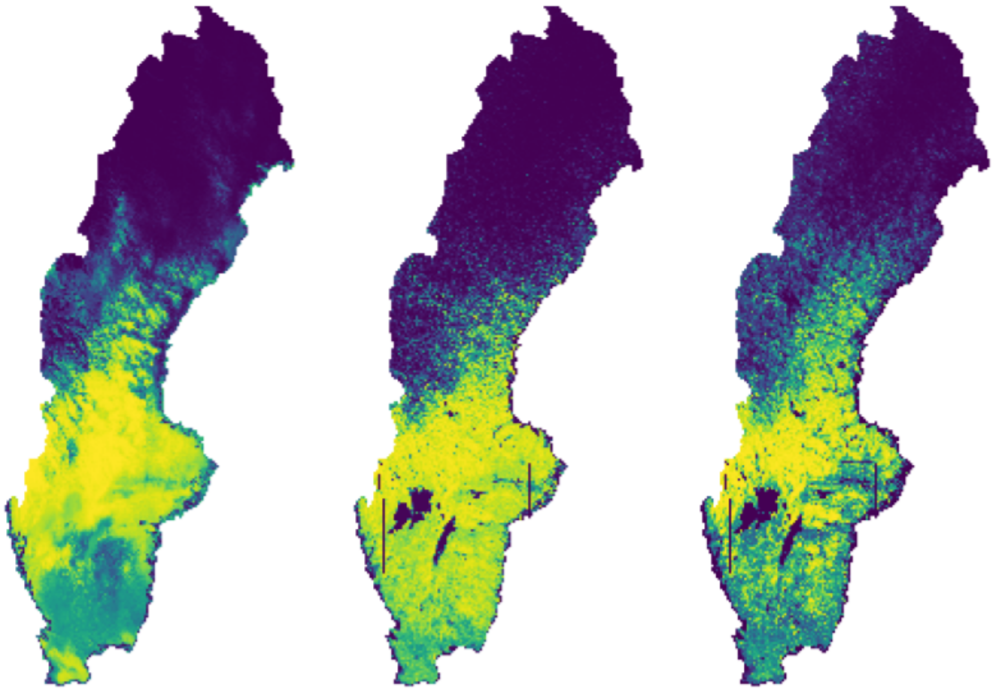


AI-based Species Distribution Modelling using Remote Sensing

Using Convolutional Neural Networks with High Resolution Remote Sensing Data to Predict Species Distributions from Observation Data

Master's thesis in Complex Adaptive Systems

MARCUS BLIXT, CARL KLANG

DEPARTMENT OF ENVIRONMENTAL AND ENERGY SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

AI-based Species Distribution Modelling using Remote Sensing

Using Convolutional Neural Networks with High Resolution Remote Sensing Data to Predict Species Distributions from Observation Data

MARCUS BLIXT
CARL KLANG

Department of Environmental and Energy Sciences
Chalmers University of Technology
Gothenburg, Sweden 2026

AI-based Species Distribution Modelling using Remote Sensing
Using Convolutional Neural Networks with High Resolution Remote Sensing Data to Predict Species Distributions from Observation Data
MARCUS BLIXT, CARL KLANG

Supervisor: Olof Mogren, RISE
Examiner: Claes Andersson, Department of Environmental and Energy Sciences

Master's Thesis 2026
Department of Environmental and Energy Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Prediction maps of wolves over Sweden from three species distribution models based on the convolutional neural network ResNet-18, using environmental input, satellite embedding input and, location + satellite fusion respectively.

Typeset in LaTeX
Gothenburg, Sweden 2026

AI-based Species Distribution Modelling using Remote Sensing
Using Convolutional Neural Networks with High Resolution Remote Sensing Data to Predict Species Distributions from Observation Data
MARCUS BLIXT, CARL KLANG
Department of Environmental and Energy Sciences
Chalmers University of Technology

Abstract

Species distribution modelling (SDM) is an essential tool for informing conservation efforts amid critical biodiversity loss driven by climate change and habitat exploitation. By pairing species observation records with environmental data, SDMs predict species ranges across geographic areas. This thesis investigates whether convolutional neural networks (CNNs) using high-resolution remote-sensing-derived satellite embeddings as environmental input can improve model performance when using opportunistic observation data from the Swedish citizen-science platform Artportalen. Because of their opportunistic nature, the observation records in Artportalen lack confirmed absences and suffer from severe spatial sampling biases. To address this, we develop a custom pseudo-absence generation method that weights sampling probabilities based on background observation density to enable model training and counteract these biases. For this investigation, we evaluate three CNN architectures (ResNet-18, U-Net, and U-Net++) using either satellite embeddings or traditional environmental data, benchmarking them against traditional-data baseline models consisting of random forest and an established multi-layer perceptron architecture. We also design fusion variants for the CNNs that incorporate raw observation coordinates into the input.

We train and evaluate all models on Artportalen observations for 34 Swedish mammal species, alongside external validation using the Rovbase dataset for large carnivores and expert-derived range maps. Satellite embeddings substantially improve CNN performance. While location–satellite fusion models achieved the best score on the Artportalen test set, external validation on Rovbase revealed that only the ResNet-18 fusion variant consistently outperformed models utilizing satellite inputs alone. Species-level results show that spatial spread, rather than observation count alone, strongly affects model difficulty. However, observer bias remains visible, especially around populated areas and under-sampled regions with few human visitors. The results suggest that remote-sensing embeddings are highly promising for high-resolution SDM, but bias-aware sampling and calibration remain central challenges.

Keywords: species distribution modelling, pseudo-absence, deep learning, convolutional neural network, citizen science, remote sensing, satellite embeddings.

Acknowledgements

We would first and foremost express our gratitude towards our supervisor at RISE, Olof Mogren. From arranging access to different data, arranging meetings with experts in the field and guiding us through the many twists and turns of the project, your support has been vital. We would also like to thank the other people at RIDR, the Deep Learning Research Group at RISE, for creating a welcoming working environment and helping out when needed, especially Ghjulia Sialelli for helping us with AlphaEarth access. A special thanks to André Osthövener, an intern who worked in parallel with the same data for the first three months of our project. You were of great help discussing ideas and testing things out in parallel, without you our work would have taken much longer.

Thank you to Naturvårdsverket and SLU Artdatabanken for providing and entrusting us with data for protected species.

We also want to thank all of the people who took time to have meetings with us and discuss the project. Oisin Mac Aodha and Mélisande Teng who have both done research in species distribution modelling, Peter Jaxgård and Robert Ekblom from Naturvårdsverket, and Henrik Thurfjell from SLU.

Lastly we would like to thank our examiner Claes Andersson for providing us with lots of ideas and enthusiasm at the start of the project, your trust in us working independently towards RISE, and your work with assessing our thesis.

Marcus Blixt and Carl Klang, Gothenburg, May 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AEF	AlphaEarth Foundations
AP	Average precision
CNN	Convolutional neural network
COG	Cloud Optimized GeoTIFF
DNN	Deep neural network
FN	False negative
FP	False positive
IUCN	International Union for Conservation of Nature
LPT	Lowest presence threshold
LPT-R	Lowest presence threshold - robust
mAP	Mean average precision
MLP	Multi-layer perceptron
RF	Random forest
SAR	Synthetic aperture radar
SDM	Species distribution modelling
SINR	Spatial implicit neural representation
SLU	Sveriges Lantbruksuniversitet (Swedish University of Agricultural Sciences)
TN	True negative
TP	True positive

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables and functions that have been used throughout this thesis.

Indices

b	Index for presence and pseudo-absence sample pairs within a batch
i, j	Spatial pixel coordinate indices within a patch
k	Index for sub-models in the U-Net++ model
n	Index for sorted threshold levels
s	Index for a species
s_b	Index of the observed species associated with sample pair b

Sets

$\mathcal{S}$	The set of species being modelled
$\mathcal{X}$	The set of environmental variables representing a spatial domain

Parameters

N	Total number of sample pairs within a batch
K	Total number of sub-models of the U-Net++ model
r	Fixed radius defining the active spatial area around a species observation
$\lvert\mathcal{S}\rvert$	Total number of species in the set $\mathcal{S}$
w_A	Loss function weighting coefficient for pseudo-absence samples
w_P	Loss function weighting coefficient for presence samples

θ	Classification threshold for converting continuous probabilities into binary results
$\lvert\Omega(r)\rvert$	Total number of active pixel elements inside the spatial mask matrix

Variables

P	Precision metric value
P_n	Precision achieved at the n-th threshold level
R	Recall metric value
R_n	Recall achieved at the n-th threshold level
V_{final}	De-quantized continuous floating-point value
V_{raw}	Raw quantized signed 8-bit integer value stored in the embedding vector
$\mathbf{x}$	Environmental variables for a location
$\mathbf{x}_{\text{loc}}$	Sinusoidal location encoding vector
$\hat{\mathbf{y}}$	Predicted probability vector across all species S
$\hat{y}_s$	Predicted probability for species $s \in S$
$\tilde{\mathbf{y}}$	Binary predicted presence-absence vector across all species S
$\tilde{y}_s$	Binary predicted presence-absence classification for species $s \in S$
$\hat{\mathbf{y}}_{1:N}$	Sequence of predicted probability vectors of presence samples in a batch
$\hat{\mathbf{y}}'_{1:N}$	Sequence of predicted probability vectors of pseudo-absence samples in a batch
$\hat{y}_{b,s_b}$	Predicted presence probability for species s_b in the presence sample of pair b
$\hat{y}'_{b,s_b}$	Predicted presence probability for species s_b in the pseudo-absence sample of pair b
$\hat{\mathbf{Y}}_{1:N}$	Sequence of predicted spatial segmentation patch tensors of presence samples in a batch
$\hat{\mathbf{Y}}'_{1:N}$	Sequence of predicted spatial segmentation patch tensors of pseudo-absence samples in a batch
$\hat{\mathbf{Y}}^k_{1:N}$	Sequence of predicted spatial segmentation patch tensors of presence samples in a batch from sub-model k in the U-Net++ model
$\hat{\mathbf{Y}}'^k_{1:N}$	Sequence of predicted spatial segmentation patch tensors of pseudo-absence samples in a batch from sub-model k in the U-Net++ model

$\hat{Y}_{b,s_b,i,j}$	Predicted probability at pixel (i, j) for species s_b in the presence sample of pair b
$\hat{Y}'_{b,s_b,i,j}$	Predicted probability at pixel (i, j) for species s_b in the pseudo-absence sample of pair b
$\tilde{\lambda}$	Normalized longitude coordinate in the range $[-1, 1]$
$\tilde{\phi}$	Normalized latitude coordinate in the range $[-1, 1]$

Functions

$f(\mathbf{x})$	Species distribution model function mapping environmental variables to a likelihood of presence
$\mathcal{L}_{\mathrm{ds}}$	Overall deep supervision loss function (U-Net++)
$\mathcal{L}_{\mathrm{mpo}}(\cdot)$	Loss function for masked patch outputs (U-Net)
$\mathcal{L}_{\mathrm{svo}}(\cdot)$	Loss function for single-value outputs (SINR and ResNet-18)
$\Omega(r)$	Spatial mask function defining the active area for the loss function based on radius r

Contents

List of Figures

List of Tables

1
Introduction

Climate change represents one of the most critical global challenges of the modern era, with the accelerating loss of biodiversity serving as a profound consequence [1]. These ecological shifts threaten not only environmental stability but, by extension, human socioeconomic systems as well. Another variable disrupting the balance in nature is human exploitation, including farming, forestry, hunting, and fishing [2]. To mitigate these effects and implement effective conservation strategies, a comprehensive understanding of species geographic distributions and the environmental factors governing their ranges is important.

A commonly used method for addressing this need is species distribution modelling (SDM) [3], which leverages species observation records and environmental data to estimate the range of the observed species across geographical areas. The development of deep learning and improved quality and resolution of remote sensing imagery, collected by satellites, drones and planes, have opened new avenues for SDM and their ability to map the distribution of multiple species at once and make high resolution predictions [4, 5]. This thesis will investigate the performance of some deep learning models, both when used with traditional environmental data and remote sensing.

The code for the project is available at:
`https://github.com/ridr-se/2026-species-distribution-modeling`

1.1 Background

The geographic distribution of a species is most commonly represented by a map of where the species is present and absent and this representation can be created in a variety of ways. Three of them are to show raw species observation data, expert-derived species range maps, or employ SDM [5].

Species observation data is the foundation for all representations of species distributions, with the simplest approach being the data itself. However, it lacks the ability to make predictions for unseen areas [5]. Observation data can be divided into two categories: presence-only data, which logs coordinate locations of occurrences of the species, and presence-absence data, which also contains locations where the species do not occur.

In contrast to using raw observation data, the expert-derived range maps are the most expensive and time consuming way to represent species distribution, but generally the most trusted as well [5]. Experts set the geographic range of the species based on observations, knowledge of species behaviour and environmental preferences, which requires extensive knowledge of each individual species. Some of the most widely known expert range maps are those from the International Union for Conservation of Nature (IUCN) Red List of vulnerable and endangered species [6]. A drawback to expert range maps is the cost of continuously updating them as habitats and species populations change because of factors like global warming and human exploitation. Additionally, expert range maps are highly scale-dependent, with many maps, including those from IUCN, being on a global or continental scale missing local variations [7].

A more efficient method is to employ SDM, which uses supervised learning, either through a statistical model or machine learning algorithms to make predictions based on localised environmental characteristics, such as temperature, precipitation, and elevation [5]. These environmental characteristics can be local measurements averaged over a time-period, such as the WorldClim bioclimatic variables [8], but can also include features derived from remote sensing data [5]. By training the model on gathered observation and environmental data, it can make predictions of the species' presence and absence into unmonitored regions, provided that the environmental characteristics are known. Many models need both positive and negative inputs, thus requiring presence-absence data. While some models utilize presence-only data it can also be augmented by generating pseudo-absence to work for all models. Some commonly used models for SDM are generalized linear models and random forest classifiers, both relying on presence-absence data, or maximum entropy modelling, which uses presence-only data.

In recent years, the use of deep learning in SDM has emerged as a powerful alternative to traditional statistical methods. Deep learning methods, or deep neural networks (DNNs), have the ability to map the distribution of multiple species simultaneously compared to most traditional models that target one species at the time. This multi-species modelling allows deep learning methods to capture species interactions, which can improve predictive performance [4]. However, to operate effectively, deep learning methods require large datasets, which can be difficult to acquire. A solution to this is to use crowd-sourced datasets, such as iNaturalist [9] or eBird [10], where the data is gathered through observations submitted by the public. This comes with the drawback of observation bias towards highly populated areas and easily spotted species, which must be taken into account.

Beyond the advantages of multi-species architectures, deep learning also fundamentally changes how environmental predictors are processed, particularly when leveraging remote sensing data. Modern remote sensing imagery achieves incredibly high spatial resolution, capturing local environmental variations. Traditional SDMs, as well as point-based deep learning architectures like multi-layer perceptrons (MLP), fail to utilize the full depth of this data because they restrict input to the environmental variables of a single pixel. In contrast, patch-based architectures like convolutional neural networks (CNNs) process entire spatial neighborhoods as inputs. By doing so, they capture the localized variations, gradients, and spatial textures that dictate species distribution, leading to improved model performance [11].

1.2 Aim

The primary aim of this thesis is to investigate the extent to which high-resolution satellite embeddings, combined with a custom-designed pseudo-absence generation method, can improve the predictive performance of SDMs and reduce spatial sampling biases inherent in crowd-sourced, presence-only data. To achieve this, two key objectives are pursued. The first is to develop a pseudo-absence generation method that weighs sampling probabilities based on the spatial density of observations in order to counteract biases. The second is to evaluate the performance of different CNN architectures, with either satellite embeddings or environmental data as their primary input, against each other and against baseline models.

1.3 Research Questions

To fulfil the aim of this thesis the following research questions are addressed:

1. To what extent can the integration of high-resolution satellite embeddings and convolutional neural networks improve the predictive performance of species distribution models compared to traditional environmental baselines?
2. How does a custom pseudo-absence generation method that weighs absence probabilities based on spatial observation density reduce the impact of inherent spatial bias in crowd-sourced data?

1.4 Scope and Limitations

The scope of this thesis will be limited to evaluating three different CNN architectures, ResNet-18, U-Net, and U-Net++, for SDM. Subsequently, the training and evaluation of the models will be constrained to the area of Sweden with species observation data from Artportalen, a Swedish crowd-sourced presence-only dataset [12]. To evaluate performance against established modelling standards, these CNN architectures are benchmarked against two baseline models: a traditional random forest classifier and the spatial implicit neural representation (SINR) MLP framework [13], which is discussed further in Section 2.2.1.

Two different sets of environmental variables will be used as model input: one with vector embeddings derived from remote-sensing data from the AlphaEarth Foundations Satellite Embeddings dataset [14] and the other with the WorldClim bioclimatic variables [8] with an additional variable for elevation. The baseline models are constrained to using the WorldClim and elevation data, with SINR having an additional input of raw geographic coordinates. Furthermore, the architectural scope includes designing hybrid, late-fusion models. These models process data through two parallel tracks: a CNN track that extracts spatial patterns from the environmental or satellite data, and the SINR model that handles raw geographic coordinates. The distinct features extracted by both tracks are combined at the final stage of the network to produce a single, joint species distribution prediction.

This set of constraints on the training of the models will limit the applicable geographical areas of the trained models to Sweden and, with limited performance, other regions with similar environments. These limitations stem both from the models only being trained on the environmental characteristics of Sweden and biases, such as hunting pressure for different species and other human interactions, not captured by the input characteristics that might differ between countries.

A broader limitation for all SDM is what to use as ground truth for evaluation of model performance. Unlike when classifying the contents of an image, there is no definite ground truth for species distribution. Instead, the performance of the model has to be evaluated using other representations of the species distribution. Hold-out sets, which are common in statistical models and machine learning, have the same biases as the training set, and separate high-quality datasets to use for evaluation can be difficult to obtain, especially regarding presence-absence data. The use of expert range maps is another alternative, but as mentioned in section 1.1, they are not perfect representations either.

1.5 Use of AI Tools

The AI tools used for this project are GitHub Copilot, Claude, ChatGPT, and Gemini. These have been used for various purposes with the main ones being coding, writing and processing results.

Our use of AI for coding includes generating, modifying and debugging code based on directed prompts. All outputs and functionality has been verified by us and checked for plagiarism and copyright using the Copilot public code matching filter. The raw data has also been kept separate from the codespace to ensure that sensitive data is not shared or published.

For writing AI has been used to suggest formulations and structure for the report. This includes generating example paragraphs based on input on what it should contain to help convey information in a better way. Thus all underlying information presented in the report are our own but specific wordings may have been generated.

AI use for processing results has mainly consisted of summarizing and analysing them for broader findings. This includes generating result tables, comparing model performance and discussing the results. The tables have been checked and verified after generation. When analysing and discussing the results AI has been used to bounce ideas about potential causes and correlations and to check our discussions for potential errors. This means that AI has been used mainly to validate our ideas and all presented findings and discussions are still our own.

2

Theory

2.1 Species Distribution Modelling

Formally, SDMs aim to predict where species are likely to occur across a spatial domain by mapping environmental predictors to occurrence probabilities [3]. Let $\mathcal{X}$ represent the set of environmental variables defining a spatial domain, and let $\mathbf{x} \in \mathcal{X}$ denote the feature vector of environmental variables at a specific location. For a set of species S, the model function $f(\mathbf{x})$ produces a continuous probability output vector, denoted as $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = f(\mathbf{x}) \in [0, 1]^S, \tag{2.1}$$

where each element $\hat{y}_s$ represents the estimated likelihood of presence for a given species $s \in S$ (where 0 denotes absolute absence and 1 denotes absolute presence). By making predictions across the entire spatial domain, the model produces a spatial probability distribution. This continuous distribution can serve directly as a continuous range map or enable the calculation of threshold-independent evaluation metrics.

However, to construct binary range maps or evaluate threshold-dependent performance metrics, the continuous predictions $\hat{\mathbf{y}}$ must be converted into a binary classifications $\tilde{\mathbf{y}}$ via a decision threshold $\theta \in [0, 1]$ [15]. The final binary prediction $\tilde{y}_s$ for a species s is defined as:

$$\tilde{y}_s = \begin{cases} 1 & \text{if } \hat{y}_s \geq \theta \\ 0 & \text{if } \hat{y}_s < \theta \end{cases} \tag{2.2}$$

This mapping allows for the determination of discrete presence-absence range maps and the computation of threshold-dependent metrics, which are detailed alongside threshold-independent evaluation metrics in section 2.3

To optimize an SDM's internal parameters during training, observational data within the spatial domain of the environmental variables $\mathcal{X}$ is required to learn the environmental preferences of the target species. As discussed in section 1.1, this observational data typically takes one of two forms, presence-absence or presence-only data, each requiring a distinct methodological approach.

Most machine learning models require both positive and negative instances to optimize their internal parameters. This makes presence-absence data the preferred formulation, as it provides explicit binary targets $y \in \{0, 1\}$ representing confirmed presence (1) and absence (0) for standard supervised optimization.

In contrast, presence-only datasets lack negative instances ($y = 0$), meaning standard supervised learning algorithms cannot inherently establish a baseline for unsuitable environments. To overcome this limitation and enable binary classification, presence-only data must be augmented through the generation of pseudo-absences. These pseudo-absences are generated by sampling environmental characteristics directly from the environmental variables $\mathcal{X}$ of the spatial domain. This sampling can be executed at random across the entire domain or guided by the specific characteristics and spatial footprint of the observation data to explicitly account for collection biases [16].

2.1.1 Citizen-science data

Producing high-quality species observation data traditionally requires professional ecological surveys, making the process both expensive and time-consuming. This constraint is particularly true for presence-absence data due to the inherent difficulty of documenting absence observations. Compared to presence observations, which only require a sighting, tracks, or droppings to confirm, absence observations require a comprehensive search to confidently determine that a species is not present at a given location. As a result, professionally surveyed observation datasets are scarce, often limited to a few specific species or restricted geographic areas. To fulfil the data demands of complex models, citizen-science data, information collected opportunistically by the general public, has seen rapidly increasing use in SDMs [17].

By leveraging crowd-sourced contributions from the public, citizen-science initiatives can generate massive, multi-species datasets. Well-known global examples include iNaturalist [9], a massive observation platform for plants and animals, and eBird [10], which focuses entirely on avian observations. Nationally focused platforms also exist, such as Swedish platform Artportalen [12].

Despite their scale, citizen-science datasets possess significant architectural drawbacks. Observations are heavily prone to spatial sampling biases, frequently clustering around highly accessible or densely populated regions, as well as taxonomic biases favouring easily spotted or charismatic species [18]. Furthermore, data quality can vary due to the logistical impossibility of professionally verifying every user submission [19].

From a modeling perspective, the most critical limitation is that platforms like iNaturalist and Artportalen are strictly presence-only. Because data collection is opportunistic, users simply log what they happen to see, the spatial gaps between sightings are left entirely unrecorded [17]. Conversely, some platforms implement structured protocols to bypass this limitation. For instance, eBird requires observers to complete exhaustive checklists of both observed and unobserved species during a session, allowing its outputs to function as true presence-absence data [10]. Failing to account for these distinct data collection modes can severely degrade model performance, particularly in undersampled regions or when modelling rare species.

2.1.2 Deep learning approaches

As described in section 1.1, deep learning has seen increased application in SDMs, leveraging multi-species modelling to better capture the complex, non-linear relationships inherent in ecological distributions [4]. Cole et al. [13] proposed spatial implicit neural representations (SINR), which model species distributions as continuous functions over geographic space using a custom DNN architecture detailed further in section 2.2.1. Their approach learns shared spatial representations across thousands of species over the entire global landmass by leveraging crowd-sourced observation data from iNaturalist [9].

To compensate for the lack of explicit absence records in the dataset, the SINR framework generates artificial pseudo-absence data dynamically during training. This generation combines random spatial sampling across the entire domain with targeted sampling at the known observation locations of other species. Because the model spans a global geographic scale, the authors operate under the assumption that the probability of a true presence and a generated pseudo-absence overlapping at the exact same location is sufficiently low. The authors investigated multiple model variants, including models that relied solely on spatial coordinates, models that incorporated environmental variables, and approaches that combined both, with models using the combined input achieving the highest performance. Their findings demonstrate that implicit neural representations can scale efficiently to global datasets while jointly modelling extensive species pools and that coordinates and environmental variables carry complementary information that can contribute to improved model performance.

While approaches like SINR rely on custom designed networks, other deep learning approaches adapt already established deep learning architectures for SDM, while also integrating high-resolution remote sensing data alongside traditional environmental variables. Teng et al. [20] introduced SatBird, a dataset and benchmarking framework designed to predict bird species encounter rates by combining satellite imagery and environmental rasters with citizen-science data from eBird [10]. Their benchmark evaluated a diverse suite of pre-existing models, including vision transformers pretrained on remote-sensing data and the CNN architecture ResNet-18 [21]. Their findings show that while raw satellite imagery didn't improve performance, combining it with environmental rasters yielded significant improvement. This demonstrates the potential of incorporating remote sensing data in the environmental variables to improve robustness and spatial coverage in SDMs.

Beyond optimizing neural architectures, resolving data-scarcity constraints remains a critical research focus. Lange et al. [22] introduced an active learning framework for species range estimation that utilizes models pre-trained on large, weakly supervised datasets like iNaturalist as initialization. Through an active sampling loop, the framework iteratively selects highly informative geographic coordinates for targeted empirical observations. This allows the model to map the distributions of previously unmapped or data-deficient species while minimizing the total number of physical field observations required, highlighting active learning as a viable strategy to enhance SDMs under severe data constraints.

2.2 Deep Neural Networks

The following section describes the structure and use of the deep neural networks used in this study and their possible benefit when using them for SDM. Each subsection will give a brief explanation of each network architecture and its benefits.

2.2.1 Spatial Implicit Neural Representation

SINR is an architecture based on a multi-layer perceptron (MLP) with residual connections, specifically tailored for spatial species distribution modelling [13]. As illustrated in Figure 2.1a, the model features a modular design consisting of a location encoder network mapped directly to a multi-label classification head.

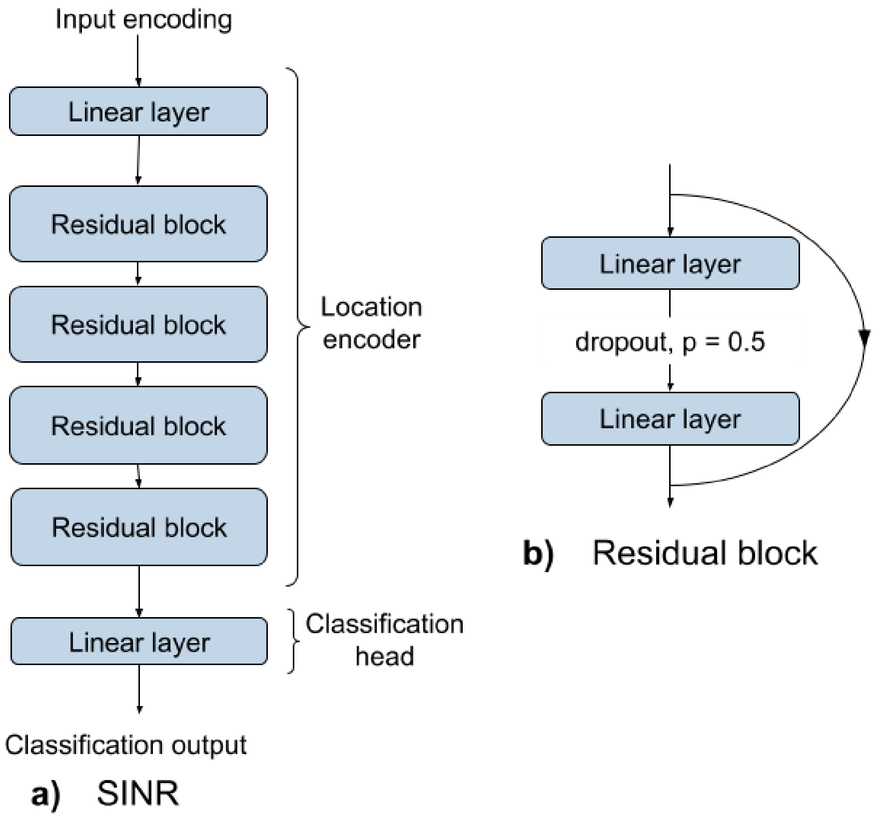


Figure 2.1: **a)** SINR model design [13] with a linear input layer, four residual blocks in its location encoding and a linear layer as classification head. **b)** The components of the residual blocks. Two linear layers with 0.5 dropout probability between them and a residual connection between block input and output.

The location encoder is constructed with a linear layer for the input connected to four sequential residual blocks. As detailed in Figure 2.1b, each individual residual block comprises two linear layers containing 256 nodes each, dropout with probability 0.5 between them and a residual connection between the input and the output of the block. The primary objective of this location encoder is to project low-dimensional spatial inputs into a non-linear, high-dimensional representation. This representation vector is subsequently passed to the classification head, a linear layer, which maps the shared feature space to individual occurrence probabilities across all targeted species. Depending on the configuration, the model accepts a sinusoidal encoding of the raw longitude and latitude coordinates of the observation, environmental variables of the location or a concatenation of the two as input.

