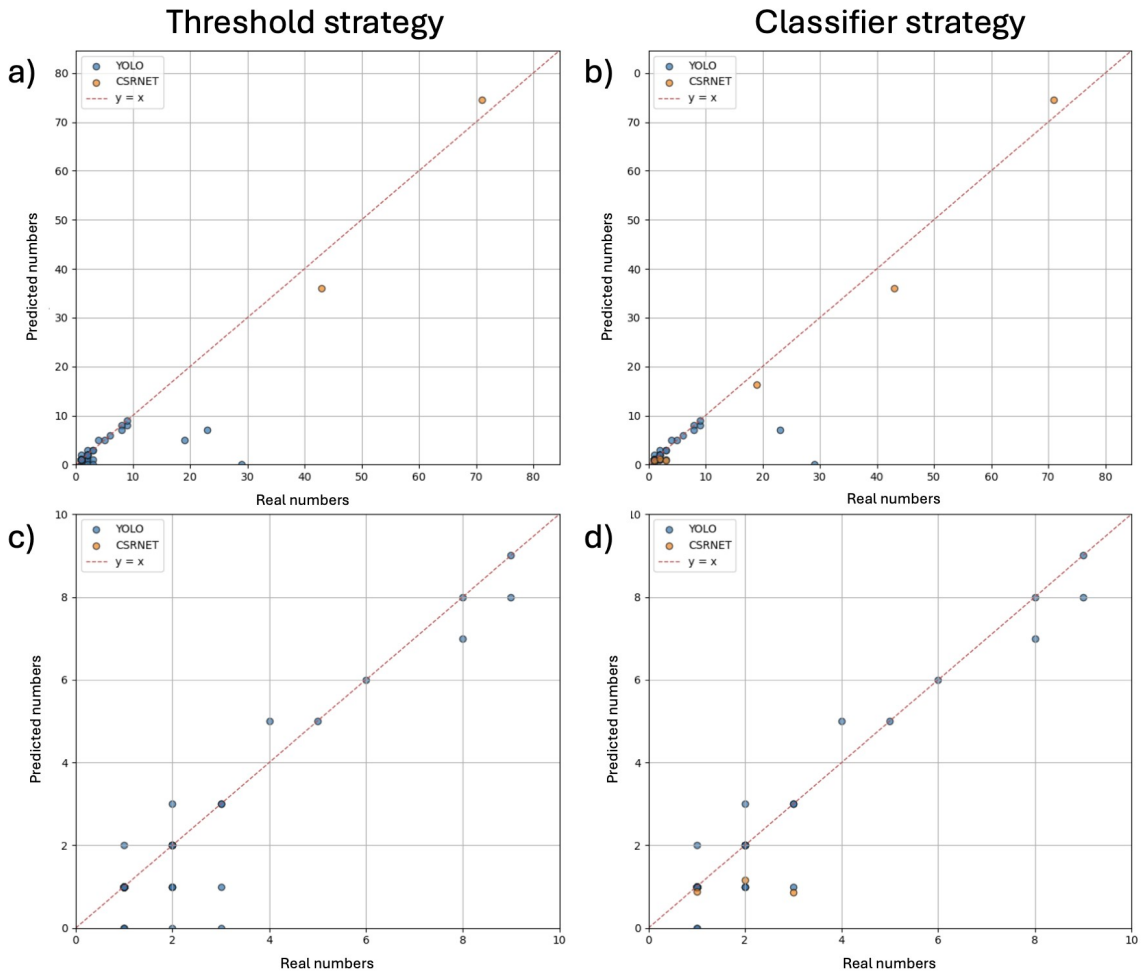


Figure 8. Predicted versus True Bat Counts under Threshold and Classifier Strategies, showing (a) threshold rule full range, (b) classifier full range, (c) threshold rule zoom 0–10, and (d) classifier zoom 0–10.

IV. Discussion

This study introduces a novel, fully automated pipeline that performs detection, counting, and species-level classification of bats from static images, leveraging the complementary strengths of three state-of-the-art deep learning models. Designed to address the methodological limitations of traditional bat monitoring approaches, this modular system provides a robust, non-invasive alternative adaptable to diverse environments and roost configurations.

The pipeline's modular design, combining YOLOv11s-seg for object detection, CSRNet for density estimation, and Swin Transformer V2 for fine-grained classification, proves particularly effective at handling the wide variability in bat colony densities.

