Multi-species distribution modeling using deep CNNs and satellite embeddings

Markus Blixt[1,*], Carl Klang[1,*], and Olof Mogren[2,3,4]

[1] Chalmers University of Technology
[2] RISE Research Institutes of Sweden
[3] Climes - Swedish Centre for Impacts of Climate Extremes
[4] Climate AI Nordics www.climateainordics.com

Abstract. Traditional Species Distribution Models (SDMs) rely on coarse bioclimatic variables, struggling to capture localized ecological niches, habitat fragmentation, and multi-species interactions. This paper presents a high-resolution, multi-species deep learning framework modeling 34 terrestrial mammals across Sweden using opportunistic crowd-sourced data. We benchmark a deep convolutional neural network architecture (ResNet-18) against a spatial MLP (SINR) and Random Forest baselines. Crucially, integrating AlphaEarth Foundations (AEF) 64-dimensional satellite embeddings yields a performance improvement over traditional climate variables, pushing mean Average Precision (mAP) from 0.924 to 0.972. To counteract spatial observation biases inherent in presence-only records, we introduce a custom distance-density pseudo-absence generation mechanism. Models are evaluated across three distinct proxies: a held-out test split, an independent apex carnivore dataset (Rovbase), and expert range maps (IUCN Red List) [10]. Our results demonstrate that high-resolution satellite embeddings are essential for capturing fine-grained local landscape variations, accounting for human impacts, and maintaining robust performance under significant spatial distribution shifts.

1 Introduction

Quantifying geographic species distributions is critical for establishing proactive conservation policies, predicting responses to climate change, and mitigating biodiversity loss driven by human exploitation [11, 14]. Species Distribution Modeling (SDM) relates observed occurrences to environmental predictors to estimate habitat suitability across spatial domains [7]. While traditional methods (e.g., Generalized Linear Models, Random Forests) typically evaluate solitary species against coarse point-source data, deep learning networks enable multi-species architectures that capture nonlinear environmental preferences and implicit biotic interactions simultaneously [2, 4].

A primary bottleneck in modern deep learning-based SDMs is the representational capacity of the environmental predictors. Historically, models have

[*] Equal contribution.

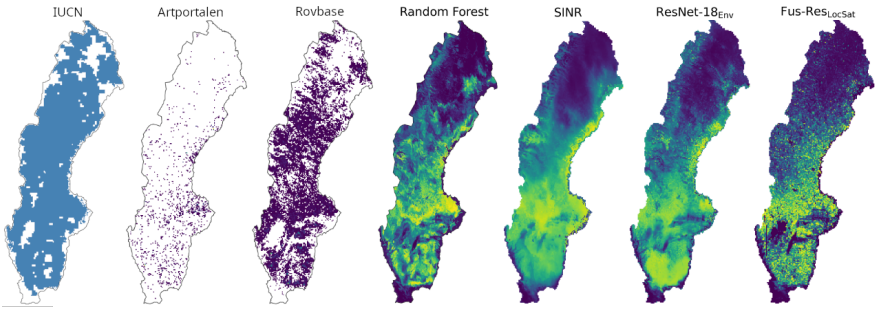

Fig. 1: Comparison of datasets and output from different models for lynx. From left to right: IUCN, Artportalen dataset, Rovbase dataset, outputs from Random Forest, SINR, ResNet-18_{Env}, Fus-$\text{Res}_{\text{LocSat}}$ respectively.

relied on interpolated bioclimatic variables (e.g., WorldClim), which fail to mirror fine-grained local landscape textures, urban boundaries, or micro-habitats. Concurrently, utilizing raw remote sensing imagery forces down-stream models to handle massive computational workloads, spatial alignment anomalies, and temporal cloud-masking variations.

This paper investigates the performance of the AlphaEarth Foundations (AEF) Satellite Embeddings dataset, a general-purpose, information-dense feature space, as a primary spatial predictor for multi-species neural SDMs. By processing entire spatial neighborhoods (128×128 pixel grids at 50 m resolution) through a patch-based CNN backbone, our approach extracts highly detailed surface characteristics directly from the de-quantized unit hypersphere space.