2.2.2 ResNet-18

ResNet-18 is a specific size, with 18 layers, of the ResNet architecture [21]. All ResNet designs follow the structure in Figure 2.2a. An input layer with a 7×7 (kernel size) convolutional layer followed by 3×3 max-pooling, both with stride 2 to sequentially reduce the spatial dimensions of the feature map by half. This is followed by four blocks with a series of convolutional layers with residual connections, with layers 2, 3, and 4 downsampling the incoming feature map by half while simultaneously doubling the number of feature channels, and an output layer composed of global average pooling followed by a fully connected layer.

Within this framework, ResNet-18 specifically implements four 3×3 convolutional layers in each block with residual connections every two convolutions, illustrated in Figure 2.2b. The downsampling in layer 2, 3 and 4 is executed by setting the stride to 2 for the first convolutional layer within the block. When the feature map is downsampled the residual connection is projected using a 1×1 convolution with a stride of 2 to match both the spatial dimensions and channel depth.

Although originally engineered for image classification tasks, ResNet can be directly adapted for SDM. This adaptation requires modifying the input layer's channel capacity to match the number environmental input features and restructuring the final fully connected layer to output predictions matching the total number of target species in $\mathcal{S}$. The fundamental advantage of utilising ResNet, and CNNs in general, over point-based models like SINR lies in its capacity to preserve and leverage spatial context. Rather than evaluating a single isolated point, the network processes a spatial patch defined by an $M \times N$ pixel grid. This structural ability to capture localized spatial variations, environmental gradients, and surrounding landscape textures has been shown to enhance predictive performance in distribution tasks [11].

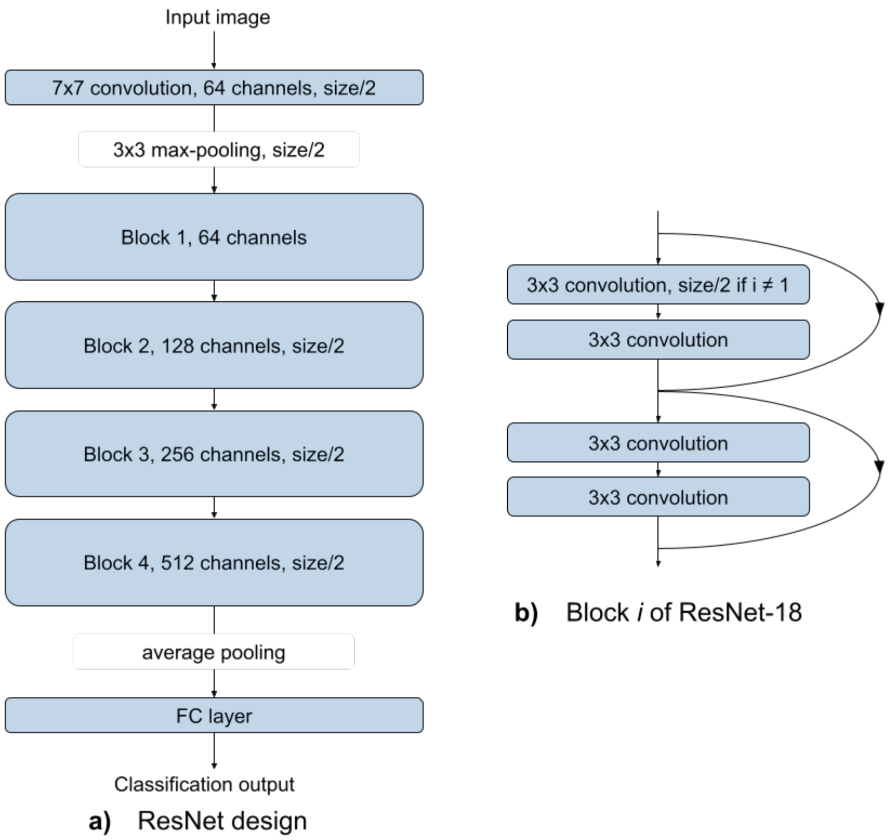


Figure 2.2: **a)** General design of ResNet [21] models with a 7×7 (kernel size) convolutional layer and 3×3 max-pooling, both with stride 2, as input layer, four blocks with a series of convolutional layers with residual connections, with block 2, 3 and 4 downsampling the feature map and doubling the channels, and an outlayer consisting of average pooling and a fully connected layer to produce the classification output. **b)** The structure of the blocks of ResNet-18 [21]. Four 3×3 convolutional layers with residual connections every two layers. For blocks 2, 3 and 4 the downsampling is done by setting stride of 2 for the first convolutional layer in the block.

2.2.3 U-Net

U-Net [23] is a fully convolutional neural network architecture characterized by its ability to output a dense spatial feature tensor rather than a single prediction vector. This architectural design enables dense, pixel-wise predictions across an entire input grid, a process mathematically framed as semantic segmentation [24].

As illustrated in Figure 2.3a, the U-Net architecture consists of a contracting path (the encoder) and an expansive path (the decoder) [23]. The encoder is composed of repetitive convolutional blocks containing two 3×3 convolutional layers. Mimicking the structural pattern of ResNet, the spatial dimension of the feature map is halved after each block while the channel depth is doubled. However, this downsampling is executed via a 2×2 max-pooling with a stride of 2. Conversely, the decoder mirrors this structure but reverses the spatial operations, progressively doubling the spatial grid size through up-convolutions (transposed convolutions) between blocks while halving the channel depth. To preserve fine-grained spatial information lost during downsampling, the architecture implements lateral skip connections that concatenate the output tensors of each encoder block directly onto the input of the corresponding decoder block. The network terminates in a final 1×1 convolutional layer that functions as a pixel-wise classification head.

The strength of the U-Net architecture is that it minimizes the trade-off between localization accuracy and the use of context. Each pixel in the output get the context of the full input map through the down- and upsampling and local relations at different resolutions from the skip connections between the encoder and decoder. The architecture was originally created for biomedical image segmentation, but U-Net has also shown good performance in geospatial applications such as land-cover classification and forest mapping [25, 26]. U-Net has not seen any extensive use in SDM of animal species, but given its success in related geospatial tasks, the architecture holds significant promise for generating high-resolution animal distribution predictions by effectively mapping subtle shifts within extracted feature maps.

2.2.4 U-Net++

U-Net++ is an extension of the standard U-Net framework that preserves the backbone encoder-decoder structure while fundamentally re-engineering the skip pathways [27]. As illustrated in Figure 2.3b, the lateral skip connections of the original U-Net are replaced with nested, dense skip pathways that incorporate intermediate convolutional blocks. This nested architecture effectively integrates a collection of multi-scale U-Net sub-models of varying depths within a single unified network structure. Furthermore, U-Net++ introduces deep supervision, an optimization technique that allows the network to operate in two distinct inference modes: an ensemble mode where the spatial predictions from all sub-models are averaged, and a pruned mode where the output of a single, computationally optimized sub-model is selected.

The aim of the U-Net++ architecture is to improve the segmentation accuracy by having the model more effectively capture fine-grained details of the target segments. The redesigned dense skip pathways is meant to reduce the semantic gap between the feature maps of the encoder and decoder to try to simplify the optimization problem, while the deep supervision enables more accurate segmentation.

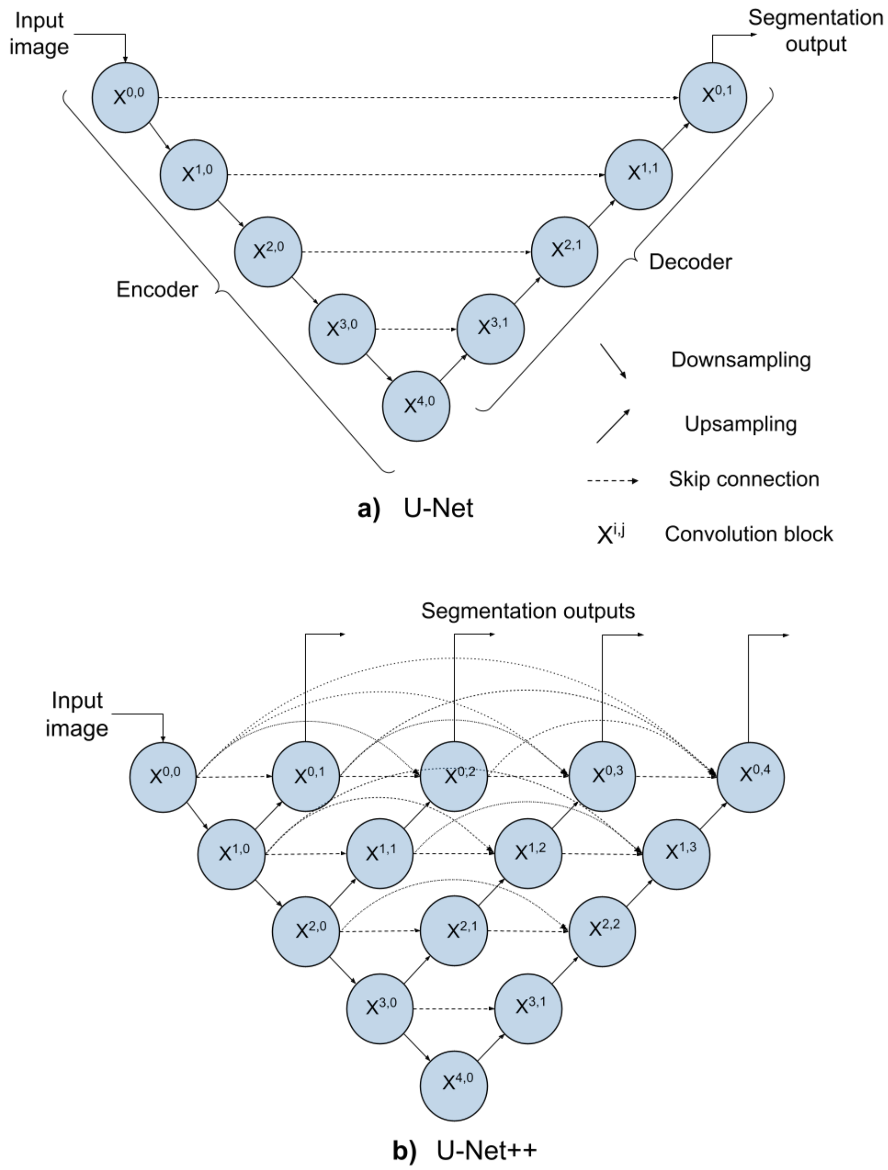


Figure 2.3: a) U-Net model design [23] with its encoder, decoder and skip connections. The convolution blocks in the encoder and decoder consists of two consecutive 3×3 (kernel size) convolutional layers, the downsampling is done using 2×2 max-pooling with stride 2 and the upsampling through up-convolutions. The skip connections concatenate the output of each level in the encoder to the input of the decoder on the same level. At the end of the decoder is a final 1×1 convolutional layer that produce the segmentation output. **b)** U-Net++ model design [27] which adds nested skip connections to the encoder-decoder structure of the U-Net model. The convolution blocks, down- and upsampling works the same as for the U-Net. Each convolution block on a level has skip connections from every preceding convolution block on the same level. A separate 1×1 convolutional layer is connected the each convolution block at the top level, except the input block, to produce segmentation outputs for all sub-models in the nested structure.

2.3 Evaluation metrics

While SDM can be evaluated as a binary classification task to determine presence and absence, the raw model outputs inhabit a continuous range $\hat{\mathbf{y}} \in [0,1]^S$. As established in section 2.1, transforming these continuous outputs into discrete states requires the application of a decision threshold θ according to the mapping defined in Equation (2.2).

The binary outputs, $\tilde{\mathbf{y}}$, allows for the compilation of a confusion matrix to compute standard threshold-dependent performance metrics, including accuracy, precision, recall, and the F1-score:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{2.3}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.4}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.5}$$

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{2.6}$$

Where TP (true positive), TN (true negative), FP (false positive), and FN (false negative) represent the correctly and incorrectly classified instances based on the chosen threshold. While a default threshold of $\theta = 0.5$ is frequently applied, it is often suboptimal for unbalanced ecological datasets, necessitating alternative threshold selection strategies [15].

Because performance metrics like accuracy and F1 are inherently tied to a single, fixed threshold θ, evaluating a model solely on one chosen threshold can obscure its true predictive capabilities. To assess a model's performance across all possible thresholds simultaneously, threshold-independent metrics such as the average precision (AP) is utilised.

AP focus on the quality of positive predictions by evaluating the continuous trade-off between precision, P, and recall, R, across different threshold levels [28]. Mathematically, it integrates the precision over the continuous recall interval:

$$AP = \int_0^1 P(R)\, dR. \tag{2.7}$$

In practice, this is approximated as the weighted mean of precisions achieved at each discrete threshold step, where the weight corresponds to the incremental increase in recall from the preceding threshold:

$$AP = \sum_n (R_n - R_{n-1}) P_n, \tag{2.8}$$

where P_n and R_n represent precision and recall at the n-th threshold, respectively.

When evaluating multi-species models that make predictions across a large set of target species $\mathcal{S}$, pooling evaluation data can cause highly abundant or easily detected species to dominate the metric. To prevent this, the mean Average Precision (mAP) is implemented to evaluate performance independently per species:

$$\text{mAP} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \text{AP}_s, \tag{2.9}$$

where AP_s represents the independent AP score calculated for species $s \in \mathcal{S}$, and $|\mathcal{S}|$ is the number of species in the set. This macro-averaging approach ensures robust validation by normalizing across species with vastly different baseline abundances and detection probabilities.

3

Methods

3.1 Geographical Representation of Sweden

To create a map of Sweden for the project we combined the country shapefile from Eurostat [29] with the administrative borders from Topografi 50 by Lantmäteriet [30]. The Eurostat shapefile contained all land masses of Sweden as separate polygons. Rather than retaining all islands, only those reachable via stepping-stone hops from the mainland were included, producing a more continuous and computationally convenient study area while excluding remote islands that are unlikely to contain meaningful observations of terrestrial mammals. Islands were included using an iterative connectivity algorithm: starting from the mainland (the largest polygon in the shapefile), any island within 5 km of an already-connected landmass was iteratively added until no further connections could be made. Narrow water gaps were then closed by applying a morphological buffer–unbuffer operation, followed by coastline simplification to produce a smooth, continuous outline. This is not expected to drastically enhance the results but is rather meant to make the geographic representation easier to work with. Lastly, the administrative border against Norway and Finland, taken from Topografi 50, was applied as a mask to clean up the land borders. The finished map was saved as a polygon in GeoPackage (.gpkg) format.

Three different grid representations of Sweden were also created for later use. These consist of square grid cells with cell sizes of 500 m $\times$ 500 m, 5 km $\times$ 5 km and 50 km $\times$ 50 km. Each grid representation was created for the full raster based on the total bounds of the Sweden polygon with indexing for which grid cells are inside the polygon.

3.2 Data

This section describes the data used in the project, covering the species observation data used for training and evaluation, the environmental predictors used as model input, and the IUCN expert range maps used for external evaluation.

3.2.1 Observation Data

Two sources of species observation data were used in this project, both being presence-only. Artportalen provided the primary training data across 34 mammal species, while Rovbase provided high-quality observations of three large carnivore species used exclusively as an independent test set.

3.2.1.1 Artportalen

The training data used has been taken from Artportalen [12]. Artportalen is a Swedish crowdsourced system for reporting and retrieving species observations. It is managed by SLU Artdatabanken in collaboration with a range of partner organizations including voluntary nature societies, county administrations and the Swedish Environmental Protection Agency (Naturvårdsverket) [31]. It is open for anyone to submit their observations and all observations are published openly except for sensitive species. Contributors are primarily knowledgeable amateurs and volunteers, alongside conservation officers and researchers, and the data is used extensively in Swedish conservation work.

Most observations were downloaded directly from Artportalen using an API provided by SLU Artdatabanken [32]. Initially, all observations between 2003-01-01 to 2024-07-31 of mammals which are permanently residing in Sweden were downloaded. The data was then filtered to only include land dwelling mammals which are not bats and that have more than 500 observations. Furthermore, species with many observations were downsampled to a maximum of 20 000 observations.

In addition to these observations, we got access to observations of three sensitive species, wolf, brown bear, and lynx. The total observational data used for training thus consists of 34 species and a total of 220 355 observations. For more information about the data used, see Appendix A.1.

3.2.1.2 Rovbase

Rovbase is a dataset containing observations of large carnivores, mainly in Sweden and Norway [33]. The observations, which varying between sightings, scat, tracks, etc., are systematically collected, confirmed by authorities and show good coverage of the Scandinavian peninsula. Thus the data can be considered to be of higher quality in comparison to citizen-science data. In this project we have access to the complete observation record of wolf, brown bear and lynx in Sweden between 2003-01-01 and 2024-07-31. Table 3.1 shows a quick summary over the number of observations in Rovbase.

When talking to experts from Naturvårdsverket working with Rovbase, it became clear that the large carnivores in Sweden are so well tracked and analysed that further distribution modelling is not necessary. The raw data in itself also gives a very accurate species distribution from just looking at the density and distribution of observations. Therefore training a model on the high quality Rovbase data is not of particular scientific interest. These qualities do however mean that the Rovbase

Table 3.1: Number of observations in Rovbase, per species and total.

Species	Observations
Lynx (*Lodjur*)	44 527
Bear (*Björn*)	40 704
Wolf (*Varg*)	37 347
Total	122 578

data is close to a ground truth for presence which makes it useful for evaluation. Thus the Rovbase data was used as a test set to see how well the model is able to learn the distribution of wolves, bears and lynx from the lower quality and more sparse Artportalen data.

3.2.2 Environmental Predictors

Two sources of environmental predictor data were used as spatial input to the models, bioclimatic variables and elevation from WorldClim, and pre-computed satellite embeddings from the AlphaEarth Foundations dataset. These capture complementary aspects of the environment the former representing climate conditions relevant to species ecology, the latter encoding rich surface properties derived from multiple remote sensing sources.

3.2.2.1 WorldClim Bioclimatic Variables and Elevation

The environmental data used is the elevation and 19 bioclimatic variables from WorldClim version 2.1 climate data at the 30 seconds resolution [8]. The bioclimatic variables are derived from monthly temperature and precipitation values to produce ecologically meaningful summaries of climate conditions. They are widely used in species distribution modelling and related ecological analyses. By extracting and combining the precipitation and temperature values in different ways for the bioclimatic variables they are able to represent annual trends, seasonality and extreme or limiting factors [34]. Specifications of all 19 bioclimatic variables can be found in table A.3.

The elevation and bioclimatic variables were downloaded as GeoTiff (.tif) files and were formatted before being used as model input. First they were cropped to an area around Sweden based on the north, south, west and east bounds of the Sweden polygon plus an additional 20 km buffer. Then nodata values were removed before normalizing to mean 0 and standard deviation 1. Lastly the data was combined into a single GeoTiff file with 20 bands, one for each of elevation and the 19 bioclimatic variables.

3.2.2.2 AlphaEarth Foundations Satellite Embeddings

The AlphaEarth Foundations (AEF) Satellite Embedding dataset is produced by Google and Google DeepMind. It consists of annual, analysis-ready embedding field layers covering Earth's terrestrial surface at a 10 m pixel resolution. Each pixel contains a compact 64-dimensional embedding vector that encodes surface properties derived from a diverse set of input sources, including optical imagery, synthetic aperture radar (SAR), LiDAR, climate reanalysis data, and elevation, all jointly processed and temporally summarised over a single calendar year. To create the embeddings, the raw collection dates of the satellite images are first converted into a standardized format. Then, a specialized model called the Space Time Precision (STP) encoder processes these images. It analyzes the spatial details (the land features), the timing (how things change over months), and the sensor measurements simultaneously. Finally, a summarizer compiles this information over a specific timeframe into the 64-dimensional embedding vector. During training, the model is penalized if it loses critical information, ensuring that this tiny digital code retains enough detail to reconstruct the original satellite observations [35].

The key advantage of using pre-computed embeddings over raw satellite imagery is that they provide a general-purpose, information-dense feature space that captures a broad range of surface properties without requiring access to or reprocessing of the original imagery for each downstream task. The embeddings are constrained to be uniformly distributed on the unit hypersphere, meaning the magnitude of each vector is normalised to Euclidean length 1 and no further normalisation is needed [35].

The data was downloaded from Google Cloud storage as Cloud Optimized GeoTIFFs (COGs), where each COG contains an image tile of 8192×8192 pixels. Each value within the embedding vector for a pixel is stored as a signed 8-bit integer in the range $[-127, 127]$ (with -128 being reserved as the no-data value), which has to be de-quantized to a float using the formula in Equation (3.1) before use.

$$V_{final} = \left(\frac{V_{raw}}{127.5}\right)^2 \times \operatorname{sign}(V_{raw})\,. \tag{3.1}$$

After de-quantization, the values are in the range $[-1, 1]$ [14].

Due to storage and memory limitations, only embeddings for a single year, 2020, were used and the resolution was downsampled from 10 m pixels to 50 m pixels. To do this the raw signed 8-bit values in the downloaded files were de-quantized to floats, the resolution was downsampled and the values were re-quantized again. To make patch extraction quicker, mappings between all presence samples and the corresponding tile and bounds were created. Each tile was also expanded by 64 pixels around the edges to allow for observations at the edge of a tile to get a full patch extraction without having to extract from multiple tiles on the fly.

3.2.3 IUCN Expert Range Maps

As mentioned in Section 1.1, expert range maps from IUCN [6] are widely accepted and used for evaluating SDMs, for example by Cole et. al. [13] and Lange et. al. [22]. As also mentioned, they are scale-dependent and therefore often miss local variations [7]. For this project, IUCN expert range maps were available for 29 of the 34 species trained on and at varying levels of detail.

3.3 Models

The models used for SDM where random forest (RF) and SINR as baselines, the three chosen CNN architectures: ResNet-18, U-Net and U-Net++, and late-fusion models combining each CNN with SINR.

3.3.1 Baseline Models

The RF classifier was used as a baseline of the performance of the traditional models and used the environmental variables as input. SINR was used as a baseline of deep learning models with single point input and used both the environmental variables and location as input.

3.3.2 ResNet-18

The ResNet-18 model was obtained from the TorchVision library [36, 37] and initialized with random weights. The fully connected layer was redefined to have 34 output neurons and connected to a sigmoid function to get the predictions of each species in the dataset. The number of input channels of the input layer were customized depending on what input features were used: 20 for the environmental input or 64 for the satellite embedding input.

3.3.3 U-Net

The U-Net model used was a slight modification to the original design with four down- and upsampling steps with the top level having 32 channels, which gives the deeper levels 64, 128, 256 and 512 channels respectively. Like the ResNet-18 model the input and output layers were modified to match the number of input features and species.

Because each input was a patch centred on a presence or pseudo-absence observation the samples only had one target value and no per-pixel target. To account for this it was assumed that a presence or absence observation in the centre of the extracted patch meant that the species was present or absent, respectively, within a set radius, r, from the observation since animals move around. This was implemented as a mask, $\Omega(r)$, with an active area defined by r as illustrated in Figure 3.1. During training, every pixel value of the segmentation output within the active area was used in the loss function with their target set to the observation type of the sample, either presence or absence, and the pixel values outside the active area were masked. When the model was used for classification only the values of the four centre pixels were averaged since that is were the actual observation was made. The radius could be set to different values for each species, but because of time constraints $r = 2$km was set for all species.

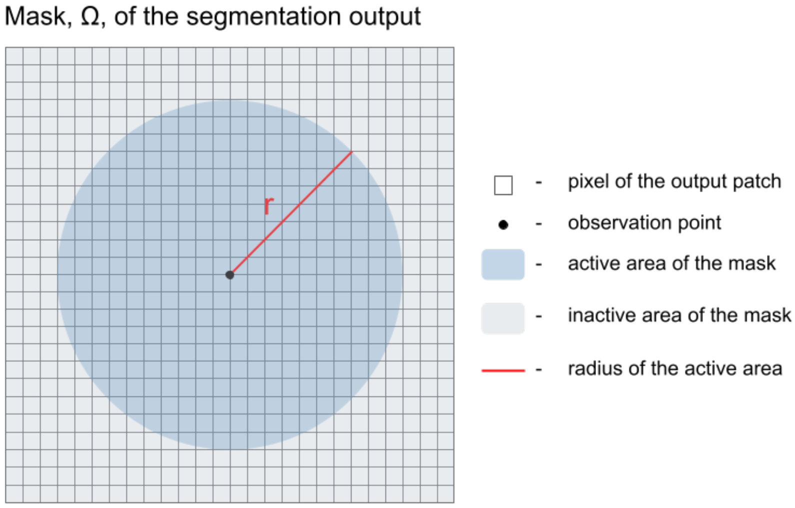


Figure 3.1: An illustration of the mask, $\Omega(r)$, applied to the segmentation output of a species to determine what pixels values were used in continued calculations and what values were masked. The active area was set by a circle with radius r around the observation point. All pixels with their centre point overlapping the active area were set as active and used in the loss function, the rest were masked. All active pixels had their target set to the observation type of the current sample.

3.3.4 U-Net++

Our U-Net++ model design follows the same modifications as the U-Net model, but with only two down- and upsampling steps with the channels per level set to 32, 64 and 128 respectively. This depth yields two sub-model outputs, with both being used for loss calculations during training, but during presence-absence classification only the output of the deepest sub-model was used.

3.3.5 Fusion Models

The fusion models used late fusion between one of the CNN models and SINR where both models ran in parallel until the classification head and then fused before making the classification. This resulted in both models having their location encoding (if we use the SINR terms) and these are then combined in a redefined classification head for the fusion model. The CNN model used either the environmental or satellite embedding input and SINR used the location input.

If ResNet-18 was used in the fusion model a modified fully connected layer was made to also take the SINR location encoding as input to make predictions for each species. If U-Net or U-Net++ was used the features in the location encoding from SINR were put through a linear layer to lower its dimension from 256 to 32, the same as the number of channels in the outlayer of the models, to save on memory usage. Each of the 32 features were expanded to a 128×128 patch to create 32 channels, with each channel having the same value in all pixels, that were concatenated with location encoding of the U-Net or U-Net++ model for a total of 64 channels. The classification head was modified to handle the increased number of input channels.

3.4 Dataset Construction

This section describes how the raw observation data was processed into training, validation and test sets ready for use in model training and evaluation. This includes the generation of pseudo-absence data, the spatial splitting strategy used to create the data subsets, and the preparation of model inputs from the different data sources.

3.4.1 Pseudo-Absence Generation

As the data is presence-only, pseudo absence data had to be generated to give the model negative input. To do this we used two different methods, distance-only and distance-density, both based on weighted sampling from the $500\,\mathrm{m} \times 500\,\mathrm{m}$ cell size grid of Sweden.

The first method, distance-only, is purely based on distance to the nearest sample. For each species, the grid cells are assigned probability weights based on distance to nearest grid cell containing an observation of the species. Pseudo-absence is then randomly sampled using these probabilities meaning that cells far from any known presence are more likely to be sampled as absences.

The second method, distance-density, is based on distance and the overall observation density for the whole dataset. The observation density acts as an indication for sampling effort and is used to counteract biases in where we have observations. A density map of the observations from Artportalen is created by assigning a value to each grid cell based on how the total observations are within a 10 km radius of it. The observations are summed with a gaussian decay using a standard deviation equal to half the sampling density radius. A maximum density cap of 200 was then applied to reduce extreme hotspots. When creating the absence weights, the density map was normalized to a $[0, 1]$ interval and multiplied with the previously mentioned distance absence weights. This means that areas that are far away from an observation of a certain species, but that is heavily surveyed otherwise, will have a high pseudo-absence probability, whereas areas that are rarely surveyed will have lower pseudo-absence probability. By generating more total pseudo-absences in highly observed areas this aims to reduce the risk of the models fitting to these areas across all species. It also lowers the risk of pseudo-absences being generated in areas where a species might exist but that is just rarely visited by people, allowing the models to generalize to these areas from other areas with more data instead.

One limitation of the distance-density method is that it does not explicitly address the presence biases in species that are present in these highly observed areas. For example, a species that in reality has equal distribution across all of southern Sweden will have more observations in certain highly surveyed areas, and since the distribution of the presences is not altered during training these biases are likely to remain in the model.

An additional issue with only using the distance-density method is that areas that actually have no or very few observations, such as big lakes and high mountains, will get very few pseudo-absences. To combat this, 10% of the pseudo-absence samples were generated using the distance-only method, with the remaining 90% using the distance-density method. This approach and ratio was decided mainly through assumptions and test on smaller datasets due to time limitations. Thus the exact ratio was not tuned further and it remains a noteworthy limitation.

3.4.2 Train, Validation and Test Sets

To create train, validation and test splits on the Artportalen data, a coarse grid over Sweden with a cell size of 50 km was used. Rather than splitting the individual observations directly, each grid cell was assigned a split, ensuring that all observations within a given cell belong to the same subset. This spatial partitioning prevents data leakage that would otherwise arise from spatially autocorrelated observations being placed in both training and test sets.

The coarse grid cells were split across 100 candidate random seeds. For each seed, the cells were first divided into a training set (70%) and a combined holdout set (30%), after which the holdout was further divided into validation (2/3 of the holdout, 20% of total) and test (1/3 of the holdout, 10% of total) subsets yielding a final 70/20/10 train/validation/test split. Note that while this was the goal split, the actual proportions between the splits are slightly different due to the constraint of having to work from the coarse grid cells. When enabled, stratification was applied at both stages. For the stratification each coarse cell was assigned a label corresponding to the dominant taxon among observations falling within it, with rare taxa (those appearing in fewer than a minimum number of cells) grouped into a single stratum to ensure sufficient class support for stratified sampling.

