Confident Naturalness Explanation (CNE): A Framework to Explain and Assess Patterns Forming Naturalness in Fennoscandia with Confidence

1 Introdiction

Unaffected by extensive human interference, protected natural areas represent regions of the Earth that maintain their original condition, largely untouched by urbanization, agriculture, logging, and other human activities. These regions host rich biodiversity and offer numerous ecological advantages. They provide unique opportunities to study natural ecosystem processes, such as water and pollination cycles.

Consequently, careful mapping and monitoring of these areas are crucial for uncovering intricate geo-ecological patterns essential for preserving their authenticity. This explains the increasing focus on monitoring and comprehending natural areas in both remote sensing and environmental research.[5, 6]. Satellite imagery enables consistent observation of remote protected areas, surpassing human accessibility challenges. It offers efficient, cost-effective data collection while minimizing disturbances to delicate ecosystems. Utilizing Machine Learning (ML) models, particularly Convolutional Neural Networks (CNNs), enables precise classification of natural regions by analyzing satellite imagery datasets. To illustrate, [2] constructs a dataset and a foundational CNN model that precisely classifies and categorizes these protected natural regions.

In their research analyzing naturalness, Stomberg et al. [7] designed an inherently explanatory classification network that generates attribution maps. These maps effectively highlight patterns indicative of protected natural areas in satellite imagery. [] also introduce an approach that generates images with highlighted naturalness patterns utilizing Activation Maximization and Generative Adversarial Networks (GANs)[4, 8]. This approach provides comprehensive and valid explanations for the authenticity of naturalness.

Nevertheless, while these methods effectively identify designating patterns that characterize the authenticity of natural regions, they face challenges in offering a quantitative metric that precisely represents the contribution of these discerning patterns. Additionally, these methods do not tackle the issue of uncertainty associated with the assigned importance of each individual pattern.

To overcome these limitations, we introduce an innovative approach that integrates explainability and uncertainty quantification. Our aim is to establish a novel metric that captures both the significance and confidence associated with each pattern. Notably, our contributions extend to developing certainty-aware segmentation masks. These masks not only yield precise segmentation outcomes but also pinpoint pixels where the model showcases high uncertainty.

2 Framework

The cornerstone of our work is the development of the Confident Naturalness Explanation (CNE) metric. This metric is utilized to prioritize and arrange the contributing patterns based on their inherent quality, thereby deepening our grasp of the concept of naturalness in satellite imagery. Through these breakthroughs, we collectively amplify the interpretability and confidence levels of the model's insights. As a result, we facilitate a more comprehensive understanding of the intricate naturalness patterns inherent in satellite imagery. The CNE framework consists of two main parts, the explainability and the uncertainty quantification part. In the first part, we use a grey box approach [1] to assign an importance value to each pattern contributing to the concept of naturalness. The grey box approach consists of a black box semantic segmentation model that lacks interpretability and a transparent model that is responsible for explaining the decision-making mechanism of the black box model. In the second part, we utilize the MC-Dropout [3] technique to quantify the uncertainty in predicting the classes contributing to naturalness. We

