Protoype-based Dataset Comparison: Supplementary Material

Nanne van Noord
University of Amsterdam
n.j.e.vannoord@uva.nl

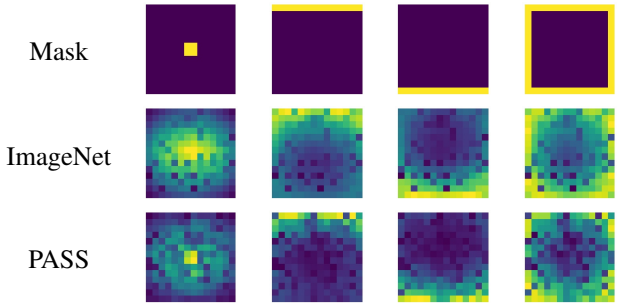

Figure 1. Visualisation of between token correlations for selected tokens marked in yellow in the first row. The second row shows ImageNet and the strong (circular) centre bias between tokens, whereas the third row shows a minimal bias for the PASS dataset. A bright (yellow) colour indicates a (correlation) value of 1.

A. Single dataset summarisation

In addition to comparison, we also used ProtoSim to summarise ImageNet and PASS independently. Nonetheless, summarising datasets independently does enable some comparison. For instance, ImageNet has a well-documented centre bias (*i.e.*, the salient objects primarily occur in the centre of the image). With ProtoSim we can demonstrate that this is indeed the case, and additionally compare to other datasets. Additionally, we explore if there are structural differences between the prototypes learned for class and patch tokens, and to what extent the prototypes represent semantics.

Centre Bias. Each of the 14×14 squares in Figure 1 represents a patch token, with its colour indicating the strength of the correlation with the pixels in the mask (first row). The first column shows the correlations between the 4 central tokens with all other tokens, for ImageNet we observe a clear circular pattern to the correlations, with strong correlations between tokens in the centre. For PASS (third row) this pattern is much less pronounced as all correlations, except those selected, are much lower. Similar observations can be made for the other columns, when selecting tokens at the top, bottom, or all edges respectively.

Class vs. Patch tokens. ProtoSim learns prototypes for patch tokens and the class token z^0, as such prototypes can represent global image information. As there is no architectural difference between how class tokens and patch tokens are treated we can find prototypes for both tokens. However, we find that there is a sharp distinction between *class prototypes* and *spatial prototypes*, with most being spatial tokens as illustrated in Figure 2. Notably, there are only a few prototypes that have proportion between 0 and 0.9.

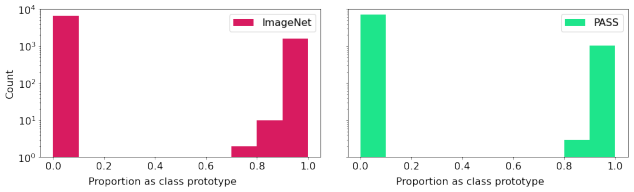

Figure 2. Histograms of the prototype's proportion as class prototype. A proportion of 1.0 indicates a prototype is always a class prototype and 0 means it is only ever a spatial prototype.

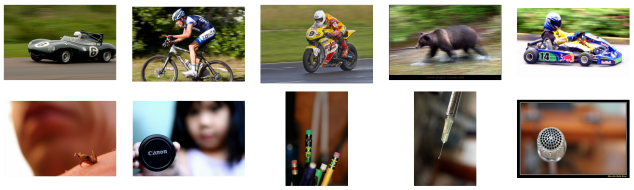

Figure 3. ImageNet prototypes that appear to represent motion blur (first row) and a shallow depth of field (last row).