Results confirm that YOLO achieves excellent accuracy in sparse scenes, where individual bats can be clearly separated and localized. However, in scenes with more than ten bats, YOLO's performance declines due to occlusion, overlapping individuals, and morphological deformations (such as folded wings), which aligns with the known limitations of instance detection in dense environments.

In contrast, CSRNet performs well in high-density scenarios where its convolutional architecture can extract spatial patterns to estimate total abundance, even when individuals are not clearly visible. However, its performance drops in low-density images because it is trained mainly on crowded scenes, highlighting the importance of aligning model specialization with data characteristics. The meta-decision mechanism, which selects the optimal model based on scene content, enables dynamic adaptation and helps overcome the limitations of each model.

The Swin Transformer V2 classifier excelled at identifying common bat species by learning subtle shape and texture cues from diverse images. It also performed well on certain rare taxa, although success varied with the distinctiveness of their visual features. These performances suggest that transformer-based architectures can learn distinctive morphological features even in limited-data regimes. However, confusions among visually similar Myotis species reflect genuine taxonomic challenges and highlights the value of additional viewpoints, such as side profiles or close-ups of diagnostic features, to boost accuracy for underrepresented classes

The evaluation of the full pipeline further underscores the value of combining complementary modules within a unified framework. The fixed-threshold strategy, although effective in sparse scenes, amplified YOLO’s underestimation bias in crowded environments, thereby reducing CSRNet’s contribution. This result aligns with the density-dependent errors seen in each model separately, where YOLO performed well at low densities but sharply declined with more than ten individuals, while CSRNet produced exaggerated errors at very low counts but matched true values more closely in high-density situations. In contrast, the Random Forest–based selection

process addressed these weaknesses, significantly boosting CSRNet's recall while maintaining YOLO's near-perfect accuracy. This adaptive mechanism lowered overall error rates, demonstrating that hybrid pipelines can outperform single-model or rule-based methods when colony density fluctuates. These results underscore the importance of adaptive decision-making in ecological monitoring, where natural variation in colony size and roosting behavior demands flexible yet dependable analytical strategies.

This work contributes to the growing field of AI-assisted biodiversity monitoring by demonstrating that hybrid AI architectures can outperform single-model approaches in complex ecological tasks. Unlike previous bat-specific models like *BatNet*, which focus solely on species identification, this pipeline captures the full monitoring process, from individual localization to abundance estimation and species classification. Furthermore, the integration of a learned decision module (Random Forest) marks a shift toward adaptive inference strategies, allowing the system to respond to image-specific characteristics in real time.

From a conservation perspective, such tools are invaluable. Bats are essential for ecosystem functioning, yet notoriously difficult to monitor due to their nocturnal, cryptic behavior. By offering a scalable, cost-efficient, and accurate monitoring solution, this pipeline aligns with conservation legislation requirements, such as the ERC framework and Habitats Directive, by enabling reliable fauna–flora inventories with minimal disturbance to the animals.

Several limitations should be acknowledged. First, training labels were verified by a single expert, which may have introduced classification noise, particularly among closely related species. Future work should integrate multi-expert consensus or crowdsourced validation with confidence scoring to improve label quality. Second, imbalanced class distribution, with some species represented by fewer than 50 images, limits generalization. While Focal Loss mitigates this partially, targeted data acquisition or generative augmentation could improve model robustness for rare taxa. Moreover, CSRNet's underperformance in low-density scenarios

reveals the need for either broader training or model ensembles. Although the meta-model improves counting accuracy, its current architecture relies on hand-crafted features (e.g., count and confidence scores) and may benefit from end-to-end learning approaches. Finally, the use of static images precludes temporal or behavioral analysis; incorporating video data or multi-frame sequences could enable tracking of emergent patterns, flight behavior, or even colony dynamics over time.

To extend this work, several promising avenues exist:

- **Dataset expansion**: Enriching the dataset with geographically and visually diverse images, including those from citizen science platforms like iNaturalist or DeepFaune, would enhance generalizability.
- **Active learning**: Implementing models that can identify uncertain predictions and query for expert validation can reduce annotation effort while improving performance.
- **Domain adaptation**: Applying transfer learning or unsupervised domain adaptation would allow the system to generalize to novel lighting conditions, camera angles, or box types without extensive retraining.
- **Multimodal data fusion**: Integrating acoustic recordings, temperature logs, or roost metadata could boost classification performance and enable ecological inference.
- **Explainability tools**: Creating a user-facing dashboard with interactive Grad-CAM visualizations, uncertainty scores, and metadata overlays would support transparent decision-making in regulatory contexts.