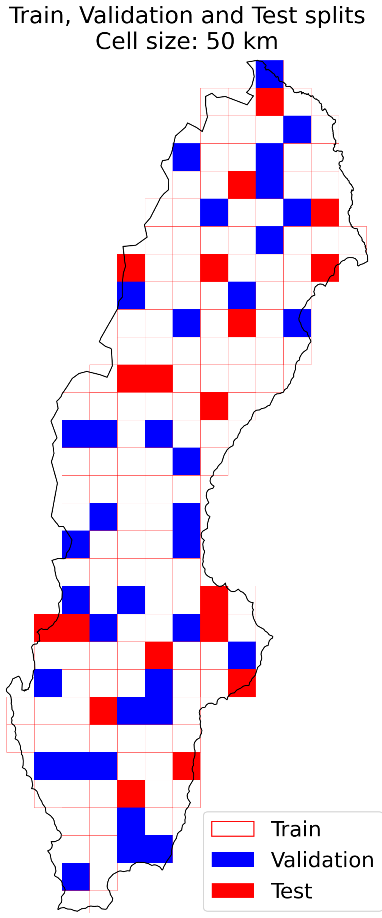


Figure 3.2: Train, validation and test split of coarse grid cells.

Each candidate split was evaluated using a combined score measuring how well the class and spatial distributions of each subset match those of the full dataset. The class balance component computes the mean absolute deviation between the distribution of dominant taxa across all cells and the distribution within each subset. The spatial balance component applies the same measure to a coarse geographic binning of cells, where Sweden is divided into a 6×6 grid of spatial bins and each cell is assigned a bin label based on its position. The two components are combined as a weighted average, with equal weight given to class and spatial balance by default. The split with the lowest total score across all candidate seeds is selected, favouring splits that are simultaneously representative in both taxonomic composition and geographic coverage.

To avoid data leakage caused by samples close to the split boundaries, all observations within a 3 km buffer zone of the split borders were removed. During training, there is also an additional check that ensures there are no missing values in the sample, removing any observations that contain NaN values for species or location. The resulting number of observations for each data set were 125 927 (65%) for train, 42 068 (22%) for validation and 26 305 (14%) for test, resulting in a total of 194 300 observations. The grid split is shown in Figure 3.2 and the full specifications of the final splits can be found in Table A.2.

In addition to these splits on the Artportalen data, a test set using the Rovbase data was generated. This was constructed from all Rovbase observations and by pre-generating pseudo-absences using the distance-only method. This was chosen as opposed to the distance-density method, or a combination of both, as the Rovbase data doesn't show the same biases as previously discussed. It should be emphasised that, because Rovbase is itself presence-only, the negative instances in this test set are generated pseudo-absences rather than confirmed absences. The Rovbase test set therefore does not constitute a true presence–absence benchmark. What distinguishes it from the Artportalen test split is not the nature of its negatives but the quality and completeness of its presence records, which are systematically collected and cover a substantially broader area of Sweden. Its primary value is thus as a test of how well a model trained on the sparse and biased Artportalen presences is able to rank and recover presence across the much wider range revealed by Rovbase. The final splits of the Rovbase species across the Artportalen and Rovbase sets are shown in Table 3.2.

To ensure consistent evaluation between different model variations and training steps, the pseudo-absences for the validation and test sets were pre-generated.

Table 3.2: Artportalen split distributions of wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

Species	**Artportalen**		**Rovbase**
	Train / Val / Test	**Total**	
Wolf (*Canis lupus*)	673 / 197 / 153	1 023	37 347
Brown bear (*Ursus arctos*)	871 / 150 / 153	1 174	40 704
Lynx (*Lynx lynx*)	1 297 / 425 / 318	2 040	44 527

3.4.3 Model Input

The CNN models, i.e. ResNet-18, U-Net and U-Net++, took input in the form of a 128×128 pixel patch with a pixel size of 50 m. The patches were taken from either the environmental data or the satellite embeddings. SINR and RF both took the environmental data at the sample location as a feature vector and SINR also took location input from the sample coordinates. The fusion models used the same patch input as the CNNs and the same location input as SINR.

3.4.3.1 Environmental Input

When the environmental data was used as input to the model, the raster was preloaded into memory at the start of training. For each sample in a batch, the 128×128 pixel patch with a pixel size of 50 m was extracted centred around the sample location. Since the resolution of the input patch was higher than the resolution of the environmental data, bilinear interpolation was used during extraction to create an accurate and continuous grid of values. The resulting model input was thus 20 128×128 pixel patches, one for each of the 19 bioclimatic variables and elevation. When using SINR or RF, the environmental data was bilinearly interpolated for a single point at the sample location instead resulting in a feature vector of length 20.

3.4.3.2 Satellite Embedding Input

When the AEF satellite embeddings was used as input to the model, the tiles referenced by the observations in the current split were preloaded into memory at the start of training or evaluation. For each sample in a batch, the 128×128 pixel patch with a pixel size of 50 m was extracted centred on the sample location. Nearest-pixel rather than bilinear extraction is used here to maximise speed, with negligible loss of accuracy given the 50 m pixel resolution. For presence samples, patch locations were looked up using the pre-computed mappings described in section 3.2.2.2. For pseudo-absence samples, the tile and patch bounds were determined on the fly. After extraction, the raw quantized values were de-quantized to floats before being fed into the model using the same de-quantization formula as before found in Equation (3.1).

3.4.3.3 Location Input

When location was used as input, the WGS84 longitude and latitude of each sample were encoded into a 4-dimensional feature vector, $\mathbf{x}_{\text{loc}}$ using a sinusoidal encoding. The coordinates were first normalised to the range $[-1, 1]$ by dividing longitude by 180° and latitude by 90°, after which sine and cosine transformations were applied to each normalised coordinate:

$$\mathbf{x}_{\text{loc}} = \left[\sin(\pi\tilde{\lambda}),\ \sin(\pi\tilde{\phi}),\ \cos(\pi\tilde{\lambda}),\ \cos(\pi\tilde{\phi})\right], \tag{3.2}$$

where $\tilde{\lambda}$ and $\tilde{\phi}$ denote the normalised longitude and latitude respectively. This encoding ensures that geographically close locations produce similar feature vectors. It also avoids discontinuities at the boundaries of the coordinate system for global models, however, this is not relevant in this project.

3.5 Training

For each model variation, the model was trained using stochastic mini-batch gradient descent with the Adam optimiser, a batch size of 128 with 64 observations from

Artportalen and 64 pseudo-absences freshly generated for each batch, and an initial learning rate of 5×10^{-4}. The learning rate was decayed after each epoch using an exponential learning rate schedule with a decay factor of $\gamma = 0.98$. During training one can set a random seed for pseudo-absence generation to reproduce runs exactly however this was not done for this project. The final models were trained once for five epochs. Repeated runs were not able to be performed due to time limitations.

The pseudo-absences in each batch were generated for the same species as the Artportalen observations in the batch, meaning that each presence and pseudo-absence pair in the batch were of the same species.

3.5.1 Loss Function

The loss function used for all neural network models was a slightly modified binary cross-entropy loss, but because of the different outputs of the models, the loss has small differences between architectures. For a given batch, let $b \in \{1, \ldots, N\}$ index the presence and pseudo-absence sample pairs, where N is the number of sample pairs. Each sample pair b is associated with a specific observed species, denoted by the index $s_b \in \mathcal{S}$, where $\mathcal{S}$ is the set of observed species.

SINR and ResNet-18 outputs a single probability vector per sample across all species. For a batch, let $\hat{\mathbf{y}}_{1:N}$ and $\hat{\mathbf{y}}'_{1:N}$ denote the sequences of predictions for presence and pseudo-absence samples, respectively. Their loss function for single-value outputs, $\mathcal{L}_{\text{svo}}$, evaluates the prediction for the selected species s_b across the entire batch:

$$\mathcal{L}_{\text{svo}}(\hat{\mathbf{y}}_{1:N}, \hat{\mathbf{y}}'_{1:N}) = -\frac{1}{(w_P + w_A)N} \sum_{b=1}^{N} \left[w_P \ln(\hat{y}_{b,s_b}) + w_A \ln(1 - \hat{y}'_{b,s_b}) \right], \tag{3.3}$$

where $\hat{y}_{b,s_b}$ and $\hat{y}'_{b,s_b}$ are the predicted probabilities for species s_b of the presence and pseudo-absence samples in b. The parameters w_P and w_A are weights for presence and pseudo-absence that decide their relative importance in the loss function.

The U-Net model outputs a spatial segmentation patch of size 128×128 for each species. For a batch, let $\hat{\mathbf{Y}}_{1:N}$ and $\hat{\mathbf{Y}}'_{1:N}$ denote the sequences of predictions for presence and pseudo-absence samples, respectively. It uses the loss function for masked patch outputs, $\mathcal{L}_{\text{mpo}}$, which selects the patch of species s_b for each sample and evaluates only the pixels within the mask $\Omega(r)$:

$$\begin{aligned}\mathcal{L}_{\text{mpo}}(\hat{\mathbf{Y}}_{1:N}, \hat{\mathbf{Y}}'_{1:N}) = -\frac{1}{(w_P + w_A)N|\Omega(r)|} \sum_{b=1}^{N} \sum_{i,j \in \Omega(r)} \Big[& w_P \ln(\hat{Y}_{b,s_b,i,j}) + \\ & + w_A \ln(1 - \hat{Y}'_{b,s_b,i,j}) \Big]\end{aligned} \tag{3.4}$$

where $\hat{Y}_{b,s_b,i,j}$ and $\hat{Y}'_{b,s_b,i,j}$ are the predicted probabilities for species s_b in pixel (i, j) of the presence and pseudo-absence samples in b. $|\Omega(r)|$ represents the total number of tensor elements (pixels) contained within the mask $\Omega(r)$.

U-Net++, because of its deep supervision, used the loss function $\mathcal{L}_{\text{ds}}$ in Equation (3.5), which takes the average of the losses $\mathcal{L}_{\text{mpo}}(\hat{\mathbf{Y}}^k_{1:N}, \hat{\mathbf{Y}}'^k_{1:N})$ from the outputs of all sub-models $k \in \{1, \ldots, K\}$ of the U-Net++ model, with $K = 2$ for the chosen U-Net++ design:

$$\mathcal{L}_{\text{ds}} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}_{\text{mpo}}(\hat{\mathbf{Y}}^{k}_{1:N}, \hat{\mathbf{Y}}'^{k}_{1:N}) \tag{3.5}$$

The presence and pseudo-absence weights were set to $w_P = 1.25$ and $w_A = 1$ for all models. This was done to prevent overfitting towards the pseudo-absences that occurred when using equal weights.

3.5.2 Validation

During the training process the model was validated using holdout validation. This was selected as opposed to k-fold cross-validation mainly due to training time constraints. Validation was run at set intervals during and at the end of each epoch. As mentioned in section 3.4.2, the pseudo-absences for the validation set were pregenerated meaning all validation was done on the exact same samples. At each validation point, the full validation set was evaluated using the same loss function as during training. The only difference was that since validation doesn't require any model weight gradients to be calculated, a larger batch size was used. To further ensure consistent validation the order of the samples was also kept the same for each validation point. The model version with the lowest validation loss was stored and used for further evaluation on the test set.

3.5.3 Threshold Selection

In order to perform binary classification for further evaluation, a per-species threshold was calculated for each species using the Lowest Presence Threshold - Robust (LPT-R) method proposed by Dorm et al. [38] for generating binary species range maps. The method collects the continuous model predictions for all presence observations of a species in the training set and sets the threshold at the fifth percentile of these values, meaning that 95% of training presence observations will receive a prediction score above the threshold. An advantage of this method is that it requires no additional pseudo-absence data making it easy to implement. It is worth noting that while LPT-R was shown to be the best-performing thresholding method for binary range map generation, it may not be optimal for individual sample classification, which is what we used it for here. However, it provides a consistent, reliable, and efficient way of setting per-species classification thresholds for further evaluation. During this process, the computed thresholds were also used to generate predictions on the training and validation sets in order to produce confusion matrices and performance metrics.

For RF the standard Lowest Presence Threshold (LPT) method, which selects the single lowest prediction value for the presences in the train set, was used. The reason for using this instead of the LPT-R method was that the work done by Dorm et al. only evaluated the methods on deep-learning models, more specifically different versions of SINR [38]. LPT on the other hand was proposed as a thresholding method for traditional SDMs [39] and we also confirmed that it was more effective for RF when evaluating on the validation set.

3.5.4 Hyperparameter Selection

The hyperparameters for the models were selected in different ways. The key parameters explored were presence loss weight (w_P), patch size and pixel size. These were tuned using parameter sweeps and grid searches using the validation method previously described. These tuning runs were run for three epochs each.

Other parameters such as species radius (r) and parameters for pseudo-absence were mainly estimated and tested on smaller versions of the dataset due to limitations in computational capacity and time.

3.6 Evaluation

After training, the best model, i.e. the one with the lowest validation loss, was evaluated on the two test sets, the test split from Artportalen and the Rovbase test set. The per-species thresholds computed during training were applied to generate binary predictions, from which confusion matrices and performance metrics were derived. The reported metrics include accuracy, precision, recall, and F1 score, computed both in total across all samples and per species individually. AP was also computed per species to provide a threshold-independent measure of performance and combined into mAP for an overall performance score of the models.

Additionally, distribution maps were generated for each species using the $5\,\text{km} \times 5\,\text{km}$ grid over Sweden, with the prediction made for the centre point of the grid cell, visualising the model's predicted presence probability across the country. The distribution maps were evaluated using the IUCN expert range maps as ground truth. Due to the discrepancies in resolution between the range maps for different species and the models, they were used to evaluate broader prediction ability across the country and compared primarily on AP and mAP. This gives a measure of general presence ranking ability across larger areas which in this case is more useful than thresholded ranges since the models are expected to produce higher detail local predictions that are not able to be evaluated using the IUCN expert range maps.

4

Results

This chapter presents the test results for each model variation on each test set. This is done aggregated over all species in the respective test sets and for certain selected species. For examples of generated prediction maps and more complete results, see Appendices B and C. When referencing a model in the text they are named according to their model architecture with the input as subscript when applicable. The fusion models between SINR and the CNN models are named Fus-Res (ResNet-18), Fus-U (U-Net) and Fus-U++ (U-Net++) and the input subscripts are Env for environmental data, Sat for satellite embeddings and Loc for location encoding. For example U-Net with environmental input is referred to as U-Net$_{\text{Env}}$ and fusion model with ResNet-18 and location and satellite embedding inputs is referred to as Fus-Res$_{\text{LocSat}}$.

4.1 Aggregated Results

The aggregated results over all 34 species in the Artportalen data and the three species in the Rovbase data are shown in Table 4.1. Accuracy and F1 are computed from the total number of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) where each of the samples has been predicted using the species specific thresholds. The mean average precision (mAP) is calculated as the mean of each species average precision (AP) without weighting based on class size.

The Artportalen test set results in Table 4.1 shows that the models with satellite embedding input (Sat and Loc+Sat) score the highest across all metrics. Fus-U$_{\text{LocSat}}$ has the highest scores overall with the single highest accuracy (0.918), F1 (0.921) and mAP (0.975). Fus-Res$_{\text{LocSat}}$ is second overall being second on accuracy (0.906) and mAP (0.972) and third on F1 (0.906). Of the pure Sat CNNs, U-Net$_{\text{Sat}}$ scores highest on F1 (0.907) while ResNet-18$_{\text{Sat}}$ leads on mAP (0.965). Fus-U++$_{\text{LocSat}}$ scores between the pure Sat CNNs and Fus-Res$_{\text{LocSat}}$. Comparing the environmental input (Env and Loc+Env) and baseline models, there are no clearly better performing models between SINR, the Env CNNs and the Env fusion models. The Env fusion models show modest improvements over their unfused counterparts for accuracy and F1 but slightly worse results for mAP on average. U-Net$_{\text{Env}}$ has the highest scores among the Env CNNs, followed by ResNet-18$_{\text{Env}}$ and U-Net++$_{\text{Env}}$. RF scores lower for accuracy and F1 but has comparable mAP to the environmental input CNNs and SINR.

Table 4.1: Aggregated test set results for the Artportalen and Rovbase datasets. The results are aggregated separately for each dataset over all species in the respective dataset. Accuracy and F1 are computed from the total number of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) where each of the samples has been predicted using the species specific thresholds. The mean average precision (mAP) is calculated as the mean of each species average precision (AP) without weighting based on class size. The highest score for each metric and dataset is displayed in boldface.

	Model	**Input**	**Accuracy**	**F1**	**mAP**
Artportalen	RF	Env	0.777	0.730	0.910
	SINR	Loc+Env	0.838	0.853	0.919
	ResNet-18	Env	0.832	0.841	0.924
	U-Net	Env	0.836	0.851	0.934
	U-Net++	Env	0.827	0.840	0.917
	ResNet-18	Sat	0.900	0.900	0.965
	U-Net	Sat	0.904	0.907	0.961
	U-Net++	Sat	0.876	0.883	0.960
	Fus-Res	Loc+Env	0.835	0.847	0.922
	Fus-U	Loc+Env	0.844	0.857	0.925
	Fus-U++	Loc+Env	0.839	0.853	0.923
	Fus-Res	Loc+Sat	0.906	0.906	0.972
	Fus-U	Loc+Sat	**0.918**	**0.921**	**0.975**
	Fus-U++	Loc+Sat	0.888	0.894	0.964
Rovbase	RF	Env	0.826	0.826	0.892
	SINR	Loc+Env	0.827	0.838	0.877
	ResNet-18	Env	0.817	0.826	0.875
	U-Net	Env	0.831	0.843	0.841
	U-Net++	Env	0.819	0.834	0.848
	ResNet-18	Sat	0.855	0.859	0.914
	U-Net	Sat	0.862	0.869	0.914
	U-Net++	Sat	0.844	0.851	0.900
	Fus-Res	Loc+Env	0.830	0.840	0.899
	Fus-U	Loc+Env	0.813	0.827	0.854
	Fus-U++	Loc+Env	0.823	0.837	0.865
	Fus-Res	Loc+Sat	**0.866**	**0.870**	**0.923**
	Fus-U	Loc+Sat	0.841	0.848	0.904
	Fus-U++	Loc+Sat	0.846	0.854	0.894

The Rovbase test results in Table 4.1 show a similar overall ranking between models but the performance gap between the satellite embedding and environmental input models narrows. Notably, Fus-U$_{\text{LocSat}}$ has a large performance drop and Fus-Res$_{\text{LocSat}}$ is now top performer across all metrics (accuracy 0.866, F1 0.870, mAP 0.923). Fus-U$_{\text{LocSat}}$ and Fus-U++$_{\text{LocSat}}$ score within the Sat CNN range on accuracy and F1 but their mAP values (0.904 and 0.894) are lower, suggesting the U-Net fusion benefit does not transfer as well to the Rovbase distribution. U-Net$_{\text{Sat}}$ and ResNet-18$_{\text{Sat}}$ remain close in performance, and U-Net++$_{\text{Sat}}$ still scores lower than both. Looking at the rest of the models and baseline there are some notable changes. Between SINR, the environmental input CNNs and Fus-Res$_{\text{LocEnv}}$, Fus-Res$_{\text{LocEnv}}$ now has the highest mAP followed by ResNet-18$_{\text{Env}}$ and SINR who have similar scores. Fus-U$_{\text{LocEnv}}$ and Fus-U++$_{\text{LocEnv}}$ show lower mAP scores than their unfused counterparts, suggesting the fusion does not improve generalisation for environmental inputs on this dataset. U-Net$_{\text{Env}}$ and U-Net++$_{\text{Env}}$ have similar but notably lower scores in mAP. Most notable however is RF which shows improved accuracy and F1 and the lowest drop in mAP. It now has higher mAP than the Env CNNs and SINR and similar accuracy and F1.

Across both test sets the satellite embedding input models score consistently higher while the interarchitecture differences are small. Fusion models lead to slight improvement on average with the best model overall being Fus-Res$_{\text{LocSat}}$.

4.2 Species Specific Results for Rovbase Species

This section breaks down the species specific results for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*) over the Artportalen and Rovbase test sets. The results are related to the differences in distribution between the sparse and biased Artportalen and the more accurate Rovbase distributions to see how well the models are able to learn shifts in the distribution.

4.2.1 Wolf Results

Table 4.2 shows the species specific results for wolf (*Canis lupus*) and includes the species specific decision threshold, accuracy, F1 and AP over both sets. Figure 4.1 shows the distribution of observations across Sweden for Artportalen and Rovbase on the 5 km grid.

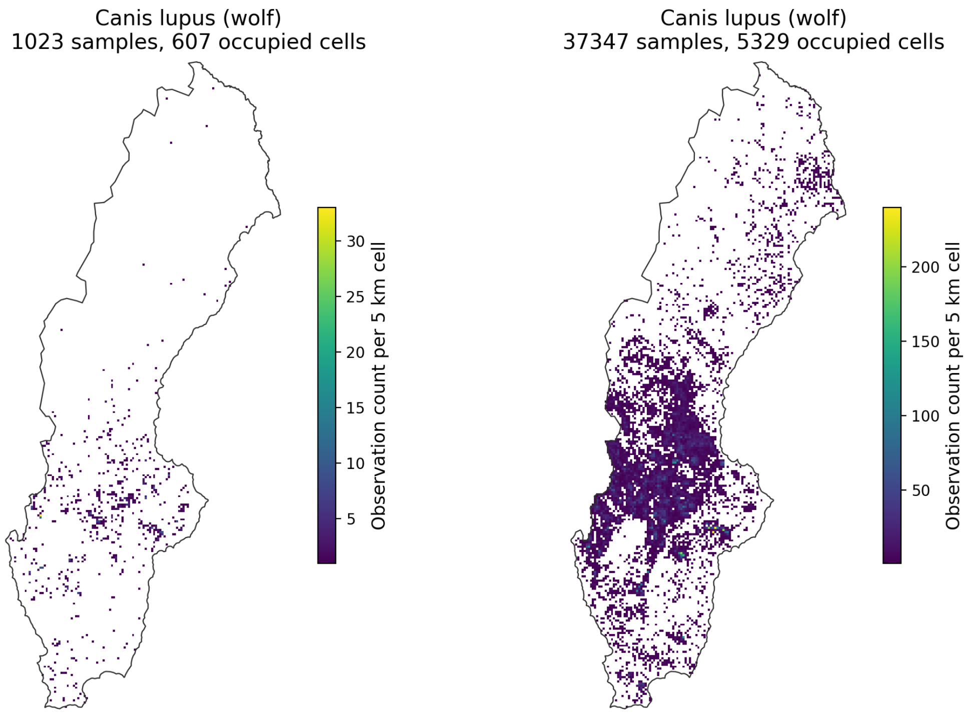
The test results for wolves in Table 4.2 show that all models score high across all metrics for both Artportalen and Rovbase. On Artportalen the AP is close for all models except RF, ranging from 0.991 to 0.996, with ResNet-18$_{\text{Sat}}$ achieving the highest AP (0.996). RF has slightly lower AP (0.969) and the lowest accuracy and F1 by a significant margin (0.826 and 0.790). Accuracy and F1 show slightly more variance between the deep learning models. The Sat CNNs, Fus-U$_{\text{LocEnv}}$, Fus-U$_{\text{LocSat}}$, Fus-U++$_{\text{LocEnv}}$, Fus-U++$_{\text{LocSat}}$, and SINR all share the highest score of 0.987 for both metrics. The rest of the deep learning models are close behind except for U-Net++$_{\text{Env}}$ (0.938 and 0.941) and Fus-Res$_{\text{LocSat}}$ (0.954 for both metrics).

Table 4.2: Test set results for wolf (*Canis lupus*) on the Artportalen and Rovbase datasets. Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). AP is the average precision for the individual species. The highest score for each metric and dataset is displayed in boldface.

Wolf (*Canis lupus*)

	Model	Input	Thr.	Acc.	F1	AP
Artportalen	RF	Env	0.540	0.826	0.790	0.969
	SINR	Loc+Env	0.432	**0.987**	**0.987**	0.993
	ResNet-18	Env	0.531	0.971	0.970	0.993
	U-Net	Env	0.486	0.984	0.984	0.993
	U-Net++	Env	0.451	0.938	0.941	0.994
	ResNet-18	Sat	0.709	**0.987**	**0.987**	**0.996**
	U-Net	Sat	0.415	**0.987**	**0.987**	0.994
	U-Net++	Sat	0.436	**0.987**	**0.987**	0.992
	Fus-Res	Loc+Env	0.495	0.984	0.984	0.994
	Fus-U	Loc+Env	0.472	**0.987**	**0.987**	0.994
	Fus-U++	Loc+Env	0.457	**0.987**	**0.987**	0.994
	Fus-Res	Loc+Sat	0.523	0.954	0.954	0.991
	Fus-U	Loc+Sat	0.263	**0.987**	**0.987**	0.994
	Fus-U++	Loc+Sat	0.429	**0.987**	**0.987**	0.993
Rovbase	RF	Env	0.540	0.853	0.852	0.930
	SINR	Loc+Env	0.432	0.867	0.872	0.931
	ResNet-18	Env	0.531	0.859	0.866	0.937
	U-Net	Env	0.486	0.864	0.870	0.899
	U-Net++	Env	0.451	0.852	0.862	0.915
	ResNet-18	Sat	0.709	0.864	0.865	0.947
	U-Net	Sat	0.415	**0.887**	**0.890**	0.927
	U-Net++	Sat	0.436	0.875	0.877	0.936
	Fus-Res	Loc+Env	0.495	0.881	0.885	0.939
	Fus-U	Loc+Env	0.472	0.870	0.878	0.912
	Fus-U++	Loc+Env	0.457	0.870	0.879	0.907
	Fus-Res	Loc+Sat	0.523	0.880	0.883	**0.958**
	Fus-U	Loc+Sat	0.263	0.865	0.867	0.885
	Fus-U++	Loc+Sat	0.429	0.869	0.873	0.906

On Rovbase the differences in model performance are clearer. The Sat CNNs score slightly higher than the Env CNNs overall and we see slight improvements when going to the respective fusion models on average. U-Net$_{\text{Sat}}$ has the highest accuracy (0.887) and F1 (0.890) while Fus-Res$_{\text{LocSat}}$ has the highest AP (0.958). The Fus-U and Fus-U++ models score in the range of the standard Env and Sat CNNs, with Fus-U$_{\text{LocEnv}}$ achieving a notably higher AP (0.912) than its base U-Net$_{\text{Env}}$ counterpart (0.899), whereas the Fus-U$_{\text{LocSat}}$ AP (0.885) is lower than U-Net$_{\text{Sat}}$ (0.927). RF and SINR both now have scores comparative to or better than the Env CNNs, with SINR reaching 0.867 accuracy and 0.872 F1. U-Net$_{\text{Env}}$ stands out with a notably lower AP (0.899) than the other Env CNNs.



(a) Artportalen distribution. **(b)** Rovbase distribution.

Figure 4.1: Distribution of *Canis lupus* (wolf) over Sweden using the 5 km × 5 km grid for the **a)** Artportalen and **b)** Rovbase datasets. The title of each map states the number of observation samples of wolf in the dataset and the number grid cells the observations occupy.

The distributions in Figure 4.1 show that the datasets share similar spatial distributions with the highest observation density in the middle of Sweden, but that Rovbase has much more complete coverage of these areas. The main difference besides the number of observations is that the high density area stretches further north for Rovbase. This suggests that all models are capable of generalizing the presence area for slight shifts in distribution but that models with satellite embedding input does this slightly better.

4.2.2 Brown Bear Results

Table 4.3 shows the species specific results for brown bear (*Ursus arctos*) and includes the species specific decision threshold, accuracy, F1 and AP over both sets. Figure 4.2 shows the distribution of observations across Sweden for Artportalen and Rovbase on the 5 km grid.

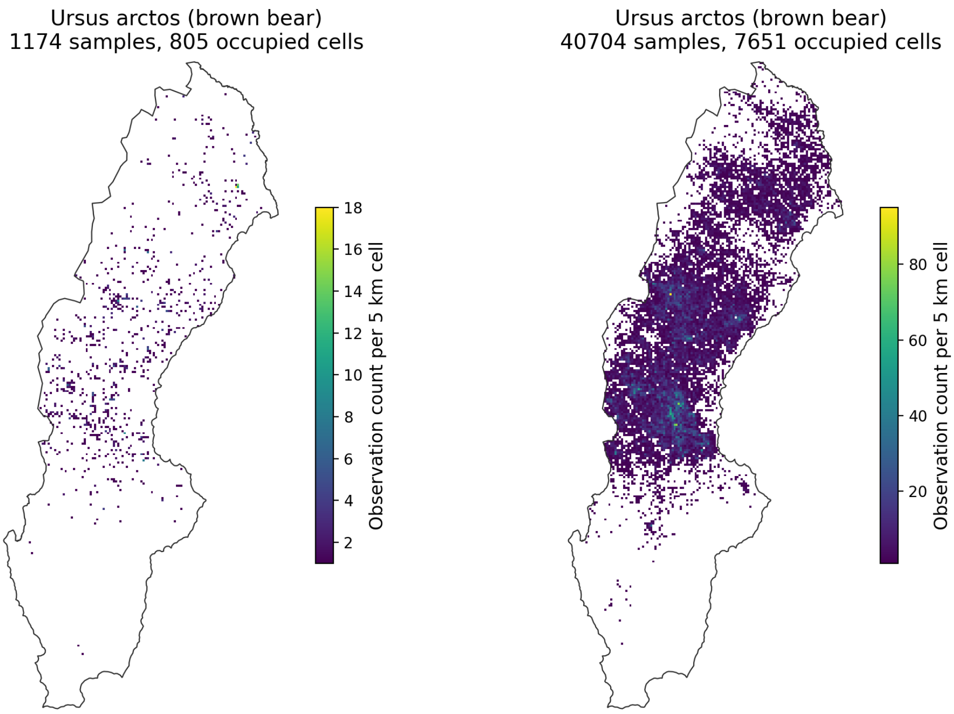
Table 4.3: Test set results for brown bear (*Ursus arctos*) on the Artportalen and Rovbase datasets. Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). AP is the average precision for the individual species. The highest score for each metric and dataset is displayed in boldface.