Semantic Prototypes. Based on the ImageNet labels we can relate prototypes to categories and explore to what extent there is alignment with the semantic categories. For 874 out of 1000 categories the most frequent prototype also has that category as its most frequent category. The remaining 126 categories that fall in the top 4 for a specific prototype mainly concern one-of a few fine-grained or highly related categories, such as: husky, Alaskan malamute, and Siberian husky; ambulance, police van, and minibus; beer bottle and soda bottle; photocopier and printer; or bikini and tank suit. We can thus conclude that on ImageNet the model learns highly semantic prototypes. In Figure 4 we experiment with zeroing out the most frequent prototypes for these classes and report their before and after accuracy, we can observe that for most classes this results in a stark

drop in performance. On average there is a difference of 22.6 percentage points between before and after zeroing, thereby highlighting how strongly class-specific these prototypes are. In addition to semantic prototypes it also learns non-semantic visual effects such as motion blur or a shallow depth of field, as shown in Figure 3.

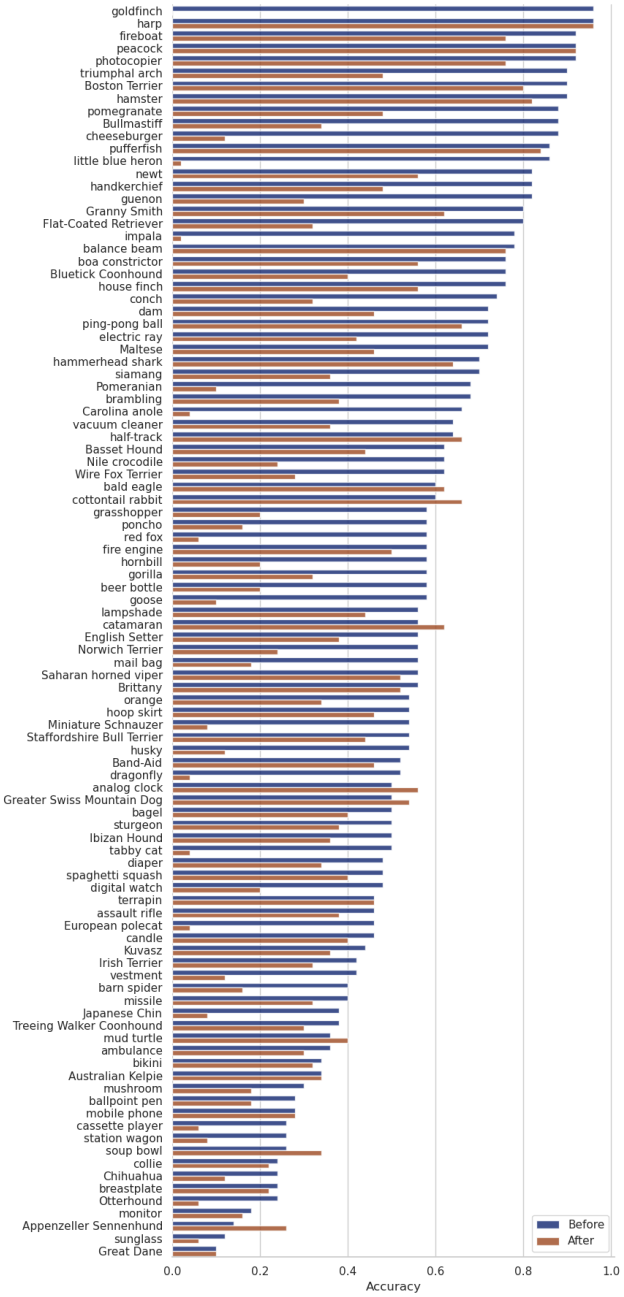

Figure 4. Before and after per-class accuracy of zeroing out the most frequent prototype for each of top 100 ImageNet classes ranked by the strength of their association to a prototype. For the majority of classes we see a stark drop in performance.