To overcome the challenges of opportunistic, citizen-science reporting networks (i.e., the presence-only nature and heavy clustering around human populations), we design a custom pseudo-absence generation routine. Furthermore, we present a hybrid, late-fusion neural architectures that combine coordinate-based implicit spatial memories with patch-based vision networks. All code, model checkpoints, and data pipelines will be open-spurced after double-blind peer-review.

2 Experimental setup

This section details the boundaries of our study area, the composition of the crowd-sourced and verified observation datasets, and the preprocessing steps implemented to de-quantize and downsample the multi-sensor satellite arrays. A schematic overview of this multi-criteria spatial data mapping and validation layout is illustrated in Fig. 3. As depicted, the data flows from opportunistic civilian inputs (Artportalen) through our distance-density filtering network into the late-fusion architecture, which is then continuously evaluated against systematic field records (Rovbase) and macro-ecological bounds (IUCN).

Study area and spatial domain mapping. The geographic scope is strictly constrained to the landmass of Sweden. To ensure topological consistency and continuous spatial bounds, the national shapefile from Eurostat [8] was unified with administrative boundary vectors from Topografi 50 (Lantmäteriet) [12].

Co-registered spatial grid mechanics. The engineering of three co-registered grid overlays was necessary to isolate different processing workflows across distinct geographic scales:

- *The* 500 m *Fine-Grained Grid* forms the core canvas for our distance-density pseudo-absence algorithm. This cell scale aligns with the spatial precision threshold of opportunistic smartphone records while preventing background data from overlapping directly with true presences.
- *The* 5 km *Intermediate Grid* acts as our nationwide inference mapping matrix. It aggregates our 50 m continuous patch predictions into scannable country-wide trend lines without overloading standard geographic information system memory allocations.
- *The* 50 km *Coarse Grid* isolates spatial autocorrelation during data partitioning. By grouping training and testing data within these large, distinct blocks, we ensure that the model is tested on its ability to generalize to new geographic regions, rather than simply memorizing nearby points.

Observation datasets. Models were optimized using a pool of 34 land-dwelling terrestrial mammals extracted from Artportalen [1], spanning observations logged between 2003-01-01 and 2024-07-31. Rare taxa with fewer than 500 records were pruned, while abundant species were downsampled to a cap of 20,000 instances to stabilize training dynamics. The training pool includes highly restricted data for three sensitive large carnivores: wolf (*Canis lupus*), brown bear (*Ursus arctos*), and Eurasian lynx (*Lynx lynx*). The resulting clean dataset contains 220,355 presence records.

For independent validation, an authoritative, professionally verified dataset was obtained from Rovbase [13], consisting of 122,578 systematically logged telemetry, track, and genetic occurrence markers for the three large carnivore species over the same time horizon *(number of observations: lynx: 44,527, brown bear: 40,704, wolf: 37,347)*. Figure 1 gives an overview of the datasets together with maps generated from the evaluated models.

Predictor variables and preprocessing. Two distinct input suites were compiled to contrast predictive accuracy:

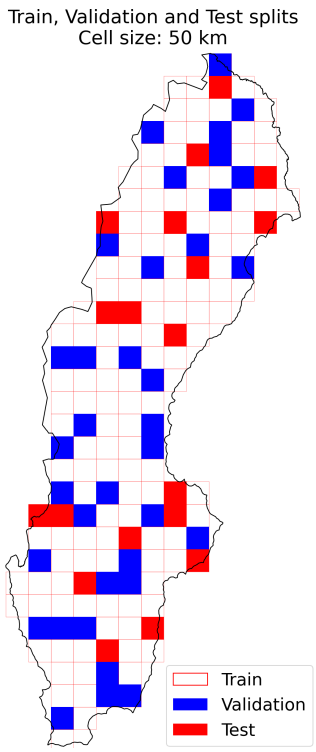

Fig. 2: Train, validation and test split of coarse grid cells.