Brown bear (*Ursus arctos*)

	Model	Input	Thr.	Acc.	F1	AP
Artportalen	RF	Env	0.570	0.840	0.861	0.957
	SINR	Loc+Env	0.750	0.843	0.855	0.982
	ResNet-18	Env	0.759	0.944	0.945	0.955
	U-Net	Env	0.607	0.879	0.886	0.978
	U-Net++	Env	0.619	0.850	0.862	0.978
	ResNet-18	Sat	0.700	0.944	0.944	0.984
	U-Net	Sat	0.741	0.915	0.917	0.984
	U-Net++	Sat	0.646	**0.967**	**0.967**	**0.990**
	Fus-Res	Loc+Env	0.738	0.863	0.873	0.899
	Fus-U	Loc+Env	0.712	0.840	0.855	0.918
	Fus-U++	Loc+Env	0.644	0.840	0.855	0.982
	Fus-Res	Loc+Sat	0.814	0.958	0.957	0.978
	Fus-U	Loc+Sat	0.578	0.863	0.863	0.958
	Fus-U++	Loc+Sat	0.817	0.938	0.938	**0.990**
Rovbase	RF	Env	0.570	0.931	0.936	0.964
	SINR	Loc+Env	0.750	0.928	0.931	0.951
	ResNet-18	Env	0.759	0.926	0.928	0.951
	U-Net	Env	0.607	0.924	0.928	0.878
	U-Net++	Env	0.619	0.926	0.929	0.946
	ResNet-18	Sat	0.700	0.935	0.937	0.968
	U-Net	Sat	0.741	**0.940**	**0.942**	0.942
	U-Net++	Sat	0.646	0.929	0.932	0.932
	Fus-Res	Loc+Env	0.738	0.926	0.931	0.964
	Fus-U	Loc+Env	0.712	0.917	0.922	0.934
	Fus-U++	Loc+Env	0.644	0.919	0.924	0.964
	Fus-Res	Loc+Sat	0.814	0.934	0.936	0.961
	Fus-U	Loc+Sat	0.578	0.925	0.926	**0.972**
	Fus-U++	Loc+Sat	0.817	0.931	0.934	0.955

The test results for brown bears in Table 4.3 show that on Artportalen, U-Net++$_{\mathrm{Sat}}$ scores the highest accuracy and F1 (both 0.967), and tied highest AP (0.990) with Fus-U++$_{\mathrm{LocSat}}$. Fus-U++$_{\mathrm{LocSat}}$ achieves a high accuracy and F1 of 0.938, placing it second after U-Net++$_{\mathrm{Sat}}$. Fus-Res$_{\mathrm{LocSat}}$ has the third highest accuracy and F1 (0.958 and 0.957) while ResNet-18$_{\mathrm{Sat}}$ and U-Net$_{\mathrm{Sat}}$ are tied at 0.984 AP. Among the Env fusion models, Fus-U++$_{\mathrm{LocEnv}}$ achieves a notably high AP (0.982), matching U-Net++$_{\mathrm{Env}}$ but at slightly lower accuracy (0.840 vs 0.850). Fus-Res$_{\mathrm{LocEnv}}$ still has the lowest AP overall at 0.899.

On Rovbase the differences between models narrow considerably with all models scoring between 0.917 and 0.940 for accuracy and 0.922 and 0.942 for F1. U-Net$_{\mathrm{Sat}}$ has the highest accuracy (0.940) and F1 (0.942) while Fus-U$_{\mathrm{LocSat}}$ has the highest AP (0.972). ResNet-18$_{\mathrm{Sat}}$ achieves the second highest AP (0.968) and RF shares the third position (0.964) with Fus-Res$_{\mathrm{LocEnv}}$ and Fus-U++$_{\mathrm{LocEnv}}$. U-Net$_{\mathrm{Env}}$ has the lowest AP of all models at 0.878, considerably below the rest. The scores are also generally higher than the wolf scores on Rovbase despite wolf having higher scores on Artportalen.



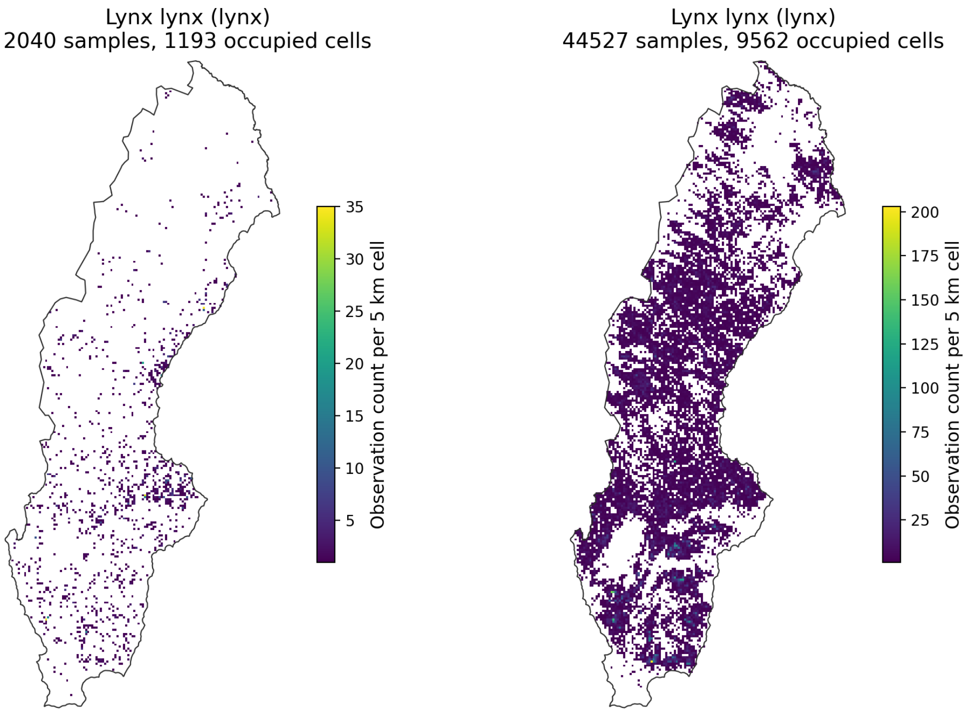
(a) Artportalen distribution. **(b)** Rovbase distribution.

Figure 4.2: Distribution of *Ursus arctos* (brown bear) over Sweden using the 5 km × 5 km grid for the **a)** Artportalen and **b)** Rovbase datasets. The title of each map states the number of observation samples of brown bear in the dataset and the number grid cells the observations occupy.

The distributions in Figure 4.2 show that the spatial distributions are nearly identical with the only significant difference being that Rovbase once again shows much more complete coverage. Relating this to the fact that the performance gap narrows when moving from the Artportalen test set to Rovbase, and that the Rovbase performance is better on brown bear than wolf, it suggests that when the spatial shift in distribution is low all models are able to accurately learn the species range and the advantage of satellite embeddings lessens.

4.2.3 Lynx Results

Table 4.4 shows the species specific results for lynx (*Lynx lynx*) and includes the species specific decision threshold, accuracy, F1 and AP over both sets. Figure 4.3 shows the distribution of observations across Sweden for Artportalen and Rovbase on the 5 km grid.



(a) Artportalen distribution. **(b)** Rovbase distribution.

Figure 4.3: Distribution of *Lynx lynx* (lynx) over Sweden using the 5 km × 5 km grid for the **a)** Artportalen and **b)** Rovbase datasets. The title of each map states the number of observation samples of lynx in the dataset and the number grid cells the observations occupy.

The test results for lynx are shown in Table 4.4. On Artportalen, RF achieves the highest accuracy and F1 (both 0.928). The AP ranking is led by Fus-U$_{\mathrm{LocSat}}$ (0.978), well ahead of SINR (0.969) and RF (0.967). Among the remaining CNN models, Fus-Res$_{\mathrm{LocEnv}}$ (0.958) and Fus-Res$_{\mathrm{LocSat}}$ (0.956) have the next highest APs, followed by ResNet-18$_{\mathrm{Sat}}$ (0.955). U-Net$_{\mathrm{Sat}}$ leads the unfused CNNs on accuracy (0.906) and F1 (0.913) but has a lower AP (0.943). The Env fusion models Fus-U$_{\mathrm{LocEnv}}$ (0.881) and Fus-U++$_{\mathrm{LocEnv}}$ (0.889) are the weakest fusion variants on AP, sitting just above U-Net++$_{\mathrm{Env}}$ (0.832) which has the lowest AP of any model.

On Rovbase all scores drop substantially across all models, and the satellite embedding models clearly separate from the rest. U-Net$_{\mathrm{Sat}}$ has the highest AP (0.874) while Fus-Res$_{\mathrm{LocSat}}$ leads on accuracy (0.793) and F1 (0.799). Fus-U$_{\mathrm{LocSat}}$ (AP 0.854) ranks second on AP, ahead of Fus-Res$_{\text{LocSat}}$ (0.851), while Fus-U++$_{\text{LocSat}}$ (0.821) and U-Net++$_{\text{Sat}}$ (0.832) score somewhat lower. RF and SINR drop sharply below the Sat models on all metrics. The Env CNNs and all three Env fusion models form the weakest group, all scoring below 0.80 on AP, with Fus-U$_{\text{LocEnv}}$ (AP 0.715) and Fus-U++$_{\text{LocEnv}}$ (AP 0.723) the lowest overall.

Table 4.4: Test set results for lynx (*Lynx lynx*) on the Artportalen and Rovbase datasets. Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). AP is the average precision for the individual species. The highest score for each metric and dataset is displayed in boldface.

Lynx (*Lynx lynx*)

	Model	**Input**	**Thr.**	**Acc.**	**F1**	**AP**
Artportalen	RF	Env	0.490	**0.928**	**0.928**	0.967
	SINR	Loc+Env	0.370	0.891	0.901	0.969
	ResNet-18	Env	0.338	0.857	0.868	0.918
	U-Net	Env	0.341	0.859	0.872	0.934
	U-Net++	Env	0.271	0.868	0.880	0.832
	ResNet-18	Sat	0.349	0.868	0.869	0.955
	U-Net	Sat	0.416	0.906	0.913	0.943
	U-Net++	Sat	0.258	0.890	0.898	0.934
	Fus-Res	Loc+Env	0.273	0.873	0.882	0.958
	Fus-U	Loc+Env	0.318	0.840	0.857	0.881
	Fus-U++	Loc+Env	0.268	0.865	0.878	0.889
	Fus-Res	Loc+Sat	0.247	0.876	0.879	0.956
	Fus-U	Loc+Sat	0.169	0.890	0.899	**0.978**
	Fus-U++	Loc+Sat	0.377	0.871	0.881	0.924
Rovbase	RF	Env	0.490	0.707	0.691	0.783
	SINR	Loc+Env	0.370	0.701	0.732	0.748
	ResNet-18	Env	0.338	0.682	0.703	0.737
	U-Net	Env	0.341	0.720	0.750	0.747
	U-Net++	Env	0.271	0.695	0.732	0.684
	ResNet-18	Sat	0.349	0.775	0.783	0.827
	U-Net	Sat	0.416	0.770	0.791	**0.874**
	U-Net++	Sat	0.258	0.741	0.761	0.832
	Fus-Res	Loc+Env	0.273	0.698	0.723	0.794
	Fus-U	Loc+Env	0.318	0.670	0.703	0.715
	Fus-U++	Loc+Env	0.268	0.695	0.730	0.723
	Fus-Res	Loc+Sat	0.247	**0.793**	**0.799**	0.851
	Fus-U	Loc+Sat	0.169	0.745	0.767	0.854
	Fus-U++	Loc+Sat	0.377	0.749	0.769	0.821

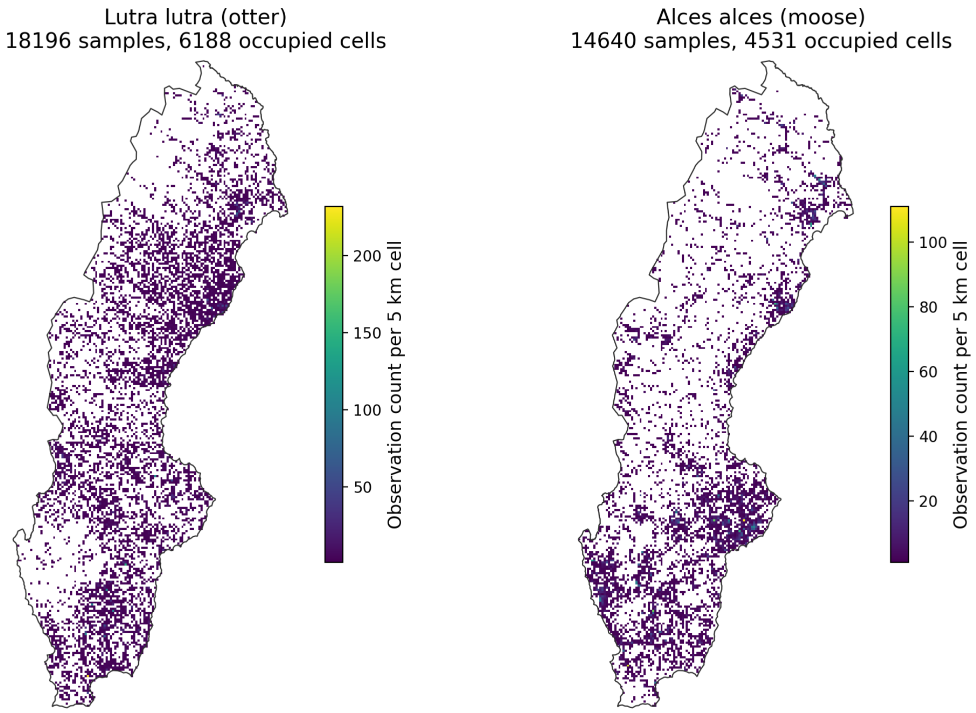
The distributions in Figure 4.3 show a larger distribution shift between Artportalen and Rovbase compared to for wolf and brown bear. The distributions are similar in the south but Artportalen severely lacks observations in the north. The distributions are also more spread out across Sweden with less coherent high density areas. Relating this to the large performance drops across all models and the fact that the Sat models retain better performance for Rovbase, it again suggests that satellite embeddings increase the models ability to generalize when the distribution shifts.

4.3 Selected Species with Notable Results

This section examines a small number of species selected to illustrate how the relationship between observation count and spatial distribution affects model performance. Unlike the previous analysis, these species were exclusively available within Artportalen, meaning no independent, external test dataset was available for evaluation.

4.3.1 Widely Distributed Species

The two most widely distributed species in the dataset are Eurasian otter (*Lutra lutra*) and moose (*Alces alces*), occupying 32.7% and 24.0% of the Swedish 5 km grid cells respectively. Both also have very large sample sizes (18 196 and 14 640 observations) and span nearly the full latitudinal range of the country. Their distribution plots are shown in Figures 4.4a and 4.4b.



(a) Distribution of otter. **(b)** Distribution of moose.

Figure 4.4: Distribution of Artportalen observations of **a)** Eurasian otter (*Lutra lutra*) and **b)** moose (*Alces alces*) over Sweden using the 5 km × 5 km grid. The title of each map states the number of observation samples of the species in the dataset and the number grid cells the observations occupy.

Table 4.5: Artportalen test set results for the widely distributed species Eurasian otter (*Lutra lutra*) and moose *Alces alces*. Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). AP is the average precision. The highest score for each metric is displayed in boldface.

	Model	Input	Thr.	Acc.	F1	AP
Lutra lutra	RF	Env	0.540	0.643	0.584	0.670
	SINR	Loc+Env	0.324	0.651	0.733	0.649
	ResNet-18	Env	0.290	0.648	0.719	0.742
	U-Net	Env	0.353	0.699	0.758	0.803
	U-Net++	Env	0.287	0.619	0.696	0.717
	ResNet-18	Sat	0.326	0.852	0.853	0.934
	U-Net	Sat	0.415	0.856	0.867	0.922
	U-Net++	Sat	0.441	0.764	0.794	0.822
	Fus-Res	Loc+Env	0.276	0.680	0.735	0.752
	Fus-U	Loc+Env	0.366	0.676	0.741	0.743
	Fus-U++	Loc+Env	0.274	0.642	0.719	0.744
	Fus-Res	Loc+Sat	0.286	0.878	0.880	0.948
	Fus-U	Loc+Sat	0.263	**0.906**	**0.910**	**0.969**
	Fus-U++	Loc+Sat	0.392	0.768	0.800	0.871
Alces alces	RF	Env	0.530	0.701	0.601	0.805
	SINR	Loc+Env	0.295	0.760	0.801	0.799
	ResNet-18	Env	0.297	0.700	0.754	0.863
	U-Net	Env	0.273	0.728	0.772	0.853
	U-Net++	Env	0.245	0.724	0.775	0.814
	ResNet-18	Sat	0.200	0.825	0.845	0.894
	U-Net	Sat	0.386	0.806	0.833	0.879
	U-Net++	Sat	0.273	0.797	0.824	0.878
	Fus-Res	Loc+Env	0.257	0.708	0.768	0.877
	Fus-U	Loc+Env	0.209	0.761	0.799	0.823
	Fus-U++	Loc+Env	0.278	0.722	0.775	0.833
	Fus-Res	Loc+Sat	0.228	0.817	0.840	**0.908**
	Fus-U	Loc+Sat	0.252	**0.827**	**0.848**	0.866
	Fus-U++	Loc+Sat	0.269	0.812	0.835	0.878

Despite the large sample sizes, both are among the hardest for the models to predict on the Artportalen test set, as shown in Table 4.5. For otter, accuracy ranges from 0.642 to 0.906, F1 from 0.584 to 0.910, and AP from 0.649 to 0.969 across the models with Fus-$\mathrm{U}_{\mathrm{LocSat}}$ scoring highest on them all. For moose the ranges are 0.700 to 0.827 for accuracy, 0.601 to 0.848 for F1 and 0.799 to 0.908 for AP with Fus-$\mathrm{U}_{\mathrm{LocSat}}$ scoring highest for accuracy and F1 while Fus-$\mathrm{Res}_{\mathrm{LocSat}}$ scores highest for AP. In both cases the Sat fusion models and pure Sat CNNs form a clearly stronger group than the rest of the models. The Env fusion models score similarly to their unfused Env CNN counterparts, and RF performs the worst by a clear margin.

This pattern suggests that for widely distributed species spatial information is the limiting factor, and that satellite embeddings provide spatial context that environmental features alone do not capture.

4.3.2 Concentrated Abundant Species

Fallow deer (*Dama dama*) is a useful contrast to the species above. It has 8 021 observations which is relatively high but occupies only 7.3% of the 5 km grid cells, almost entirely in southern Sweden, as shown in Figure 4.5. Despite the smaller sample size compared to otter and moose, all models achieve much higher scores on this species, as shown in Table 4.6.

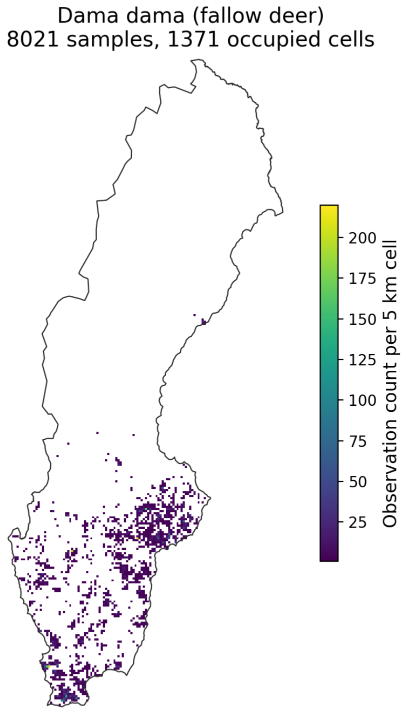


Figure 4.5: Distribution of Artportalen observations of fallow deer (*Dama dama*) over Sweden using the 5 km × 5 km grid. In the title of the map the number of observation samples of fallow deer in the dataset and the number grid cells the observations occupy is stated.

The lowest accuracy and F1 is 0.865 and 0.844 achieved by ResNet-18$_{\text{Env}}$, and AP is at or above 0.995 (RF) for every model. SINR, U-Net$_{\text{Env}}$, U-Net++$_{\text{Env}}$, U-Net++$_{\text{Sat}}$, Fus-Res$_{\text{LocEnv}}$, Fus-U$_{\text{LocEnv}}$, Fus-U$_{\text{LocSat}}$, Fus-U++$_{\text{LocEnv}}$ and Fus-U++$_{\text{LocSat}}$ all reach a perfect AP of 1.000. U-Net$_{\text{Env}}$ also scores best for accuracy and F1 with 0.995 for both.

The narrow geographic niche makes fallow deer easy to learn from environmental features alone which is shown by the fact that the environmental input models actually outperform the satellite embedding ones on average, suggesting that when a species occupies a well-defined niche, the additional spatial context provided by the satellite embeddings is not necessary.

Table 4.6: Artportalen test set results for the concentrated abundant species fallow deer (*Dama dama*). Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). The highest score for each metric is displayed in boldface.

Model	**Input**	**Thr.**	**Acc.**	**F1**	**AP**
RF	Env	0.550	0.881	0.856	0.995
SINR	Loc+Env	0.807	0.914	0.906	**1.000**
ResNet-18	Env	0.856	0.865	0.844	0.997
U-Net	Env	0.900	**0.995**	**0.995**	**1.000**
U-Net++	Env	0.887	0.971	0.970	**1.000**
ResNet-18	Sat	0.910	0.920	0.913	0.997
U-Net	Sat	0.960	0.895	0.883	0.998
U-Net++	Sat	0.948	0.874	0.856	**1.000**
Fus-Res	Loc+Env	0.916	0.916	0.908	**1.000**
Fus-U	Loc+Env	0.871	0.963	0.961	**1.000**
Fus-U++	Loc+Env	0.789	**0.989**	**0.989**	**1.000**
Fus-Res	Loc+Sat	0.913	0.877	0.861	0.996
Fus-U	Loc+Sat	0.904	0.940	0.936	**1.000**
Fus-U++	Loc+Sat	0.865	0.936	0.932	**1.000**

4.3.3 Concentrated Rare Species

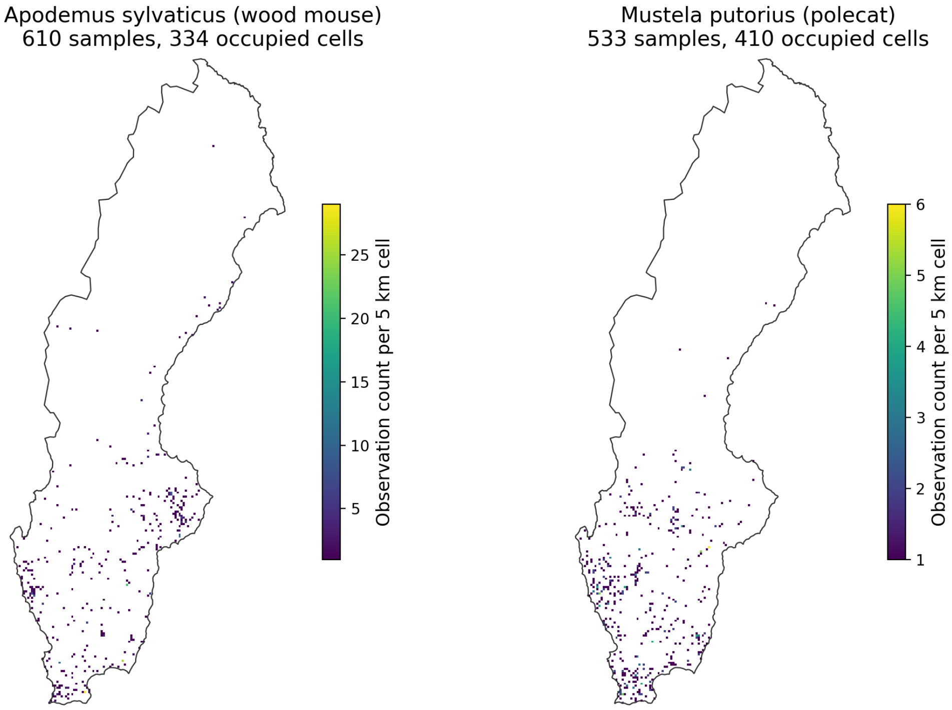
Wood mouse (*Apodemus sylvaticus*) and European polecat (*Mustela putorius*) provide further support for this pattern. Both have very few observations, 610 and 533 respectively, but small geographic ranges of only 1.8% and 2.2% of the 5 km grid cells. Their distribution plots are shown in Figures 4.6a and 4.6b.

The test results for both species are shown in Table 4.7. For wood mouse all models reach accuracy above 0.948, F1 above 0.946, and AP above 0.962, with U-Net$_{\text{Sat}}$ and Fus-U++$_{\text{LocSat}}$ being tied for highest accuracy and F1 at 0.986 and Fus-U++$_{\text{LocSat}}$ having a perfect AP of 1.000. For polecat, the lowest accuracy, F1 and AP are 0.796, 0.766 and 0.937. U-Net++$_{\text{Env}}$ achieves the highest accuracy and F1 (0.963 and 0.964), closely followed by Fus-Res$_{\text{LocEnv}}$ (both 0.963) and RF (both 0.962) which has the highest AP (0.996). These scores are significantly better than the scores for otter and moose (Table 4.5) and there is also no clear advantage for satellite embeddings.

These results show that the models can still learn accurate distributions from limited data when the geographic range is tight, and that the choice of model architecture and input matters less in this regime. Rather, spatial spread, not sample count, appears to be the binding constraint on model performance.

Table 4.7: Artportalen test set results for the concentrated rare species wood mouse (*Apodemus sylvaticus*) and European polecat (*Mustela putorius*). Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). The highest score for each metric is displayed in boldface.

	Model	Input	Thr.	Acc.	F1	AP
Apodemus sylvaticus	RF	Env	0.530	0.956	0.956	0.978
	SINR	Loc+Env	0.681	0.967	0.967	0.962
	ResNet-18	Env	0.604	0.948	0.946	0.993
	U-Net	Env	0.437	0.962	0.963	0.996
	U-Net++	Env	0.622	0.967	0.967	0.988
	ResNet-18	Sat	0.687	0.957	0.957	0.994
	U-Net	Sat	0.755	**0.986**	**0.986**	0.996
	U-Net++	Sat	0.619	0.971	0.971	0.997
	Fus-Res	Loc+Env	0.407	0.952	0.952	0.985
	Fus-U	Loc+Env	0.605	0.971	0.971	0.987
	Fus-U++	Loc+Env	0.624	0.967	0.967	0.977
	Fus-Res	Loc+Sat	0.594	0.971	0.971	0.992
	Fus-U	Loc+Sat	0.549	0.971	0.971	0.995
	Fus-U++	Loc+Sat	0.505	**0.986**	**0.986**	**1.000**
Mustela putorius	RF	Env	0.600	0.962	0.962	**0.996**
	SINR	Loc+Env	0.694	0.907	0.902	0.994
	ResNet-18	Env	0.620	0.870	0.857	0.990
	U-Net	Env	0.787	0.796	0.766	0.940
	U-Net++	Env	0.680	**0.963**	**0.964**	0.962
	ResNet-18	Sat	0.780	0.926	0.926	0.975
	U-Net	Sat	0.850	0.907	0.906	0.937
	U-Net++	Sat	0.821	0.833	0.816	0.929
	Fus-Res	Loc+Env	0.644	**0.963**	0.963	0.986
	Fus-U	Loc+Env	0.752	0.926	0.926	0.941
	Fus-U++	Loc+Env	0.719	0.907	0.906	0.984
	Fus-Res	Loc+Sat	0.721	0.889	0.885	0.956
	Fus-U	Loc+Sat	0.633	0.926	0.923	0.982
	Fus-U++	Loc+Sat	0.631	0.889	0.880	0.986



(a) Distribution of wood mouse. **(b)** Distribution of polecat.

Figure 4.6: Distribution of Artportalen observations of **a)** wood mouse (*Apodemus sylvaticus*) and **b)** European polecat (*Mustela putorius*) over Sweden using the 5 km × 5 km grid. In the title of each map the number of observation samples of the species in the dataset and the number grid cells the observations occupy is stated.

4.3.4 Threshold Sensitivity

Mouflon (*Ovis aries musimon*) illustrates a behaviour that is hidden when looking only at thresholded metrics. This is the most geographically concentrated species in the dataset, occupying just 0.48% of the 5 km grid cells, all in small concentrated areas mainly in southern Sweden as shown in Figure 4.7. It also has very few observations (673).

The test results are shown in Table 4.8. Every model achieves an AP of 1.000 for this species, meaning that the underlying probability rankings perfectly separate presence from absence in all cases. However, the thresholded F1 ranges from 0.222 (RF) to 1.000 (SINR, U-Net$_{\text{Sat}}$, Fus-U$_{\text{LocEnv}}$, and Fus-U$_{\text{LocSat}}$), a spread of 0.778. ResNet-18$_{\text{Sat}}$ and Fus-Res$_{\text{LocSat}}$ also score poorly on F1 (0.308 and 0.533) despite perfect AP, while the remaining models land in the middle around 0.840–0.952.