B. Additional PassNet Prototypes

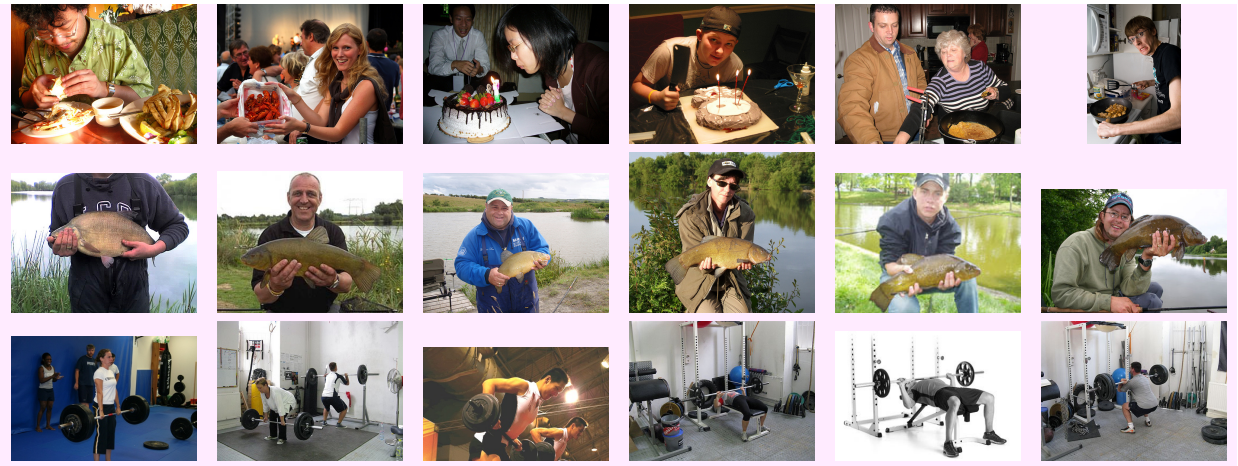

Figure 5. Three prototypes only found in ImageNet: *Persons eating food*, *Person holding fish*, and *Weightlifting* respectively.

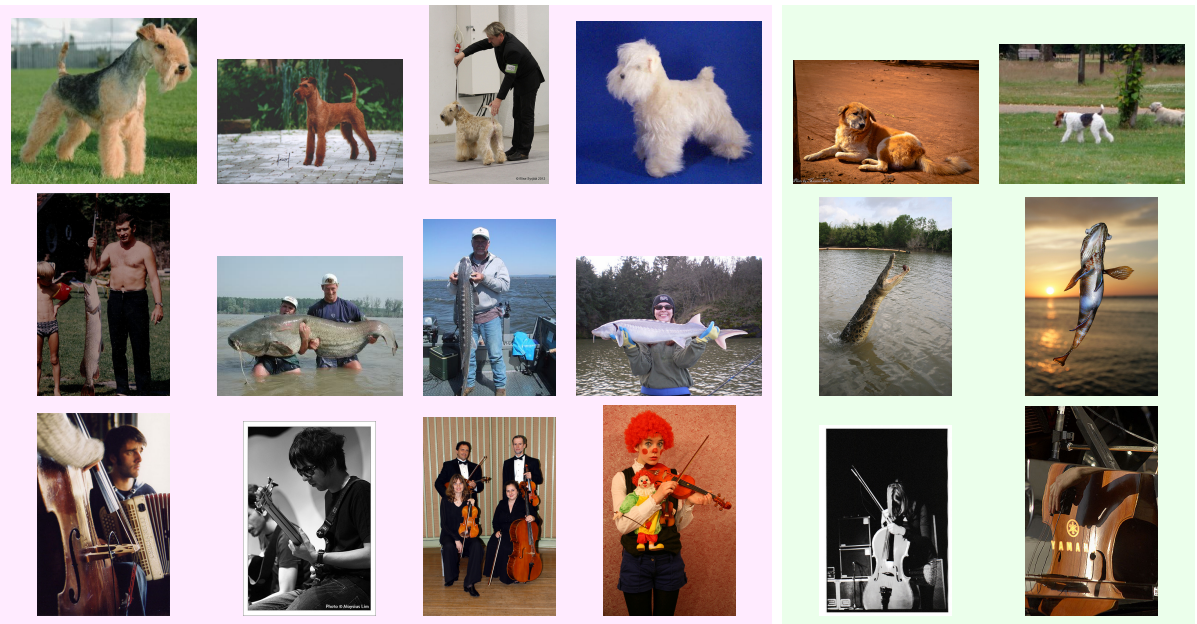

Figure 6. Three predominantly ImageNet prototypes: *Dogs*, *Persons holding large fish*, and *Persons with instrument* respectively. Last two columns are PASS images.

