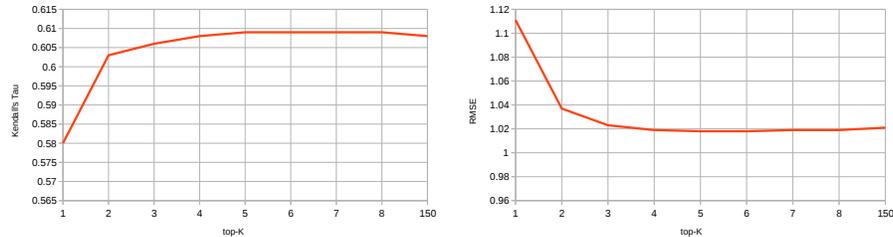


## Contextual Semantic Interpretability

### A Sensitivity to top- $K$

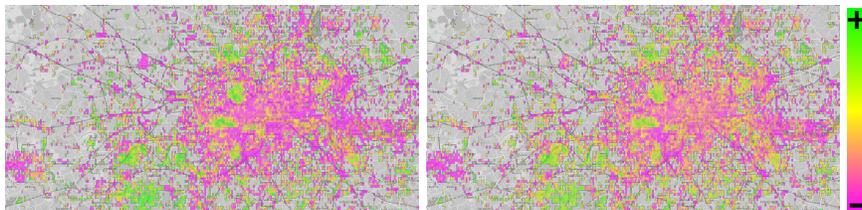
Fig. 1 shows the relationship between CSIB model performance and  $K$  sparsity. We can see how the performance saturates for values higher than  $K=5$ .



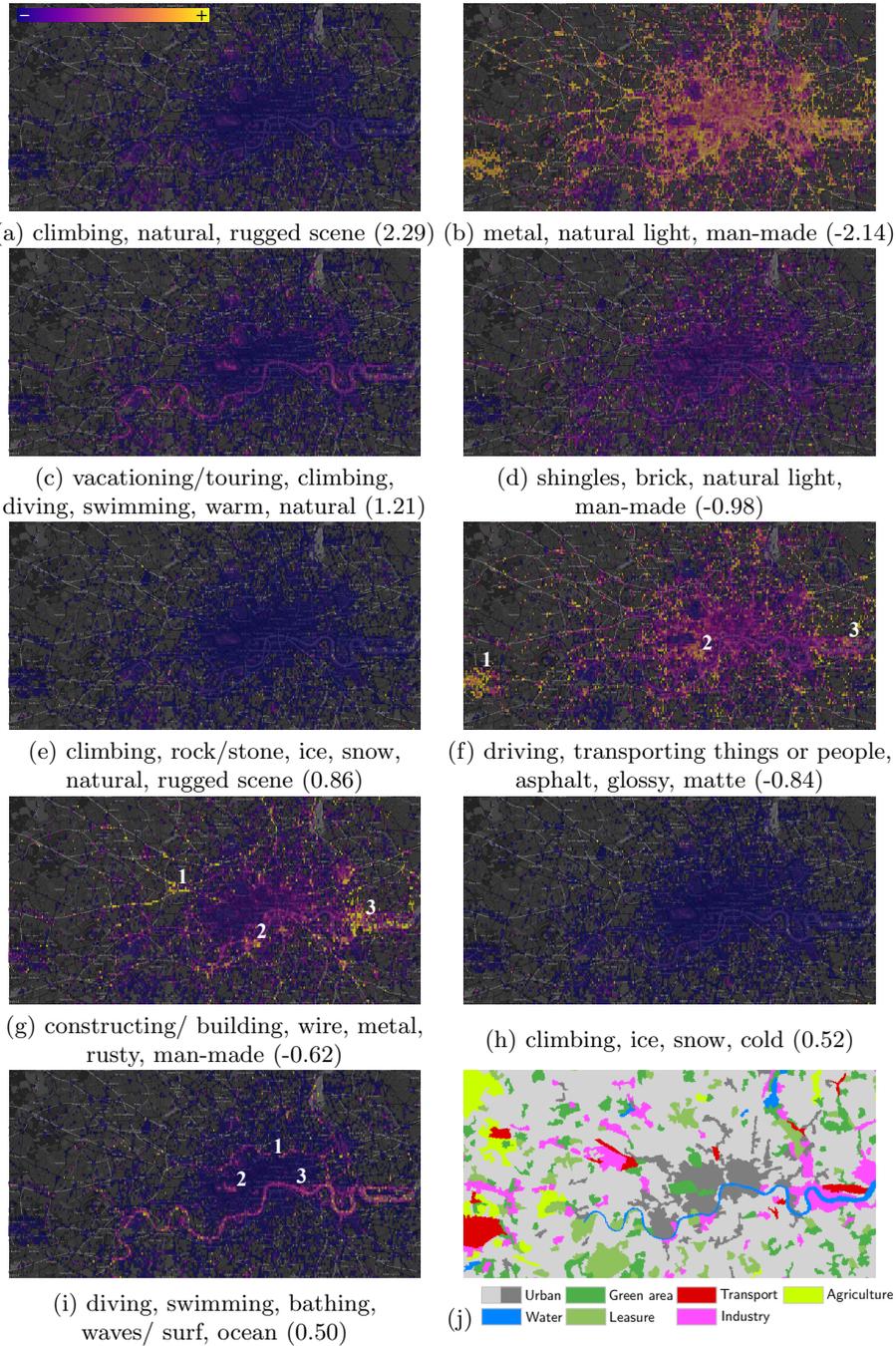
**Fig. 1.** Performance for different values of  $K$  in the top- $K$  operation, in terms of Kendall's  $\tau$  (left) and RMSE (right). A top- $K$  of 150 represents the dense model.

### B Map interpretability over London

Fig. 2 shows the scenicness maps computed with the Flickr images over London using the baseline and CSIB. Both maps show an overall agreement, with CSIB being visibly more conservative. Fig. 3 depicts the maps of CSIB group presence for the same area, showing how it allows, not just to map scenicness, but also to distinguish between different kinds thereof. For instance, unscenic areas dominated by group (g) are mostly related to industry and construction sites (g1 is to the area around Park Royal business park, g2 Nine Elms construction sites and g3 the Greenwich Peninsula), while those dominated by (f) tend to be associated to transport infrastructure (f1 and f3 correspond to the airports and f2 to Victoria station). At the same time, we are able to map elements, such as water surfaces (i1 is the Regent's canal, i2 the lake in Hyde park and i3 the Thames) even though their influence on scenicness is limited.



**Fig. 2.** Map of scenicness over London from the Flickr images obtained using the baseline ResNet-50 (left) and the proposed interpretable CSIB (right).



**Fig. 3.** CSBI group activation maps on the Flickr images in the London area (a - i), with color bar in (a), and CORINE land-cover map (j).