This is a clear example of how a poorly calibrated threshold can severely degrade thresholded metrics even when the model itself ranks predictions correctly. This suggests that for species with very narrow distributions and few observations, AP is a more reliable indicator of model quality than F1 or accuracy.

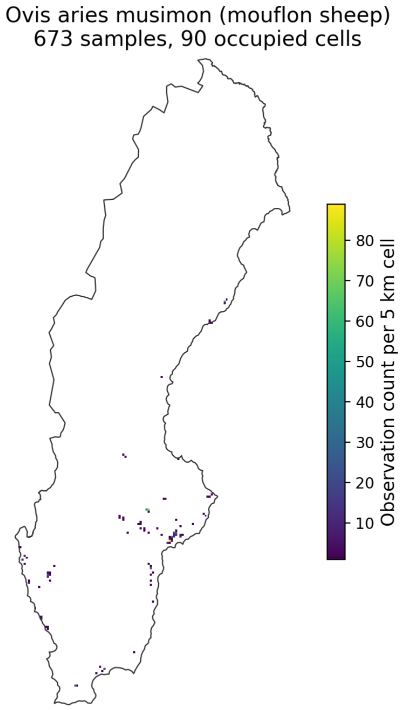


Figure 4.7: Distribution of Artportalen observations of mouflon (*Ovis aries musimon*) over Sweden using the 5 km × 5 km grid. In the title of the map the number of observation samples of mouflon in the dataset and the number grid cells the observations occupy is stated.

Table 4.8: Artportalen test set results for the highly concentrated species Mouflon (*Ovis aries musimon*). Accuracy (Acc.) and F1 are reported at the listed threshold (Thr.). The highest score for each metric is displayed in boldface.

Model	**Input**	**Thr.**	**Acc.**	**F1**	**AP**
RF	Env	0.640	0.632	0.222	**1.000**
SINR	Loc+Env	0.484	**1.000**	**1.000**	**1.000**
ResNet-18	Env	0.774	0.909	0.900	**1.000**
U-Net	Env	0.732	0.909	0.900	**1.000**
U-Net++	Env	0.703	0.955	0.952	**1.000**
ResNet-18	Sat	0.918	0.591	0.308	**1.000**
U-Net	Sat	0.794	**1.000**	**1.000**	**1.000**
U-Net++	Sat	0.570	0.955	0.952	**1.000**
Fus-Res	Loc+Env	0.644	0.864	0.842	**1.000**
Fus-U	Loc+Env	0.408	**1.000**	**1.000**	**1.000**
Fus-U++	Loc+Env	0.643	0.955	0.952	**1.000**
Fus-Res	Loc+Sat	0.853	0.682	0.533	**1.000**
Fus-U	Loc+Sat	0.616	**1.000**	**1.000**	**1.000**
Fus-U++	Loc+Sat	0.462	0.909	0.900	**1.000**

4.4 IUCN Results

This section compares the models predicted species distribution with the expert range maps from IUCN. Table 4.9 shows the mAP of all predicted species with an available IUCN range map and the AP of three selected species, hedgehog (*Erinaceus europaeus*), lynx and brown bear chosen for their different distribution within Sweden. All deep learning models have similar scores, while RF has significantly lower scores. Overall the models using environmental inputs scored the highest, with ResNet-18$_{\text{Env}}$ having top mAP (0.885), SINR has the top AP for hedgehogs (0.880) and U-Net++$_{\text{Env}}$ has the top AP for brown bears (0.964). In contrast, the models with satellite embedding input performed better on lynx than those with environmental inputs with Fus-Res$_{\text{LocSat}}$ having the highest AP (0.911). The fusion models also do not provide a general improvement over the pure CNNs.

Table 4.9: Model average precision (AP) on IUCN expert range maps for selected species and mean over all species with an available range map. Mean is taken as the mean of the AP for each species with an available IUCN range map. The highest scores are displayed in boldface.

Model	Input	Average precision (AP)			
		Mean	Hedgehog	Lynx	Brown bear
RF	Env	0.753	0.658	0.769	0.767
SINR	Loc+Env	0.875	**0.880**	0.874	0.952
ResNet-18	Env	**0.885**	0.861	0.879	0.952
U-Net	Env	0.868	0.849	0.885	0.950
U-Net++	Env	0.880	0.852	0.893	**0.964**
ResNet-18	Sat	0.876	0.830	0.889	0.958
U-Net	Sat	0.865	0.797	0.896	0.954
U-Net++	Sat	0.872	0.848	0.902	0.958
Fus-Res	Loc+Env	0.879	0.870	0.889	0.947
Fus-U	Loc+Env	0.869	0.867	0.861	0.945
Fus-U++	Loc+Env	0.871	0.844	0.878	0.940
Fus-Res	Loc+Sat	0.873	0.831	**0.911**	0.954
Fus-U	Loc+Sat	0.869	0.816	0.889	0.943
Fus-U++	Loc+Sat	0.879	0.852	0.906	0.962

To further analyse the results the IUCN range maps for the selected species are shown in Figure 4.8. These can be compared with the prediction maps for the same species made by ResNet-18$_{\text{Env}}$, shown in Figure 4.9a, and Fus-Res$_{\text{LocSat}}$, shown in Figure 4.9b.

Looking at the IUCN range map for hedgehog it is mostly uniform and limited to a defined part of Sweden. Comparing the prediction maps for hedgehogs made by ResNet-18$_{\text{Env}}$ and Fus-Res$_{\text{LocSat}}$ to the range map, both of them capture the

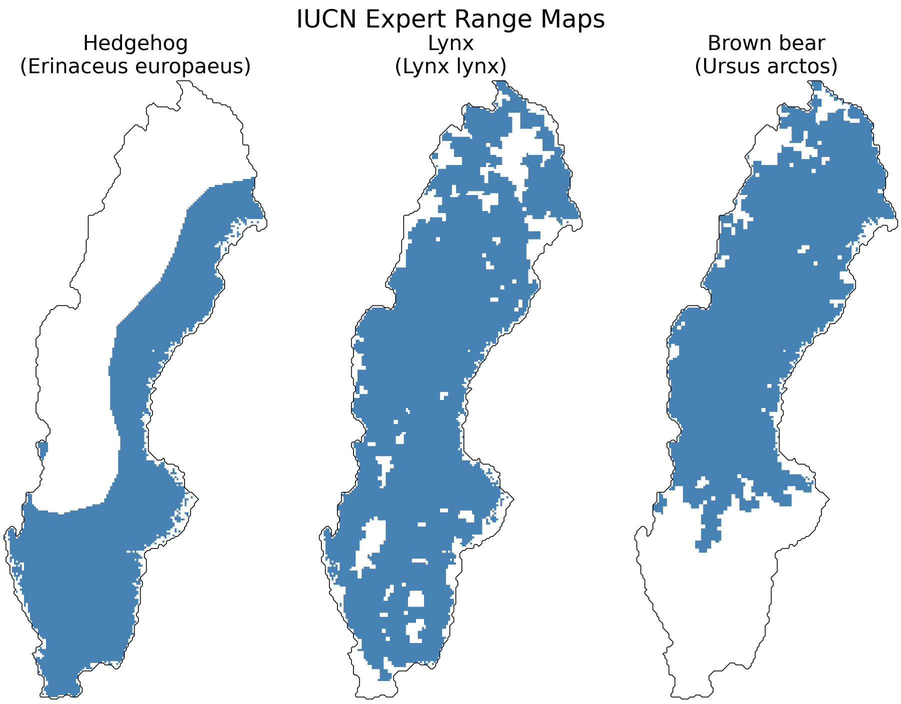


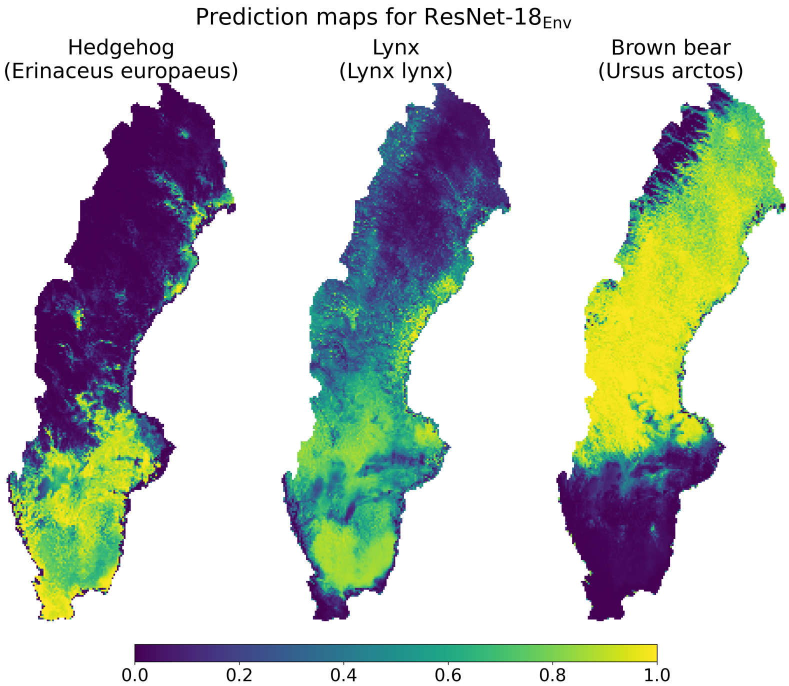
Figure 4.8: IUCN expert range maps for hedgehog (*Erinaceus europaeus*), lynx (*Lynx lynx*) and brown bear (*Ursus arctos*) [6]. The estimated presence ranges for each species are displayed by the blue colour on the maps.

main range, particularly in the south, but also somewhat along the coast. However ResNet-18$_{\text{Env}}$ shows a much smoother distribution within the southern range while Fus-Res$_{\text{LocSat}}$ shows much more local variability in the prediction probability. This means that ResNet-18$_{\text{Env}}$ will rank the IUCN range as presence more evenly than Fus-Res$_{\text{LocSat}}$ which is reflected in the difference in AP (0.861 against 0.831).

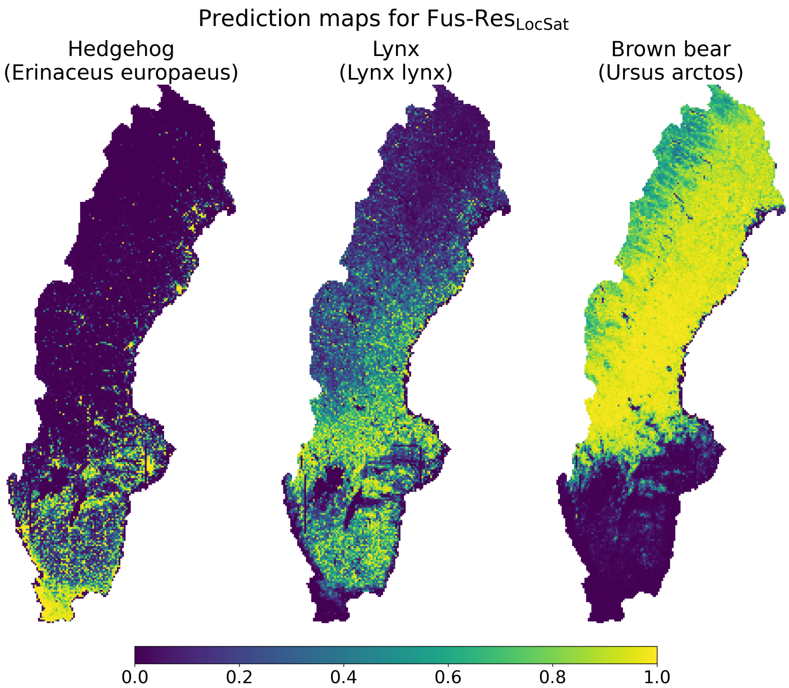
Moving on to lynx the IUCN range map is less uniform, containing many gaps, and covers most of Sweden. Comparing the prediction maps made by ResNet-18$_{\text{Env}}$ and Fus-Res$_{\text{LocSat}}$ both of them capture the southern half of Sweden well but lack somewhat in the north. However Fus-Res$_{\text{LocSat}}$ is better able to capture the areas of absence in the south while retaining roughly equivalent coverage in the north compared to ResNet-18$_{\text{Env}}$. This is again reflected in the AP with Fus-Res$_{\text{LocSat}}$ having a score of 0.911 compared to 0.879 for ResNet-18$_{\text{Env}}$.

Lastly, looking at the brown bear IUCN range map it is in-between hedgehog and lynx in terms of how uniform it is and it is also limited to a defined part of Sweden. The prediction maps made by ResNet-18$_{\text{Env}}$ and Fus-Res$_{\text{LocSat}}$ for brown bear are similar and both cover the entire range well and accurately. Both also have nearly identical and high AP with 0.952 for ResNet-18$_{\text{Env}}$ and 0.954 for Fus-Res$_{\text{LocSat}}$.

This suggests that in general environmental data is sufficient or even preferred for broader range map predictions, yielding smoother prediction maps. However satellite embeddings may still be better for harder to predict species such as lynx with more irregular ranges.



(a) ResNet-18$_{\mathrm{Env}}$ prediction maps.



(b) Fus-Res$_{\mathrm{LocSat}}$ prediction maps.

Figure 4.9: Prediction maps for hedgehog (*Erinaceus europaeus*), lynx (*Lynx lynx*), and brown bear (*Ursus arctos*) of the **a)** ResNet-18$_{\mathrm{Env}}$ and **b)** Fus-Res$_{\mathrm{LocSat}}$ models. The predicted presence probabilities are displayed by the colour grading shown below the maps. The maps were created using the centre of the grid cells in the 5 km $\times$ 5 km grid as inputs.

5

Discussion

5.1 Environmental Variables vs. Satellite Embeddings

Across the Artportalen and Rovbase test sets, CNNs with satellite embedding input consistently outperform their environmental-input counterparts (Table 4.1). On Artportalen, the jump in mAP from the best environmental CNN (U-Net$_{\text{Env}}$, 0.934) to the weakest satellite CNN (U-Net++$_{\text{Sat}}$, 0.960) is larger than the spread across all environmental models combined, and the fusion model Fus-U$_{\text{LocSat}}$ reaches the highest overall scores (0.918 accuracy, 0.921 F1, 0.975 mAP) closely followed by Fus-Res$_{\text{LocSat}}$. A similar ranking holds on Rovbase, although the absolute gap between satellite and environmental models narrows by roughly half. Fus-U$_{\text{LocSat}}$ also drops significantly more in performance compared to Fus-Res$_{\text{LocSat}}$ which now has the highest scores (0.866 accuracy, 0.870 F1, 0.923 mAP).

The narrowing performance gap on Rovbase reflects the composition of that evaluation rather than a general reduction in the utility of satellite embeddings. The Rovbase test set covers only wolf, brown bear and lynx, and for these three species the systematic surveying captured by Rovbase shows a far broader presence than the Artportalen data ever exposes the model to during training as shown in Figures 4.1, 4.2 and 4.3. The Rovbase test set is therefore best understood not as an in-distribution generalisation test but as a distribution-shift test in which the model is asked to predict presence across an area substantially broader than what its training data labelled. This reading is reinforced by the construction of the set itself. Since Rovbase supplies only presence records, its negatives are distance-only pseudo-absences, so what the evaluation actually probes is the recovery and ranking of presence across a broadened range, rather than the correct identification of true absence. The value Rovbase adds over the Artportalen split therefore lies mainly on the presence side, in its systematic, near-complete coverage, and the retained performance of all models under this shift should be read in that light. The fact that all models retain reasonable performance under this shift is encouraging and shows how they are able to learn about species distributions outside the support of their training distribution. The species-level breakdown supports this reading (Tables 4.2, 4.3 and 4.4). Lynx, with the broadest Rovbase range and the largest discrepancy between datasets, shows the clearest retained advantage for the satellite

CNNs, whereas wolf and bear, whose Rovbase distributions remain more contained, show much smaller gaps between input types.

The dominance of satellite embeddings is largely reversed when the predicted distributions are compared against the IUCN expert range maps. Here the environmental models score marginally higher on average (Table 4.9), with ResNet-18$_{\text{Env}}$ achieving the top mean AP of 0.885. This is not a contradiction of the test-set results but a consequence of resolution. The satellite embeddings are constructed from 10 m imagery downsampled to 50 m pixels, and the 128×128 patches give the CNNs access to fine-grained local texture. The environmental layers, by contrast, are bilinearly interpolated from the original 30 seconds (∼1 km) WorldClim resolution and therefore vary smoothly across the patch. When predictions are evaluated against the IUCN polygons, which themselves represent smoothed expert opinion of broad geographic range, the lower-resolution environmental predictions are structurally closer to the target than the high-resolution satellite predictions. This is visible in Figures 4.9b and 4.9a, the Fus-Res$_{\text{LocSat}}$ predictions show sharp local variation while the ResNet-18$_{\text{Env}}$ predictions are smoother. Despite this, satellite embeddings still might be preferred for harder to predict species. Going back to Table 4.9 the satellite embedding models all outscored their environmental counterpart and baseline for lynx which we have found to be a difficult species to predict due to the large distribution shift between the training data and its actual distribution.

A practical implication follows. If the goal of the SDM is binary classification of fine-grained presence, or the identification of within-range habitat variation, the satellite embeddings provide a clear advantage. If the goal is to reproduce a broad expert range, lower-resolution environmental variables may be preferred, or the satellite predictions should be spatially smoothed before evaluation. More generally, the three evaluation sets used here, Artportalen, Rovbase, and the IUCN range maps, disagree about which model is best. This is not a flaw of any one of them but a reminder of the limitation already raised in Section 1.4, that there is no single ground truth for species distribution, and a model's apparent quality depends on which proxy is used.

5.2 Comparison of Performance Between Model Architectures

Among the CNN architectures, ResNet-18 and U-Net perform similarly across both input types and both test sets, while U-Net++ trails both for our tests (Table 4.1). On the Artportalen set with environmental input, U-Net$_{\text{Env}}$ achieves the highest mAP among the CNN variants (0.934), followed by ResNet-18$_{\text{Env}}$ (0.924), with U-Net++$_{\text{Env}}$ falling slightly behind (0.917). With satellite input ResNet-18$_{\text{Sat}}$ now has the highest mAP (0.965) followed by U-Net$_{\text{Sat}}$ (0.961) and U-Net++$_{\text{Sat}}$ (0.961). Important to note is that since the models were only trained once for the final version, small differences in performance between architectures are not necessarily statistically significant.

The slight underperformance of U-Net++ relative to U-Net is noteworthy given that U-Net++ was designed to improve upon U-Net through nested skip connections and deep supervision. A likely contributing factor is the reduced depth used in our U-Net++ implementation, only two down- and upsampling steps compared to four in U-Net, which was necessary to keep the model tractable. This means U-Net++ might have less representational capacity rather than more, which may explain why the additional architectural complexity did not translate into improved performance. It is possible that a deeper U-Net++ configuration would behave differently, but this was not explored due to computational constraints. However the fact that the difference is still relatively small speaks to a strength in how U-Net++ is able to achieve similar performance despite less depth.

A notable result on the Rovbase test set is the relatively sharp drop in mAP for U-Net$_{\text{Env}}$ (from 0.934 on Artportalen to 0.841 on Rovbase, Table 4.1), which is larger than for any other model including U-Net++$_{\text{Env}}$ and ResNet-18$_{\text{Env}}$. This may indicate that U-Net$_{\text{Env}}$ overfits more strongly to the spatial and observational structure of the Artportalen training data when using environmental input, and that the segmentation output and wider spatial context it uses amplifies this bias rather than alleviating it when generalising to the systematically collected Rovbase data. The fact that U-Net++ does not show the same indications of over-fitting is most likely due to the smaller network size since fewer parameters results in reduced ability to fit exactly to the training data.

Comparing the CNN architectures to the baseline models, SINR achieves accuracy and F1 scores on Artportalen that are broadly comparable to the environmental CNNs, despite using only a point input rather than a patch. This suggests that SINR's location encoding captures some of the same spatial signal as the patch-based CNNs when both rely on environmental variables. On Rovbase, the differences between SINR and the environmental CNNs are similarly small, and RF closes the gap further, even outscoring them in mAP. This indicates that for species with a clear and consistent geographic range, simpler models can perform competitively when evaluated against a high-quality test set.

In general adding SINR's location encoding to the CNNs via the fusion models does not always result in improvement. The improvement is most noticeable when using satellite embedding input and specifically for Fus-U$_{\text{LocSat}}$ and Fus-Res$_{\text{LocSat}}$ on Artportalen. However these differences show significant variation between species showing that the potential improvements might be dependent on the species range it self. Fus-U$_{\text{LocSat}}$ also has a large performance drop when moving to Rovbase while Fus-Res$_{\text{LocSat}}$ retained improved performance. This indicates that the fusion implementation for ResNet-18 was more successful which is reasonable given its simpler and more robust implementation described in Section 3.3.5. There the SINR location encoding is concatenated directly with ResNet-18's own feature vector in the classification head, a single vector-level operation. The U-Net and U-Net++ fusion has to bridge a larger architectural gap, SINR's 256-dimensional encoding is first projected down to 32 dimensions via a linear layer and then tiled across the full 128×128 patch so that it can be concatenated with the decoder's spatially varying output.

The projection acts as a strong bottleneck that discards much of the SINR representation before it can interact with the CNN features, and the tiled channels carry no spatial variation by construction, which might conflict with the per-pixel reasoning of U-Net and U-Net++. This combination plausibly explains why Fus-$\mathrm{U}_{\mathrm{LocSat}}$ fits the Artportalen training distribution well but drops more sharply on Rovbase, where the underlying species ranges are broader than the location-correlated observer bias of the training data would imply. The lack of improvement for environmental input also suggests that when a CNN already receives a spatial patch as input the additional location encoding doesn't always provide valuable complementary information. This in turn suggest that the information carried by the coordinate location overlaps more with the broad spatial signals already present in the coarse environmental variables, while it is more complementary to the local texture captured by the high-resolution satellite embeddings.

5.3 Best Model Overall

The best overall model for our tests is Fus-$\mathrm{Res}_{\mathrm{LocSat}}$. It achieves the second highest mAP and accuracy and the third highest F1 on Artportalen, and the highest scores across all metrics on Rovbase (Table 4.1). Although Fus-$\mathrm{U}_{\mathrm{LocSat}}$ scores highest on Artportalen, its performance drops significantly on Rovbase as previously discussed making Fus-$\mathrm{Res}_{\mathrm{LocSat}}$ the better choice across both sets. The improvement over the pure satellite CNNs is, however, relatively modest. On Artportalen, Fus-$\mathrm{Res}_{\mathrm{LocSat}}$ reaches an mAP of 0.972 compared to 0.965 for ResNet-18_{Sat} and 0.961 for U-$\mathrm{Net}_{\mathrm{Sat}}$. On Rovbase the margins are similarly narrow, 0.923 mAP compared to 0.914.

If model simplicity is a priority, U-$\mathrm{Net}_{\mathrm{Sat}}$ and ResNet-18_{Sat} offer competitive performance without the added complexity of the fusion architecture. Between these two, U-$\mathrm{Net}_{\mathrm{Sat}}$ has a slight edge on thresholded metrics while ResNet-18_{Sat} performs marginally better on mAP. The choice between them may therefore depend on whether threshold-dependent classification or probability ranking is the primary use case.

Across all models, the choice of input type has a larger effect on performance than the choice of architecture. The consistent advantage of satellite embeddings over environmental variables, and the comparatively small differences between ResNet-18 and U-Net within each input type, suggest that further gains are more likely to come from improvements to the input representation, pseudo-absence generation, or training procedure than from changes to the CNN architecture alone.

5.4 Effect of Data Volume and Distribution

As previously discussed and shown by the patterns in Section 4.3, the geographic spread of a species is a stronger predictor of model difficulty than its sample count, and this interacts with the pseudo-absence sampling in ways that help explain some of the more striking results.

For widely distributed species like otter and moose, performance remains low despite large observation counts (Table 4.5). A key reason is that the distance-based pseudo-absence sampling struggles when a species occupies a large portion of the study area. When presences are spread across most of Sweden, pseudo-absences generated by distance weighting are pushed to increasingly marginal and remote areas, leaving the model with little informative signal about where the species boundary actually lies. The density-weighted component mitigates this to some extent by targeting well-surveyed regions far from the species' observations, but for a broadly distributed species this distinction collapses since heavily surveyed areas tend to overlap with the species range. The result is a noisier training signal, and one where satellite embeddings, by providing fine-grained local spatial context, offer more discriminative power than environmental features alone can.

The opposite pattern holds for geographically concentrated species with abundant data, illustrated by fallow deer (Table 4.6). Here, the clear spatial boundary makes pseudo-absence placement straightforward: the distance weighting naturally generates absences well outside the species range, giving the model clean and informative negative examples. In this regime environmental features alone are sufficient to characterise the niche, and the addition of satellite embeddings provides little further benefit. The near perfect AP scores across all models confirm that the classification task itself is tractable, not merely that certain architectures handle it well.

The concentrated rare species, wood mouse and polecat, reinforce this interpretation (Table 4.7). Despite very few observations, both have tight geographic ranges and both are modelled accurately by nearly all architectures. Because the presences are clustered, pseudo-absences are placed cleanly outside the range regardless of model architecture, and the model receives a consistent training signal. This supports the conclusion that spatial spread, rather than sample count, is the primary binding constraint on model performance. The pseudo-absence generation method functions well when there is a well-defined boundary to learn from, and poorly when that boundary is diffuse.

This problem of geographical coverage having a large impact on modelling performance is especially pronounced when working with a smaller closed area such as a single country. When making global models the ranges of each animal will typically be much smaller than the entire modelling area meaning that broad, well-defined boundaries will almost always exist and pseudo-absences will not be generated close to presences at a large scale. This shows the limitations of a general pseudo-absence generation method when working with local SDMs on a closed area and suggest that species specific adaptations may be beneficial.

The mouflon results in Table 4.8 add a further nuance around threshold calibration that applies more generally. With only a small area of Sweden covered, the pseudo-absences will be generated far from the small clusters of presences, and every model achieves a perfect AP of 1.000, indicating that the underlying probability rankings separate presence from absence cleanly. Yet the thresholded F1 ranges from 0.222 to 1.000, and the selected thresholds themselves vary from 0.484 to 0.918. The cause is the LPT-R method itself. LPT-R sets the threshold at the fifth percentile of the model's predicted probabilities for the training presences and never observes the model's behaviour on absences. For species with few training presences the resulting percentile estimator is high-variance, and even small differences in how each model assigns probabilities to a handful of presence samples translate into substantially different thresholds. The ranking the model has learned is correct, but the cut-off applied to that ranking is poorly calibrated.

This is a general limitation rather than a mouflon-specific quirk. The same mechanism affects every species with a small training set or a concentrated distribution, and it also affects evaluation on the Rovbase test set. The thresholds applied there are calibrated on the Artportalen training presences, so a model that ranks Rovbase observations correctly can still produce inflated false negatives if its Artportalen presence distribution is shifted relative to its Rovbase one. This is consistent with the pattern noted earlier for wolf, brown bear and lynx, where AP is preserved much more cleanly across the dataset shift than F1 is. The implication is that threshold-independent metrics, in this case AP, should be treated as the primary indicators of model quality in this study, with thresholded metrics reported alongside them but interpreted with the calibration caveat in mind. As already noted in Section 3.5.3, LPT-R was designed for generating binary range maps and is not necessarily optimal for individual-sample classification. Future work would benefit from a dedicated calibration step, for example a threshold tuned on a held-out validation set using both presences and pseudo-absences, treated as a problem separate from learning the species distribution itself.

5.5 Observation Data Bias

As described in Section 2.1.1 the crowd-sourced data from Artportalen has a strong spatial bias towards populated areas. The goal with the custom pseudo-absence generation was to try to mitigate that bias, but that only accounts for the pseudo-absence generation in unsampled areas. The bias some species have because of excessive sampling in certain areas, mainly around cities, is still prevalent in the training data causing the models to adapt to them. In Figures 4.9b and 4.9a, showing the predicted species distributions of three different species (hedgehogs, lynx and brown bears), this is especially clear for the distribution of hedgehogs. Both the models using environmental and remote sensing input get much stronger signal around cities in the south and up the east coast of Sweden. The lower resolution of the environmental variables smooths the predictions out, making the bias less visible compared to the high resolution remote sensing embeddings.

The pseudo-absence generation doesn't completely solve the problem of the unsampled regions either, particularly for species like lynx, which exists in the entire country. The north of Sweden is much less populated than the south, causing more observations of the lynx to be in the south. As a result the predicted presence in the north is significantly lower than the actual presence. In comparison bears only exist in the north, which mitigates the effect of this kind of bias.

5.6 Potential Improvements

One potential improvement is a more sophisticated and optimized pseudo-absence generation method and sampling in general. Due to time and computational limitations, we were not able to rigorously optimize our pseudo-absence sampling method and had to rely on smaller tests to choose our method. It is possible that further tuning of parameters such as the proportion of distance-only to distance-density, the density radius used for the sampling effort, the maximum density value, etc. could improve performance. Furthermore, negatively weighting the presence sampling based on overall observation density could reduce the positive spatial biases we see in species that have presence near observation hot spots like cities. Implementations of more advanced learning methods such as the active learning proposed by Lange et al. [22] could also lead to improvements.