IV) Conclusion

This study introduces a new, modular deep learning pipeline that combines detection, counting, and species-level classification of bats in a single unified framework. Unlike previous bat-specific approaches that focus solely on classification or abundance estimation, this system combines YOLOv11s-seg, CSRNet, and Swin Transformer V2 to adapt dynamically to varying colony densities while ensuring fine-grained species identification.

The results highlight the complementarity of the three models: YOLO excels in sparse conditions, CSRNet offers more accurate estimates in dense colonies, and the Swin Transformer classifier provides strong recognition even for certain rare taxa. The adaptive selection mechanism based on a Random Forest classifier clearly outperformed the simple threshold rule, providing more accurate and reliable switching between YOLO and CSRNet across density conditions. By moving beyond fixed heuristics, this learned strategy represents a central innovation of the pipeline, enabling context-aware adaptation that effectively mitigates the limitations of individual models and strengthens ecological robustness.

Beyond its technical advances, the system offers a scalable, non-invasive solution that aligns with European and French conservation regulations, providing a practical alternative to traditional monitoring methods that are often costly, labor-intensive, and less accurate. By covering the entire monitoring process (localization, counting, and classification), this study takes a step forward in AI-assisted biodiversity assessment and paves the way for wider applications in ecological research and conservation. Future developments will expand this framework with larger, more diverse datasets, multimodal inputs, and improved adaptive strategies, strengthening its potential as a promising tool for bat population monitoring and beyond.

References

Bakana, S.R., Zhang, Y., Twala, B., 2024. WildARe-YOLO: A lightweight and efficient wild animal recognition model. Ecol. Inform. 80, 102541. https://doi.org/10.1016/j.ecoinf.2024.102541

Barnosky, A.D., Hadly, E.A., Bascompte, J., Berlow, E.L., Brown, J.H., Fortelius, M., Getz, W.M., Harte, J., Hastings, A., Marquet, P.A., Martinez, N.D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J.W., Gillespie, R., Kitzes, J., Marshall, C., Matzke, N., Mindell, D.P., Revilla, E., Smith, A.B., 2012. Approaching a state shift in Earth's biosphere. Nature 486, 52–58. https://doi.org/10.1038/nature11018

Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., Nguyen, H., Benjamin, C., Menzel, A., 2023. Automated wildlife image classification: An active learning tool for ecological applications. Ecol. Inform. 77, 102231. https://doi.org/10.1016/j.ecoinf.2023.102231

Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2005. Bundesartenschutzverordnung (BArtSchV) [Federal Species Protection Ordinance].

Cao, Z., Li, C., Wang, K., He, K., Wang, X., Yu, W., 2023. A fast and accurate identification model for Rhinolophus bats based on fine-grained information. Sci. Rep. 13, 16375. https://doi.org/10.1038/s41598-023-42577-1

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Sci. Adv. 1, e1400253. https://doi.org/10.1126/sciadv.1400253

Chen, L., Li, G., Zhang, S., Mao, W., Zhang, M., 2024. YOLO-SAG: An improved wildlife object detection algorithm based on YOLOv8n. Ecol. Inform. 83, 102791. https://doi.org/10.1016/j.ecoinf.2024.102791

Council of Europe, 1979. Convention on the Conservation of European Wildlife and Natural Habitats (Bern Convention).

Council of the European Communities, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora, OJ L.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B., Collen, B., 2014. Defaunation in the Anthropocene. Science 345, 401–406. https://doi.org/10.1126/science.1251817

Ferreira, D.F., Jarrett, C., Jules Atagana, P., Powell, L.L., Rebelo, H., 2024. Are bat mist nets ideal for capturing bats? From ultrathin to bird nets, a field test. ResearchGate. https://doi.org/10.1093/jmammal/gyab109

Froidevaux, J.S.P., Boughey, K.L., Hawkins, C.L., Jones, G., Collins, J., 2020. Evaluating survey methods for bat roost detection in ecological impact assessment. Anim. Conserv. 23, 597–606. https://doi.org/10.1111/acv.12574

Huang, Y.-P., Basanta, H., 2021. Recognition of Endemic Bird Species Using Deep Learning Models. IEEE Access 9, 102975–102984. https://doi.org/10.1109/ACCESS.2021.3098532

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Service (Summary for Policymakers), Global assessment report on biodiversity and ecosystem services. IPBES Secretariat, Bonn, Germany.