Bioclimatic baselines (Env): 19 WorldClim v2.1 variables at 30-arcsecond resolution ($\sim$ 1 km) paired with a digital elevation model (DEM) [9]. The layers were cropped to a 20 km buffer zone surrounding Sweden, standardized to zero mean and unit variance, and stacked into a 20-band GeoTIFF tensor.

AlphaEarth Foundations Satellite Embeddings (Sat): Compact, pre-computed 64-dimensional neural embedding vectors summarizing annual optical imagery, SAR, LiDAR, and climate reanalysis data across 2020 [3]. Raw 8-bit quantized integers spanning $[-127, 127]$ were fetched from Google Cloud Storage. Individual pixel arrays were downsampled to 50 m cells, dynamically de-quantized, and mapped to a standard $[-1, 1]$ continuous unit hypersphere boundary.

Taxonomic breakdown and sample densities. To provide a clear view of the modeling domain, Table 1 breaks down the 34 modeled terrestrial mammal species. It categorizes them into three major ecological groups based on their total records in Artportalen and their subsequent sampling parameters.

Table 1: Taxonomic composition, clean training sample counts, and threshold parameters.

Ecological Guild	Common Name / Species Tag	Counts	LPT-R
Large Carnivores	Eurasian Lynx (*Lynx lynx*)	11,432	0.342
	Brown Bear (*Ursus arctos*)	8,114	0.298
	Grey Wolf (*Canis lupus*)	3,211	0.412
Ubiquitous Ungulates	Moose (*Alces alces*)	20,000*	0.115
	Roe Deer (*Capreolus capreolus*)	20,000*	0.095
	Wild Boar (*Sus scrofa*)	18,412	0.187
Range-Restricted	Fallow Deer (*Dama dama*)	7,311	0.512
	Mouflon (*Ovis aries musimon*)	612	0.684

*Records artificially capped at the 20,000 downsampling threshold bound.

Data partitioning and model evaluation. To prevent data leakage from spatially autocorrelated observations, spatial partitioning was enforced using a 50 km grid. As illustrated in Fig. 2, continuous grid cells across Sweden were split using a stratified selection algorithm, yielding a representative geographic split across Train (125,927 observations / 65%), Validation (42,068 observations / 22%), and Test (26,305 observations / 14%) subsets. Model selection used the lowest validation loss across 5 training epochs. Final binary maps were generated by implementing per-species thresholding via the Lowest Presence Threshold - Robust (LPT-R) method at the 5th percentile [6].

Hyperparameter profiles and optimization protocol. All neural models were implemented using PyTorch 2.3 and optimized on single NVIDIA A100 GPU nodes. The ResNet-18 and hybrid late-fusion networks were trained using the AdamW optimizer with a base learning rate of $\eta = 10^{-4}$ and a weight decay factor of 10^{-2} to prevent extreme gradient explosions in the final classification layer. Mini-batch sizes were fixed at 128 elements, composed of 64 true presences and 64 pseudo-absences.

To maintain numerical stability across multi-task iterations, we utilized a cosine annealing learning rate scheduler with a linear warmup phase spanning the first epoch. Gradient clipping was enforced at a threshold of 1.0. Model checkpoints were verified at 100-stride step intervals against the stratified validation grid blocks to catch early overfitting before termination at the fifth epoch.