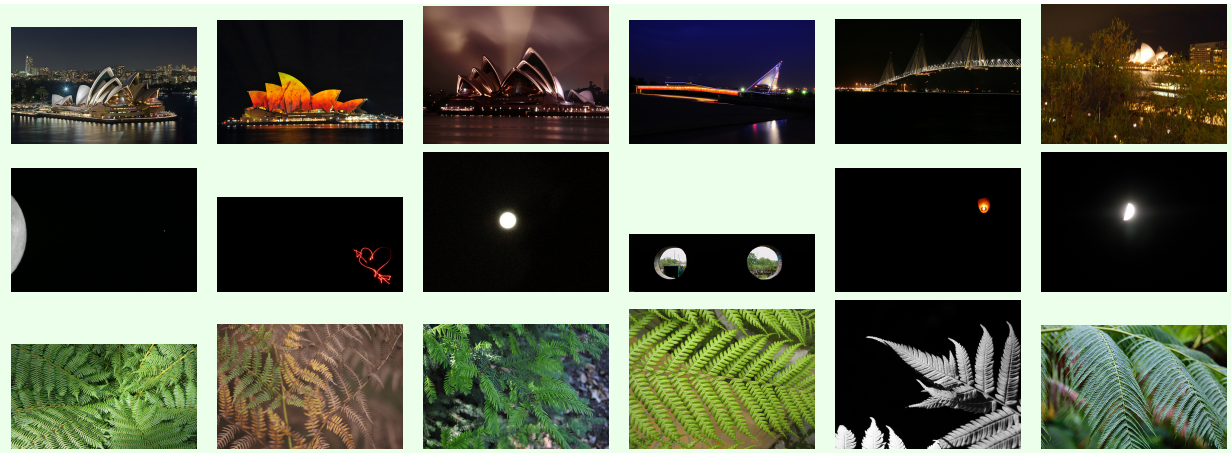

Figure 7. Three prototypes only found in PASS: *Illuminated structure*, *Darkness with bright area/celestial body*, and *Palm fronds* respectively.

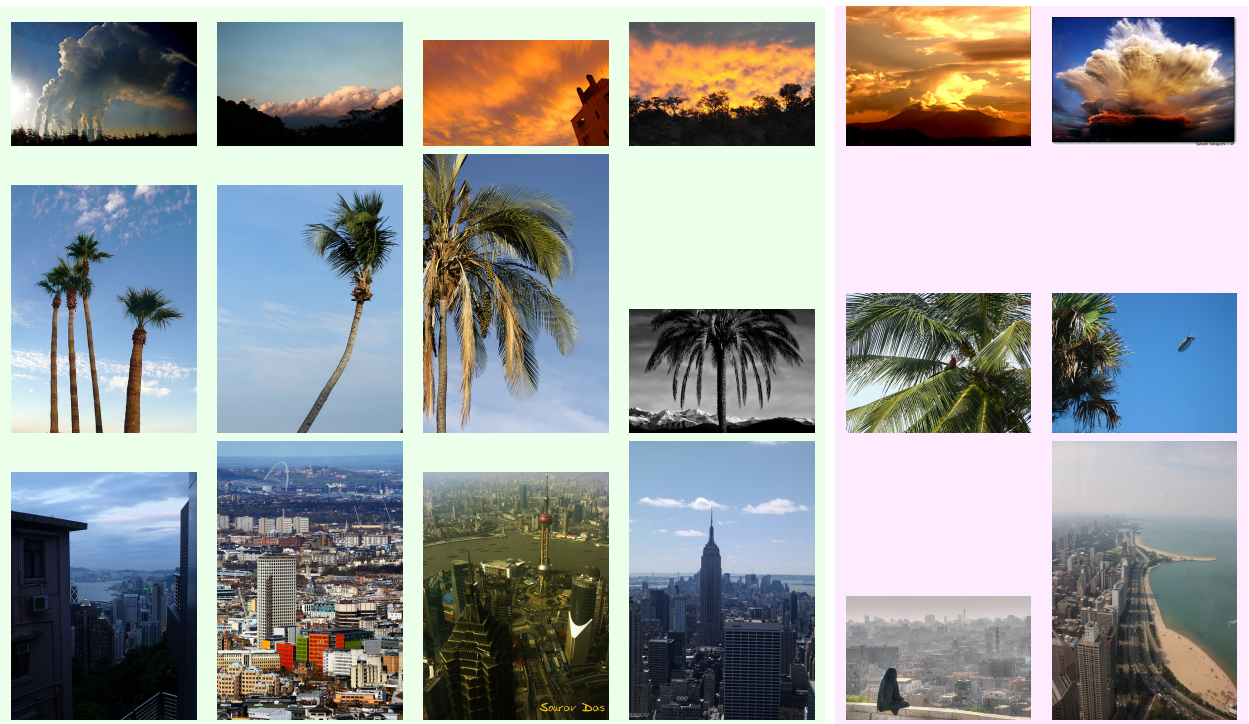

Figure 8. Three prototypes predominantly found in PASS: *Sunlit clouds*, *Palm trees*, and *Cityscape* respectively. The last two columns are images found in ImageNet.

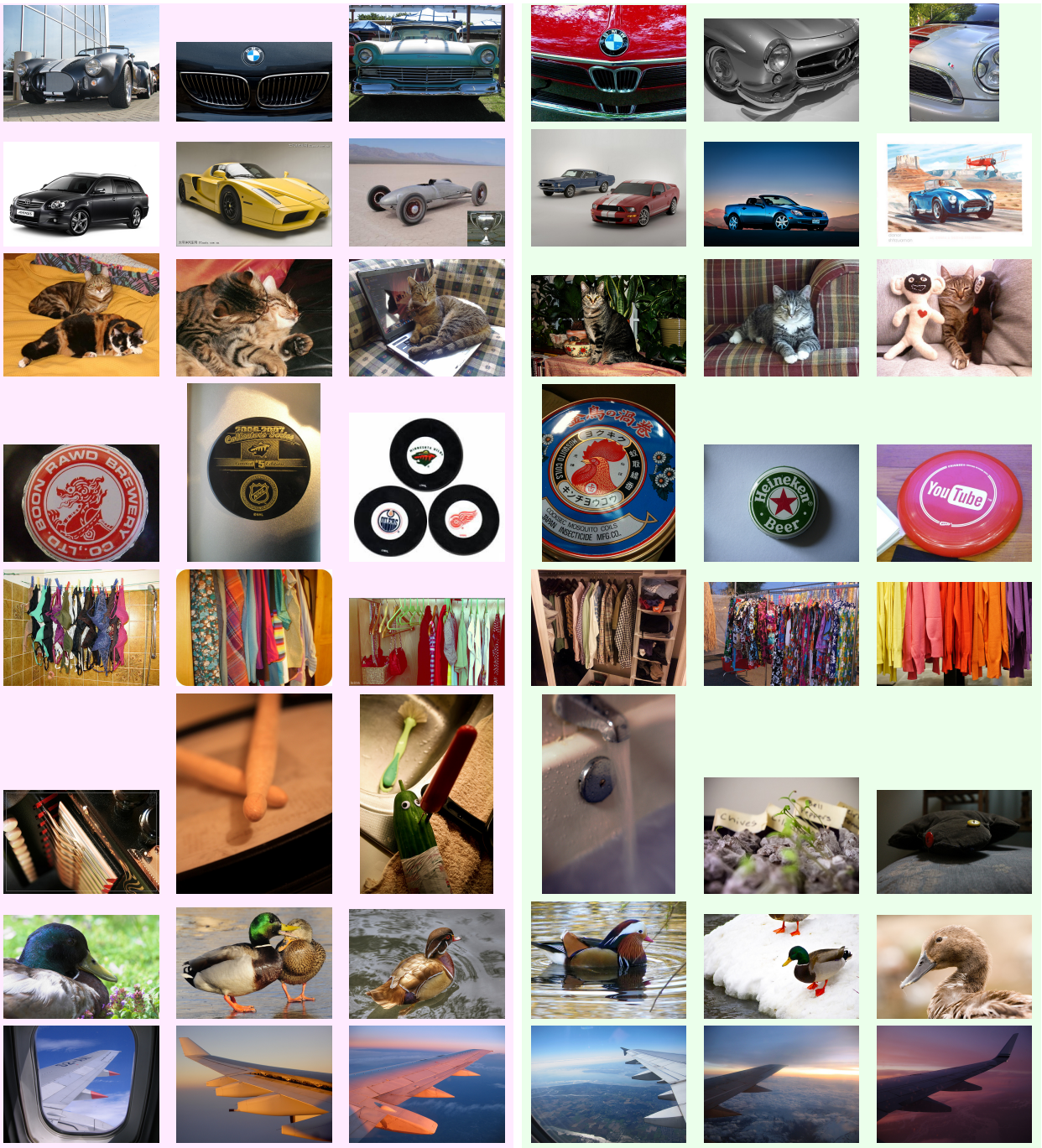

Figure 9. Eight shared prototypes commonly found in both ImageNet and PASS: *Car details*, *Cars*, *Cats*, *Caps/lids*, *Clothing rack*, *Close-ups*, *Ducks*, and *Airplane wings* respectively. The first three columns show ImageNet images and the last three columns are PASS images.