Jung, T.S., Kukka, M., 2013. Conserving and monitoring little brown bat (Myotis lucifugus) colonies in Yukon: 2012 annual report. (No. PR-13-03). Yukon Fish and Wildlife Branch, Whitehorse, Yukon, Canada.

Khanam, R., Hussain, M., 2024. YOLOv11: An Overview of the Key Architectural

Enhancements. https://doi.org/10.48550/arXiv.2410.17725
Krivek, Gabriella, Gillert, A., Harder, M., Fritze, M., Frankowski, K., Timm, L., Meyer-Olbersleben, L., von Lukas, U.F., Kerth, G., van Schaik, J., 2023. BatNet: a deep learning-based tool for automated bat species identification from camera trap images. Remote Sens. Ecol. Conserv. 9, 759–774. https://doi.org/10.1002/rse2.339
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. Anim. Conserv. 26, 701–713. https://doi.org/10.1111/acv.12856
Larsen, R., 2007. Mist Net Interaction, Sampling Effort, and Species of Bats Captured on Montserrat, British West Indies. Electron. Theses Diss.
Li, Y., Zhang, X., Chen, D., 2018. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. https://doi.org/10.48550/arXiv.1802.10062
Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B., 2022. Swin Transformer V2: Scaling Up Capacity and Resolution, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11999–12009. https://doi.org/10.1109/CVPR52688.2022.01170
Ministère de la Transition écologique et solidaire, 2019. séquence Éviter, Réduire et Compenser (ERC) – doctrine et recommandations. (Code de l'environnement).
Noa Rigoudy, Noa Rigoudy, Abdelbaki Benyoub, Abdelbaki Benyoub, Aurélien Besnard, Aurélien Besnard, Carole Birck, Carole Birck, Yoann Bollet, Yoann Bollet, Yoann Bunz, Yoann Bunz, Nina De Backer, Nina De Backer, Gérard Caussimont, Gérard Caussimont, Anne Delestrade, Anne Delestrade, Lucie Dispan, Lucie Dispan, Jean-François Elder, Jean-François Elder, Jean-Baptiste Fanjul, Jean-Baptiste Fanjul, Jocelyn Fonderflick, Jocelyn Fonderflick, Mathieu Garel, Mathieu Garel, William Gaudry, William Gaudry, Agathe Gérard, Agathe Gérard, Olivier Gimenez, Olivier Giménez, Arzhela Hemery, Arzhela Hemery, Audrey Hemon, Audrey Hemon, Jean-Michel Jullien, Jérôme Jullien, Maden Le Barh, Maden Le Barh, Isabelle Malafosse, Isabelle Malafosse, Malory Randon, Malory Randon, Romain Ribière, Romain Ribière, Sandrine Ruette, Sandrine Ruette, Guillaume Terpereau, Guillaume Terpereau, Wilfried Thuiller, Wilfried Thuiller, Valentin Vautrain, Valentin Vautrain, Bruno Spataro, B. Spataro, Vincent Miele, Vincent Miele, Simon Chamaillé-Jammes, Simon Chamaillé-Jammes, 2022. The DeepFaune initiative: a collaborative effort towards the automatic identification of the French fauna in camera-trap images. https://doi.org/10.1101/2022.03.15.484324
Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12, 150–161. https://doi.org/10.1111/2041-210X.13504
Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. U. S. A. 115, E5716–E5725. https://doi.org/10.1073/pnas.1719367115
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 779–788.
République française, 2021. Article L.411-2 du Code de l'environnement, Code de l'environnement.
République française, 2016. Article L.411-1 du Code de l'environnement, Code de l'environnement.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Lukacs, P.M., Moeller, A.K., Mandeville, E.G., Clune, J., Miller, R.S., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods Ecol. Evol. 10, 585–590. https://doi.org/10.1111/2041-210X.13120