Another potential improvement when it comes to pseudo-absence generation is varying the number of pseudo-absences generated, both in total and per species. Global models like SINR [13] use oversampling of pseudo-absences and subsequent weighting of presences since the area of absence is greater than the area of presence for most species globally. A similar approach could potentially be applied for our model, but with potential modifications. As discussed in Section 5.4, the models struggle with species that have a particularly small or large spatial distribution over Sweden. Varying the number of pseudo-absence samples generated based on the overall distribution of the species could therefore be useful as it would fill in the large absence areas for species with a small spatial spread and not force a large number of pseudo-absences to be generated within the presence areas for species with large spatial distributions.

Lastly, the model parameters could likely have been better tuned and the training process improved in general. Doing hyperparameter tuning using k-fold cross-validation and a more sophisticated grid search could potentially lead to a more accurate and finely tuned model. We also had training functionalities which we were not able to implement in the end. For example we implemented location jittering during training which randomly jittered the observation samples between training epochs with the aim of reducing overfitting to stationary observations. This worked well when using environmental input but lead to greatly increased training time when using satellite embeddings due to not being able to use the quick lookup for presence samples. Thus it was ultimately abandoned for fair comparisons between models. These limitations were once again caused by time constraints and computational capacity.

5.7 Future Research

Suggestions for future research include expanding the model in terms of number of species and area. This project can be seen as a proof-of-concept that convolutional models using local, high resolution data are able to learn about species distributions at a scale applicable to Sweden. It would therefore be interesting to see if this could be expanded further to create a model that is able to do high resolution species distribution modelling at a global scale. Including more species in the model would also be interesting as it would perhaps be able to capture more interspecies correlations in a more complete ecosystem, perhaps leading to better performance and the ability to find relationships between species.

Another idea for future research is taking advantage of the fact that UNet and UNet++ are able to make predictions across the entire input patch. Thus it should be possible to train the model on multiple samples at different locations within the patch simultaneously instead of focusing on a single centred sample at a time. This would allow for detailed predictions to be made for multiple species over an entire patch in one go. Important to note is that this would likely require more and higher quality data and modifications to pseudo-absence sampling. It would also require more care to be taken into how far different species roam and interact.

6

Conclusion

This thesis investigated the viability of using convolutional neural networks (CNNs) paired with high-resolution remote sensing data to model species distributions from imbalanced, citizen-science data from Artportalen. This work sought to determine whether spatial context from remote sensing could outperform traditional environmental predictors and established baselines. To achieve this, a standard image classification architecture (ResNet-18) was evaluated against two semantic segmentation networks (U-Net and U-Net++), testing them both as standalone models and combined with the SINR location encoder in a late-fusion configuration. The models were primarily evaluated on two observational test sets: a test split from Artportalen and the complete Rovbase test set containing high-quality observations of wolf, brown bear, and lynx. In addition to these, the models were evaluated against IUCN expert range maps to test for broader range prediction. This chapter summarises the findings of this work by directly addressing the two primary research questions established at the outset of this thesis.

In addressing the first research question, to what extent the integration of high-resolution satellite embeddings and convolutional neural networks can improve the predictive performance of species distribution models compared to traditional environmental baselines, the choice of input data proved to be the defining factor. With traditional environmental variables, the CNNs offered no consistent advantage over the much simpler baselines. This was especially evident on the Rovbase test set, where the random forest baseline even surpassed them on mAP. The integration of high-resolution remote sensing data, on the other hand, unlocked a clear advantage for the deep learning architectures. With satellite embedding inputs, the CNNs consistently outperformed both random forest and SINR, demonstrating that deep spatial features can meaningfully improve predictive performance when paired with the right architecture.

Within each input type, the performance differences between architectures were minor. U-Net and ResNet-18 performed comparably, with U-Net++ trailing slightly, most plausibly due to its reduced depth. However, these inter-architecture gaps are small enough that they could fall within the expected variation between distinct training runs. Because each model was trained from a single random initialisation, the architectural rankings should be interpreted as indicative rather than definitive; the robust performance leap is fundamentally driven by the choice of input type rather than the specific network backbone.

The late-fusion approach, which incorporated an additional input of raw geographical coordinates, yielded the highest Artportalen scores for both ResNet-18 and U-Net backbones when paired with satellite embedding inputs. However, only the ResNet-18 fusion retained its advantage on Rovbase, making Fus-Res$_{\mathrm{LocSat}}$ the most robust performer across test sets. This discrepancy is possibly due to less robust fusion methods implemented for U-Net and U-Net++. Notably, fusion did not result in any clear improvement when paired with traditional environmental inputs. This suggests that fusion approaches lead to performance gains primarily when implemented with robust architectures and when the input channels provide highly complementary information, though further research is needed to isolate these dynamics.

The three evaluation proxies did not agree on which model was superior, a variance that is itself a core finding rather than an experimental flaw. The Rovbase set, whose negatives are distance-only pseudo-absences, is best understood as a presence-ranking test under distribution shift rather than a true presence–absence benchmark. All satellite models lost three to five points of accuracy moving to it, with no clearly superior generaliser. Conversely, evaluation against the IUCN range maps reversed the test-set ranking, with the environmental models scoring marginally higher on average (top mAP of 0.885 for ResNet-18$_{\mathrm{Env}}$). This indicates a resolution effect rather than a structural contradiction. The practical implication for the first research question is that the ideal input choice is highly context-dependent: satellite embeddings excel at fine-grained habitat mapping and local binary classification, while traditional environmental variables remain superior when the goal is to produce broad, macro-level range maps.

Regarding the second research question, how a custom pseudo-absence generation method that weighs absence probabilities based on spatial observation density can reduce the impact of inherent spatial bias in crowd-sourced data, the findings reveal both the strengths and the fundamental limitations of data-driven bias mitigation. The underlying spatial and reporting biases of the Artportalen data were not fully eliminated by either input type or the multi-modal architectures. For species with strong urban association in their observation records, such as hedgehog (*Erinaceus europaeus*), the models captured a disproportionate signal clustered around population centres, reflecting observer density rather than true ecological limits. This artifact was most visible in the high-resolution satellite predictions.

Conversely, the systematic under-representation of observations in the sparsely populated north of Sweden constrained the models' ability to map species with truly nationwide distributions, such as lynx (*Lynx lynx*). This resulted in significantly higher prediction accuracy in the well-sampled southern regions compared to the data-scarce north. In these cases, the CNNs partially mapped the distribution of observer effort rather than the actual species range. These outcomes demonstrate that while a density-weighted pseudo-absence generation method is a vital tool, it can only mitigate bias effectively where the underlying sampling-density proxy itself remains informative.

Overall, the findings of this thesis support the integration of remote sensing in species distribution modelling. Point-based models like random forest and SINR remain valuable baselines when only standard environmental grids are available, and they can be surprisingly robust under distribution shift precisely because they do not overfit to local spatial structures. However, CNNs leveraging high-resolution remote sensing data offer a meaningfully more accurate and scalable approach for high-resolution mapping, and the small additional benefit observed from late fusion with a location encoder suggests that further performance gains are likely to come from richer multi-modal inputs. Future work should focus on hybrid architectures that explicitly combine macro-environmental constraints with deep remote sensing features, on pseudo-absence sampling strategies that better account for species range size and observer-effort structure, and on training procedures that more directly counteract the spatial reporting biases inherent to citizen-science registries.

Bibliography

[1] K. R. Shivanna, "Climate change and its impact on biodiversity and human welfare," *Proceedings of the Indian National Science Academy*, vol. 88, pp. 160–171, 2022. DOI: 10.1007/s43538-022-00073-6.

[2] F. Keck et al., "The global human impact on biodiversity," *Nature*, vol. 641, no. 8062, pp. 395–400, May 2025. DOI: 10.1038/s41586-025-08752-2.

[3] J. Elith and J. R. Leathwick, "Species distribution models: Ecological explanation and prediction across space and time," *Annual review of ecology, evolution, and systematics*, vol. 40, no. 1, pp. 677–697, 2009.

[4] D. Chen, Y. Xue, S. Chen, D. Fink, and C. Gomes, *Deep multi-species embedding*, 2017. arXiv: 1609.09353 [cs.LG].

[5] S. Beery, E. Cole, J. Parker, P. Perona, and K. Winner, *Species distribution modeling for machine learning practitioners: A review*, 2021. arXiv: 2107.10400 [cs.LG].

[6] IUCN. "The IUCN red list of threatened species. version 2025-2," [Online]. Available: https://www.iucnredlist.org (Accessed: 2025-04-22).

[7] A. H. Hurlbert and W. Jetz, "Species richness, hotspots, and the scale dependence of range maps in ecology and conservation," *Proceedings of the National Academy of Sciences*, vol. 104, no. 33, pp. 13 384–13 389, 2007. DOI: 10.1073/pnas.0704469104, PMCID: PMC1948922.

[8] S. E. Fick and R. J. Hijmans, "Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas," *International Journal of Climatology*, vol. 37, no. 12, pp. 4302–4315, 2017. DOI: 10.1002/joc.5086.

[9] iNaturalist, [Online]. Available: https://www.inaturalist.org/ (Accessed: 2026-05-07).

[10] B. L. Sullivan, C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling, "Ebird: A citizen-based bird observation network in the biological sciences," *Biological Conservation*, vol. 142, no. 10, pp. 2282–2292, 2009. DOI: 10.1016/j.biocon.2009.05.006.

[11] B. Deneu, M. Servajean, P. Bonnet, C. Botella, F. Munoz, and A. Joly, "Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment," *PLOS Computational Biology*, vol. 17, no. 4, e1008856, Apr. 2021. DOI: 10.1371/journal.pcbi.1008856.

[12] S. Artdatabanken. "Artportalen," [Online]. Available: https://www.artportalen.se/ (Accessed: 2026-05-11).

[13] E. Cole et al., *Spatial implicit neural representations for global-scale species mapping*, 2023. arXiv: 2306.02564 [cs.LG].

[14] W. Rucklidge. "AlphaEarth foundations satellite embeddings: Now available on Google Cloud Storage," [Online]. Available: https://medium.com/google-earth/alphaearth-foundations-satellite-embeddings-now-available-on-google-cloud-storage-f9ab0f7252d6 (Accessed: 2026-05-12).

[15] R. G. Pearson, T. P. Dawson, and C. Liu, "Modelling species distributions in britain: A hierarchical integration of climate and land-cover data," *Ecography*, vol. 27, no. 3, pp. 285–298, 2004. DOI: https://doi.org/10.1111/j.0906-7590.2004.03740.x.

[16] S. J. Phillips et al., "Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data," *Ecological Applications*, vol. 19, no. 1, pp. 181–197, Jan. 2009. DOI: 10.1890/07-2153.1.

[17] A. Johnston, N. Moran, A. Musgrove, D. Fink, and S. R. Baillie, "Estimating species distributions from spatially biased citizen science data," *Ecological Modelling*, vol. 422, p. 108 927, 2020. DOI: https://doi.org/10.1016/j.ecolmodel.2019.108927.

[18] J. Geldmann et al., "What determines spatial bias in citizen science? exploring four recording schemes with different proficiency requirements," *Diversity and Distributions*, vol. 22, no. 11, pp. 1139–1149, 2016. DOI: https://doi.org/10.1111/ddi.12477.

[19] M. Kosmala, A. Wiggins, A. Swanson, and B. Simmons, "Assessing data quality in citizen science," *Frontiers in Ecology and the Environment*, vol. 14, no. 10, pp. 551–560, 2016. DOI: https://doi.org/10.1002/fee.1436.

[20] M. Teng et al., *Satbird: Bird species distribution modeling with remote sensing and citizen science data*, 2023. arXiv: 2311.00936 [cs.LG].

[21] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: 1512.03385 [cs.CV].

[22] C. Lange, E. Cole, G. V. Horn, and O. M. Aodha, *Active learning-based species range estimation*, 2023. arXiv: 2311.02061 [cs.LG].

[23] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, 2015. arXiv: 1505.04597 [cs.CV].

[24] J. Long, E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, 2015. arXiv: 1411.4038 [cs.CV].

[25] P. Kramarczyk and B. Hejmanowska, "Unet neural network in agricultural land cover classification using sentinel-2," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-1/W3-2023, pp. 85–90, 2023. DOI: 10.5194/isprs-archives-XLVIII-1-W3-2023-85-2023.

[26] F. H. Wagner et al., "Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images," *Remote Sensing in Ecology and Conservation*, vol. 5, no. 4, pp. 360–375, 2019. DOI: https://doi.org/10.1002/rse2.111.

[27] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, *Unet++: A nested u-net architecture for medical image segmentation*, 2018. arXiv: 1807.10165 [cs.CV].

[28] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, Mar. 2024. DOI: 10.1038/s41598-024-56706-x.

[29] Eurostat – GISCO. "Countries." © EuroGeographics for the administrative boundaries, [Online]. Available: https://ec.europa.eu/eurostat/web/gisco/geodata/administrative-units/countries (Accessed: 2026-03-02).

[30] Lantmäteriet. "Topografi 50 nedladdning, vektor," [Online]. Available: https://geotorget.lantmateriet.se/geodataprodukter/topografi-50-nedladdning-vektor.

[31] SLU Artdatabanken. "Om artportalen," [Online]. Available: https://www.slu.se/artdatabanken/rapportering-och-fynd/artportalen/om/ (Accessed: 2026-05-07).

[32] S. Artdatabanken. "Slu artdatabankens utvecklarportal," [Online]. Available: https://api-portal.artdatabanken.se/ (Accessed: 2026-05-11).

[33] Miljødirektoratet and Naturvårdsverket. "Rovbase," [Online]. Available: https://www.rovbase.se/ (Accessed: 2026-05-07).

[34] WorldClim. "Bioclimatic variables," [Online]. Available: https://www.worldclim.org/data/bioclim.html (Accessed: 2026-05-07).

[35] C. F. Brown et al., *Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data*, 2025. arXiv: 2507.22291 [cs.CV].

[36] S. Marcel and Y. Rodriguez, *Torchvision: Pytorch's computer vision library*, 2016. [Online]. Available: https://github.com/pytorch/vision.

[37] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035.

[38] F. Dorm, C. Lange, S. Loarie, and O. M. Aodha, *Generating binary species range maps*, 2024. arXiv: 2408.15956 [q-bio.QM].

[39] R. G. Pearson, C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson, "Original article: Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in madagascar," *Journal of Biogeography*, vol. 34, no. 1, pp. 102–117, 2007. DOI: https://doi.org/10.1111/j.1365-2699.2006.01594.x.

A

Data specifications

A.1 Artportalen

This section describes the Artportalen data used in greater detail. Table A.1 shows the total observation count of the filtered and downsampled Artportalen data but before being split into train, validation and test sets. Table A.2 shows the final train, validation and test sets from Artportalen and represent the actual observations that are fed into the model for training and evaluation.

Table A.1: Number of observations per species in the filtered and downsampled Artportalen dataset.

Scientific name	Common name	Observations
Vulpes vulpes	Red fox	20 000
Capreolus capreolus	Roe deer	20 000
Sciurus vulgaris	Red squirrel	20 000
Lutra lutra	Otter	20 000
Lepus europaeus	European hare	19 274
Alces alces	Moose	16 168
Erinaceus europaeus	Hedgehog	14 153
Neovison vison	Mink	11 620
Castor fiber	Beaver	10 340
Lepus timidus	Mountain hare	9 200
Dama dama	Fallow deer	8 945
Sus scrofa	Wild boar	7 785
Meles meles	Badger	5 614
Oryctolagus cuniculus	European rabbit	4 248
Martes martes	Pine marten	3 842
Muscardinus avellanarius	Hazel dormouse	3 707
Cervus elaphus	Red deer	3 458
Rattus norvegicus	Brown rat	2 363
Lynx lynx	Lynx	2 188
Mustela nivalis	Least weasel	1 969
Sorex araneus	Common shrew	1 938

Continued on next page

Table A.1 – *continued*

Scientific name	Common name	Observations
Myodes glareolus	Bank vole	1 912
Mustela erminea	Stoat	1 715
Apodemus flavicollis	Yellow-necked mouse	1 502
Ursus arctos	Brown bear	1 329
Arvicola amphibius	Water vole	1 112
Canis lupus	Wolf	1 101
Microtus agrestis	Field vole	999
Apodemus sylvaticus	Wood mouse	717
Ovis aries musimon	Mouflon sheep	698
Ondatra zibethicus	Muskrat	686
Mus musculus	House mouse	653
Mustela putorius	Polecat	577
Lemmus lemmus	Norway lemming	542
Total		220 355

Table A.2: Number of presence observations per species and dataset split for the Artportalen data. Percentages indicate each split's share of the species total.

Species	Train	Validation	Test	Total
Alces alces (moose)	9 776 (67%)	2 587 (18%)	2 277 (16%)	14 640
Apodemus flavicollis (yellow-necked mouse)	730 (58%)	285 (23%)	243 (19%)	1 258
Apodemus sylvaticus (wood mouse)	334 (55%)	171 (28%)	105 (17%)	610
Arvicola amphibius (water vole)	606 (61%)	193 (19%)	193 (19%)	992
Canis lupus (wolf)	673 (66%)	197 (19%)	153 (15%)	1 023
Capreolus capreolus (roe deer)	11 088 (64%)	3 464 (20%)	2 756 (16%)	17 308
Castor fiber (beaver)	5 461 (57%)	2 409 (25%)	1 719 (18%)	9 589
Cervus elaphus (red deer)	2 367 (76%)	469 (15%)	272 (9%)	3 108
Dama dama (fallow deer)	5 970 (74%)	1 303 (16%)	748 (9%)	8 021
Erinaceus europaeus (hedgehog)	7 246 (60%)	3 389 (28%)	1 379 (11%)	12 014
Lemmus lemmus (Norway lemming)	415 (81%)	78 (15%)	21 (4%)	514

Continued on next page

(continued)

Species	Train	Validation	Test	Total
Lepus europaeus (European hare)	9 841 (59%)	4 556 (28%)	2 165 (13%)	16 562
Lepus timidus (mountain hare)	5 649 (71%)	893 (11%)	1 428 (18%)	7 970
Lutra lutra (otter)	12 923 (71%)	3 297 (18%)	1 976 (11%)	18 196
Lynx lynx (lynx)	1 297 (64%)	425 (21%)	318 (16%)	2 040
Martes martes (pine marten)	2 195 (64%)	778 (23%)	451 (13%)	3 424
Meles meles (badger)	3 247 (65%)	1 087 (22%)	672 (13%)	5 006
Microtus agrestis (field vole)	643 (71%)	132 (15%)	133 (15%)	908
Mus musculus (house mouse)	313 (65%)	111 (23%)	54 (11%)	478
Muscardinus avellanarius (hazel dormouse)	2 256 (64%)	1 241 (35%)	40 (1%)	3 537
Mustela erminea (stoat)	1 098 (71%)	235 (15%)	215 (14%)	1 548
Mustela nivalis (least weasel)	1 134 (67%)	368 (22%)	187 (11%)	1 689
Mustela putorius (polecat)	353 (66%)	153 (29%)	27 (5%)	533
Myodes glareolus (bank vole)	1 052 (61%)	435 (25%)	231 (13%)	1 718
Neovison vison (mink)	7 162 (70%)	1 896 (18%)	1 192 (12%)	10 250
Ondatra zibethicus (muskrat)	594 (91%)	10 (2%)	50 (8%)	654
Oryctolagus cuniculus (European rabbit)	2 144 (72%)	818 (27%)	29 (1%)	2 991
Ovis aries musimon (mouflon sheep)	614 (91%)	48 (7%)	11 (2%)	673
Rattus norvegicus (brown rat)	778 (40%)	1 019 (53%)	142 (7%)	1 939
Sciurus vulgaris (red squirrel)	10 346 (58%)	4 267 (24%)	3 276 (18%)	17 889
Sorex araneus (common shrew)	1 333 (74%)	265 (15%)	198 (11%)	1 796
Sus scrofa (wild boar)	4 535 (63%)	1 582 (22%)	1 113 (15%)	7 230

Continued on next page

(continued)

Species	Train	Validation	Test	Total
Ursus arctos (brown bear)	871 (74%)	150 (13%)	153 (13%)	1 174
Vulpes vulpes (red fox)	10 883 (64%)	3 757 (22%)	2 378 (14%)	17 018
Total	125 927 **(65%)**	42 068 **(22%)**	26 305 **(14%)**	194 300

A.2 WorldClim Bioclimatic variables

Table A.3 below containes the definitions of the 19 bioclimatic variables from WorldClim [34]. These are also visualized together with the elevation in Figure A.1.

Table A.3: The 19 bioclimatic variables and their definitions.

Code	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (mean of monthly max – min temperature)
BIO3	Isothermality (BIO2 / BIO7 × 100)
BIO4	Temperature Seasonality (standard deviation × 100)
BIO5	Maximum Temperature of Warmest Month
BIO6	Minimum Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5 – BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (coefficient of variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

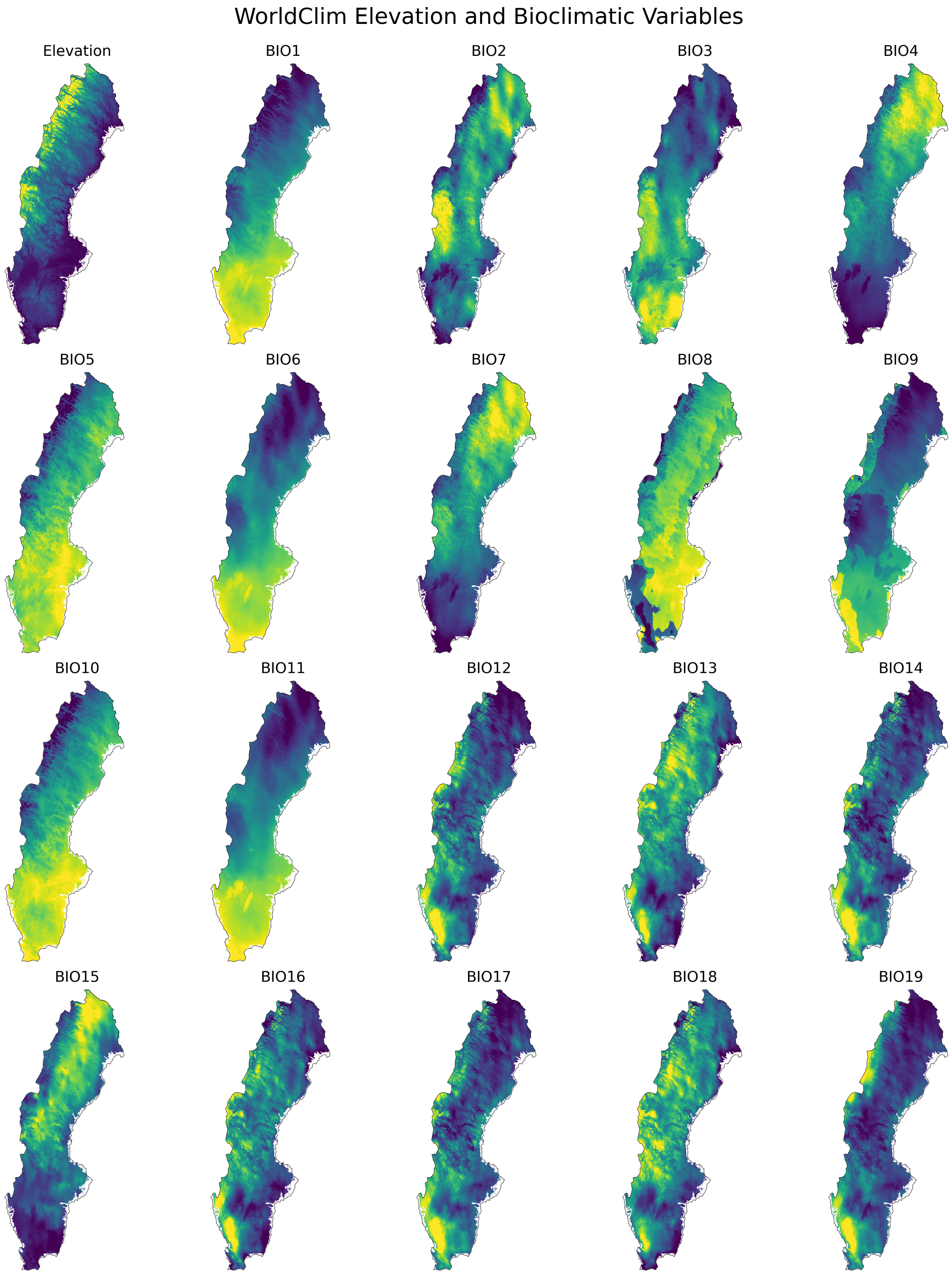


Figure A.1: Visualization of WorldClim elevation and bioclimatic variables.

B

Prediction Map Examples

The following appendix presents examples of prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*) for selected models. These are the two baseline models RF (Figure B.1) and SINR (Figure B.2), one environmental input CNN in ResNet-18$_{\mathrm{Env}}$ (Figure B.3), one satellite embedding input CNN in U-Net$_{\mathrm{Sat}}$ (Figure B.4) and one fusion model in Fus-Res$_{\mathrm{LocSat}}$ (Figure B.5).