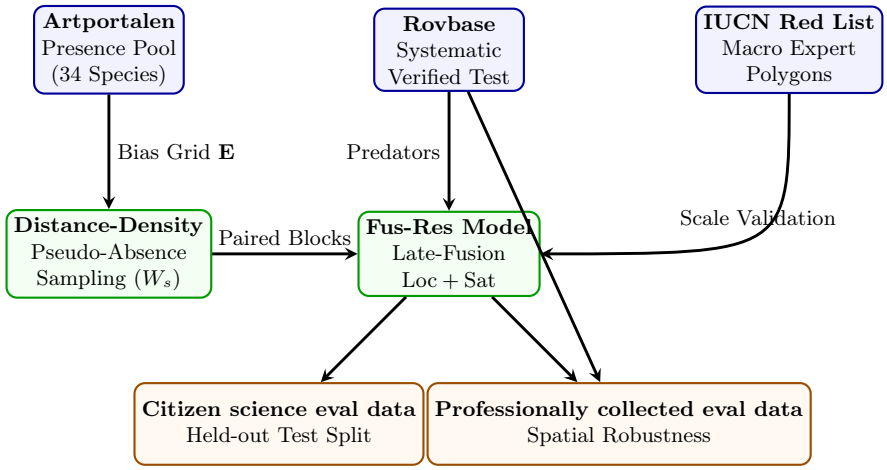

Fig. 3: Schematic pipeline of the multi-criteria spatial data mapping and validation layout. Data flows from opportunistic civilian inputs (Artportalen) through our distance-density filtering network into the late-fusion architecture, verified continuously against systematic field records (Rovbase) and macro-ecological bounds (IUCN).

3 High-resolution neural multi-species modeling

To map fine-grained species distributions while accounting for the limitations of crowd-sourced data, we implement a multi-species neural network pipeline. This section details our density-aware pseudo-absence sampling framework, the architectural variations of the vision backbones, and the design of the hybrid late-fusion models that integrate spatial coordinates with landscape patches.

Density-aware pseudo-absence generation. To counterbalance the geographic biases inherent in opportunistic citizen-science data, we implement a two-step distance-density sampling routine on the 500 m grid overlay.

First, we estimate spatial sampling effort across Sweden by computing a global density map. For each cell, we sum all observations across all 34 species within a 10 km radius using a Gaussian decay function ($\sigma = 5$ km). To prevent hyper-dense urban hotspots from dominating the probability fields, this sum is capped at a threshold of 200 and normalized to a relative interval $[0, 1]$, yielding the standardized collection effort matrix, $\mathbf{E}$.

Second, to isolate the specific range boundaries of a target species s, we compute a distance matrix, $\mathbf{D}_s$, representing the spatial proximity of each grid cell to the nearest confirmed presence record of species s. The final sampling weights (W_s) for generating pseudo-absences are then defined as the element-wise product of these two matrices:

$$W_s = \mathbf{E} \odot \mathbf{D}_s \tag{1}$$

This formulation mathematically forces the pseudo-absence sampling probability to zero at known presence locations while favoring intensively surveyed zones where the target species was explicitly not observed. To maintain robust baseline coverage in unmonitored or uninhabited landscapes (e.g., alpine ranges), 90% of the final pseudo-absences are drawn using these distance-density weights (W_s), while the remaining 10% are populated using a pure distance-only algorithm ($\mathbf{D}_s$).

Model architecture and loss formulation. Formally, SDMs aim to map environmental predictors $\mathbf{x} \in \mathcal{X}$ to continuous occurrence probabilities $\hat{\mathbf{y}} = f(\mathbf{x}) \in [0, 1]^S$ for a set of species S. Continuous predictions are converted into binary classifications $\tilde{y}_s$ via a decision threshold $\theta \in [0, 1]$. Three architectural frameworks were evaluated:

Random Forest (RF baseline): Fitted on single point vectors interpolated from the 20-band bioclimatic baseline.

Spatial Implicit Neural Representation (SINR baseline): An MLP containing 4 sequential residual blocks (256 nodes per layer, dropout $p = 0.5$) utilizing point-based environmental data concatenated with a 4D sinusoidal coordinate encoder [5].

ResNet-18: An 18-layer residual architecture customized to process input grid dimensions of 128×128 pixels ($\approx 6.4 \times 6.4$ km) with either 20 channels (Env) or 64 channels (Sat). Parameter optimization utilized a modified single-value binary cross-entropy loss tracking observed species s_b with loss weights w_P and w_A for presence and absence observations, respectively:

$$\mathcal{L} = -\frac{1}{(w_P + w_A)N} \sum_{b=1}^{N} \left[w_P \ln(\hat{y}_{b,s_b}) + w_A \ln(1 - \hat{y}'_{b,s_b}) \right] \tag{2}$$

Late-Fusion Hybrid Layout (Fus-Res): The ResNet-18 backbone processed spatial patches in parallel with a coordinate-based SINR stream. The distinct visual features extracted by the CNN track and the location encodings from the SINR stream were combined at the final stage via a modified fully connected layer to produce a joint prediction for each species.