C. Additional Art Dataset Prototypes

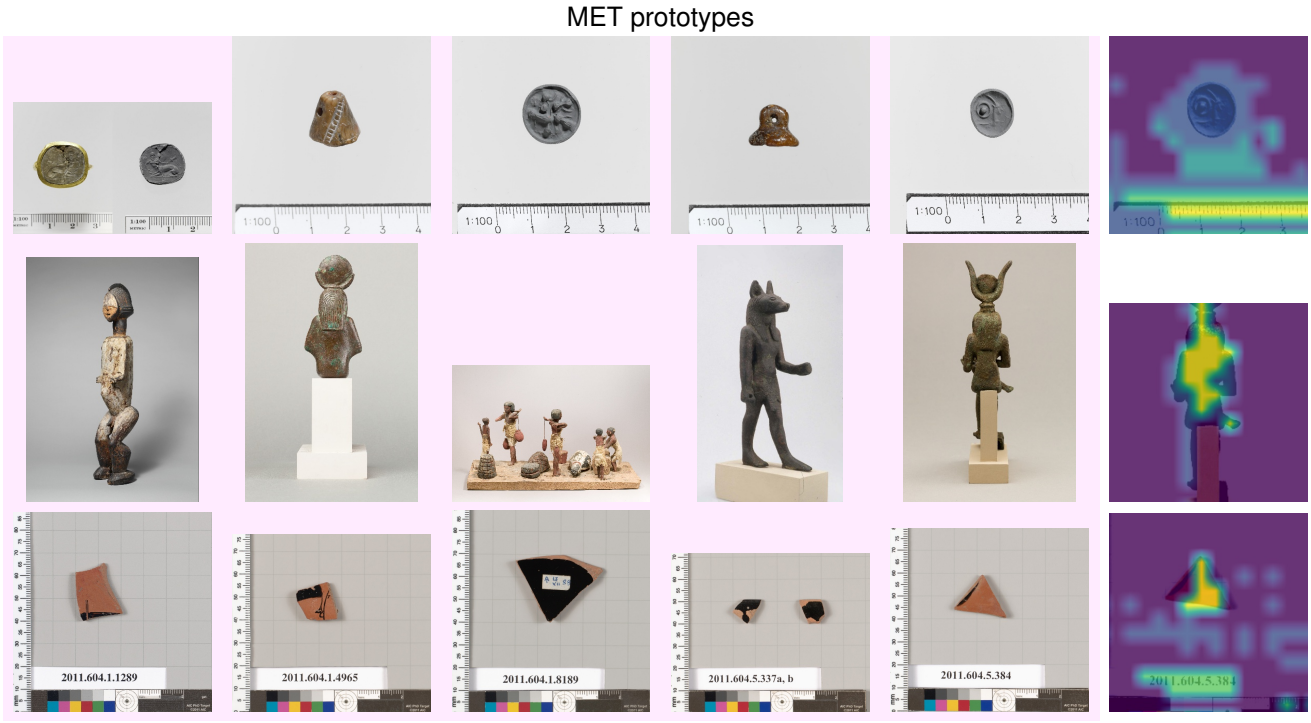

Figure 10. Three predominantly MET prototypes: *Tape measure*, *Statues*, and *Fragments with colour chart*. Per row five example images are shown, the sixth column contains the square-cropped attention mask of the prototype for the image in the fifth column.

Figure 11. Three predominantly Rijksmuseum prototypes: *Crowds in prints*, *Model boats*, and *Scene with tower(s)*. Per row five example images are shown, the sixth column contains the square-cropped attention mask of the prototype for the image in the fifth column.