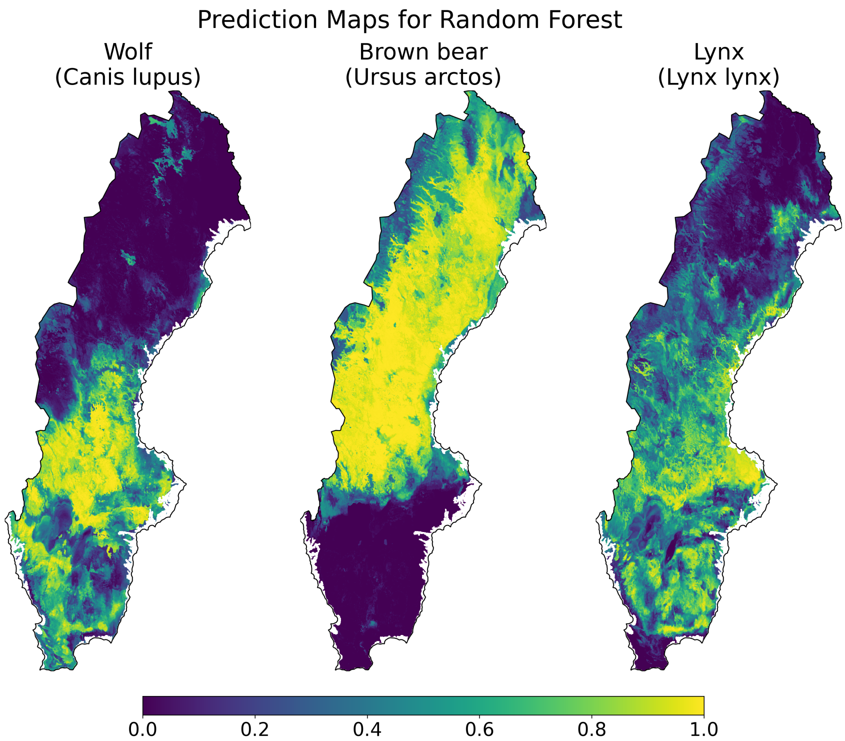


Figure B.1: RF prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

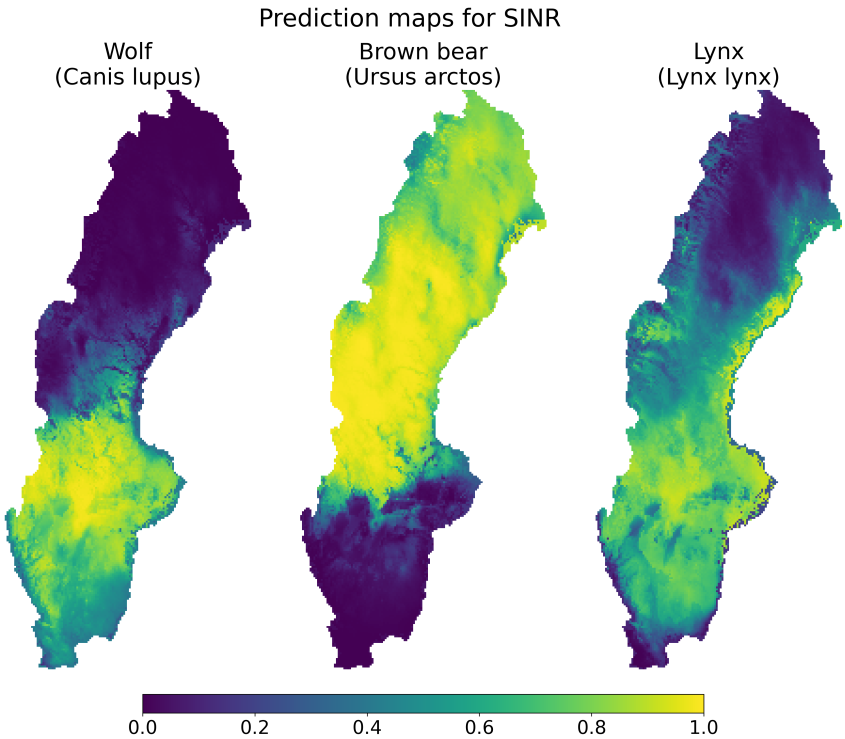


Figure B.2: SINR prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

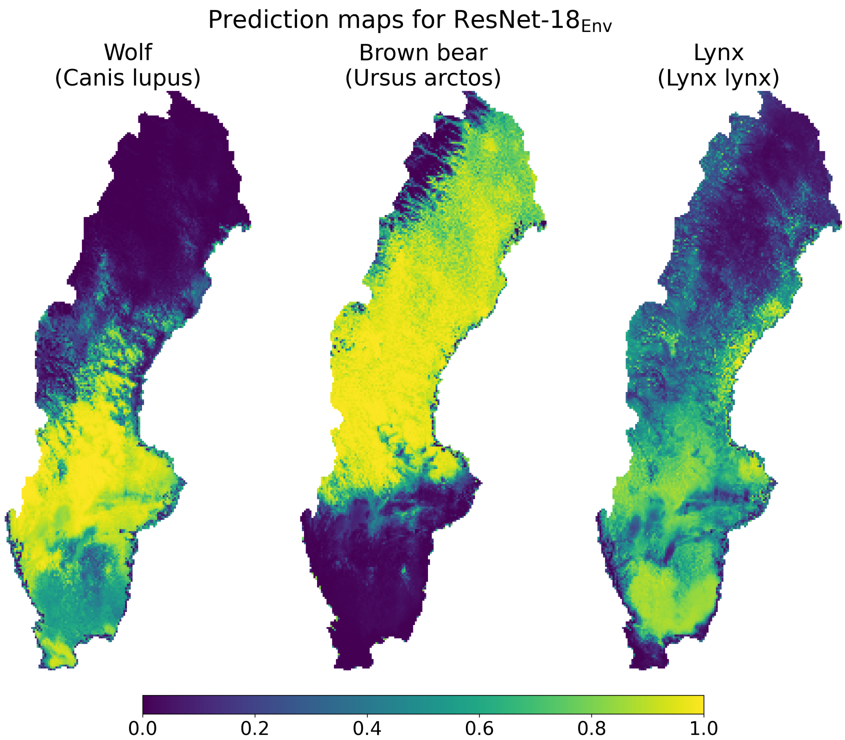


Figure B.3: ResNet-18$_{\mathrm{Env}}$ prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

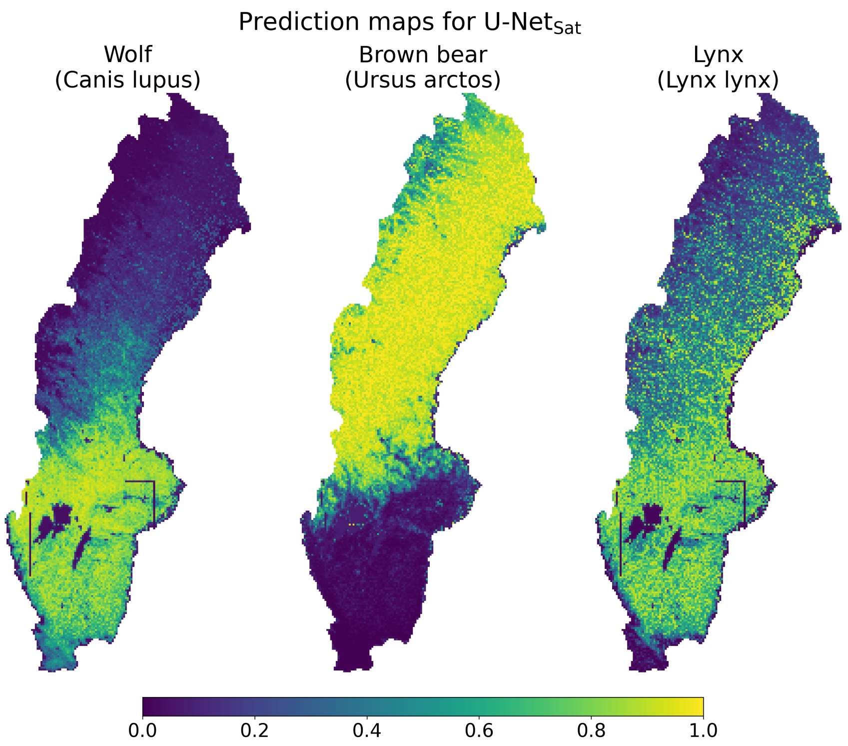


Figure B.4: U-Net$_{\mathrm{Sat}}$ prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

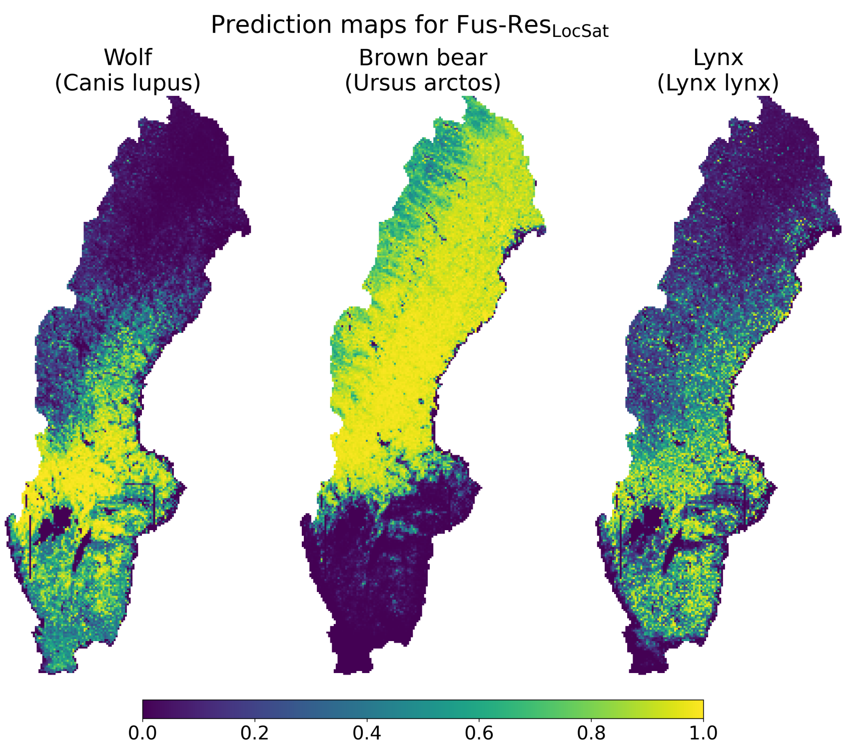


Figure B.5: Fus-Res$_{\mathrm{LocSat}}$ prediction maps for wolf (*Canis lupus*), brown bear (*Ursus arctos*) and lynx (*Lynx lynx*).

C

Complete results

C.1 RF (Env)

Table C.1: RF (Env) — Test set: Artportalen.

Species	**Thr.**	**Acc.**	**Prec.**	**Rec.**	**F1**	**AP**
Combined	—	0.777	0.889	0.620	0.730	0.909
Erinaceus europaeus (hedgehog)	0.540	0.824	0.981	0.614	0.755	0.976
Lynx lynx (lynx)	0.490	0.928	0.947	0.910	0.928	0.967
Lutra lutra (otter)	0.540	0.643	0.711	0.495	0.584	0.670
Muscardinus avellanarius (hazel dormouse)	0.590	0.731	0.913	0.525	0.667	0.954
Ursus arctos (brown bear)	0.570	0.840	0.794	0.941	0.861	0.957
Sciurus vulgaris (red squirrel)	0.530	0.570	0.880	0.099	0.177	0.883
Castor fiber (beaver)	0.580	0.909	0.902	0.923	0.912	0.934
Sorex araneus (common shrew)	0.550	0.730	0.916	0.481	0.630	0.850
Lepus europaeus (European hare)	0.520	0.985	0.995	0.976	0.986	0.998
Lepus timidus (mountain hare)	0.530	0.680	0.698	0.557	0.620	0.671
Oryctolagus cuniculus (European rabbit)	0.580	0.593	1.000	0.214	0.353	0.973
Rattus norvegicus (brown rat)	0.600	0.717	0.944	0.415	0.576	0.934
Apodemus flavicollis (yellow-necked mouse)	0.550	0.963	0.970	0.958	0.964	0.984

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Apodemus sylvaticus (wood mouse)	0.530	0.956	0.942	0.970	0.956	0.978
Mus musculus (house mouse)	0.570	0.714	0.808	0.477	0.600	0.785
Microtus agrestis (field vole)	0.580	0.849	0.937	0.736	0.824	0.951
Myodes glareolus (bank vole)	0.600	0.725	0.980	0.451	0.617	0.909
Arvicola amphibius (water vole)	0.590	0.902	0.939	0.856	0.896	0.962
Lemmus lemmus (Norway lemming)	0.580	0.946	0.952	0.952	0.952	0.993
Vulpes vulpes (red fox)	0.510	0.738	0.889	0.536	0.669	0.884
Mustela erminea (stoat)	0.560	0.618	0.642	0.544	0.589	0.639
Mustela nivalis (least weasel)	0.580	0.712	0.920	0.458	0.611	0.909
Neovison vison (mink)	0.500	0.657	0.650	0.601	0.625	0.666
Mustela putorius (polecat)	0.600	0.962	1.000	0.926	0.962	0.996
Martes martes (pine marten)	0.540	0.758	0.926	0.543	0.685	0.893
Meles meles (badger)	0.520	0.897	0.943	0.833	0.884	0.955
Cervus elaphus (red deer)	0.520	0.805	0.988	0.631	0.770	0.974
Dama dama (fallow deer)	0.550	0.881	0.996	0.750	0.856	0.995
Capreolus capreolus (roe deer)	0.490	0.865	0.960	0.754	0.845	0.962
Alces alces (moose)	0.530	0.701	0.856	0.464	0.601	0.805
Ondatra zibethicus (muskrat)	0.610	0.892	0.862	0.862	0.862	0.948
Ovis aries musimon (mouflon sheep)	0.640	0.632	1.000	0.125	0.222	1.000
Canis lupus (wolf)	0.540	0.826	0.980	0.662	0.790	0.969
Sus scrofa (wild boar)	0.600	0.995	0.992	0.999	0.995	0.998

Table C.2: RF (Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.826	0.864	0.791	0.826	0.892
Lynx lynx (lynx)	0.490	0.707	0.773	0.624	0.691	0.783
Ursus arctos (brown bear)	0.570	0.931	0.922	0.950	0.935	0.964
Canis lupus (wolf)	0.540	0.854	0.889	0.817	0.852	0.930

C.2 SINR (Loc+Env)

Table C.3: SINR (Loc+Env) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.839	0.782	0.939	0.853	0.919
Erinaceus europaeus (hedgehog)	0.540	0.917	0.956	0.873	0.913	0.966
Lynx lynx (lynx)	0.370	0.891	0.830	0.984	0.901	0.969
Lutra lutra (otter)	0.324	0.651	0.593	0.960	0.733	0.649
Muscardinus avellanarius (hazel dormouse)	0.903	0.650	1.000	0.300	0.462	0.963
Ursus arctos (brown bear)	0.750	0.843	0.793	0.928	0.855	0.982
Sciurus vulgaris (red squirrel)	0.273	0.915	0.878	0.965	0.919	0.941
Castor fiber (beaver)	0.573	0.902	0.845	0.985	0.909	0.949
Sorex araneus (common shrew)	0.332	0.942	0.935	0.950	0.942	0.955
Lepus europaeus (European hare)	0.739	0.952	0.976	0.927	0.951	0.988
Lepus timidus (mountain hare)	0.286	0.578	0.544	0.973	0.698	0.834
Oryctolagus cuniculus (European rabbit)	0.742	0.603	1.000	0.207	0.343	0.973
Rattus norvegicus (brown rat)	0.526	0.912	1.000	0.824	0.903	0.998
Apodemus flavicollis	0.488	0.977	0.968	0.988	0.978	0.952

Continued on next page

(continued)

Species	**Thr.**	**Acc.**	**Prec.**	**Rec.**	**F1**	**AP**
(yellow-necked mouse)						
Apodemus sylvaticus (wood mouse)	0.681	0.967	0.971	0.962	0.967	0.962
Mus musculus (house mouse)	0.496	0.815	0.827	0.796	0.811	0.867
Microtus agrestis (field vole)	0.209	0.906	0.860	0.970	0.912	0.905
Myodes glareolus (bank vole)	0.302	0.816	0.792	0.857	0.823	0.870
Arvicola amphibius (water vole)	0.299	0.834	0.768	0.959	0.853	0.898
Lemmus lemmus (Norway lemming)	0.959	0.952	1.000	0.905	0.950	0.996
Vulpes vulpes (red fox)	0.239	0.807	0.737	0.955	0.832	0.915
Mustela erminea (stoat)	0.289	0.553	0.531	0.907	0.670	0.621
Mustela nivalis (least weasel)	0.345	0.845	0.774	0.973	0.863	0.899
Neovison vison (mink)	0.321	0.636	0.582	0.965	0.726	0.631
Mustela putorius (polecat)	0.694	0.907	0.958	0.852	0.902	0.994
Martes martes (pine marten)	0.392	0.807	0.759	0.900	0.824	0.875
Meles meles (badger)	0.333	0.931	0.919	0.945	0.932	0.966
Cervus elaphus (red deer)	0.426	0.958	0.977	0.938	0.957	0.972
Dama dama (fallow deer)	0.807	0.914	1.000	0.829	0.906	1.000
Capreolus capreolus (roe deer)	0.354	0.931	0.922	0.941	0.932	0.978
Alces alces (moose)	0.295	0.760	0.684	0.966	0.801	0.799
Ondatra zibethicus (muskrat)	0.749	0.910	0.887	0.940	0.913	0.985
Ovis aries musimon (mouflon sheep)	0.484	1.000	1.000	1.000	1.000	1.000
Canis lupus (wolf)	0.432	0.987	0.987	0.987	0.987	0.993
Sus scrofa	0.904	0.907	0.989	0.822	0.898	0.994

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
(wild boar)						

Table C.4: SINR (Loc+Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.827	0.787	0.897	0.838	0.877
Lynx lynx (lynx)	0.370	0.701	0.664	0.815	0.732	0.748
Ursus arctos (brown bear)	0.750	0.928	0.891	0.975	0.931	0.951
Canis lupus (wolf)	0.432	0.867	0.838	0.909	0.872	0.931

C.3 ResNet18 (Env)

Table C.5: ResNet18 (Env) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.832	0.800	0.887	0.841	0.924
Erinaceus europaeus (hedgehog)	0.740	0.945	0.967	0.921	0.943	0.986
Lynx lynx (lynx)	0.338	0.857	0.804	0.943	0.868	0.918
Lutra lutra (otter)	0.290	0.648	0.598	0.899	0.719	0.742
Muscardinus avellanarius (hazel dormouse)	0.881	0.775	0.958	0.575	0.719	0.982
Ursus arctos (brown bear)	0.759	0.944	0.936	0.954	0.945	0.955
Sciurus vulgaris (red squirrel)	0.308	0.915	0.913	0.918	0.915	0.921
Castor fiber (beaver)	0.636	0.875	0.846	0.917	0.880	0.960
Sorex araneus (common shrew)	0.301	0.899	0.876	0.929	0.902	0.954
Lepus europaeus (European hare)	0.727	0.923	0.975	0.868	0.918	0.990
Lepus timidus (mountain hare)	0.283	0.674	0.612	0.951	0.745	0.850

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Oryctolagus cuniculus (European rabbit)	0.631	0.603	1.000	0.207	0.343	0.976
Rattus norvegicus (brown rat)	0.574	0.799	1.000	0.599	0.749	0.991
Apodemus flavicollis (yellow-necked mouse)	0.448	0.967	0.963	0.971	0.967	0.988
Apodemus sylvaticus (wood mouse)	0.604	0.948	0.980	0.914	0.946	0.993
Mus musculus (house mouse)	0.750	0.778	0.917	0.611	0.733	0.855
Microtus agrestis (field vole)	0.183	0.891	0.861	0.932	0.895	0.933
Myodes glareolus (bank vole)	0.265	0.775	0.741	0.844	0.789	0.839
Arvicola amphibius (water vole)	0.184	0.860	0.817	0.927	0.869	0.949
Lemmus lemmus (Norway lemming)	0.914	0.905	1.000	0.809	0.895	0.998
Vulpes vulpes (red fox)	0.217	0.825	0.780	0.905	0.838	0.905
Mustela erminea (stoat)	0.196	0.660	0.604	0.935	0.734	0.655
Mustela nivalis (least weasel)	0.262	0.874	0.833	0.936	0.882	0.899
Neovison vison (mink)	0.283	0.698	0.637	0.921	0.753	0.747
Mustela putorius (polecat)	0.620	0.870	0.955	0.778	0.857	0.990
Martes martes (pine marten)	0.381	0.784	0.759	0.832	0.794	0.838
Meles meles (badger)	0.339	0.897	0.912	0.878	0.895	0.955
Cervus elaphus (red deer)	0.452	0.912	0.979	0.842	0.905	0.953
Dama dama (fallow deer)	0.856	0.865	1.000	0.730	0.844	0.997
Capreolus capreolus (roe deer)	0.361	0.908	0.964	0.848	0.902	0.978
Alces alces (moose)	0.297	0.700	0.638	0.921	0.754	0.863
Ondatra zibethicus (muskrat)	0.832	0.780	0.868	0.660	0.750	0.893

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Ovis aries musimon (mouflon sheep)	0.774	0.909	1.000	0.818	0.900	1.000
Canis lupus (wolf)	0.531	0.971	0.987	0.954	0.970	0.993
Sus scrofa (wild boar)	0.880	0.872	0.988	0.752	0.854	0.988

Table C.6: ResNet18 (Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.817	0.788	0.867	0.826	0.875
Lynx lynx (lynx)	0.338	0.682	0.659	0.755	0.703	0.737
Ursus arctos (brown bear)	0.759	0.926	0.905	0.953	0.928	0.952
Canis lupus (wolf)	0.531	0.859	0.826	0.909	0.866	0.937

C.4 ResNet18 (Sat)

Table C.7: ResNet18 (Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.900	0.896	0.904	0.900	0.966
Erinaceus europaeus (hedgehog)	0.798	0.958	0.991	0.923	0.956	0.998
Lynx lynx (lynx)	0.349	0.868	0.863	0.874	0.869	0.955
Lutra lutra (otter)	0.326	0.852	0.847	0.859	0.853	0.934
Muscardinus avellanarius (hazel dormouse)	0.849	0.863	0.939	0.775	0.849	0.945
Ursus arctos (brown bear)	0.700	0.944	0.947	0.941	0.944	0.984
Sciurus vulgaris (red squirrel)	0.351	0.932	0.944	0.918	0.931	0.981
Castor fiber (beaver)	0.649	0.942	0.950	0.932	0.941	0.987
Sorex araneus	0.192	0.886	0.909	0.859	0.883	0.973

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
(common shrew)						
Lepus europaeus (European hare)	0.876	0.940	0.992	0.888	0.937	0.998
Lepus timidus (mountain hare)	0.224	0.810	0.747	0.938	0.832	0.931
Oryctolagus cuniculus (European rabbit)	0.697	0.724	1.000	0.448	0.619	1.000
Rattus norvegicus (brown rat)	0.878	0.919	1.000	0.838	0.912	1.000
Apodemus flavicollis (yellow-necked mouse)	0.373	0.990	0.992	0.988	0.990	0.998
Apodemus sylvaticus (wood mouse)	0.687	0.957	0.971	0.943	0.957	0.994
Mus musculus (house mouse)	0.670	0.870	0.976	0.759	0.854	0.983
Microtus agrestis (field vole)	0.503	0.940	0.961	0.917	0.939	0.967
Myodes glareolus (bank vole)	0.420	0.818	0.821	0.814	0.817	0.846
Arvicola amphibius (water vole)	0.273	0.917	0.889	0.953	0.920	0.948
Lemmus lemmus (Norway lemming)	0.847	0.905	0.870	0.952	0.909	0.985
Vulpes vulpes (red fox)	0.331	0.878	0.857	0.908	0.882	0.926
Mustela erminea (stoat)	0.392	0.837	0.811	0.879	0.844	0.922
Mustela nivalis (least weasel)	0.267	0.904	0.879	0.936	0.907	0.961
Neovison vison (mink)	0.403	0.864	0.861	0.869	0.865	0.931
Mustela putorius (polecat)	0.780	0.926	0.926	0.926	0.926	0.975
Martes martes (pine marten)	0.300	0.817	0.766	0.913	0.833	0.893
Meles meles (badger)	0.408	0.924	0.923	0.926	0.924	0.970
Cervus elaphus (red deer)	0.487	0.888	0.969	0.801	0.877	0.986
Dama dama (fallow deer)	0.910	0.920	0.998	0.841	0.913	0.997
Capreolus capreolus	0.585	0.929	0.975	0.881	0.926	0.982

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
(roe deer)						
Alces alces (moose)	0.200	0.825	0.759	0.954	0.845	0.894
Ondatra zibethicus (muskrat)	0.873	0.940	0.940	0.940	0.940	0.988
Ovis aries musimon (mouflon sheep)	0.918	0.591	1.000	0.182	0.308	1.000
Canis lupus (wolf)	0.709	0.987	0.987	0.987	0.987	0.996
Sus scrofa (wild boar)	0.952	0.955	0.991	0.917	0.953	0.998

Table C.8: ResNet18 (Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.855	0.837	0.883	0.859	0.914
Lynx lynx (lynx)	0.349	0.775	0.757	0.812	0.783	0.827
Ursus arctos (brown bear)	0.700	0.935	0.906	0.971	0.937	0.968
Canis lupus (wolf)	0.709	0.864	0.857	0.874	0.865	0.947

C.5 U-Net (Env)

Table C.9: U-Net (Env) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.837	0.781	0.936	0.851	0.934
Erinaceus europaeus (hedgehog)	0.747	0.848	0.949	0.737	0.829	0.975
Lynx lynx (lynx)	0.341	0.859	0.797	0.962	0.872	0.934
Lutra lutra (otter)	0.353	0.699	0.634	0.940	0.758	0.803
Muscardinus avellanarius (hazel dormouse)	0.765	0.787	0.926	0.625	0.746	0.909
Ursus arctos (brown bear)	0.607	0.879	0.837	0.941	0.886	0.978

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Sciurus vulgaris (red squirrel)	0.326	0.917	0.895	0.946	0.920	0.971
Castor fiber (beaver)	0.454	0.816	0.739	0.976	0.841	0.948
Sorex araneus (common shrew)	0.432	0.922	0.920	0.924	0.922	0.941
Lepus europaeus (European hare)	0.874	0.981	0.985	0.977	0.981	0.995
Lepus timidus (mountain hare)	0.286	0.521	0.512	0.924	0.659	0.833
Oryctolagus cuniculus (European rabbit)	0.764	0.966	1.000	0.931	0.964	0.998
Rattus norvegicus (brown rat)	0.652	0.926	0.984	0.866	0.921	0.997
Apodemus flavicollis (yellow-necked mouse)	0.589	0.982	0.968	0.996	0.982	0.991
Apodemus sylvaticus (wood mouse)	0.437	0.962	0.945	0.981	0.963	0.996
Mus musculus (house mouse)	0.658	0.824	0.818	0.833	0.826	0.920
Microtus agrestis (field vole)	0.262	0.876	0.833	0.940	0.883	0.942
Myodes glareolus (bank vole)	0.342	0.823	0.799	0.862	0.829	0.818
Arvicola amphibius (water vole)	0.276	0.842	0.775	0.964	0.859	0.935
Lemmus lemmus (Norway lemming)	0.797	0.881	0.833	0.952	0.889	0.974
Vulpes vulpes (red fox)	0.273	0.797	0.730	0.944	0.823	0.920
Mustela erminea (stoat)	0.294	0.560	0.534	0.944	0.682	0.699
Mustela nivalis (least weasel)	0.281	0.861	0.792	0.979	0.876	0.939
Neovison vison (mink)	0.412	0.709	0.639	0.958	0.767	0.772
Mustela putorius (polecat)	0.787	0.796	0.900	0.667	0.766	0.940
Martes martes (pine marten)	0.361	0.850	0.780	0.976	0.867	0.934
Meles meles (badger)	0.445	0.921	0.899	0.949	0.923	0.962

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Cervus elaphus (red deer)	0.497	0.949	0.959	0.938	0.948	0.980
Dama dama (fallow deer)	0.900	0.995	1.000	0.989	0.995	1.000
Capreolus capreolus (roe deer)	0.435	0.911	0.906	0.917	0.912	0.977
Alces alces (moose)	0.273	0.728	0.663	0.923	0.772	0.853
Ondatra zibethicus (muskrat)	0.737	0.940	0.958	0.920	0.939	0.976
Ovis aries musimon (mouflon sheep)	0.732	0.909	1.000	0.818	0.900	1.000
Canis lupus (wolf)	0.486	0.984	0.987	0.980	0.984	0.993
Sus scrofa (wild boar)	0.929	0.990	0.991	0.989	0.990	0.968

Table C.10: U-Net (Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.831	0.788	0.907	0.843	0.841
Lynx lynx (lynx)	0.341	0.720	0.676	0.842	0.750	0.747
Ursus arctos (brown bear)	0.607	0.924	0.884	0.976	0.928	0.878
Canis lupus (wolf)	0.486	0.864	0.833	0.910	0.870	0.899

C.6 U-Net (Sat)

Table C.11: U-Net (Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.903	0.875	0.942	0.907	0.961
Erinaceus europaeus (hedgehog)	0.906	0.967	0.992	0.941	0.966	0.998
Lynx lynx (lynx)	0.416	0.906	0.849	0.987	0.913	0.943
Lutra lutra	0.415	0.856	0.805	0.938	0.867	0.922

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
(otter)						
Muscardinus avellanarius (hazel dormouse)	0.917	0.900	0.944	0.850	0.895	0.940
Ursus arctos (brown bear)	0.741	0.915	0.894	0.941	0.917	0.984
Sciurus vulgaris (red squirrel)	0.586	0.935	0.912	0.964	0.937	0.989
Castor fiber (beaver)	0.689	0.957	0.938	0.979	0.958	0.990
Sorex araneus (common shrew)	0.585	0.927	0.925	0.929	0.927	0.984
Lepus europaeus (European hare)	0.943	0.962	0.992	0.931	0.961	0.998
Lepus timidus (mountain hare)	0.308	0.734	0.662	0.957	0.782	0.930
Oryctolagus cuniculus (European rabbit)	0.888	0.759	1.000	0.517	0.682	1.000
Rattus norvegicus (brown rat)	0.903	0.968	1.000	0.937	0.967	1.000
Apodemus flavicollis (yellow-necked mouse)	0.809	0.979	0.968	0.992	0.980	0.998
Apodemus sylvaticus (wood mouse)	0.755	0.986	0.972	1.000	0.986	0.996
Mus musculus (house mouse)	0.868	0.954	0.930	0.982	0.955	0.978
Microtus agrestis (field vole)	0.469	0.925	0.895	0.962	0.927	0.964
Myodes glareolus (bank vole)	0.514	0.887	0.835	0.965	0.896	0.928
Arvicola amphibius (water vole)	0.491	0.899	0.857	0.959	0.905	0.953
Lemmus lemmus (Norway lemming)	0.772	0.786	0.714	0.952	0.816	0.765
Vulpes vulpes (red fox)	0.447	0.890	0.853	0.942	0.895	0.962
Mustela erminea (stoat)	0.417	0.828	0.758	0.963	0.848	0.950
Mustela nivalis (least weasel)	0.312	0.861	0.787	0.989	0.877	0.918
Neovison vison (mink)	0.560	0.901	0.865	0.949	0.905	0.927
Mustela putorius	0.850	0.907	0.923	0.889	0.906	0.937

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
(polecat)						
Martes martes (pine marten)	0.319	0.856	0.788	0.973	0.871	0.924
Meles meles (badger)	0.555	0.949	0.953	0.944	0.948	0.985
Cervus elaphus (red deer)	0.714	0.888	0.924	0.846	0.883	0.971
Dama dama (fallow deer)	0.960	0.895	0.998	0.791	0.883	0.998
Capreolus capreolus (roe deer)	0.714	0.953	0.975	0.929	0.952	0.990
Alces alces (moose)	0.386	0.806	0.731	0.969	0.833	0.879
Ondatra zibethicus (muskrat)	0.901	0.920	0.938	0.900	0.918	0.973
Ovis aries musimon (mouflon sheep)	0.794	1.000	1.000	1.000	1.000	1.000
Canis lupus (wolf)	0.415	0.987	0.987	0.987	0.987	0.994
Sus scrofa (wild boar)	0.947	0.926	0.982	0.868	0.921	0.997

Table C.12: U-Net (Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.862	0.826	0.918	0.869	0.914
Lynx lynx (lynx)	0.416	0.770	0.726	0.868	0.791	0.874
Ursus arctos (brown bear)	0.741	0.940	0.910	0.976	0.942	0.943
Canis lupus (wolf)	0.415	0.887	0.866	0.914	0.890	0.927

C.7 U-Net++ (Env)

Table C.13: U-Net++ (Env) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.827	0.781	0.909	0.840	0.917