For all neural configurations, loss weights were fixed at $w_P = 1.25$ and $w_A = 1$ to prevent model overfitting toward pseudo-absences.

4 Results

This section presents a comparative evaluation of the baseline models, patch CNNs, and hybrid fusion networks across our testing proxies. The systematic data flow and validation architecture, tracing records from opportunistic civilian inputs (Artportalen) through our distance-density filtering network into the late-fusion model, and continuously verifying them against systematically collected field records (Rovbase) and macro-ecological bounds (IUCN), is schematically detailed in Fig. 3. We report aggregated metrics on the held-out spatial partitions and assess our continuous probability outputs against broad macro-ecological range maps.

Table 2: Aggregated test set results for the Artportalen dataset.

Model	Input Type	Accuracy	F1	mAP
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
ResNet-18	Sat	0.900	0.900	0.965
Fus-Res	Loc+Env	0.835	0.847	0.922
Fus-Res	**Loc+Sat**	**0.906**	**0.906**	**0.972**

Models were evaluated globally across both the Artportalen test split and the Rovbase dataset to assess performance under shifting sampling bias. Fig. 1 shows maps generated from the evaluated models for lynx and indicates that models utilizing the Alpha Earth embeddings show more detail. Accuracy, F1-score, and macro-averaged Mean Average Precision (mAP) are reported in Table 2 and Table 3.

The experimental evaluations indicate that architectures using satellite embedding inputs (*Sat* and *Loc+Sat*) outperform environmental alternatives across all metrics. On both the Artportalen test split and the Rovbase dataset, the hybrid late-fusion *Fus-Res*$_{LocSat}$ strategy proved to be the top performer, delivering 0.972 mAP on Artportalen and leading all benchmarking categories on Rovbase with 0.866 accuracy, 0.870 F1, and 0.923 mAP.

Table 3: Aggregated test set results for the Rovbase dataset.

Model	Input Type	Accuracy	F1	mAP
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
ResNet-18	Sat	0.855	0.859	0.914
Fus-Res	Loc+Env	0.830	0.840	0.899
Fus-Res	**Loc+Sat**	**0.866**	**0.870**	**0.923**

Table 4 reports the macro-averaged and species-specific Average Precision (AP) scores evaluated against expert-derived IUCN Red List range maps. Because these authoritative references consist of highly coarse, simplified polygons designed for continental-scale consensus, they lack the spatial granularity required to capture localized habitat fragmentation or fine-grained urban boundaries (see Fig. 1 for an example map for lynx). Consequently, this validation proxy does not measure absolute localized predictive accuracy, but rather serves as a macro-ecological baseline to verify whether the models successfully reconstruct broad, regional macro-climatic ranges without producing extensive geographic overpredictions.

Spatial prediction artifacts and fine-grained resolution. A granular examination of the model outputs across Sweden highlights stark visual and structural differences between point-based baselines and our patch-based late-fusion framework. As illustrated in the Lynx prediction panels (Fig. 1), the Random Forest and SINR baselines yield diffuse, smoothed suitability gradients. Because these models rely on interpolated WorldClim tiles, they project broad, continuous climatic bands across central and northern Sweden.

In contrast, models utilizing the AlphaEarth Foundations (AEF) satellite embeddings exhibit sharp, localized spatial boundaries across diverse ecological profiles. The patch-based vision backbone identifies distinct geographical features such as major river valleys, forestry clear-cuts, and sharp transitions where agricultural plains meet boreal forests. This effect is visible when comparing the climate-only baseline (ResNet-18_{Env}) to our late-fusion network (Fus-$\text{Res}_{\text{LocSat}}$) for the brown bear (Fig. 4) and the West European hedgehog (Fig. 5).

