

On the relation between landscape beauty and land cover: A case study in the U.K. at Sentinel-2 resolution with interpretable AI

Alex Levering^{a,*}, Diego Marcos^a, Devis Tuia^{a,b}

^a Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, the Netherlands

^b Environmental Computational Science and Earth Observation Laboratory, EPFL, Switzerland

ARTICLE INFO

Keywords:

Landscape aesthetics
Deep learning
Interpretable AI
Corine land cover
Sentinel-2

ABSTRACT

The environment where we live and recreate can have a significant effect on our well-being. More beautiful landscapes have considerable benefits to both health and quality of life. When we chose where to live or our next holiday destination, we do so according to some perception of the environment around us. In a way, we value nature and assign an ecosystem service to it. Landscape aesthetics, or scenicness, is one such service, which we consider in this paper as a collective perceived quality. We present a deep learning model called ScenicNet for the large-scale inventorisation of landscape scenicness from satellite imagery. We model scenicness with an interpretable deep learning model and learn a landscape beauty estimator based on crowdsourced scores derived from more than two hundred thousand landscape images in the United Kingdom. Our ScenicNet model learns the relationship between land cover types and scenicness by using land cover prediction as an interpretable intermediate task to scenicness regression. It predicts landscape scenicness and land cover from the Corine Land Cover product concurrently, without compromising the accuracy of either task. In addition, our proposed model is interpretable in the sense that it learns to express preferences for certain types of land covers in a manner that is easily understandable by an end-user. Our *semantic bottleneck* also allows us to further our understanding of crowd preferences for landscape aesthetics.

1. Introduction

In a time where increasing urbanization is a constant factor across the world, we sometimes need a break from the busy and tiring reality of the modern city to enjoy greener and relaxing landscapes. Landscape beauty, also referred to as scenicness, is indeed a driver for tourism (Krippendorf, 1984), while it is also a driver for the creation of cultural value (Havinga et al., 2020; Daniel et al., 2012). Beyond providing tourists and artists a place to seek out, landscape scenicness has also been found to improve people's quality of life. Velarde et al. (2007) reviewed literature covering the relationship between health and landscape beauty, and found that observing scenic landscapes is associated with a reduction in stress, improved attention capacity, better recovery from illnesses, a feeling of general well-being, and positive improvements to one's mood. Grinde and Patil (Sep. 2009) conducted a literature study on the relationship between plants and quality of life and found that the absence of plants is associated with a lower quality of life and health. Seresinhe et al. (2015) quantified the relationship between scenicness and self-reported health, and found that scenic environments

are associated with an increase in self-reported health. In a later study, they also considered the relationship between self-reported happiness and landscape beauty, and found that people are happier in scenic environments (Seresinhe et al., 2019). As such, there is a significant incentive to knowing where scenic landscapes are located, as well as to understand the factors which contribute to landscape scenicness.

Much research has been devoted into determining landscape scenicness. Theoretical research on the topic stems back to the 1960s through the 1980s, when major theories about human-landscape interactions were formed, as summarized by Schroeder and Daniel (Schroeder and Daniel, 1981). A popular measure for landscape beauty at the time was the *Scenic Beauty Estimate*, which depended on crowdsourced ratings based on images of the landscape (Daniel, 1976). As scenicness is a subjective quality (since 'beauty is in the eye of the beholder'), accessing such information directly from the observer was (and still is) the only possible way, in the hope that the individual subjective views would then converge to a set of collective rules of perceived beauty. The practice of estimating landscape beauty then adopted digital means by the time that computers and geo-information

* Corresponding author.

E-mail address: alex.levering@wur.nl (A. Levering).

<https://doi.org/10.1016/j.isprsjprs.2021.04.020>

Received 19 January 2021; Received in revised form 27 April 2021; Accepted 28 April 2021

Available online 24 May 2021

0924-2716/© 2021 The Author(s). Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an

open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

systems became widely available, such as relating crowdsourced scenicness beauty estimates to land cover types through geo-information systems (Palmer, 2004). Recent efforts in data collection (Seresinhe et al., 2017) led to the distillation of the first large scale crowd-sourced dataset of landscape preferences, called ScenicOrNot¹, consisting of 217,000 ground-level images with scenicness scores from three or more annotators. This dataset is of sufficient size and diversity to allow for the emergence of machine learning research aimed at the automatic estimation of landscape scenicness, which was mostly tackled by means of convolutional neural networks (Marcos et al., 2019; Seresinhe et al., 2017; Workman et al., 2017). However, it may be difficult to acquire ground-based images of remote regions, such as those from the Geo-graph project², on which the ScenicOrNot dataset is based. For such regions, it could be beneficial to use remote sensing imagery, which are available globally and are frequently updated, to provide the scenicness assessment. Furthermore, remote sensing imagery is not affected by ground-based image biases such as weather patterns such as cloudy versus blue skies or the presence of rainbows, or photographers' biases on which scenes or objects to photograph. In this respect remote sensing imagery could be considered more objective than ground based images. The question remains to know if it is possible to predict scenicness directly from remote sensing images: in other words, we formulate the hypothesis that the characteristics visible in satellite images (e.g. land cover), allow us to estimate the beauty of the landscape. To verify this hypothesis, we resort to a deep learning approach.

In recent years, Convolutional Neural Networks (CNNs) have become a popular tool for image analysis in the remote sensing domain (Zhu et al., 2017). CNNs are commonly applied to typical remote sensing tasks such as land classification (Sumbul et al., 2019; Demir et al., 2018), or precise objects delineation at very high resolution (Campos-Taberner et al., 2016; Maggiori et al., 2017; Volpi and Tuia, 2017). While they are traditionally applied to RGB and multispectral data, there nowadays exists a wide corpus of literature about the use of deep learning for other modalities, such as hyperspectral remote sensing (Audebert et al., 2019). As a result, deep learning is becoming increasingly popular in the geosciences community, where the technology is used to tackle a wide range of problems, such as weather prediction, snow pack modeling or climate change monitoring (Camps-Valls et al., 2021).

However, their superior performance on a variety of tasks comes at a price of interpretability, since CNNs offer less transparency in their predictions compared to other machine learning models. Researchers in machine learning are therefore increasingly stressing the importance of interpretability in deep learning systems (Samek and Müller, 2019; Miller, 2017) in order to be able to challenge the assumptions of deep neural networks and to assess whether a model is trustworthy. Additionally, interpretability can be used to discover meaningful patterns to further our understanding of which learned patterns matter most (Lapuschkin et al., 2019).

While interpretability as a means of improving trust in deep learning models has picked up considerably in computer vision, it is still in its infancy for remote sensing tasks, and traditional machine learning methods have proven to be easier to interpret (Huysmans et al., 2011). In particular, understanding how variables contribute to predictions has been heavily studied with tree-based and kernel methods. Tree-based methods allow for interpretability by ranking input variables according to their influence on the final prediction, such as mode impurity and mean decrease in accuracy for Random Forest models (Biau and Scornet, 2016). Gaussian Processes allow for model inversion and parameter retrieval through their confidence intervals (Svendsen et al., 2020). Linear combinations of multiple kernels can be used to obtain variable importance estimates for kernel methods (Tuia et al., 2010). But when it comes to deep learning methods, the ranking of inputs importance is less

straightforward, and the one of the inner feature needs extra engineering steps. Instead, Post-hoc input attribution methods such as Class Attention Mapping (Zhou et al., 2016) are frequently considered as a solution to the interpretation problem for deep neural networks trained on remote sensing imagery. These methods are used to highlight which regions of the image contribute the most to the output of the model. They are commonly used in various object retrieval tasks, such as locating solar panels (Imamoglu et al., 2017), structures of interest (Vasu et al., 2018), or airplanes (Fu et al., 2019). Attribution methods such as Class Activation Maps (Zhou et al., 2016) work well when there is a clear right or wrong answer visible in the image. For instance, an airplane can be clearly identified by a human in a very high resolution satellite image, making the correctness of a pixel attribution method easy to verify. However, attribution methods are less effective when a task is subjective or when it depends on the coalescence of multiple patterns, which cannot easily be highlighted in the image. Scenicness is one such task, as landscape beauty can be the result of the interplay between visible elements of the landscapes, and such interplays cannot easily be highlighted in the input images. We therefore have to consider alternative interpretation methods to explain our predictions.