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Erinaceus europaeus (hedgehog)	0.512	0.850	0.978	0.717	0.827	0.975
Lynx lynx (lynx)	0.271	0.868	0.806	0.969	0.880	0.832
Lutra lutra (otter)	0.287	0.619	0.579	0.872	0.696	0.717
Muscardinus avellanarius (hazel dormouse)	0.788	0.662	0.882	0.375	0.526	0.901
Ursus arctos (brown bear)	0.619	0.850	0.796	0.941	0.862	0.978
Sciurus vulgaris (red squirrel)	0.255	0.908	0.911	0.905	0.908	0.969
Castor fiber (beaver)	0.484	0.864	0.806	0.958	0.876	0.955
Sorex araneus (common shrew)	0.312	0.924	0.947	0.899	0.922	0.947
Lepus europaeus (European hare)	0.824	0.969	0.992	0.946	0.969	0.988
Lepus timidus (mountain hare)	0.215	0.517	0.509	0.917	0.655	0.826
Oryctolagus cuniculus (European rabbit)	0.712	0.741	1.000	0.483	0.651	0.989
Rattus norvegicus (brown rat)	0.497	0.919	0.992	0.845	0.912	0.995
Apodemus flavicollis (yellow-necked mouse)	0.528	0.975	0.960	0.992	0.976	0.976
Apodemus sylvaticus (wood mouse)	0.622	0.967	0.962	0.971	0.967	0.988
Mus musculus (house mouse)	0.562	0.759	0.769	0.741	0.755	0.836
Microtus agrestis (field vole)	0.207	0.902	0.869	0.947	0.906	0.926
Myodes glareolus (bank vole)	0.280	0.814	0.780	0.875	0.825	0.830
Arvicola amphibius (water vole)	0.232	0.824	0.747	0.979	0.848	0.933
Lemmus lemmus (Norway lemming)	0.862	0.929	0.909	0.952	0.930	0.978
Vulpes vulpes (red fox)	0.240	0.831	0.770	0.943	0.848	0.924
Mustela erminea (stoat)	0.260	0.539	0.525	0.819	0.640	0.644

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela nivalis (least weasel)	0.245	0.866	0.804	0.968	0.879	0.912
Neovison vison (mink)	0.305	0.642	0.591	0.924	0.721	0.722
Mustela putorius (polecat)	0.680	0.963	0.931	1.000	0.964	0.962
Martes martes (pine marten)	0.247	0.802	0.743	0.922	0.823	0.909
Meles meles (badger)	0.314	0.908	0.902	0.917	0.909	0.960
Cervus elaphus (red deer)	0.444	0.958	0.966	0.949	0.957	0.951
Dama dama (fallow deer)	0.887	0.971	1.000	0.943	0.970	1.000
Capreolus capreolus (roe deer)	0.371	0.904	0.939	0.864	0.900	0.977
Alces alces (moose)	0.245	0.724	0.655	0.948	0.775	0.814
Ondatra zibethicus (muskrat)	0.711	0.870	0.878	0.860	0.869	0.881
Ovis aries musimon (mouflon sheep)	0.703	0.955	1.000	0.909	0.952	1.000
Canis lupus (wolf)	0.451	0.938	0.899	0.987	0.941	0.994
Sus scrofa (wild boar)	0.915	0.979	0.998	0.961	0.979	0.998

Table C.14: U-Net++ (Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.819	0.770	0.910	0.834	0.848
Lynx lynx (lynx)	0.271	0.695	0.653	0.833	0.732	0.684
Ursus arctos (brown bear)	0.619	0.925	0.882	0.982	0.929	0.946
Canis lupus (wolf)	0.451	0.852	0.807	0.925	0.862	0.915

C.8 U-Net++ (Sat)

Table C.15: U-Net++ (Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.876	0.837	0.935	0.883	0.960
Erinaceus europaeus (hedgehog)	0.730	0.959	0.983	0.935	0.958	0.997
Lynx lynx (lynx)	0.258	0.890	0.835	0.972	0.898	0.934
Lutra lutra (otter)	0.441	0.764	0.703	0.913	0.794	0.822
Muscardinus avellanarius (hazel dormouse)	0.904	0.900	0.944	0.850	0.895	0.931
Ursus arctos (brown bear)	0.646	0.967	0.980	0.954	0.967	0.990
Sciurus vulgaris (red squirrel)	0.331	0.913	0.884	0.950	0.916	0.983
Castor fiber (beaver)	0.568	0.927	0.885	0.981	0.931	0.985
Sorex araneus (common shrew)	0.341	0.944	0.944	0.944	0.944	0.986
Lepus europaeus (European hare)	0.878	0.952	0.992	0.912	0.950	0.998
Lepus timidus (mountain hare)	0.303	0.691	0.622	0.971	0.758	0.914
Oryctolagus cuniculus (European rabbit)	0.842	0.759	1.000	0.517	0.682	0.986
Rattus norvegicus (brown rat)	0.467	0.972	0.993	0.951	0.971	0.998
Apodemus flavicollis (yellow-necked mouse)	0.535	0.979	0.964	0.996	0.980	0.994
Apodemus sylvaticus (wood mouse)	0.619	0.971	0.971	0.971	0.971	0.997
Mus musculus (house mouse)	0.665	0.880	0.887	0.870	0.878	0.970
Microtus agrestis (field vole)	0.273	0.921	0.900	0.947	0.923	0.977
Myodes glareolus (bank vole)	0.256	0.866	0.807	0.961	0.877	0.921
Arvicola amphibius (water vole)	0.344	0.886	0.831	0.969	0.895	0.969
Lemmus lemmus (Norway lemming)	0.637	0.809	0.760	0.905	0.826	0.967
Vulpes vulpes (red fox)	0.267	0.865	0.815	0.945	0.875	0.946

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.327	0.672	0.616	0.912	0.736	0.921
Mustela nivalis (least weasel)	0.252	0.858	0.789	0.979	0.874	0.947
Neovison vison (mink)	0.354	0.759	0.692	0.933	0.795	0.852
Mustela putorius (polecat)	0.821	0.833	0.909	0.741	0.816	0.929
Martes martes (pine marten)	0.298	0.856	0.786	0.978	0.872	0.934
Meles meles (badger)	0.384	0.924	0.905	0.948	0.926	0.972
Cervus elaphus (red deer)	0.532	0.952	0.962	0.941	0.952	0.982
Dama dama (fallow deer)	0.948	0.874	1.000	0.749	0.856	1.000
Capreolus capreolus (roe deer)	0.414	0.942	0.959	0.923	0.941	0.984
Alces alces (moose)	0.273	0.797	0.726	0.953	0.824	0.878
Ondatra zibethicus (muskrat)	0.670	0.970	0.980	0.960	0.970	0.990
Ovis aries musimon (mouflon sheep)	0.570	0.955	1.000	0.909	0.952	1.000
Canis lupus (wolf)	0.436	0.987	0.987	0.987	0.987	0.992
Sus scrofa (wild boar)	0.914	0.944	0.991	0.897	0.942	0.997

Table C.16: U-Net++ (Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.844	0.815	0.890	0.851	0.900
Lynx lynx (lynx)	0.258	0.741	0.706	0.824	0.761	0.832
Ursus arctos (brown bear)	0.646	0.929	0.901	0.965	0.931	0.932
Canis lupus (wolf)	0.436	0.875	0.866	0.888	0.877	0.936

C.9 Fusion ResNet18 (Loc+Env)

Table C.17: Fusion ResNet18 (Loc+Env) — Test set: Artportalen.

Species	**Thr.**	**Acc.**	**Prec.**	**Rec.**	**F1**	**AP**
Combined	—	0.835	0.789	0.915	0.847	0.922
Erinaceus europaeus (hedgehog)	0.678	0.860	0.933	0.777	0.848	0.959
Lynx lynx (lynx)	0.273	0.873	0.821	0.953	0.882	0.958
Lutra lutra (otter)	0.276	0.680	0.628	0.887	0.735	0.752
Muscardinus avellanarius (hazel dormouse)	0.946	0.625	0.857	0.300	0.444	0.942
Ursus arctos (brown bear)	0.738	0.863	0.814	0.941	0.873	0.899
Sciurus vulgaris (red squirrel)	0.316	0.919	0.895	0.949	0.921	0.898
Castor fiber (beaver)	0.502	0.876	0.837	0.933	0.882	0.962
Sorex araneus (common shrew)	0.390	0.924	0.924	0.924	0.924	0.975
Lepus europaeus (European hare)	0.718	0.950	0.984	0.915	0.948	0.994
Lepus timidus (mountain hare)	0.216	0.530	0.516	0.968	0.673	0.837
Oryctolagus cuniculus (European rabbit)	0.757	0.586	1.000	0.172	0.294	0.983
Rattus norvegicus (brown rat)	0.529	0.828	0.979	0.669	0.795	0.982
Apodemus flavicollis (yellow-necked mouse)	0.470	0.979	0.964	0.996	0.980	0.971
Apodemus sylvaticus (wood mouse)	0.407	0.952	0.970	0.933	0.952	0.985
Mus musculus (house mouse)	0.656	0.796	0.900	0.667	0.766	0.894
Microtus agrestis (field vole)	0.238	0.880	0.853	0.917	0.884	0.919
Myodes glareolus (bank vole)	0.236	0.812	0.790	0.849	0.818	0.844
Arvicola amphibius (water vole)	0.327	0.803	0.735	0.948	0.828	0.855
Lemmus lemmus (Norway lemming)	0.747	0.952	1.000	0.905	0.950	0.991
Vulpes vulpes (red fox)	0.191	0.835	0.776	0.942	0.851	0.897

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.350	0.539	0.526	0.786	0.631	0.607
Mustela nivalis (least weasel)	0.300	0.850	0.783	0.968	0.866	0.921
Neovison vison (mink)	0.272	0.724	0.664	0.910	0.768	0.745
Mustela putorius (polecat)	0.644	0.963	0.963	0.963	0.963	0.986
Martes martes (pine marten)	0.235	0.836	0.798	0.900	0.846	0.925
Meles meles (badger)	0.359	0.921	0.916	0.927	0.922	0.965
Cervus elaphus (red deer)	0.387	0.895	0.974	0.812	0.886	0.964
Dama dama (fallow deer)	0.916	0.916	1.000	0.832	0.908	1.000
Capreolus capreolus (roe deer)	0.359	0.928	0.923	0.934	0.929	0.978
Alces alces (moose)	0.257	0.708	0.638	0.964	0.768	0.877
Ondatra zibethicus (muskrat)	0.800	0.750	0.879	0.580	0.699	0.911
Ovis aries musimon (mouflon sheep)	0.644	0.864	1.000	0.727	0.842	1.000
Canis lupus (wolf)	0.495	0.984	0.987	0.980	0.984	0.994
Sus scrofa (wild boar)	0.926	0.948	0.990	0.906	0.946	0.982

Table C.18: Fusion ResNet18 (Loc+Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.830	0.794	0.891	0.840	0.899
Lynx lynx (lynx)	0.273	0.698	0.668	0.788	0.723	0.794
Ursus arctos (brown bear)	0.738	0.926	0.881	0.986	0.930	0.964
Canis lupus (wolf)	0.495	0.881	0.860	0.911	0.885	0.939

C.10 Fusion ResNet18 (Loc+Sat)

Table C.19: Fusion ResNet18 (Loc+Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.906	0.910	0.901	0.906	0.972
Erinaceus europaeus (hedgehog)	0.891	0.955	0.991	0.918	0.953	0.998
Lynx lynx (lynx)	0.247	0.876	0.855	0.906	0.879	0.956
Lutra lutra (otter)	0.286	0.878	0.870	0.890	0.880	0.948
Muscardinus avellanarius (hazel dormouse)	0.924	0.812	0.931	0.675	0.783	0.932
Ursus arctos (brown bear)	0.814	0.958	0.980	0.935	0.957	0.978
Sciurus vulgaris (red squirrel)	0.588	0.941	0.975	0.905	0.939	0.990
Castor fiber (beaver)	0.659	0.947	0.972	0.921	0.946	0.992
Sorex araneus (common shrew)	0.415	0.891	0.933	0.843	0.886	0.981
Lepus europaeus (European hare)	0.827	0.946	0.992	0.898	0.943	0.998
Lepus timidus (mountain hare)	0.230	0.812	0.750	0.936	0.833	0.941
Oryctolagus cuniculus (European rabbit)	0.940	0.707	1.000	0.414	0.585	0.998
Rattus norvegicus (brown rat)	0.857	0.908	1.000	0.817	0.899	1.000
Apodemus flavicollis (yellow-necked mouse)	0.806	0.901	1.000	0.802	0.890	0.997
Apodemus sylvaticus (wood mouse)	0.594	0.971	0.981	0.962	0.971	0.992
Mus musculus (house mouse)	0.671	0.870	0.955	0.778	0.857	0.987
Microtus agrestis (field vole)	0.325	0.921	0.938	0.902	0.919	0.973
Myodes glareolus (bank vole)	0.216	0.838	0.871	0.792	0.830	0.938
Arvicola amphibius (water vole)	0.315	0.914	0.925	0.902	0.913	0.972
Lemmus lemmus (Norway lemming)	0.960	0.952	1.000	0.905	0.950	0.992
Vulpes vulpes (red fox)	0.320	0.907	0.900	0.915	0.908	0.950

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.375	0.809	0.848	0.753	0.798	0.888
Mustela nivalis (least weasel)	0.333	0.893	0.862	0.936	0.897	0.955
Neovison vison (mink)	0.451	0.906	0.911	0.899	0.905	0.960
Mustela putorius (polecat)	0.721	0.889	0.920	0.852	0.885	0.956
Martes martes (pine marten)	0.207	0.865	0.836	0.907	0.870	0.949
Meles meles (badger)	0.438	0.933	0.952	0.912	0.932	0.987
Cervus elaphus (red deer)	0.504	0.904	0.966	0.838	0.898	0.964
Dama dama (fallow deer)	0.913	0.877	0.991	0.761	0.861	0.996
Capreolus capreolus (roe deer)	0.544	0.934	0.974	0.892	0.931	0.985
Alces alces (moose)	0.228	0.817	0.746	0.960	0.840	0.908
Ondatra zibethicus (muskrat)	0.861	0.920	0.938	0.900	0.918	0.989
Ovis aries musimon (mouflon sheep)	0.853	0.682	1.000	0.364	0.533	1.000
Canis lupus (wolf)	0.523	0.954	0.954	0.954	0.954	0.991
Sus scrofa (wild boar)	0.928	0.941	0.982	0.898	0.938	0.995

Table C.20: Fusion ResNet18 (Loc+Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.866	0.849	0.892	0.870	0.923
Lynx lynx (lynx)	0.247	0.793	0.776	0.823	0.799	0.851
Ursus arctos (brown bear)	0.814	0.934	0.916	0.956	0.936	0.962
Canis lupus (wolf)	0.523	0.880	0.863	0.904	0.883	0.958

C.11 Fusion U-Net (Loc+Env)

Table C.21: Fusion U-Net (Loc+Env) — Test set: Artportalen.

Species	**Thr.**	**Acc.**	**Prec.**	**Rec.**	**F1**	**AP**
Combined	—	0.844	0.790	0.938	0.857	0.925
Erinaceus europaeus (hedgehog)	0.723	0.896	0.938	0.847	0.890	0.968
Lynx lynx (lynx)	0.318	0.840	0.773	0.962	0.857	0.881
Lutra lutra (otter)	0.366	0.676	0.618	0.926	0.741	0.743
Muscardinus avellanarius (hazel dormouse)	0.794	0.675	0.889	0.400	0.552	0.939
Ursus arctos (brown bear)	0.712	0.840	0.783	0.941	0.855	0.918
Sciurus vulgaris (red squirrel)	0.306	0.910	0.882	0.947	0.913	0.953
Castor fiber (beaver)	0.625	0.896	0.851	0.959	0.902	0.970
Sorex araneus (common shrew)	0.363	0.952	0.945	0.960	0.952	0.941
Lepus europaeus (European hare)	0.803	0.975	0.984	0.965	0.975	0.997
Lepus timidus (mountain hare)	0.282	0.534	0.519	0.931	0.666	0.847
Oryctolagus cuniculus (European rabbit)	0.624	0.828	1.000	0.655	0.792	0.988
Rattus norvegicus (brown rat)	0.657	0.880	0.982	0.775	0.866	0.994
Apodemus flavicollis (yellow-necked mouse)	0.686	0.979	0.964	0.996	0.980	0.985
Apodemus sylvaticus (wood mouse)	0.605	0.971	0.971	0.971	0.971	0.987
Mus musculus (house mouse)	0.537	0.843	0.814	0.889	0.850	0.904
Microtus agrestis (field vole)	0.250	0.910	0.876	0.955	0.914	0.931
Myodes glareolus (bank vole)	0.269	0.807	0.775	0.866	0.818	0.777
Arvicola amphibius (water vole)	0.255	0.811	0.736	0.969	0.837	0.959
Lemmus lemmus (Norway lemming)	0.685	0.857	0.800	0.952	0.870	0.963
Vulpes vulpes (red fox)	0.270	0.810	0.747	0.937	0.831	0.904

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.284	0.586	0.550	0.954	0.697	0.643
Mustela nivalis (least weasel)	0.256	0.869	0.800	0.984	0.882	0.939
Neovison vison (mink)	0.353	0.693	0.629	0.941	0.754	0.795
Mustela putorius (polecat)	0.752	0.926	0.926	0.926	0.926	0.941
Martes martes (pine marten)	0.332	0.861	0.801	0.962	0.874	0.941
Meles meles (badger)	0.429	0.922	0.899	0.951	0.924	0.958
Cervus elaphus (red deer)	0.430	0.947	0.952	0.941	0.946	0.957
Dama dama (fallow deer)	0.871	0.963	1.000	0.925	0.961	1.000
Capreolus capreolus (roe deer)	0.371	0.924	0.915	0.934	0.924	0.966
Alces alces (moose)	0.209	0.761	0.689	0.952	0.799	0.823
Ondatra zibethicus (muskrat)	0.785	0.920	0.889	0.960	0.923	0.954
Ovis aries musimon (mouflon sheep)	0.408	1.000	1.000	1.000	1.000	1.000
Canis lupus (wolf)	0.472	0.987	0.987	0.987	0.987	0.994
Sus scrofa (wild boar)	0.922	0.962	0.982	0.941	0.961	0.989

Table C.22: Fusion U-Net (Loc+Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.813	0.770	0.892	0.827	0.854
Lynx lynx (lynx)	0.318	0.670	0.639	0.780	0.703	0.715
Ursus arctos (brown bear)	0.712	0.917	0.872	0.977	0.922	0.934
Canis lupus (wolf)	0.472	0.870	0.829	0.933	0.878	0.912

C.12 Fusion U-Net (Loc+Sat)

Table C.23: Fusion U-Net (Loc+Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.918	0.893	0.950	0.921	0.975
Erinaceus europaeus (hedgehog)	0.697	0.968	0.984	0.951	0.967	0.999
Lynx lynx (lynx)	0.169	0.890	0.830	0.981	0.899	0.978
Lutra lutra (otter)	0.263	0.906	0.871	0.952	0.910	0.969
Muscardinus avellanarius (hazel dormouse)	0.660	0.938	0.949	0.925	0.937	0.941
Ursus arctos (brown bear)	0.578	0.863	0.863	0.863	0.863	0.958
Sciurus vulgaris (red squirrel)	0.481	0.945	0.939	0.952	0.946	0.993
Castor fiber (beaver)	0.601	0.974	0.966	0.982	0.974	0.995
Sorex araneus (common shrew)	0.381	0.934	0.926	0.944	0.935	0.985
Lepus europaeus (European hare)	0.795	0.968	0.992	0.945	0.968	0.999
Lepus timidus (mountain hare)	0.198	0.759	0.685	0.961	0.799	0.935
Oryctolagus cuniculus (European rabbit)	0.741	0.776	1.000	0.552	0.711	1.000
Rattus norvegicus (brown rat)	0.578	0.979	1.000	0.958	0.978	1.000
Apodemus flavicollis (yellow-necked mouse)	0.535	0.986	0.976	0.996	0.986	0.999
Apodemus sylvaticus (wood mouse)	0.549	0.971	0.971	0.971	0.971	0.995
Mus musculus (house mouse)	0.734	0.907	0.940	0.870	0.904	0.983
Microtus agrestis (field vole)	0.307	0.913	0.872	0.970	0.918	0.973
Myodes glareolus (bank vole)	0.429	0.903	0.866	0.952	0.907	0.946
Arvicola amphibius (water vole)	0.327	0.930	0.899	0.969	0.933	0.984
Lemmus lemmus (Norway lemming)	0.764	0.881	0.833	0.952	0.889	0.982
Vulpes vulpes (red fox)	0.321	0.899	0.865	0.945	0.903	0.957

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.314	0.777	0.698	0.977	0.814	0.956
Mustela nivalis (least weasel)	0.208	0.901	0.844	0.984	0.909	0.954
Neovison vison (mink)	0.471	0.903	0.869	0.949	0.907	0.962
Mustela putorius (polecat)	0.633	0.926	0.960	0.889	0.923	0.982
Martes martes (pine marten)	0.172	0.874	0.814	0.969	0.885	0.943
Meles meles (badger)	0.509	0.956	0.961	0.951	0.956	0.968
Cervus elaphus (red deer)	0.383	0.963	0.947	0.982	0.964	0.983
Dama dama (fallow deer)	0.904	0.940	1.000	0.880	0.936	1.000
Capreolus capreolus (roe deer)	0.526	0.958	0.975	0.940	0.957	0.988
Alces alces (moose)	0.252	0.827	0.756	0.964	0.848	0.866
Ondatra zibethicus (muskrat)	0.760	0.940	0.940	0.940	0.940	0.984
Ovis aries musimon (mouflon sheep)	0.616	1.000	1.000	1.000	1.000	1.000
Canis lupus (wolf)	0.263	0.987	0.987	0.987	0.987	0.994
Sus scrofa (wild boar)	0.834	0.949	0.991	0.907	0.947	0.994

Table C.24: Fusion U-Net (Loc+Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.841	0.812	0.888	0.848	0.904
Lynx lynx (lynx)	0.169	0.745	0.706	0.840	0.767	0.854
Ursus arctos (brown bear)	0.578	0.925	0.907	0.946	0.926	0.972
Canis lupus (wolf)	0.263	0.865	0.852	0.882	0.867	0.885

C.13 Fusion U-Net++ (Loc+Env)

Table C.25: Fusion U-Net++ (Loc+Env) — Test set: Artportalen.

Species	**Thr.**	**Acc.**	**Prec.**	**Rec.**	**F1**	**AP**
Combined	—	0.839	0.786	0.932	0.853	0.922
Erinaceus europaeus (hedgehog)	0.574	0.860	0.979	0.736	0.840	0.968
Lynx lynx (lynx)	0.268	0.865	0.799	0.975	0.878	0.889
Lutra lutra (otter)	0.274	0.642	0.592	0.918	0.719	0.744
Muscardinus avellanarius (hazel dormouse)	0.835	0.787	0.926	0.625	0.746	0.959
Ursus arctos (brown bear)	0.644	0.840	0.783	0.941	0.855	0.982
Sciurus vulgaris (red squirrel)	0.292	0.909	0.876	0.953	0.913	0.945
Castor fiber (beaver)	0.538	0.939	0.900	0.987	0.941	0.985
Sorex araneus (common shrew)	0.355	0.876	0.860	0.899	0.879	0.872
Lepus europaeus (European hare)	0.842	0.976	0.992	0.960	0.976	0.997
Lepus timidus (mountain hare)	0.284	0.523	0.513	0.905	0.655	0.821
Oryctolagus cuniculus (European rabbit)	0.708	0.724	1.000	0.448	0.619	0.963
Rattus norvegicus (brown rat)	0.417	0.887	0.991	0.782	0.874	0.998
Apodemus flavicollis (yellow-necked mouse)	0.503	0.977	0.960	0.996	0.978	0.964
Apodemus sylvaticus (wood mouse)	0.624	0.967	0.971	0.962	0.967	0.977
Mus musculus (house mouse)	0.563	0.806	0.837	0.759	0.796	0.870
Microtus agrestis (field vole)	0.262	0.838	0.781	0.940	0.853	0.820
Myodes glareolus (bank vole)	0.314	0.803	0.780	0.844	0.811	0.818
Arvicola amphibius (water vole)	0.225	0.850	0.790	0.953	0.864	0.924
Lemmus lemmus (Norway lemming)	0.816	0.952	0.952	0.952	0.952	0.991
Vulpes vulpes (red fox)	0.221	0.814	0.752	0.936	0.834	0.906

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.311	0.558	0.534	0.921	0.676	0.611
Mustela nivalis (least weasel)	0.236	0.842	0.767	0.984	0.862	0.925
Neovison vison (mink)	0.272	0.702	0.634	0.956	0.762	0.788
Mustela putorius (polecat)	0.719	0.907	0.923	0.889	0.906	0.984
Martes martes (pine marten)	0.190	0.828	0.758	0.965	0.849	0.939
Meles meles (badger)	0.340	0.923	0.905	0.945	0.924	0.966
Cervus elaphus (red deer)	0.410	0.969	0.978	0.960	0.969	0.986
Dama dama (fallow deer)	0.789	0.989	1.000	0.979	0.989	1.000
Capreolus capreolus (roe deer)	0.334	0.922	0.934	0.907	0.920	0.976
Alces alces (moose)	0.278	0.722	0.650	0.960	0.775	0.833
Ondatra zibethicus (muskrat)	0.847	0.900	0.900	0.900	0.900	0.973
Ovis aries musimon (mouflon sheep)	0.643	0.955	1.000	0.909	0.952	1.000
Canis lupus (wolf)	0.457	0.987	0.987	0.987	0.987	0.994
Sus scrofa (wild boar)	0.898	0.983	0.992	0.973	0.982	0.997

Table C.26: Fusion U-Net++ (Loc+Env) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.823	0.773	0.913	0.837	0.865
Lynx lynx (lynx)	0.268	0.695	0.655	0.823	0.730	0.723
Ursus arctos (brown bear)	0.644	0.919	0.873	0.981	0.924	0.964
Canis lupus (wolf)	0.457	0.870	0.822	0.945	0.879	0.907

C.14 Fusion U-Net++ (Loc+Sat)

Table C.27: Fusion U-Net++ (Loc+Sat) — Test set: Artportalen.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.888	0.851	0.942	0.894	0.964
Erinaceus europaeus (hedgehog)	0.625	0.967	0.984	0.950	0.966	0.997
Lynx lynx (lynx)	0.377	0.871	0.816	0.959	0.881	0.924
Lutra lutra (otter)	0.392	0.768	0.703	0.930	0.800	0.871
Muscardinus avellanarius (hazel dormouse)	0.643	0.950	0.950	0.950	0.950	0.921
Ursus arctos (brown bear)	0.817	0.938	0.935	0.941	0.938	0.990
Sciurus vulgaris (red squirrel)	0.309	0.929	0.916	0.946	0.931	0.988
Castor fiber (beaver)	0.591	0.935	0.908	0.968	0.937	0.981
Sorex araneus (common shrew)	0.265	0.929	0.934	0.924	0.929	0.990
Lepus europaeus (European hare)	0.824	0.972	1.000	0.945	0.972	0.999
Lepus timidus (mountain hare)	0.324	0.726	0.653	0.962	0.778	0.928
Oryctolagus cuniculus (European rabbit)	0.803	0.759	1.000	0.517	0.682	0.988
Rattus norvegicus (brown rat)	0.441	0.958	0.992	0.922	0.956	0.999
Apodemus flavicollis (yellow-necked mouse)	0.425	0.971	0.949	0.996	0.972	0.996
Apodemus sylvaticus (wood mouse)	0.505	0.986	0.981	0.991	0.986	1.000
Mus musculus (house mouse)	0.535	0.907	0.893	0.926	0.909	0.962
Microtus agrestis (field vole)	0.260	0.929	0.907	0.955	0.930	0.979
Myodes glareolus (bank vole)	0.290	0.879	0.846	0.926	0.884	0.935
Arvicola amphibius (water vole)	0.241	0.886	0.831	0.969	0.895	0.967
Lemmus lemmus (Norway lemming)	0.790	0.905	0.905	0.905	0.905	0.961
Vulpes vulpes (red fox)	0.247	0.874	0.834	0.934	0.881	0.949

Continued on next page

(continued)

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Mustela erminea (stoat)	0.305	0.763	0.692	0.949	0.800	0.936
Mustela nivalis (least weasel)	0.259	0.874	0.802	0.995	0.888	0.958
Neovison vison (mink)	0.385	0.756	0.689	0.936	0.793	0.857
Mustela putorius (polecat)	0.631	0.889	0.957	0.815	0.880	0.986
Martes martes (pine marten)	0.296	0.853	0.797	0.947	0.865	0.932
Meles meles (badger)	0.260	0.938	0.925	0.952	0.938	0.973
Cervus elaphus (red deer)	0.420	0.939	0.944	0.934	0.939	0.976
Dama dama (fallow deer)	0.865	0.936	1.000	0.873	0.932	1.000
Capreolus capreolus (roe deer)	0.340	0.941	0.951	0.930	0.941	0.986
Alces alces (moose)	0.269	0.812	0.744	0.951	0.835	0.878
Ondatra zibethicus (muskrat)	0.732	0.950	0.959	0.940	0.950	0.988
Ovis aries musimon (mouflon sheep)	0.462	0.909	1.000	0.818	0.900	1.000
Canis lupus (wolf)	0.429	0.987	0.987	0.987	0.987	0.993
Sus scrofa (wild boar)	0.886	0.959	0.991	0.925	0.957	0.993

Table C.28: Fusion U-Net++ (Loc+Sat) — Test set: Rovbase.

Species	Thr.	Acc.	Prec.	Rec.	F1	AP
Combined	—	0.846	0.812	0.900	0.854	0.894
Lynx lynx (lynx)	0.377	0.749	0.713	0.834	0.769	0.821
Ursus arctos (brown bear)	0.817	0.931	0.896	0.976	0.934	0.955
Canis lupus (wolf)	0.429	0.869	0.849	0.897	0.873	0.906

DEPARTMENT OF ENVIRONMENTAL AND ENERGY SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY