Contrastive Pretraining for Visual Concept Explanations of Socioeconomic Outcomes

Supplementary Material

7. Training Procedure

For model training, we performed image transformations in the following order:

1. Resizing the images to a size of 500 x 500
2. Random Augmentation [10]
3. Random Erasure [33]
4. Min-max image normalization

For the task-specific pretraining with the Rank-N-Contrast approach, we used the default hyperparameters suggested by the authors [32]. In detail, we set the temperature τ to 2, use negative L_2 norm as feature similarity metric, and L_1 distance as label distance. Further, this encoder was trained with the Adam optimizer and cosine rate annealing. Finally, we used the Adam optimizer with an exponential learning rate scheduler for the supervised training of the standard Resnet-50 encoder (as also implemented in [21]) and the linear layer that probes the output of the task-specific pre-trained encoder.

8. Concept Definitions

Table 2 provides a detailed overview of the used land cover percentages for each of the concepts and visualizes examples of concept images.

9. Conceptual Sensitivities

Figures 6, 7, 8, 9, 10 and 11 visualize the conceptual sensitivities of the Rank-N-Contrast pre-trained encoder in both datasets for the concepts of water, agriculture, impervious, sparse residential, medium residential and dense residential, respectively.

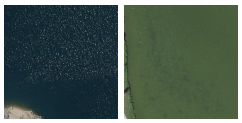
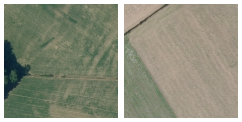
Concept	FLAIR Class	Percent	Examples
Water	Water	90-100	
Vegetation	Vegetation	90-100	
Agriculture	Agriculture	90-100	
Impervious Surface	Impervious Surface	90-100	
Dense Residential	Buildings	90-100	
Medium Residential	Buildings Vegetation, Agriculture	40-60 40-60	
Sparse Residential	Buildings Vegetation, Agriculture	10-30 70-90	

Table 2. Concept dataset composition from FLAIR Classes

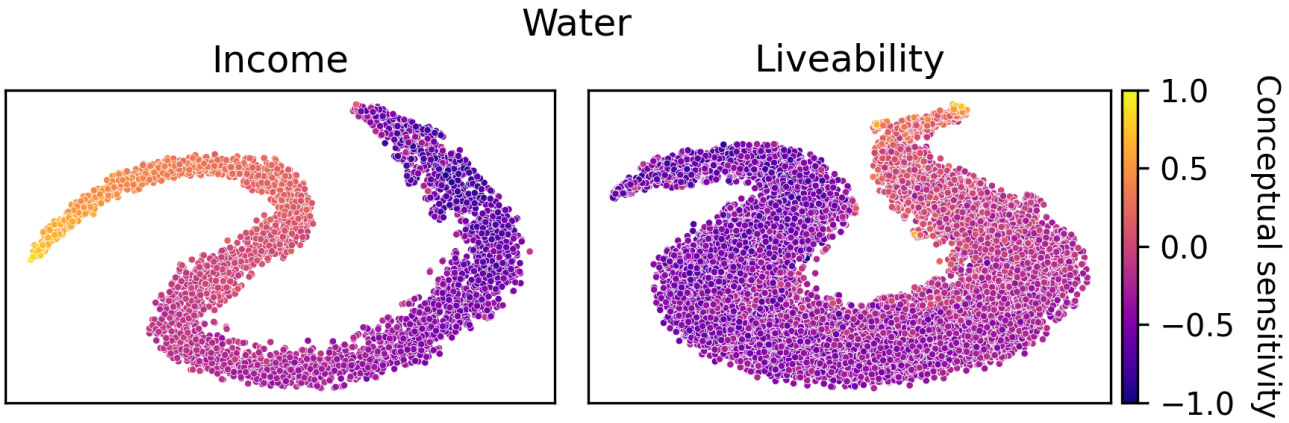


Figure 6. **The TCAV sensitivity of the water concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.

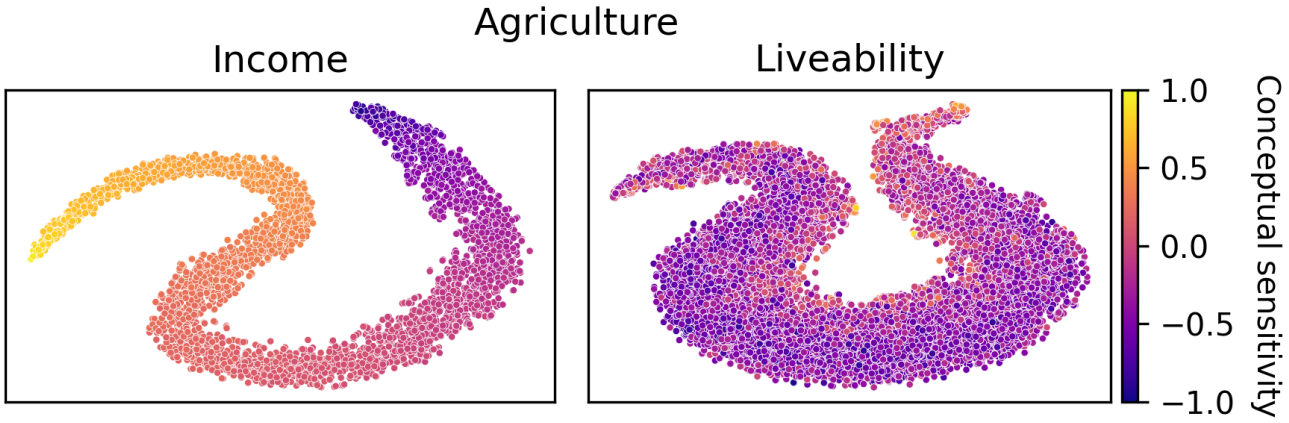


Figure 7. **The TCAV sensitivity of the agriculture concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.

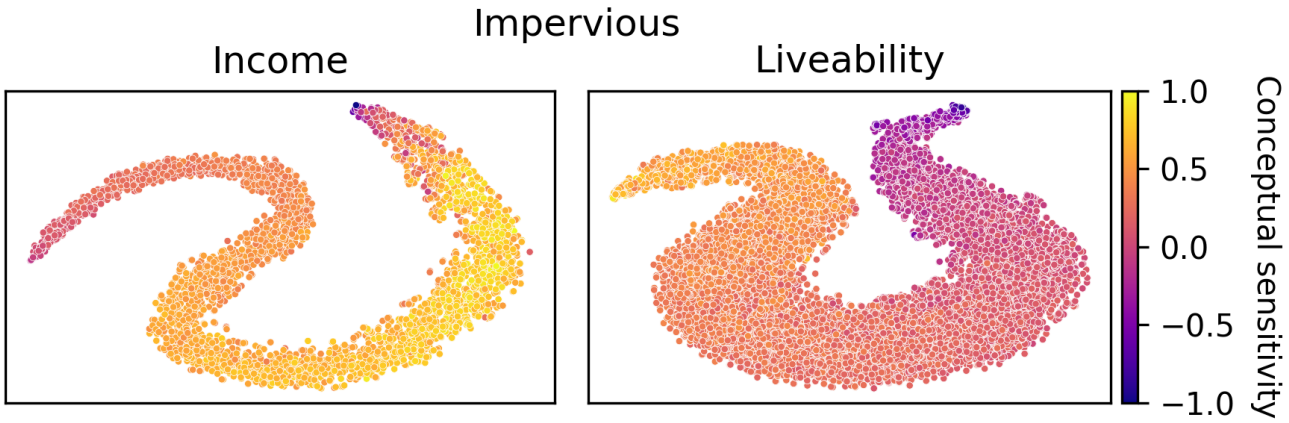


Figure 8. **The TCAV sensitivity of the impervious surface concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.

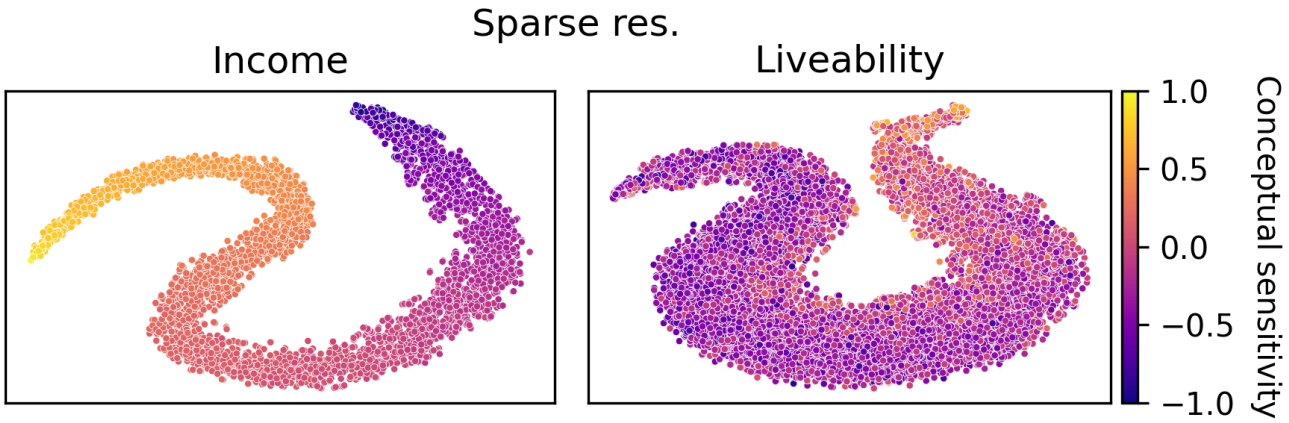


Figure 9. **The TCAV sensitivity of the sparse residential concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.