For the brown bear, ResNet-18_{Env} generates a uniform envelope across northern wilderness regions, whereas Fus-$\text{Res}_{\text{LocSat}}$ restricts suitability to distinct core corridors. Conversely, for the hedgehog, the satellite embedding model concentrates suitability values inside suburban nodes and agricultural boundaries, whereas the climate baseline spreads predictions smoothly across surrounding uninhabited forested regions.

Model resilience under shifting sampling bias. Evaluating our architectures against the independent Rovbase dataset serves as a rigorous stress-test for generaliza-

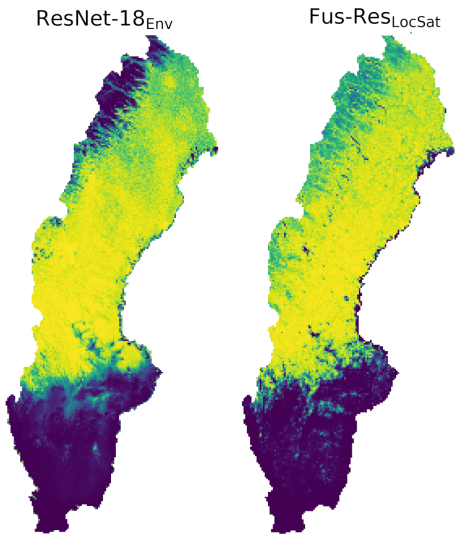

Fig. 4: Comparison output from ResNet-18$_{\text{Env}}$ (left) and Fus-Res$_{\text{LocSat}}$ (right) for brown bear.

tion under shifting sampling bias. The opportunistic crowd-sourced data from Artportalen is heavily subject to human population bias, with records heavily clustered near major roads, cities, and popular hiking trails. Conversely, Rovbase records are systematically collected via standardized DNA sampling, snow-transect tracking, and field monitoring conducted by certified wildlife wardens. While this agency-driven approach is not completely free from bias, as monitoring effort remains naturally higher within known carnivore territory boundaries and wolf management zones, it represents a highly structured collection methodology that is vastly less confounded by human demographic distributions than civilian networks.

Table 4: Model Average Precision (AP) evaluated against IUCN expert range maps.

Model Layout	**Input Suite**	**Mean mAP**	**Hedgehog**	**Lynx**	**Bear**
RF Baseline	Env	0.753	0.658	0.769	0.767
SINR Baseline	Loc+Env	0.875	**0.880**	0.874	0.952
ResNet-18	**Env**	**0.885**	0.861	0.879	0.952
ResNet-18	Sat	0.876	0.830	0.889	0.958
Fus-Res	**Loc+Sat**	0.873	0.831	**0.911**	0.954

When transitioning from Artportalen to Rovbase, point-based models like SINR suffer a severe performance drop (with mAP falling from 0.919 to 0.877).

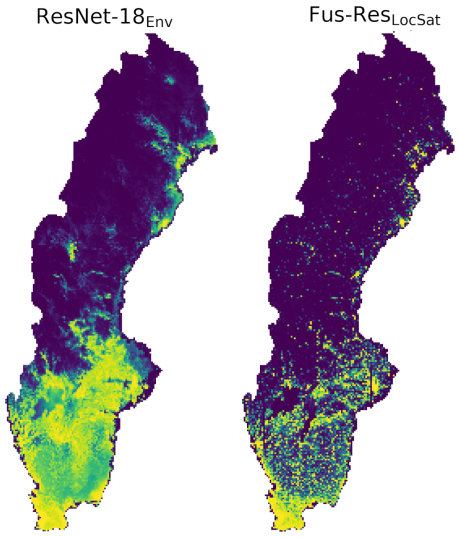

Fig. 5: Comparison output from ResNet-18_{Env} (left) and Fus-$\text{Res}_{\text{LocSat}}$ (right) for hedgehog.

