

Concept Splatters: Exploration of Latent Spaces based on Human Interpretable Concepts — Appendix —

In the supplementary document, we present further examples qualitatively supporting the two underlying hypotheses of concept splatters, as well as further results of visual confirmatory and exploratory analyses of latent spaces using concept splatters.

Contents

A Fashion MNIST (FMNIST)	2
A.1 Inter-Class Variability: Shoes	2
A.2 Intra-Class Variability: Fashion	3
A.3 UMAP Recomputation	5
A.4 UMAP Parameters	6
B Inception-V1 and ImageNet	7
B.1 Inter-Class and Intra-Class Variability: Organisms & Artifacts	7
B.2 Rare Category: Tench	10
B.3 Inter-Class Variability: Screen – Monitor and Tusker – Elephant	15
B.4 Intra-Class Variability: Edible Fruit	18
C Inception-V1 and Oxford Flowers	21
C.1 Inter-Class Variability: Coltsfoot and Dandelion	23
C.2 Intra-Concept Variation: Coltsfoot	24
C.3 Rare Categories: Fritillary	25
D word2vec (Google News) and WordNet	26
D.1 Inter-Class Variability: Organisms & Artifacts	27
D.2 Intra-Class Variability: “Topics”	29
E User Discoveries	31
References	36

A Fashion MNIST (FMNIST)

A.1 Inter-Class Variability: Shoes

As additional example demonstrating the effect of visual splatter overlaps on classification results, we show a heatmap of the three classes of shoes in FMNIST. As for the shirts (cf., Figure 4 (main paper)), we can observe most mis-classifications in the overlap regions.

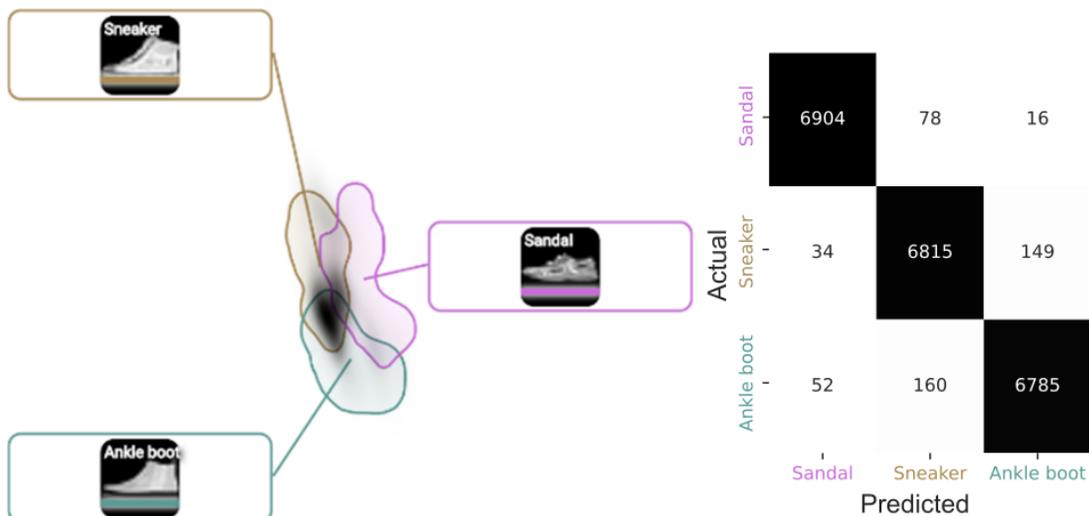


Figure A.1: Overlap of concept splatters of the three classes of shoes in FMNIST overlaid by a heatmap of mis-classified instances (left) and a confusion matrix of the classification results (right).

A query splatter around the overlap region reveals that the probably most problematic cases are high sandals and sneakers, which cannot be properly separated from ankle boots (Figure A.2).