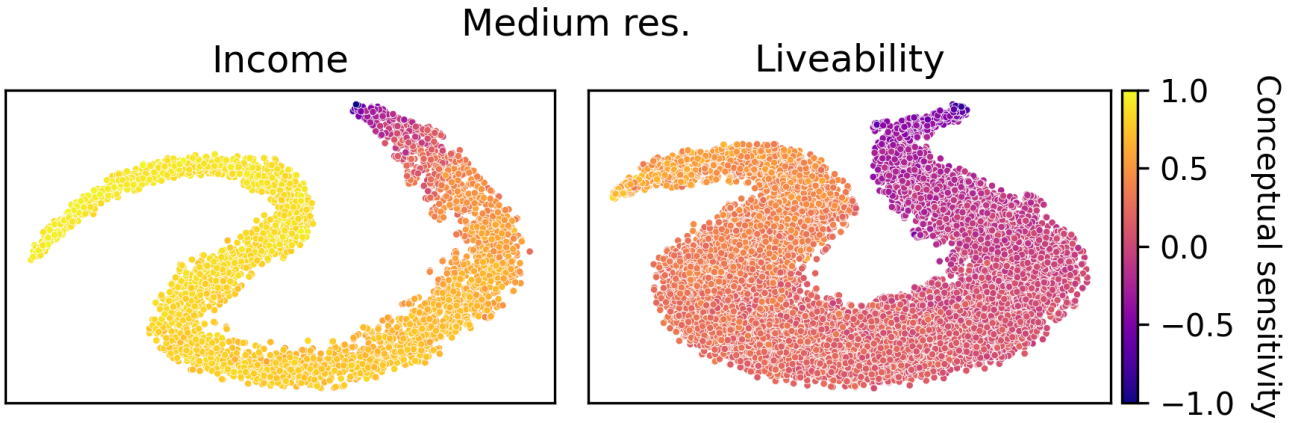


Figure 10. **The TCAV sensitivity of the medium residential concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.

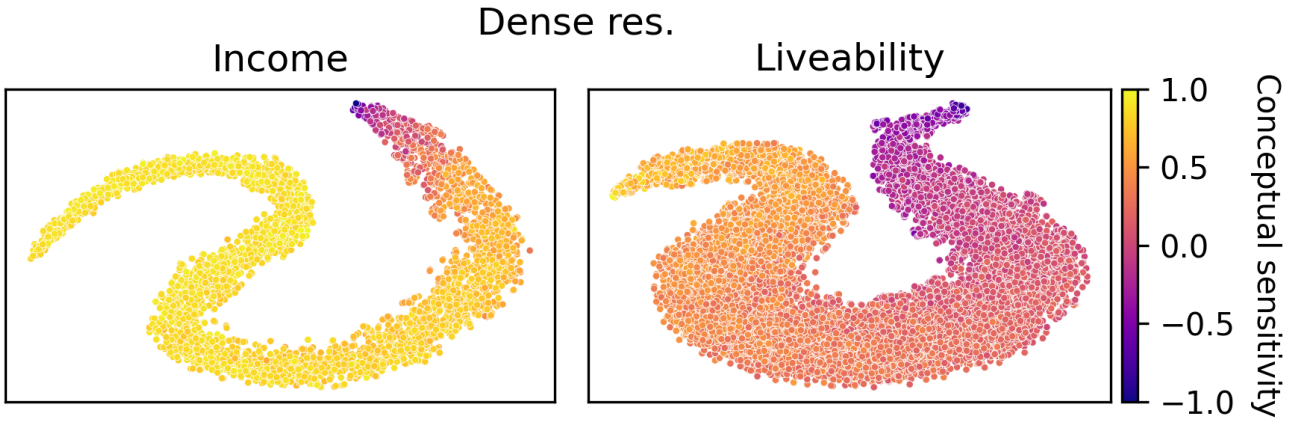


Figure 11. **The TCAV sensitivity of the dense residential concept for the income (left) and liveability (right) datasets.** The magnitude values are normalized in the range [-1, 1] with min-max normalization.