This indicates that point-based coordinate models partially overfit to the human sampling footprint inherent in citizen science data. Conversely, our proposed Fus-$\text{Res}_{\text{LocSat}}$ framework maintains a highly robust 0.923 mAP on Rovbase. This demonstrates that by processing independent 50 m landscape patches through a CNN alongside coordinate streams, the model detaches from human footprint clustering and captures genuine, underlying biophysical properties that dictate large carnivore habitats.

5 Related work

Contemporary neural SDM methodologies have been proposed focusing on either implicit spatial coordinates, bioclimatic variables, and in some cases remote sensing data. The Spatial Implicit Neural Representation (SINR) framework proposed by Cole et al. modeled global species ranges by mapping low-dimensional coordinate inputs directly to multi-label classification heads, using point-sampled environmental features as secondary indicators [5]. While computationally efficient, point-based MLPs cannot leverage geographical context.

Conversely, frameworks that integrate remote sensing imagery directly often encounter substantial computational and preprocessing challenges. The SatBird benchmarking framework evaluated deep vision networks (such as ResNet-18) against structured avian occurrence counts to predict encounter rates using multi-spectral satellite rasters [15]. The authors noted that raw satellite imagery alone did not improve baseline performance unless explicitly combined with smoothly interpolated bioclimatic variables. Our work demonstrates that

utilizing pre-computed foundation embeddings (such as AEF) addresses this limitation, providing an information-dense, general-purpose feature space that captures structural landscape context without requiring raw image reprocessing loops.

6 Discussion and conclusions

Our evaluation demonstrates that patch-based CNNs using satellite embedding inputs consistently outperform environmental baselines across both evaluation splits (Tables 2 and 3). On the Artportalen test partition, the performance leap from the environmental baseline model (ResNet-18_{Env}, 0.924 mAP) to the satellite model (ResNet-18_{Sat}, 0.965 mAP) highlights that high-resolution, multi-sensor foundation embeddings capture localized human landscape patterns, agricultural transitions, and micro-climatic variations that are missing from interpolated bioclimatic variables.

The performance advantage was partially inverted when evaluated against IUCN expert range maps (Table 4). The environmental architecture ResNet-18_{Env} achieved the highest mean score (0.885 mAP) compared to high-resolution satellite alternatives. This variation is a direct consequence of spatial scale: the smooth, continuous gradients of interpolated WorldClim variables match the broad, simplified polygons derived from continental-scale expert consensus. High-resolution satellite embeddings introduce sharp, localized landscape variations that the generalized IUCN range templates classify as false negatives. Nevertheless, for ecologically complex or poorly mapped species such as the Lynx, the satellite models retained a significant advantage (0.911 mAP for Fus-$\text{Res}_{\text{LocSat}}$), confirming their superior generalizability under dataset discrepancy constraints.

Ecological interpretation of visual patch variations. The visual behavior observed in the projections for the brown bear (Fig. 4) and the hedgehog (Fig. 5) demonstrates the ecological necessity of foundation satellite embeddings over traditional climate layers. For the brown bear, Fus-$\text{Res}_{\text{LocSat}}$ refines the broad macro-range generated by climate inputs by carving out specific river networks and deep forest core habitats while excluding localized infrastructure corridors. This directly mirrors large carnivore avoidance behavior toward anthropogenically altered environments.

Conversely, for the hedgehog, a small mammal heavily associated with human-dominated matrices, the climate baseline over-smooths suitable boundaries into surrounding uninhabited woodlands where the species is ecologically absent. By extracting high-resolution features directly from the 50 m unit hypersphere space, Fus-$\text{Res}_{\text{LocSat}}$ successfully maps micro-habitats, habitat fragmentation, and anthropogenic barriers. This prevents the geographic overprediction loops that compromise point-source bioclimatic architectures.

Architectural performance and structural bottlenecks. Integrating coordinate-based implicit spatial representations via late-fusion models produced stable

improvements. On both the Artportalen test split and the Rovbase dataset (with different sampling bias), Fus-$\text{Res}_{\text{LocSat}}$ proved highly resilient, achieving optimal performance markers across the board. This success is due to the simplicity of the fusion mechanism: by concatenating the 1D satellite feature vector with the 1D location encoding right at the final classification layer, the network processes landscape textures and geographic coordinates independently. This decoupling prevents the convolutional layers from overfitting to specific training coordinates, allowing the model to retain strong generalization properties when evaluated on an entirely separate observation system.

Data dynamics and distribution range constraints. Our species-specific analysis indicates that *the spatial distribution of a species is a stronger predictor of model performance than its absolute observation count*. For highly localized species like the Fallow Deer (*Dama dama*) or Mouflon (*Ovis aries musimon*), all tested deep neural architectures achieved near-perfect scores, regardless of whether they used environmental or satellite predictors. This confirms that when a species occupies a well-defined geographic boundary, distance-weighted pseudo-absence generation functions optimally, providing clean and informative negative samples.

Conversely, widely distributed species like the Eurasian otter (*Lutra lutra*) or moose (*Alces alces*) proved more challenging to model, despite having large training pools ($> 14,000$ records). Because these species occur across nearly all latitudes in Sweden, distance-weighted sampling fields collapse, forcing pseudo-absences into marginal or uninformative areas. In this diffuse boundary scenario, the high-resolution texture information provided by the AlphaEarth embeddings becomes essential, allowing the patch-based vision backbone to identify suitable micro-habitats where coarse climate data fails.

References

1. Artdatabanken, S.: Artportalen (2026), https://www.artportalen.se/
2. Beery, S., Cole, E., Parker, J., Perona, P., Winner, K.: Species distribution modeling for machine learning practitioners: A review (2021), https://arxiv.org/abs/2107.10400
3. Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., Rucklidge, W.J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L.L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P.: Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data (2025), https://arxiv.org/abs/2507.22291
4. Chen, D., Xue, Y., Chen, S., Fink, D., Gomes, C.: Deep multi-species embedding (2017), https://arxiv.org/abs/1609.09353

5. Cole, E., Horn, G.V., Lange, C., Shepard, A., Leary, P., Perona, P., Loarie, S., Aodha, O.M.: Spatial implicit neural representations for global-scale species mapping (2023), https://arxiv.org/abs/2306.02564
6. Dorm, F., Lange, C., Loarie, S., Aodha, O.M.: Generating binary species range maps (2024), https://arxiv.org/abs/2408.15956
7. Elith, J., Leathwick, J.R.: Species distribution models: ecological explanation and prediction across space and time. Annual review of ecology, evolution, and systematics **40**(1), 677–697 (2009)
8. Eurostat – GISCO: Countries (2024), https://ec.europa.eu/eurostat/web/gisco/geodata/administrative-units/countries, © EuroGeographics for the administrative boundaries
9. Fick, S.E., Hijmans, R.J.: Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**(12), 4302–4315 (2017). https://doi.org/10.1002/joc.5086
10. IUCN: The IUCN red list of threatened species. version 2025-2 (2025), https://www.iucnredlist.org
11. Keck, F., Peller, T., Alther, R., Barouillet, C., Blackman, R., Capo, E., Chonova, T., Couton, M., Fehlinger, L., Kirschner, D., Knüsel, M., Muneret, L., Oester, R., Tapolczai, K., Zhang, H., Altermatt, F.: The global human impact on biodiversity. Nature **641**(8062), 395–400 (05 2025). https://doi.org/10.1038/s41586-025-08752-2
12. Lantmäteriet: Topografi 50 nedladdning, vektor (2022), https://geotorget.lantmateriet.se/geodataprodukter/topografi-50-nedladdning-vektor
13. Miljødirektoratet, Naturvårdsverket: Rovbase, https://www.rovbase.se/
14. Shivanna, K.R.: Climate change and its impact on biodiversity and human welfare. Proceedings of the Indian National Science Academy **88**, 160–171 (2022). https://doi.org/10.1007/s43538-022-00073-6
15. Teng, M., Elmustafa, A., Akera, B., Bengio, Y., Abdelwahed, H.R., Larochelle, H., Rolnick, D.: Satbird: Bird species distribution modeling with remote sensing and citizen science data (2023), https://arxiv.org/abs/2311.00936