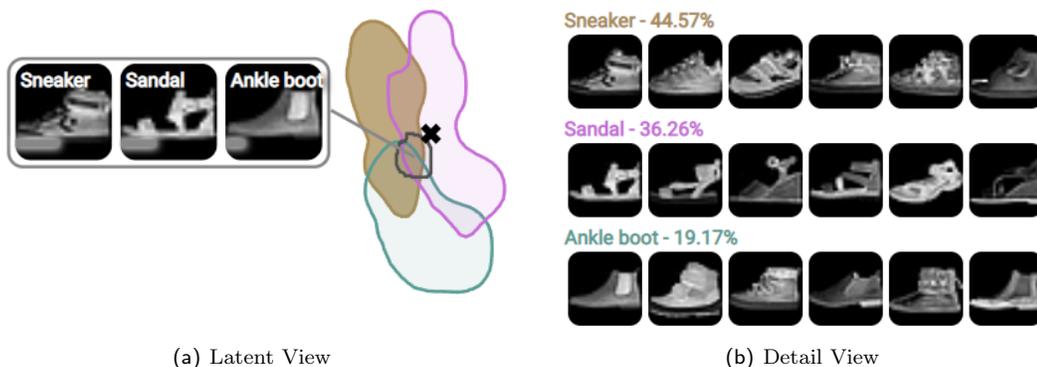


Figure A.2: Query splatter of the three-fold splatter overlap region: the inset and the detail view reveal high shoes across all three categories.

A.2 Intra-Class Variability: Fashion

We add an example to the qualitative validation of intra-class variability (cf., Section 4.3 (main paper)) expressed through the basic categorization in the latent space. In contrast to Figure 3 (main paper), we therefore decrease the bandwidth in the latent space view to reveal intra-class variability through an increasing number of more fine-grained splatters (Figure A.3).

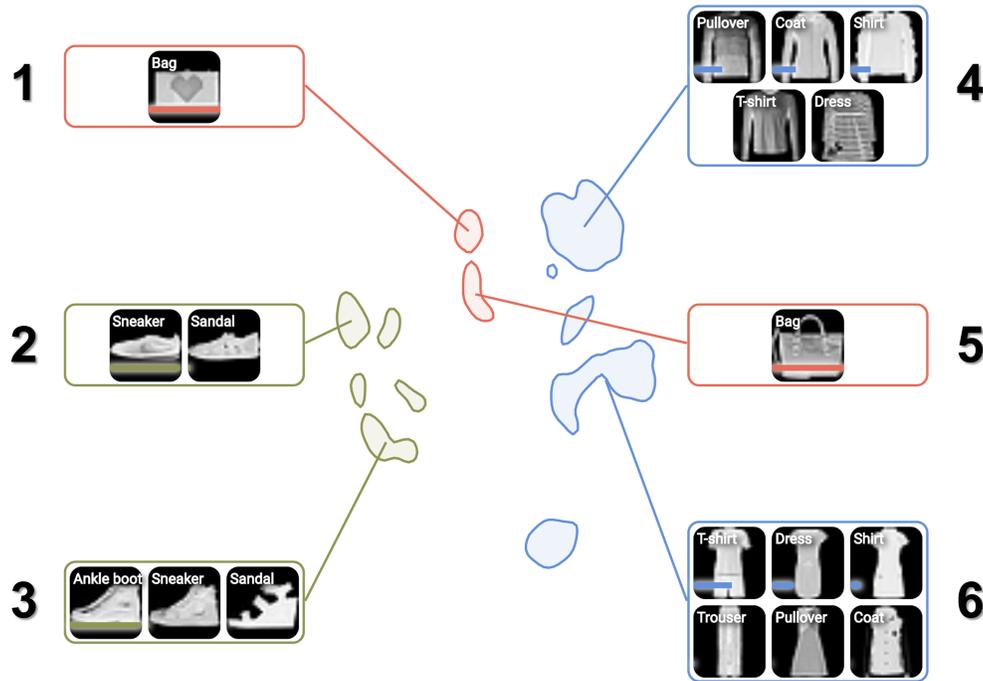


Figure A.3: Prototypical samples for six splatters.