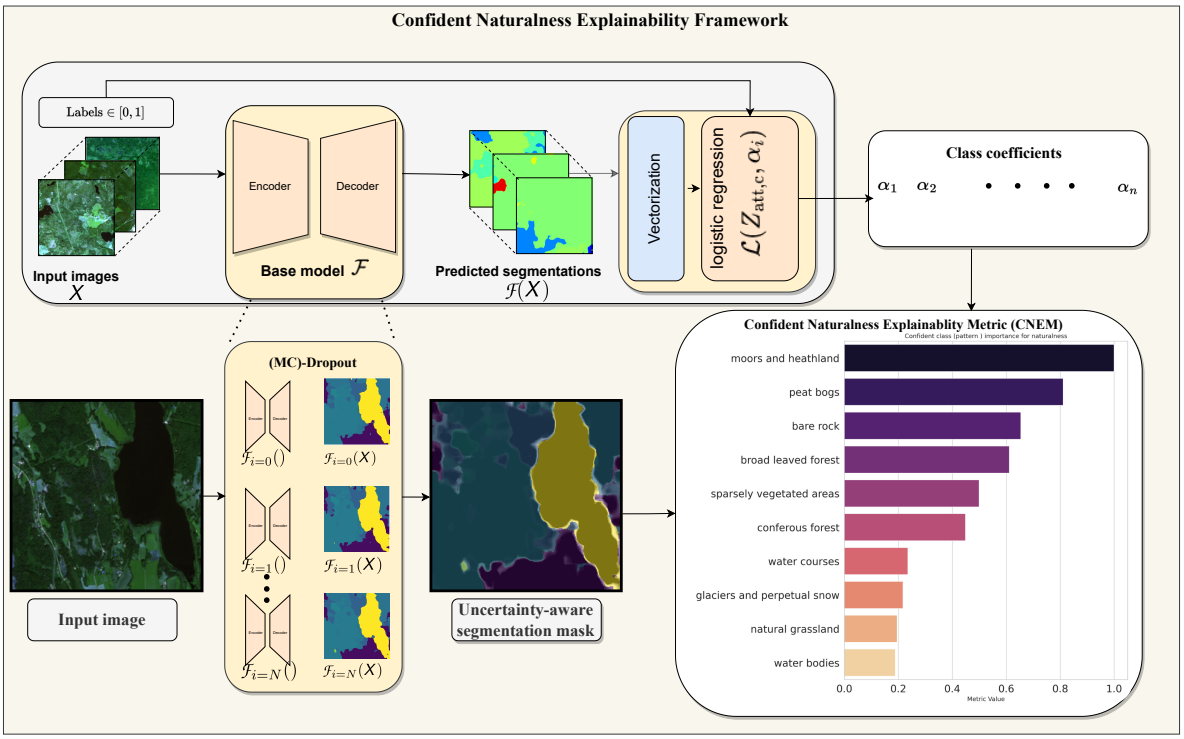

Figure 1: **An Illustration for (CNE) framework.** In the grey box, the input images are fed to the segmentation model resulting in predicted segmentation masks; they are fed to logistic regression $\mathcal{L}(\boldsymbol{Z}_{\text{att},c}, \alpha_i)$ with ground truth labels of the input images to form a vector the represents the abundance of each pattern in the predicted segmentation mask $\boldsymbol{Z}_{\text{att}}$ and classifies the input into protected natural areas and non-protected areas. For the MC-Dropout, multiple sampled models are used to quantify the uncertainty of each pattern in the input image. In the lower right corner, we combine the knowledge gained from the upper and lower part to calculate the CNE metric and assign a quantifiable metric value to each pattern, reflecting its confident contribution to the concept of naturalness.

Class	Metric
moors and heathland	1.000000
peat bogs	0.811245
bare rock	0.653868
broad leaved forest	0.611532
sparsely vegetated areas	0.499592
conferous forest	0.448674
water courses	0.235142
glaciers and perpetual snow	0.217038
natural grassland	0.195522
water bodies	0.188863

Table 1: CNE metric values for different patterns contributing to the concept of naturalness

integrate the gained knowledge to create the CNE metric, which assigns confident importance to the patterns forming naturalness in Fennoscandia.

3 Results and Discussion

Our investigation unveiled that various wetland patterns possess notably high CNE metric values, ranging from 0.8 to 1. These scores signify the existence of high-quality patterns that significantly contribute to the concept of naturalness with high certainty. Wetlands are pivotal ecosystems renowned for their roles in carbon storage, safeguarding biodiversity, regulating water resources, and providing niches for unique plant and animal species finely adapted to their specific surroundings.

In contrast, Glaciers, Grasslands, and water bodies exhibit relatively low-quality patterns, with an approximate metric value of 0.2. These values indicate patterns with a diminished contribution to the naturalness concept, accompanied by heightened uncertainty. This insight will be further expanded upon in the table 1, detailing various patterns alongside their respective metric values.

References

[1] A. Bennetot, G. Franchi, J. Del Ser, R. Chatila, and N. Diaz-Rodriguez. Greybox XAI: a Neural-Symbolic learning framework to produce interpretable predictions for image classification, Sept. 2022. URL http://arxiv.org/abs/2209.14974. arXiv:2209.14974 [cs].

[2] B. Ekim, T. T. Stomberg, R. Roscher, and M. Schmitt. MapInWild: A Remote Sensing Dataset to Address the Question What Makes Nature Wild, Dec. 2022. URL http://arxiv.org/abs/2212.02265. arXiv:2212.02265 [cs].

[3] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *ArXiv*, abs/1506.02142, 2015.

[4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks, June 2014. URL http://arxiv.org/abs/1406.2661. arXiv:1406.2661 [cs, stat].

[5] R. A. Mittermeier, C. G. Mittermeier, T. M. Brooks, J. D. Pilgrim, W. R. Konstant, G. A. B. da Fonseca, and C. Kormos. Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences*, 100(18): 10309–10313, Sept. 2003. doi: 10.1073/pnas.1732458100. URL https://www.pnas.org/doi/10.1073/pnas.1732458100. Publisher: Proceedings of the National Academy of Sciences.

[6] R. J. Smith and A. N. Gray. Strategic monitoring informs wilderness management and socioecological benefits. *Conservation Science and Practice*, 3(9):e482, 2021. ISSN 2578-4854. doi: 10.1111/csp2.482. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/csp2.482. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/csp2.482.

[7] T. T. Stomberg, T. Stone, J. Leonhardt, I. Weber, and R. Roscher. Exploring Wilderness Characteristics Using Explainable Machine Learning in Satellite Imagery, July 2022. URL http://arxiv.org/abs/2203.00379. arXiv:2203.00379 [cs].

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Aug. 2020. URL http://arxiv.org/abs/1703.10593. arXiv:1703.10593 [cs].