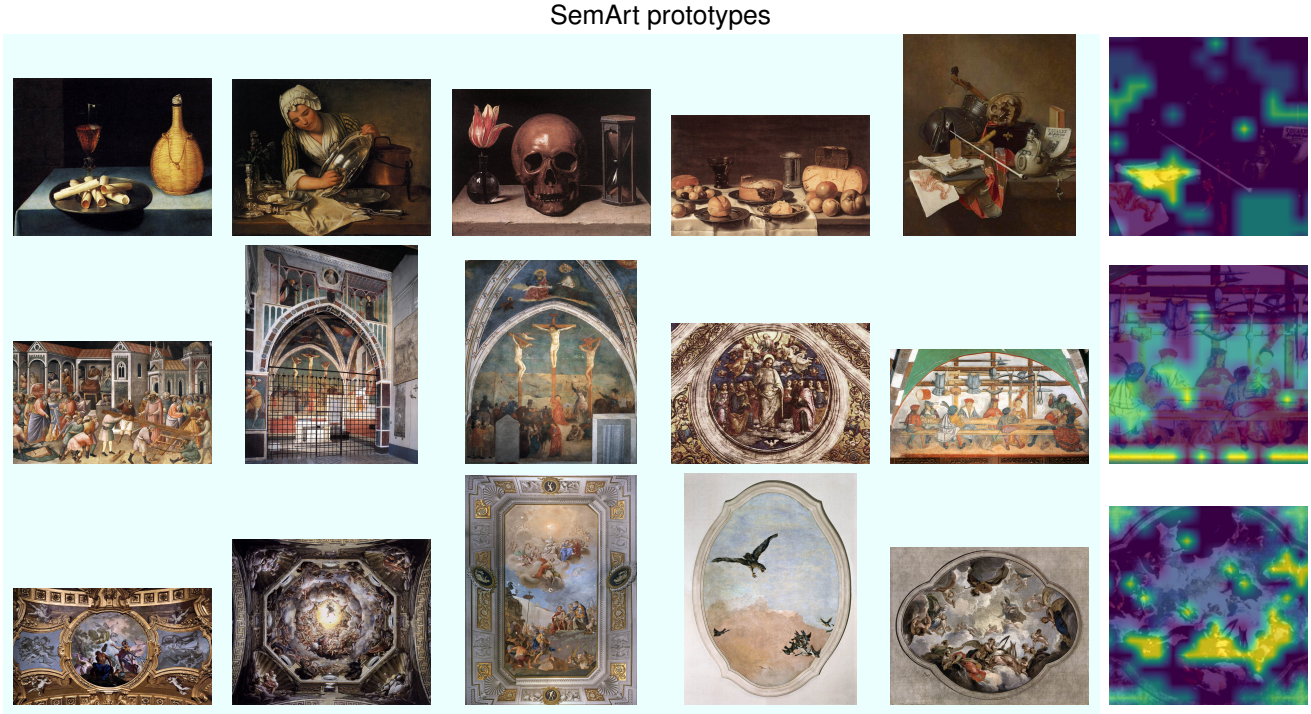

Figure 12. Three predominantly SemArt prototypes: *Still life painting*, *Panel painting*, and *Ceiling Fresco*. Per row five example images are shown, the sixth column contains the square-cropped attention mask of the prototype for the image in the fifth column.

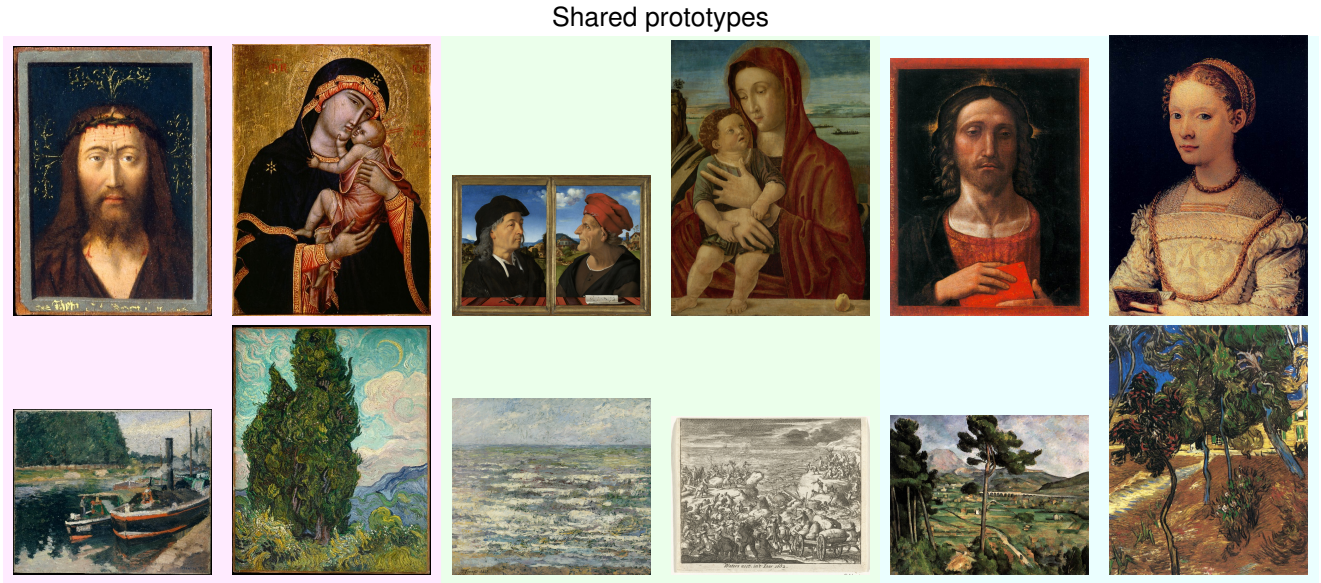

Figure 13. Two shared prototypes: *Icon paintings*, *Impressionist paintings*. The first two columns contain examples from MET, the second two from Rijksmuseum, and the last two columns from SemArt. The second prototype is rarely found in the Rijksmuseum dataset, mostly activating for drawings (as shown in second row, fourth image).