We use the terminology by Mautz et al. [4] (supplementary document), who derive a hierarchy of visual attributes in FMNIST from hierarchical clusters, to uniquely describe the six labeled splatters in Figure A.3 (associated concepts from the concept space are in brackets):

1. Bags without visible handles (accessories),
2. flat and ankle-free shoes (shoes),
3. heeled or ankle boots (shoes),
4. long-sleeved upper body clothes (clothing),
5. bags with handles or visible straps (accessories), and
6. short-sleeved or sleeve-less upper body clothes (clothing).

Comparing the descriptions with Figure A.4, we can confirm that the visual attributes fit all 100 randomly selected samples of splatters 1 to 5. In splatter 6, we can find one instance with 3/4-sleeves (last image in last row).

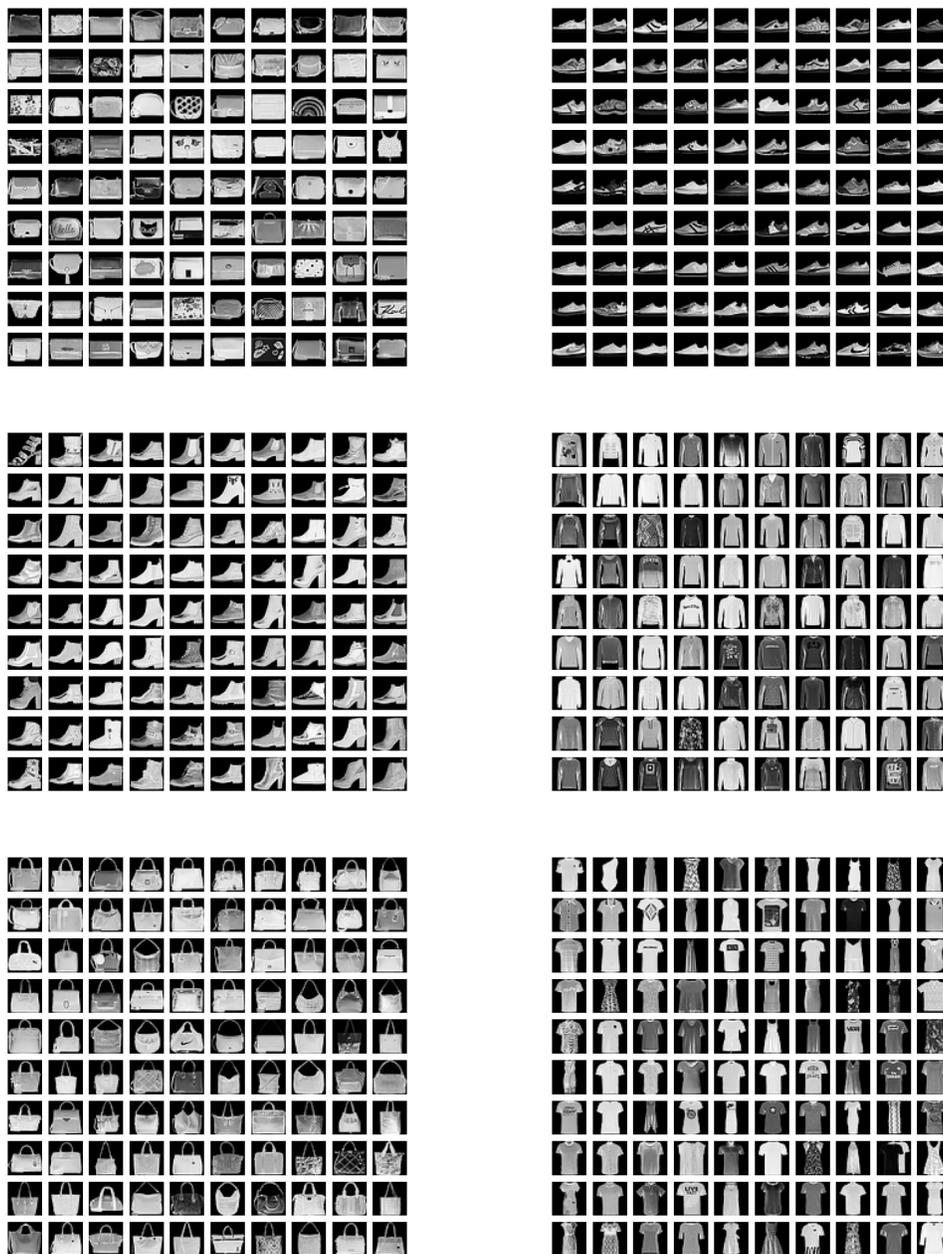


Figure A.4: 100 random samples for each of the six splatters shown in Figure A.3. Note that the FMNIST image size is only 28×28 pixels.

A.3 UMAP Recomputation

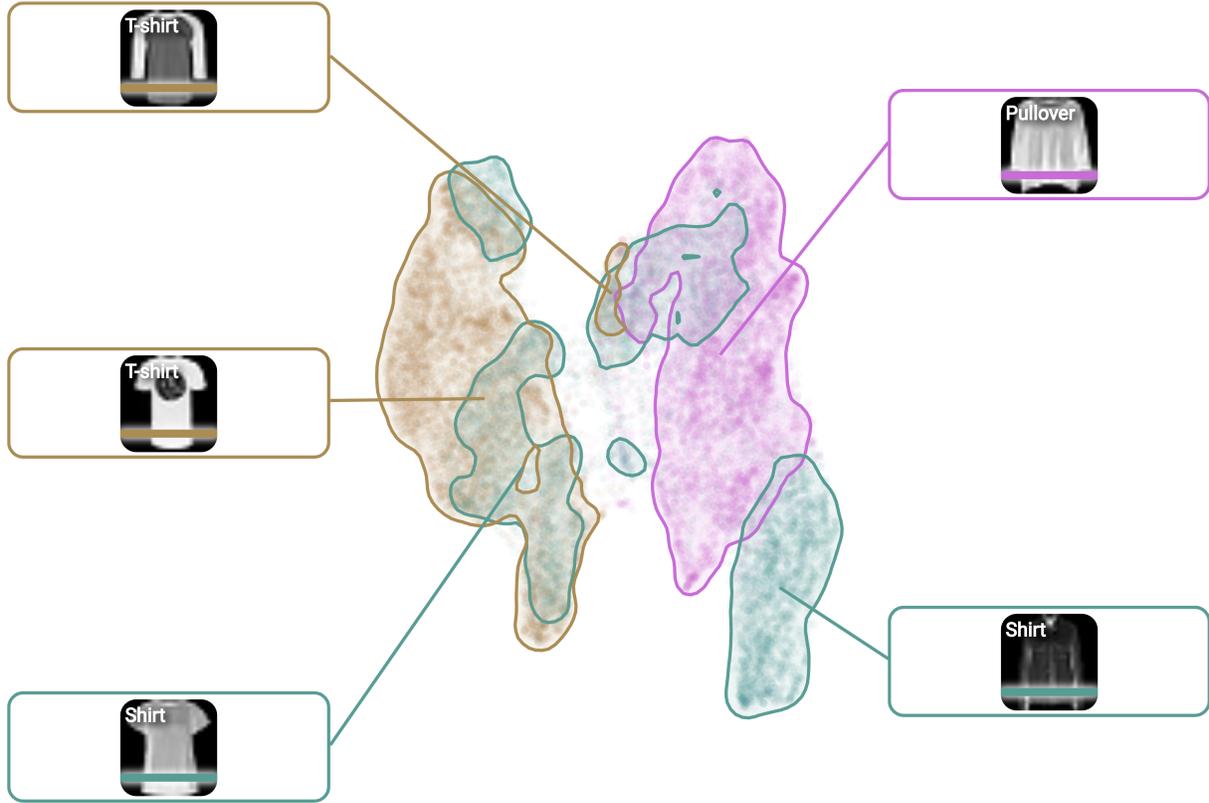


Figure A.5: Concept splatter plot of a filtered and re-calculated latent space only consisting of *shirt*, *t-shirt* and *pullover* samples.