To help us understand the drivers of landscape scenicness using deep learning, we adopt semantic bottlenecks (Marcos et al., 2020; Marcos et al., 2019), which use the prediction results of an intermediate task, ideally objective and made of human-understandable concepts, to predict the target task, while still allowing models to be trained in an end-to-end fashion. Such models have previously been applied for scenicness estimation from ground based images. As proposed in (Marcos et al., 2019), the prediction of image scenicness may depend on its content, such as the presence of snow, clouds, or roads. The presence of each object or concept may then be used to create a scoring vector for the prediction of scenicness. In that case, the semantic bottleneck was therefore made of a series of scene class objects and, to each object, a positive (this objects impacts scenicness positively) or negative (this object impacts scenicness negatively) weight was assigned. The final score was made of a bias (average scenicness) plus the combination of the single detected object scores. We build on this concept for ground-based images and adapt it to the task of scenicness prediction from remote sensing imagery while using land cover as an interpretable intermediate task. In doing so, we improve on our preliminary study (Levering et al., 2020) by adapting our model to accommodate differing scenicness scores within the same land cover class, since depending on the context one land cover type can impact positively, negatively, or not at all the beauty score. In addition to the estimation of the scenicness of landscapes, our model therefore also allows us to study the relationship between landscape scenicness and land cover types.

In this paper we conceptualize an interpretable deep learning model for remote sensing imagery which uses land cover prediction as an intermediate task for landscape scenicness regression (Section 2). We train our model to reproduce average ScenicOrNot beauty score at the level of single patches extracted from Sentinel-2 images over the United Kingdom. We implement a semantic bottleneck forcing predictions to be explicit in the land cover classes that the model is observing and explicitly using to predict the scenicness. To do so, we use the Copernicus CORINE land cover inventory (EU Copernicus Program, 2018) and predict intermediate land cover multilabel maps. Our results (Section 4) show that we can extend existing scenicness prediction models with an interpretable bottleneck without experiencing any loss of accuracy, neither in the scenicness nor land cover prediction task. In return, our model provides explanations about what it is observing and what leads it to decide for a certain beauty prediction. As such, it becomes simple to challenge the decisions of the model and analyze errors.

2. Methods

We propose an interpretable model for landscape scenicness estimation that uses a semantic bottleneck (Marcos et al., 2019). We design

¹ <http://scenicornot.datasciencelab.co.uk/>

² <https://m.geograph.org.uk>

the semantic bottleneck such that it uses the outputs of a land cover prediction task to estimate the scenicness of a given satellite image. We refer to our model as **ScenicNet**.

Our model is summarized in Fig. 1. It uses a standard CNN backbone tasked with feature extraction, a multi-label land cover classifier intermediate head and a scenicness regressor, which depends linearly on the output of the land cover classifier. Since it comprises two separate prediction heads considering different tasks learned from different datasets (see Section 3), it can be seen as a multitask model such as in (Marmanis et al., 2018; Volpi and Tuia, 2018).

Our main contribution is a method to disambiguate intra-class scenicness differences by allowing the model to discover sub-classes with different scenicness values associated to them. We call these sub-classes *modes*. Each mode corresponds to a neuron within a group of neurons associated to the same land cover class. Each mode is also connected to the scenicness head (via the weights w described below). Each mode therefore contributes to both the detection of land cover and to the estimation of beauty. The number of modes per class is defined by a hyperparameter, M , manually set. Summing up, for each land cover class $c \in C$, the model has M outputs, each with an associated learned scenicness weight. This means that a land cover class can influence scenicness positively when in a given association of classes, and then negatively when associated to others. Depending on the specific association, one or the other mode of the class will be activated.

2.1. Land cover head

Our model first has to predict C land cover classes from the feature extractor. The feature extractor produces $C \times M$ scores $Z \in \mathbb{R}^{C \times M}$ for each mode input $m \in \{1, \dots, M\}$ belonging to a given class $c \in \{1, \dots, C\}$, where $z_{c,m}$ corresponds to the features of mode m in class c . These scores are then normalized, Eq. (2) and summed for each class, Eq. (3), to obtain the C land-cover class scores as a vector $\hat{y} \in \mathbb{R}^C$. As depicted in Fig. 1, the land cover prediction problem is casted as multi-label, i.e. every class is considered separately and can be detected simultaneously with others. We use a binary cross entropy loss for every land cover class $c \in \{1 \dots C\}$ and compare predictions \hat{y} with the ground truth $y \in \{0, 1\}^C$.

For the purposes of scenicness prediction, we want to force the model to choose which mode to use for a given sample to reduce ambiguity on which modes contributed to each prediction. In order to keep the scenicness prediction layer interpretable, we also want the model to only keep the modes that have a meaningful contribution to the prediction process active. To do so, we first calculate a Softmax non-linearity for

each mode input $m \in \{1, \dots, M\}$ belonging to a given class $c \in \{1, \dots, C\}$:

$$\text{softmax}(z_{c,m}) = \frac{e^{z_{c,m}}}{\sum_{j=1}^M e^{z_{c,j}}} \quad (1)$$

For each element $z_{c,m}$ we then multiply their respective softmax scores with a sigmoid over the mode input to compute the mode presence probability for a given mode $r_{c,m}$ of matrix R :

$$r_{c,m} = \text{sigmoid}(z_{c,m}) \cdot \text{softmax}(z_{c,m}) \quad (2)$$

The softmax ensures that only one mode is dominantly active as all class-specific contributions sum to one. Through direct multiplication with the sigmoid non-linearity we allow the model to indicate which modes are active, if any. We can then use this mode presence matrix R to obtain class presence scores by summing all mode presence scores $r_{c,m}$ belonging to a given class c :

$$\hat{y}_c = \sum_{m=1}^M r_{c,m} \quad (3)$$

We can use these class-wise land cover presence scores in the following sum over c binary cross-entropy functions (one per land use class) (Fig. 2a):

$$\mathcal{L}_{CLC}(y, \hat{y}) = - \sum_c \hat{y}_c \log(y_c) + (1 - y_c) \log(1 - \hat{y}_c) \quad (4)$$

Where y is the ground truth for a single sample from the land cover dataset.

The gradients learned from the land cover prediction (in pink to purple colors in Fig. 1) are then backpropagated into the main body of the CNN through the class-specific multi-mode land cover bottleneck. The updated mode presence scores R will therefore impact the scenicness prediction described in the next section.

2.2. Scenicness prediction head

The second head of our model is responsible of predicting the landscape scenicness as a regression problem. In order to regress a scenicness value, our model multiplies a learnable weighted matrix $W \in \mathbb{R}^{C \times M}$ elementwise with the mode presence scores matrix R to create a matrix V with mode-specific scenicness contributions, where $v_{c,m}$ represents the contributions of a single mode:

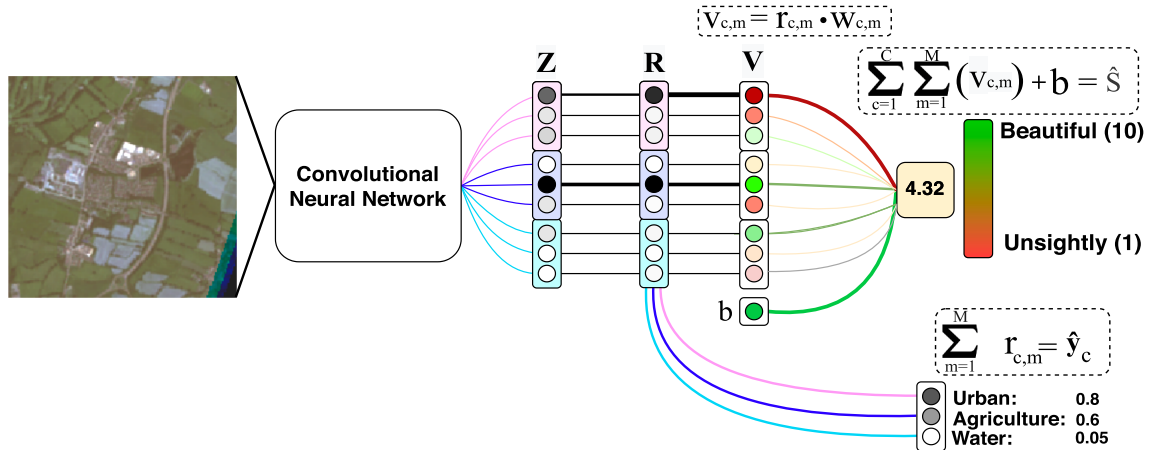


Fig. 1. Architecture overview of our semantic bottleneck. The model first extracts a matrix of Z features from a satellite image. Over these Z features it then multiplies a classwise softmax with a sigmoid non-linearity (Eq. (2)) to extract mode presence scores R . Land cover presence is predicted from these features by summing the resulting matrix (Eq. (3)). We multiply this presence matrix with one learned weight per mode to derive their scenicness contribution for a given sample (Eq. (5)). The sum of all modes is added together with a bias term to create the final scenicness prediction (Eq. (6)).

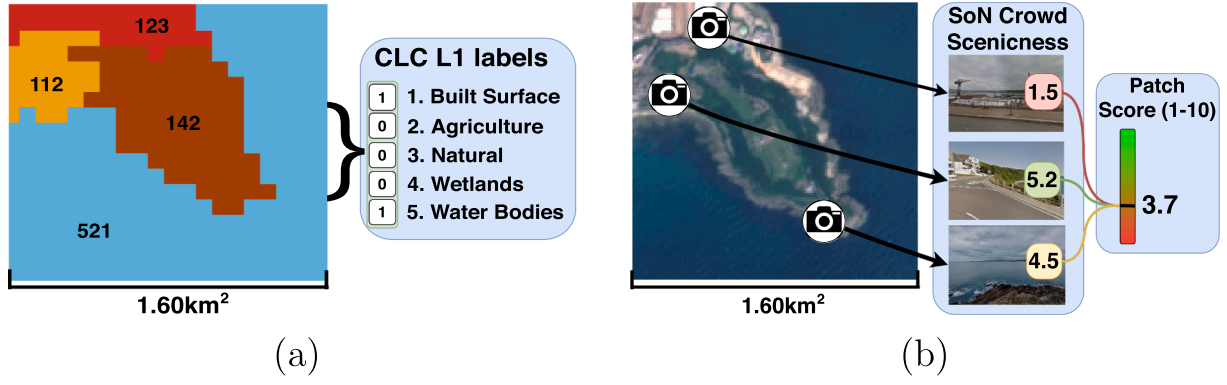


Fig. 2. Ground truth creation; (a) CORINE values are aggregated to their 1st digit, then assigned a binary present/not-present label. (b) SoN image scores within the patch boundary are averaged, which gives us the patch scenicsness score.

$$v_{c,m} = r_{c,m} \cdot w_{c,m} \quad (5)$$

The sum of all mode contributions is then added together with a bias term $b \in \mathbb{R}$ in order to compute the predicted scenicsness value:

$$\hat{s} = \left(\sum_{c=1}^C \sum_{m=1}^M v_{c,m} \right) + b \quad (6)$$

We then use this predicted scenicsness score to compute the following squared error loss function:

$$\mathcal{L}_{SoN}(s, \hat{s}) = (s - \hat{s})^2 \quad (7)$$

Where s is the crowdsourced scenicsness score for a single sample. During training, we backpropagate the mean squared error of each batch.

With a choice of $M > 1$, our model can learn more than one representation for each $c \in \{1 \dots C\}$ classes. However, we want to encourage the model to use the minimum number of modes needed for the prediction task to stop the model from forming complex non-linear interactions between multiple modes. We encourage this through the softmax in Eq. (2), through which we limit the activation budget of the model. The softmax rescales the contributions of each mode relative to all mode activations within a class. Therefore the model cannot activate all modes equally, forcing it to make deliberate choices on which modes to use for each training example.

2.3. Combined loss function

Each one of the two processing heads of the model backpropagates gradients related to a loss specific either to the land cover task (\mathcal{L}_{CLC} , Eq. (4)) or to the scenicsness estimation task (\mathcal{L}_{SoN} , Eq. (7)). The final loss of our explainable model is obtained by a weighted combination of the two terms:

$$\mathcal{L} = \mathcal{L}_{SoN} + \lambda \mathcal{L}_{CLC} \quad (8)$$

where λ is a weighting term set empirically.

3. Data and setup

3.1. Data

Our model is concurrently trained on two tasks, namely land cover prediction and scenicsness regression. In order to generate the training data for both tasks, we lay out a regular grid of 1.60 km by 1.60 km across the entirety of Great Britain as a common prediction grid. For each grid cell we then collect three data sources (Fig. 2): 1) A land cover inventory, 2) a landscape scenicsness dataset with location information, and 3) satellite imagery with a maximum of 1% cloud coverage across Great Britain.

- **Land cover.** For the land cover prediction we make use of the CORINE land cover inventory of 2018 (EU Copernicus Program, 2018). The CORINE Land Cover (CLC) is a pan-European dataset created from a combination of Sentinel-2 imagery and national land cover products. It consists of a hierarchy of three levels. CLC Level 1 consists of five land cover classes; 1) Urban, 2) Agriculture, 3) Forests and natural areas, 4) Wetlands, and 5) Water. CLC level 3 contains fine-grained land cover classes, such as 111) Continuous Urban Fabric, and 421) Salt Marshes. For our experiments, we use CLC Level 1 as training labels, and we use the L3 labels for a qualitative assessment of the modes of our model in the discussion Section. We opt for a more simplistic land cover classification task to ensure that the model is able to learn an accurate representation of land cover classes. For each grid cell we create a binary vector where 0 and 1 denote absence and presence for each class. We show this process as well as the land cover classes of the first-level hierarchy of CLC in Fig. 2a.

- **Landscape scenicsness.** We derive our landscape aesthetics score from ground-based image evaluations from the ScenicOrNot dataset. ScenicOrNot (SoN) is a crowdsourced dataset consisting of 215,000 ground-level images across Great Britain obtained from the Geograph UK project. Each image is rated with a score between 1 (not scenic) to 10 (most scenic) for their landscape aesthetic beauty by one or more volunteers on an openly accessible online platform. Moreover, each image is stored with their geolocation, and as such they can be analyzed spatially. For each grid cell in our regular grid we assign the average scenicsness score of the geotagged images within its bounds. We display this process in Fig. 2b. Fig. 3 illustrates the final ground truth, as well as the histogram of its distribution across the U.K.

- **Remote sensing data.** As input to our model we use Sentinel-2 satellite imagery. We download atmospherically corrected (L2A) satellite tiles with at most 1% cloud coverage across Great Britain, which have been taken between 2018 and 2019. We retain the 10-and 20 meter resolution bands of each satellite tile. We upsample the 20 m-resolution bands to 10 m using nearest neighbour interpolation. We remove any image patches which have all-zero values in the red, green, or blue colour bands. In total, we collect 121,067 patches of size 160×160 pixels, corresponding to an extent of 1.60×1.60 kilometers each. Land cover information is available for all of these patches, while scenicsness scores are available for 83,374 patches. We randomly sample splits of 75/15/10% for training, validation, and testing. We sample without geographical stratification to maximize the opportunities for the model to learn meaningful scenicsness differences for each class.

The scripts for creating our ground truth dataset can be found in the following Zenodo repository: <https://zenodo.org/record/4762134>. This

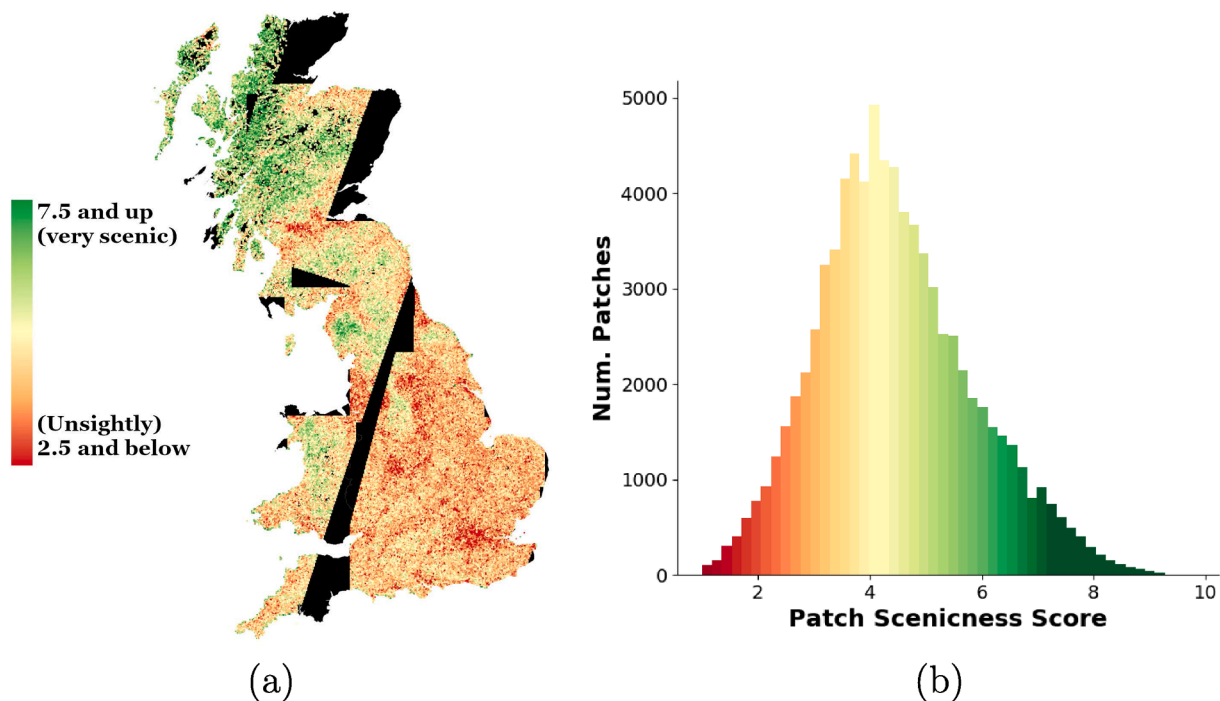


Fig. 3. Ground truth creation; (a) Map of the ground truth scores of every patch in our dataset. (b) Histogram of all patch scores.

repository also contains a PyTorch implementation of our model architecture.

3.2. Set-up

As feature extractor we use a ResNet-50 (He et al., 2016), which has not been pre-trained as we use multi-spectral imagery. We set the number of class-specific modes M to 3 and we initialize the weights in W for the class-specific modes to 0.5, 0.01, and -0.5 respectively, so that the model is encouraged to develop non-symmetrical scenicness contributions for each mode. The scenicness prediction of our model is dependent on the land cover prediction task, but during training both tasks compete for signal. To avoid that the model learns a bottleneck optimized for scenicness and that is not aligned with the CLC semantics, we set a larger weight λ for the land cover loss in Eq. (8), to a factor 10.

We explore the benefit of having multiple modes by running a baseline experiment with M , the number of sub-class nodes per class, set to 1, which makes it functionally equivalent to a linear regression dependent on the class prediction score. We train both models for 15 epochs with the ADAM optimizer (Kingma and Ba, 2015). We set the initial learning rate to 0.0005 and we add a weight decay factor of 0.0001. We use 16 samples per batch. During training we weight each loss by the inverse square root of their frequency so that we can train on a balanced number of samples.

For every training iteration, we sample one batch to compute Eqs. (4) and (7). If both labels (CLC and SoN) are available for a given patch, then we compute both losses for the sample. When processing a sample only having land cover information (and no ScenicOrNot label), we set the loss of Eq. (7) to 0. We combine and backpropagate the losses according to Eq. (8). We repeat this procedure until the smallest dataset (SoN) is exhausted, at which point the epoch ends.

We compare our models against unconstrained ResNet-50 models trained on each task separately. For the land cover prediction task, we set the number of outputs of the final fully-connected layer to 5 to equal the number of CORINE classes in the level-1 hierarchy. For the task of scenicness regression we set the number of outputs of the fully-connected layer to 1 such that the model regresses one single scenicness, as in (Workman et al., 2017; Levering et al., 2020). We also test the

performance of our model with $M = [2, 5]$ using the same training settings, but with a random initialization of W . We evaluate the land cover prediction performance of our model using the average F1-score (van Rijsbergen, 1979) for each class. The F1-score gives the harmonic mean between the precision and the recall of a given class. A value of 1 indicates perfect precision and recall. To assess the scenicness performance of our model we use the root mean squared error (RMSE) across all examples. We also calculate Kendall's τ (Kendall, 1938) over the predicted scenicness scores, which is a ranking correlation coefficient which tests whether two arrays have similarly-ranked values. For Kendall's τ , 1 indicates a perfect relationship between the predicted scores and the ground truth, and -1 indicates the inverse.

Finally, we compare the results of our scenicness regression to models which directly regress the scenicness score from the CORINE ground truth labels. We train a linear model using the level-1 hierarchy of CORINE to compare to our 1-mode linear bottleneck. We then train a random forest regressor (Breiman, 2001) with 50 trees, a maximum depth of 25, and a minimum of 5 samples per split on the L1 and L3 CORINE ground truth labels to test the performance of our multi-mode models against.

4. Results and discussion

4.1. Numerical scores

In Table 1 we display the numerical performances of the four

Table 1
F1-score, RMSE, and Kendall's τ of each model on the test set.

	land cover F1-score	Scenicness	
RMSE	Scenicness τ		
Only CORINE	0.846	–	–
Only SON	–	1.027	0.452
ScenicNet (1 mode)	0.859	1.080	0.435
ScenicNet (2 modes)	0.867	1.053	0.441
ScenicNet (3 modes)	0.872	1.038	0.456
ScenicNet (5 modes)	0.872	1.036	0.457

considered models. Each of our ScenicNet models outperform an unconstrained network on the land cover prediction task. Our 3-mode and 5-mode ScenicNet models also match the scenicness regression baseline on the Kendall's τ . Our results show that our ScenicNet model is able to leverage its modes to learn complex land cover class representations which relate to scenicness in varying ways, rather than the single learnable pattern for the 1-mode model. The numerical improvements of our multi-mode ScenicNet models on the land cover F1-score also indicates that the land cover prediction task seems to benefit from the scenicness prediction task, which is an underlying assumption of multi-task learning (Caruana, 1997).

For the baseline and the 3-mode ScenicNet model we also present the precision, recall, and F1-score for each land cover class, which can be found in Table 2. Our 3-mode ScenicNet model improves on the baseline for land cover prediction on all land cover classes. In the cases of urban and wetlands our model particularly improves the number of recalled samples.

To test the relationship between land cover and scenicness, we compare our models against a linear regressor and a random forest regressor which use the land cover ground truth labels to directly regress the scenicness score. We show our results in Table 3. Remarkably, our linear 1-mode model outperforms the score-to-score regression models. We hypothesize that our model is able to provide better performances in predicting scenicness from LC classes by allowing for subtle modifications to the LC probability maps that help with scenicness regression. Our results also show that these subtle modification not only do not degrade the LC prediction performance, but actually provide a substantial boost due to the synergy between the two tasks. By contrast, both a linear model and a random forest regressor use only the binary label present in the ground truth, without the possibility of tweaking it to improve the scenicness prediction performance.

4.2. Mode Activity

While our model is initialized with M modes, the Softmax function of Eq. (2) lets the model spend an activation budget across its modes. Through this activation budget, the model develops the tendency to allocate the vast majority of the signal on a single mode. By doing so, we encourage the model to learn a specific mode only if it needs to account for classes with contrasting scenicness values, such as forests near a city compared to forests in a scenic highland. As a result, it can occur that modes for some classes become inactive (i.e. the sigmoid + softmax combination never activates above 0.5), as there are too few intra-class contradictions to account for. In the case of $M = 3$, we found that the model eventually converges to use 2 modes per class at most, while the inactive modes can be pruned without affecting the performance of the model. Setting M to 2 resulted in a solution that is slightly worse than $M = 3$, while $M = 5$ resulted in a model with similar performance. As $M = 3$ gives a model with similar performance but less complexity, we chose this model for our experiments and discussion. We list the active modes of our 3-mode model for each class in Table 4. The choice of M should therefore be determined through experimentation, as it should

Table 2

Class-wise performance metrics of the CORINE baseline and ScenicNet with 3 modes. In each column we display the performance of the baseline on the left, and our model on the right.

	Precision		Recall		F1	
	Base- line	Ours	Base- line	Ours	Base- line	Ours
Urban	0.859	0.865	0.701	0.740	0.772	0.798
Agriculture	0.971	0.974	0.936	0.946	0.954	0.960
Forests and Natural	0.848	0.974	0.821	0.946	0.835	0.960
Wetlands	0.805	0.781	0.617	0.775	0.699	0.778
Water	0.973	0.965	0.968	0.979	0.970	0.972

Table 3

RMSE, and Kendall's τ of models trained to regress scenicness from the land cover ground truth labels.

RMSE	Scenicness	
	Scenicness τ	
Linear (L1)	1.150	0.417
Random Forest (L1)	1.081	0.425
Random Forest (L3)	1.061	0.444

capture the latent dynamics between the two tasks.


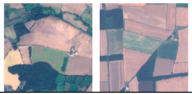



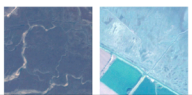
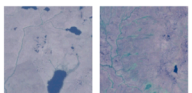

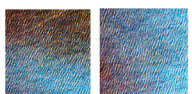
4.3. Visual evaluations

In this section we evaluate the performance of our 3-mode model, as well as its activation patterns and the behaviour of its mode. Fig. 4 illustrates the scenicness predictions of our 3-mode ScenicNet model alongside the ground truth. As can be seen, the model picks up on the major patterns of the ground truth scenicness labels. It is visible that major cities such as London and Manchester are considered unsightly, while Scotland and Wales are considerably more scenic than England. Our model also captures the relationship between elevated areas and scenicness, where higher areas typically correlate with greater scenicness. However, it is also apparent that the model is unable to approximate extreme values, such as those found in downtown London and the Scottish highlands, which we suspect to be caused by an under-representation of these values in the ground truth, as suggested by the histogram in Fig. 4.

We explore the latent space of our 3-mode ScenicNet model to understand which patches our model considers visually similar. Our main interest with this experiment is to discover whether visually-related areas and concepts are similar in the high-dimensional latent space of the CNN model. We reduce the 2'048 outputs of the feature extractor to 100 principal components using a Principal Component Analysis (Pearson, 1901), which are then reduced to 2 dimensions using t-SNE dimensionality reduction (Maaten and Hinton, 2008). t-SNE is a non-linear visualization technique that performs dimensionality reduction by learning an embedding that preserves neighborhood structures, i.e. samples that are neighbors in the high dimensional space must remain neighbors after projection. For the t-SNE hyperparameters, we use a perplexity of 300, a learning rate of 200, and we set the number of iterations to 1'000. We show the resulting plots for the predicted scenicness and class labels in Fig. 5. We find that the latent space of our model is organized by the predictions made through the class-specific modes. Predictions routed through each mode relate to strongly differing land cover archetypes, which are grouped by their relative scenicness. This organizes the latent space into an arrangement where both similarity in land cover visuals (e.g. "bare rocks" and "sparsely vegetated") as well as their relative scenicness are important. An example of this behaviour can be seen on the overlap between modes 3+ and 4+: activations of both of these modes are neighbours in the latent space, while they both have a considerably high learned scenicness score. From Table 4, we can infer that these modes are activated by a similar set of fine-grained land cover concepts, namely highland and plains environments. These findings are encouraging as they indicate that the model is consistent in the concepts it considers to be scenic between different but related land cover classes. The plots of the modes also reveal a gradual transition in visual similarity from man-made land cover classes to natural areas. The large cluster in the center is dominated by un-scenic agriculture and urban land covers, which corresponds to England's countryside. To the right, it is connected with and slowly transitions into a cluster dominated by mixed agriculture and woodland environments typically found in Wales, the north of England, and Scotland. From this transition we infer that the model considers natural areas to be more scenic. This pattern is reflected in the gradient of the top-left cluster. It sees urban areas on the far-left of the cluster

Table 4

Modes for each class with their learned scenicness score, and their most-recalled level-3 CORINE labels. We renamed modes according to their scenicness score and removed inactive modes from the table. While our model is trained with the coarse 5-class first-level hierarchy ground truth of CORINE, the two modes of each class (except urban) are associated with differing fine-grained land cover concepts.

Mode	Weight	Top L3 class by recall	Most activating
Bias	4.65	-	-
1	-0.938	111 cont. urban fabric (0.966) 141 green urban areas (0.948) 121 industrial/commercial (0.688)	
2 -	-1.080	244 agro-forestry (1.0) 222 fruit trees (0.611) 211 non-irrigated (0.561)	
2 +	0.068	313 mixed forests (0.621) 243 agriculture with nature (0.594) 311 broad-leaved forests (0.562)	
3 -	-0.172	312 coniferous forests (0.586) 324 woodland-scrub transition (0.576) 313 mixed forests (0.523)	
3 +	1.391	332 bare rock (0.757) 333 sparsely vegetated (0.756) 334 burnt areas (0.667)	
4 -	-0.678	421 inland marshes (0.512) 423 intertidal flats (0.405) 522 estuaries (0.404)	
4 +	1.178	412 peat bogs (0.620) 333 sparsely vegetated (0.496) 332 bare rock (0.314)	
5 -	0.105	331 beaches, dunes, sands (0.614) 522 estuaries (0.590) 123 ports (0.55)	
5 +	1.193	523 sea/ocean (0.818) 521 coastal lagoons (0.5) 331 beaches, dunes, sands (0.181)	

transition into very scenic natural areas and wetlands at the other edge of the cluster, which suggests that there is a similar transition of scenicness from man-made to natural areas for coastal environments.

Effect of multiple modes to the final prediction. The learned bias of our model is 4.65, which corresponds to an average value of scenicness for the whole region. Deviations from this value are related to the land cover-related weights. We further assess these deviations by analyzing the most-recalled level-3 CORINE classes per mode in Table 4, as well as their weights. We find that each class has at most two active modes with a large difference in scenicness scores between both modes. Each mode tends to recall different thematic clusters, such as mode 4- (the minus sign represent here the negative influence this mode has on scenicness) recalling flat coastal wetland environments, while mode 4 + tends to recall elevated boglands, Scottish highlands and loch environments, which impact landscape beauty positively. This spatial binning effect of the positive and negative modes can be seen in Fig. 6 for all classes, except for the *Urban* and *Agriculture* classes. The *Urban* class defaults to one single un-scenic mode, while the *Agriculture* class experiences strong mixing between its two modes, as both semantic clusters tend to be widespread throughout the country. The presence of human influences on the landscape can be seen in the water and wetland classes. While the coastline of England is predicted to be very scenic (mode 5+),

its rivers and estuaries are only mildly positively associated with scenicness (mode 4-). This pattern is visible for all of the major estuaries in England. However, the inlets and open waters connected to the ocean in Scotland are considered strongly positive. While inland waters are only mildly positively associated with scenicness as in England, its presence strongly correlates with natural areas (mode 3+) and wetlands (mode 4+). The weights of these modes indicate that our model considers these land cover classes to be very scenic in the Scottish highlands. This indicates that people value inland water environments, but mostly for their nature and wetland environments. The validation of such observations, for example via interviews, could be the topic of further studies.

The learned weights of our modes can be related to three previously quantified observations. Firstly, our model supports the notion that the presence of human influences and structures in a landscape reduce the beauty of it (de Vries et al., 2012; Lindemann-Matthies et al., 2010; Palmer, 2004; Hodgson and Thayer, 1980), as visible by the scenicness weights of modes 1, 2- and 5-. However, not all classes with human influences are considered un-scenic, such estuaries and beaches in mode 5-. Secondly, landscape beauty is greater in natural areas where there is an open canopy (Schirpke et al., 2013; Hill et al., 2007), which can be inferred from the differences in scenicness values of modes 2+, 3-, 3+, and 5+. These results are corroborated by the spatial patterns of modes

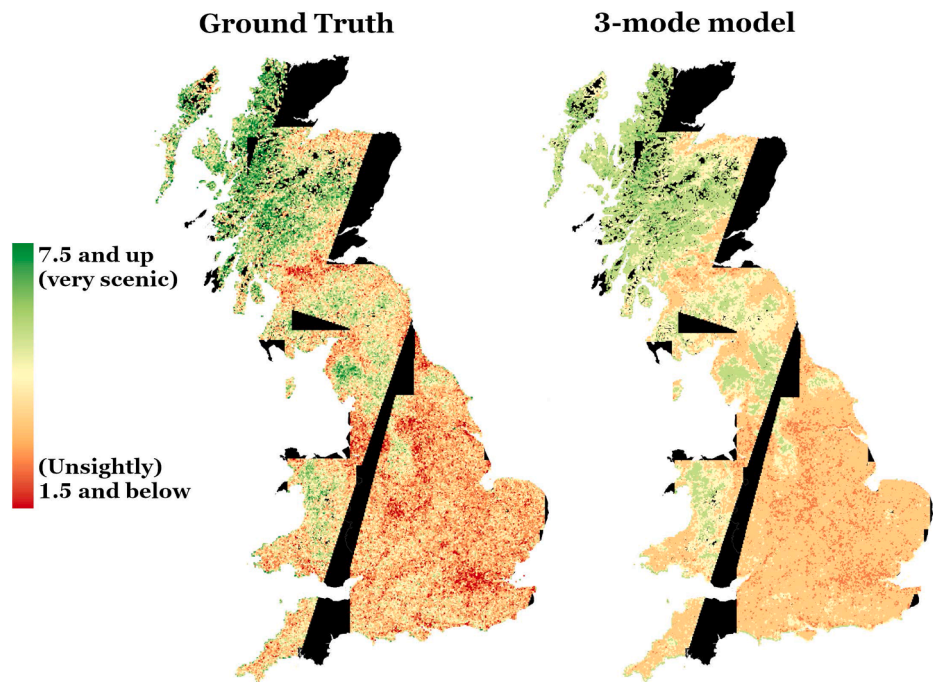


Fig. 4. Left: Geotagged scenicness scores from the ScenicOrNot project. Right: Scenicness values predicted by 3-mode ScenicNet model. Scenicness is clipped between 1.5 and 7.5 to show more variation in the 3-mode model.

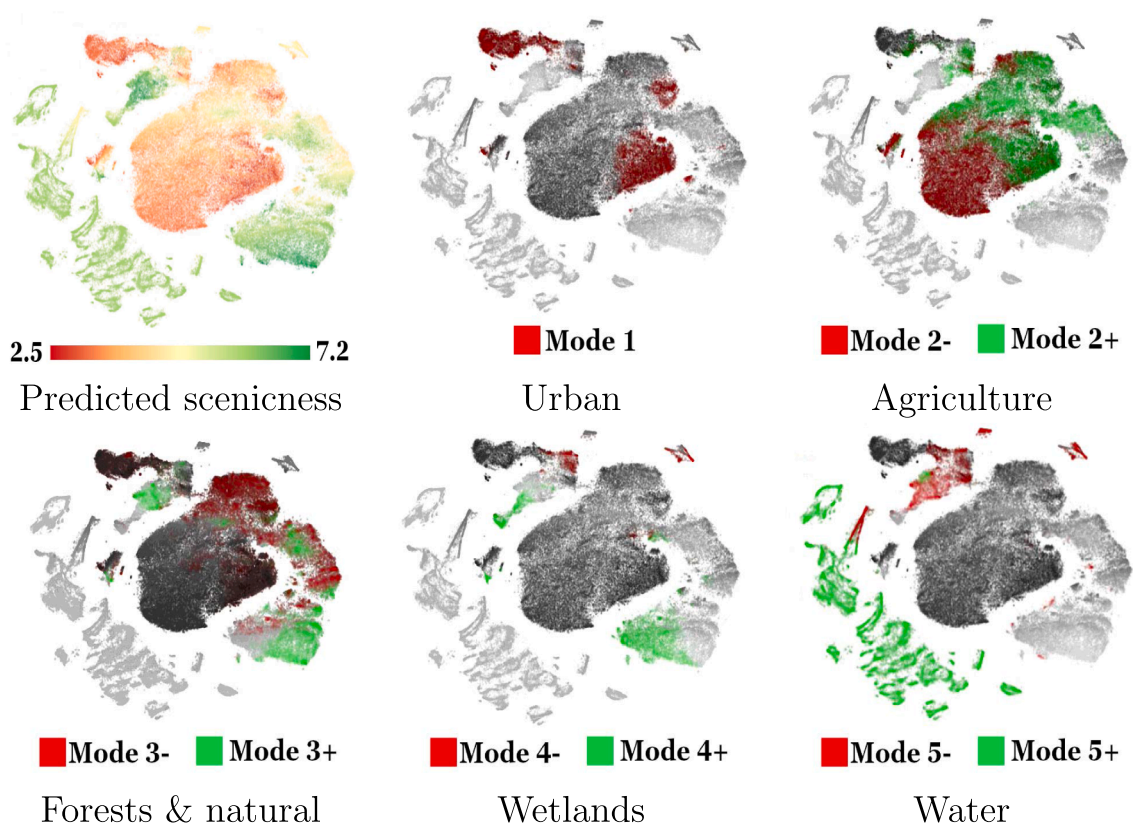


Fig. 5. Low-dimensional representation of the prediction outputs of our 3-mode ScenicNet model, visualized using t-SNE. We display the predicted scenicness scores for each datapoint as well as the predictions of the active modes of each class. The red colors of each mode refers to modes with the most negative weight within each class, while green is used for the mode with the most positive weight within each class.

3 + and 4+, which can be seen in the forest map of Fig. 6. These modes are considered very scenic, while their recalled geo-located patches often correspond with hilly and mountainous regions. Lastly, our

learned weights for the agriculture class do not directly support survey data which indicates that the British public enjoys the British countryside for its landscape beauty (Hall et al., 2004). It should be noted that

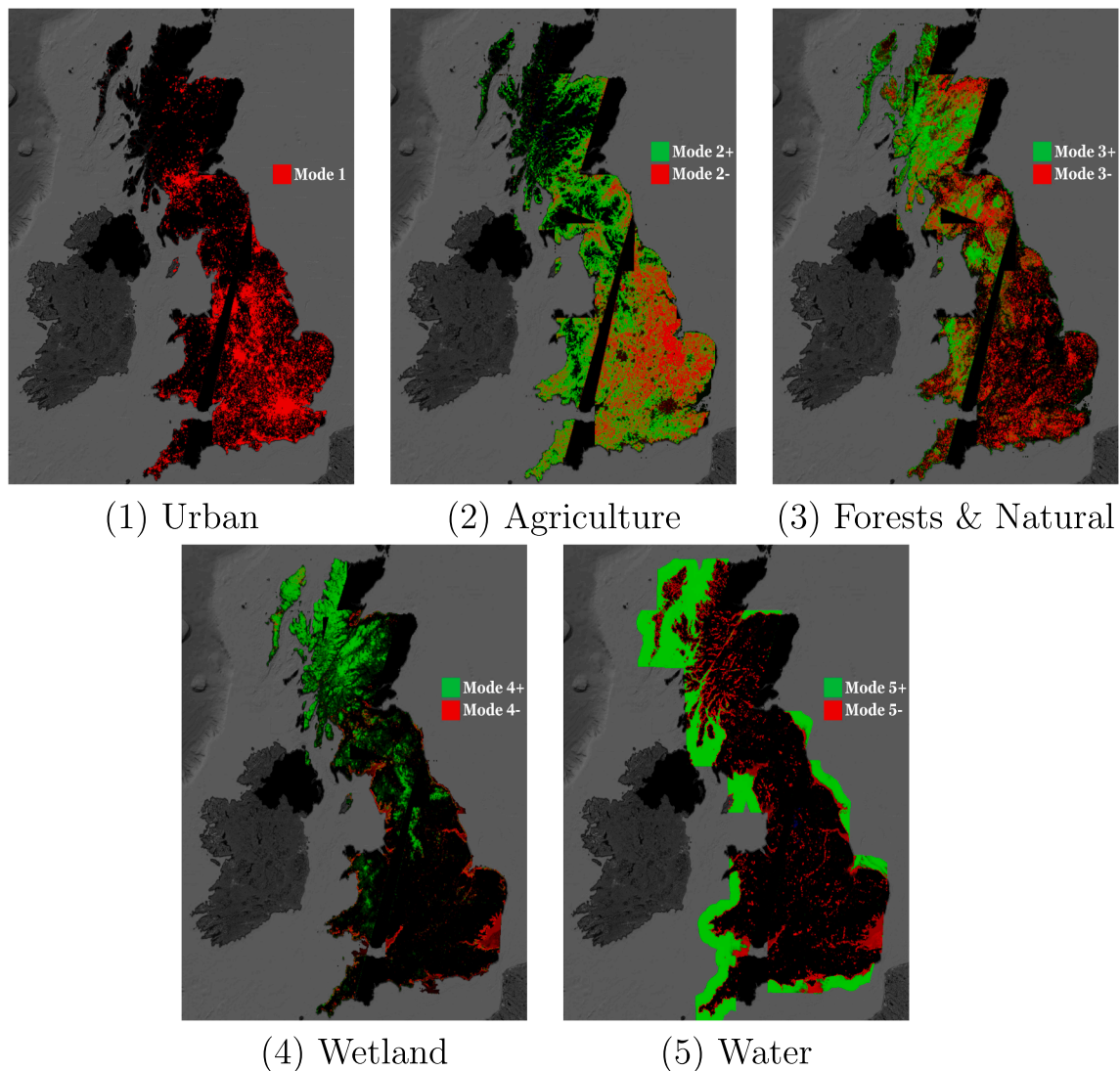


Fig. 6. Plots of predictions for each patch made by our 3-mode ScenicNet model. Areas coloured in red are predicted by the negative mode of a given class, while green areas represent the positive mode. Areas with blended colours have activations in both modes, resulting in a scenicness score that is in between both weights.

these quantified patterns are difficult to relate to our research, as they cover different countries or regions, as well as using different measurement techniques. Further research may attempt to learn patterns on a local scale to see whether local patterns in the United Kingdom extend across regions.

5. Conclusions

In this paper we present and test a novel method for large-scale inventorization of landscape scenicness, which uses land cover prediction as an interpretable intermediate task. Our model is able to learn scenic and un-scenic representations of the same land cover type by being able to choose which of several land cover-specific weights to use for the scenicness regression task. Our model outperforms an unconstrained model on the task of land cover prediction while matching an unconstrained model on scenicness regression. Furthermore, our model is able to express preferences for fine-grained land cover types while being trained on just five coarse land cover concepts, which allows us to study the relationship between landscape beauty and land cover types. Our work also opens up possibilities for knowledge and sub-class discovery. We note that our findings are still subject to the fact that all data come from the U.K. and only apply to landscape preferences in the U.K., and most probably provided by British citizens. Expanding these

findings to global measures of landscape aesthetics would require a larger corpus of crowdsourced data, as well as images coming from all over the world. Creating such dataset would open the possibility for cultural and global studies about human preferences and appreciations of nature.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Audebert, N., Le Saux, B., Lefèvre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geosci. Remote Sens. Mag.* 7 (2), 159–173.
- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25 (2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Campos-Taberner, M., Romero-Soriano, A., Gatta, C., Camps-Valls, G., Lagrange, A., Saux, B.L., Beaupère, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H., Ferecatu, M., Shimoni, M., Moser, G., Tuia, D., 2016. Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS

- Data Fusion Contest. Part A: 2D contest. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (12), 5547–5559.
- Camps-Valls, G., Tuia, D., Zhu, X.X., Reichstein, M., 2021. Deep learning for Earth Sciences - A comprehensive approach to remote sensing, climate science and geosciences. Wiley and Sons.
- Caruana, R., 1997. Multitask learning. *Mach. Learn.* 28, 41–75.
- Daniel, T.C., 1976. Measuring landscape esthetics: The scenic beauty estimation method. Rocky Mountain Forest and Range Experiment Station.
- Daniel, T.C., Muhar, A., Arnberger, A., Aznar, O., Boyd, J.W., Chan, K.M.A., Costanza, R., Elmqvist, T., Flint, C.G., Gobster, P.H., Gret-Regamey, A., Lave, R., Muhar, S., Penker, M., Ribe, R.G., Schauppenlehner, T., Sikor, T., Soloviy, I., Spierenburg, M., Taczanowska, K., Tam, J., Dunk, A. v. d., 2012. Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences*. 109(23), 8812–8819. 109 (23), 8812–8819, number: 23. <https://www.fs.usda.gov/treesearch/pubs/54464>.
- de Vries, S., de Groot, M., Boers, J., 2012. Eyesores in sight: Quantifying the impact of man-made elements on the scenic beauty of Dutch landscapes. *Landscape Urban Plan.* 105 (1), 118–127 <http://www.sciencedirect.com/science/article/pii/S0169204611003562>.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- EU Copernicus Program, 2018. CLC 2018 - Copernicus Land Monitoring Service. URL <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>.
- Fu, K., Dai, W., Zhang, Y., Wang, Z., Yan, M., Sun, X., Jan. 2019. MultiCAM: Multiple Class Activation Mapping for Aircraft Recognition in Remote Sensing Images. *Remote Sensing* 11 (5), 544, number: 5 Publisher: Multidisciplinary Digital Publishing Institute. <https://www.mdpi.com/2072-4292/11/5/544>.
- Grinde, B., Patil, G.G., Sep. 2009. Biophilia: Does Visual Contact with Nature Impact on Health and Well-Being? *International Journal of Environmental Research and Public Health* 6 (9), 2332–2343, number: 9 Publisher: Molecular Diversity Preservation International. URL <https://www.mdpi.com/1660-4601/6/9/2332>.
- Hall, C., McVittie, A., Moran, D., 2004. What does the public want from agriculture and the countryside? A review of evidence and methods. *J. Rural Stud.* 20 (2), 211–225 <http://www.sciencedirect.com/science/article/pii/S0743016703000536>.
- Havinga, I., Bogaart, P., Hein, L., Tuia, D., 2020. Defining and modelling cultural ecosystem services using user-generated geographic information. *Ecos. Serv.* 43, 101091.
- He, K., Zhang, X., Ren, S., Sun, J., Jun. 2016. Deep Residual Learning for Image Recognition. In: *CVPR*. pp. 770–778.
- Hill, D., Daniel, T.C., Dec. 2007. Foundations for an Ecological Aesthetic: Can Information Alter Landscape Preferences? *Society & Natural Resources* 21 (1), 34–49, publisher: Routledge. eprint: doi: 10.1080/08941920701655700. <https://doi.org/10.1080/08941920701655700>.
- Hodgson, R.W., Thayer, R.L., 1980. Implied human influence reduces landscape beauty. *Landscape Plan.* 7 (2), 171–179 <http://www.sciencedirect.com/science/article/pii/0304392480900143>.
- Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., Baesens, B., 2011. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decis. Support Syst.* 51 (1), 141–154 <https://www.sciencedirect.com/science/article/pii/S0167923610002368>.
- Imamoglu, N., Kimura, M., Miyamoto, H., Fujita, A., Nakamura, R., 2017. Solar Power Plant Detection on Multi-Spectral Satellite Imagery using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion. In: *BMVC*.
- Kendall, M.G., 1938. A New Measure for Rank Correlation. *Biometrika* 30 (1–2), 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>.
- Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization. *ICLR*.
- Krippendorf, J., 1984. Die Ferienmenschen. Für ein neues Verständnis von Freizeit und Reisen, Orell Füssli.
- Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., Müller, K.-R., Mar. 2019. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications* 10 (1), 1096, number: 1 Publisher: Nature Publishing Group. <https://www.nature.com/articles/s41467-019-08987-4>.
- Levering, A., Marcos, D., Lobry, S., Tuia, D., 2020. Interpretable Scenicness from Sentinel-2 Imagery. In: *Proceedings of the 2020 International Geoscience and Remote Sensing Symposium, Hawaii*, p. 4.
- Lindemann-Matthies, P., Briegel, R., Schüpbach, B., Junge, X., Nov. 2010. Aesthetic preference for a Swiss alpine landscape: The impact of different agricultural land-use with different biodiversity. *Landscape Urban Plan.* 98 (2), 99–109 <http://www.sciencedirect.com/science/article/pii/S0169204610001830>.
- Maaten, L.v.d., Hinton, G., 2008. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9 (86), 2579–2605 <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- Maggiore, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* 55 (12), 7092–7103.
- Marcos, D., Fong, R., Lobry, S., Flamary, R., Courty, N., Tuia, D., Sep. 2020. Contextual Semantic Interpretability. *arXiv:2009.08720 [cs]* [ArXiv: 2009.08720. http://arxiv.org/abs/2009.08720](https://arxiv.org/abs/2009.08720).
- Marcos, D., Lobry, S., Tuia, D., 2019. Semantically Interpretable Activation Maps: what-where-how explanations within CNNs. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 4207–4215 ISSN: 2473–9944.
- Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS. J. Int. Soc. Photo. Remote Sens.* 135, 158–172.
- Miller, T., 2017. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* 267, 1–38 [arXiv: 1706.07269](https://arxiv.org/abs/1706.07269).
- Palmer, J.F., 2004. Using spatial metrics to predict scenic perception in a changing landscape: Dennis, Massachusetts. *Landscape and Urban Planning* 69 (2), 201–218 <http://www.sciencedirect.com/science/article/pii/S0169204603001968>.
- Pearson, K., Nov. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11), 559–572, publisher: Taylor & Francis. eprint: doi: 10.1080/14786440109462720. <https://doi.org/10.1080/14786440109462720>.
- Samek, W., Müller, K.-R., 2019. Towards Explainable Artificial Intelligence. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 5–22. doi: 10.1007/978-3-030-28954-6_1.
- Schirpke, U., Tasser, E., Tappeiner, U., 2013. Predicting scenic beauty of mountain regions. *Landscape and Urban Planning* 111, 1–12 <http://www.sciencedirect.com/science/article/pii/S0169204612003271>.
- Schroeder, H., Daniel, T.C., Mar. 1981. Progress in Predicting the Perceived Scenic Beauty of Forest Landscapes. *Forest Science* 27 (1), 71–80, publisher: Oxford Academic. <https://academic.oup.com/forestscience/article/27/1/71/4656458>.
- Seresinhe, C.I., Preis, T., MacKerron, G., Moat, H.S., 2019. Happiness is Greater in More Scenic Locations. *Scientific Reports* 9 (1), 1–11 <https://www.nature.com/articles/s41598-019-40854-6>.
- Seresinhe, C.I., Preis, T., Moat, H.S., 2015. Quantifying the Impact of Scenic Environments on Health. *Scientific Reports* 5 (1), 1–9 <https://www.nature.com/articles/srep16899>.
- Seresinhe, C.I., Preis, T., Moat, H.S., 2017. Using deep learning to quantify the beauty of outdoor places. *Royal Society Open Science* 4 (7), 170170, publisher: Royal Society. <https://royalsocietypublishing.org/doi/full/10.1098/rsos.170170>.
- Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2019. Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. In: *IGARSS*, pp. 5901–5904.
- Svendsen, D.H., Morales-Álvarez, P., Ruescas, A.B., Molina, R., Camps-Valls, G., 2020. Deep Gaussian processes for biogeophysical parameter retrieval and model inversion. *ISPRS Journal of Photogrammetry and Remote Sensing* 166, 68–81 <http://www.sciencedirect.com/science/article/pii/S0924721620301118>.
- Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M., Oct. 2010. Learning relevant image features with multiple-kernel classification. *IEEE Transactions on Geoscience and Remote Sensing* 48 (10), 3780–3791, publisher: Institute of Electrical and Electronics Engineers. <https://research.wur.nl/en/publications/learning-relevant-image-features-with-multiple-kernel-classificat>.
- van Rijsbergen, C., 1979. Information Retrieval. *J. Am. Soc. Inform. Sci.* 30 (6), 374–375. eprint: <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630300621>. URL <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630300621>.
- Vasu, B., Rahman, F.U., Savakis, A., Jun. 2018. Aerial-CAM: Salient Structures and Textures in Network Class Activation Maps of Aerial Imagery. In: *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. pp. 1–5.
- Velarde, M.D., Fry, G., Tveit, M., 2007. Health effects of viewing landscapes - Landscape types in environmental psychology. *Urban Forestry & Urban Greening* 6 (4), 199–212 https://www.academia.edu/8672414/Health_effects_of_viewing_landscapes_Landscape_types_in_environmental_psychology.
- Volpi, M., Tuia, D., 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* 55 (2), 881–893.
- Volpi, M., Tuia, D., 2018. Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. *ISPRS. J. Int. Soc. Photo. Remote Sens.* 144, 48–60.
- Workman, S., Souvenir, R., Jacobs, N., Oct. 2017. Understanding and Mapping Natural Beauty. In: *ICCV. IEEE, Venice*, pp. 5590–5599. <http://ieeexplore.ieee.org/document/8237858/>.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., Jun. 2016. Learning Deep Features for Discriminative Localization. In: *CVPR. IEEE, Las Vegas, NV, USA*, pp. 2921–2929. <http://ieeexplore.ieee.org/document/7780688/>.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., Dec. 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine* 5 (4), 8–36, conference Name: IEEE Geoscience and Remote Sensing Magazine.