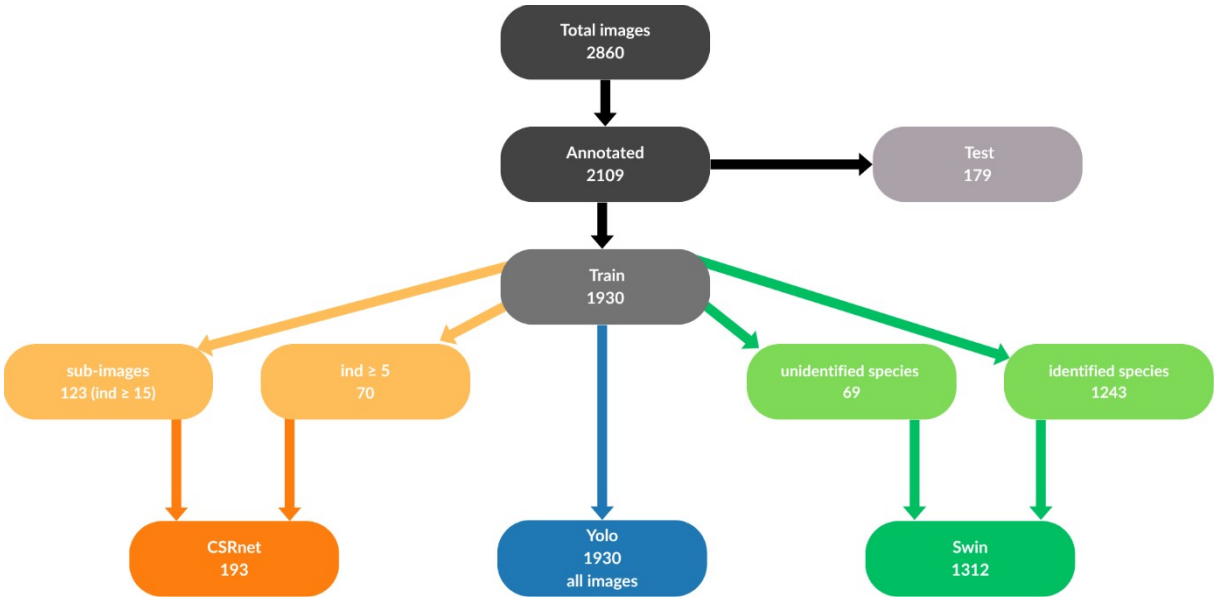
United Nations Environment Programme (UNEP), Convention on Migratory Species Secretariat, 1991. Agreement on the Conservation of Populations of European Bats (EUROBATS).

Whiting, J.C., Doering, B., Aho, K., 2022. Can acoustic recordings of cave-exiting bats in winter estimate bat abundance in hibernacula? Ecol. Indic. 137, 108755. https://doi.org/10.1016/j.ecolind.2022.108755

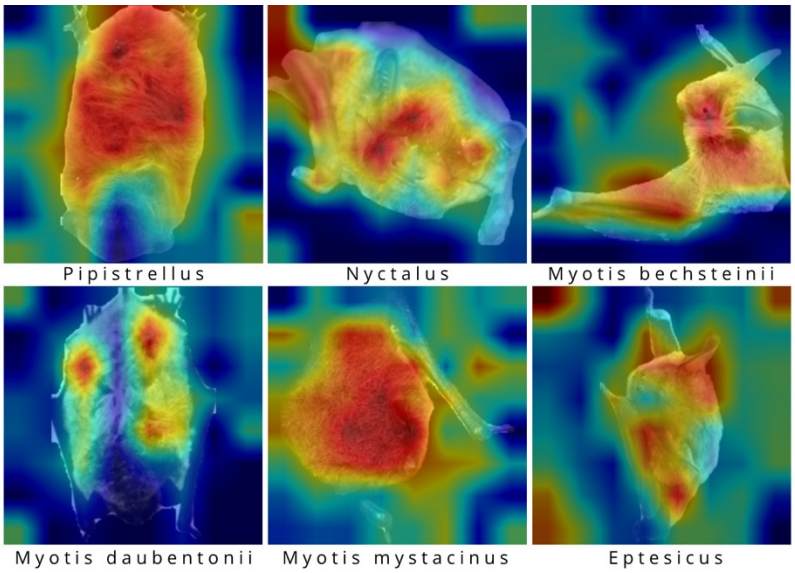
Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M., Bahaa-el-din, L., Boekee, K., Brittain, S., Cardoso, A.W., Henschel, P., Lehmann, D., Momboua, B., Kiebou Opepa, C., Orbell, C., Pitman, R.T., Robinson, H.S., Abernethy, K.A., 2021. Robust ecological analysis of camera trap data labelled by a machine learning model. Methods Ecol. Evol. 12, 1080–1092. https://doi.org/10.1111/2041-210X.13576

Supplementary Figure S2 : Grad-CAM visualizations of Swin Transformer attention across bat species

Supplementary Material



Supplementary Figure S1. Summary of Image Counts and Bat Abundance Statistics by Species


Pipistrellus
Nyctalus
Myotis bechsteinii
Myotis daubentonii
Myotis mystacinus
Eptesicus