Like Figure 4 (main paper), Figure A.5 shows the latent view of tops, but after recomputation of the dimensionality reduction. Overall, we can observe the same characteristics (i.e., short-sleeved vs. long-sleeved shirts) as in the original space, but more fine-grained intra-class variability becomes visible, such as several instances of long-sleeved t-shirts.

A.4 UMAP Parameters

Here, we investigate the effects of the two main parameters of UMAP – `n_neighbors` and `min_dist` – both of which control the balance between the preservation of local and global structures. The first parameter controls how many nearest neighbors are considered for each point during the projection, while the second one determines the minimum distance between points in the lower-dimensional space, i.e., how dense the projection ends up. As shown in Figure A.6, while there are some visual changes in the projection, there seem to be no considerable topological changes.

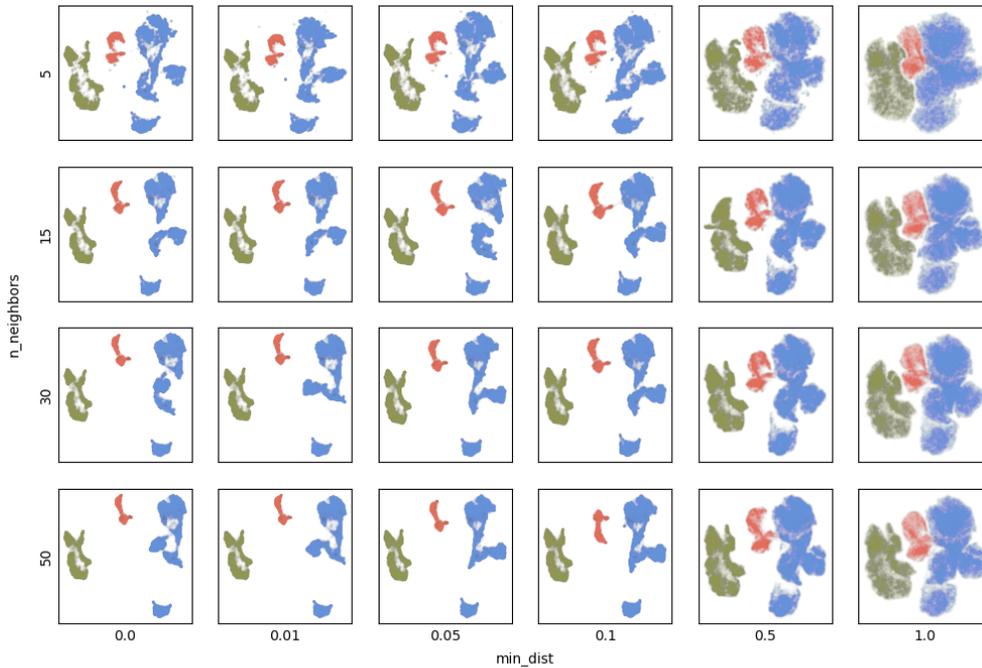


Figure A.6: Different parameter settings for UMAP.

For all our examples and the web-based implementation, we chose a parameter pre-set of (`n_neighbors`: 30 and `min_dist`: 0.2) as it seems to provide a good balance between the preservation of structure, without packing samples too tightly.

A comparison between t-SNE and UMAP for FMNIST can be found, for instance, in the Distill article by Li et al. [3].

B Inception-V1 and ImageNet

B.1 Inter-Class and Intra-Class Variability: Organisms & Artifacts

In the main paper, we analyze the separation between organism and artifact images, as well as splatters overlapping with the respective other category. Figure B.1 shows the concept view of the parent synset “whole”.

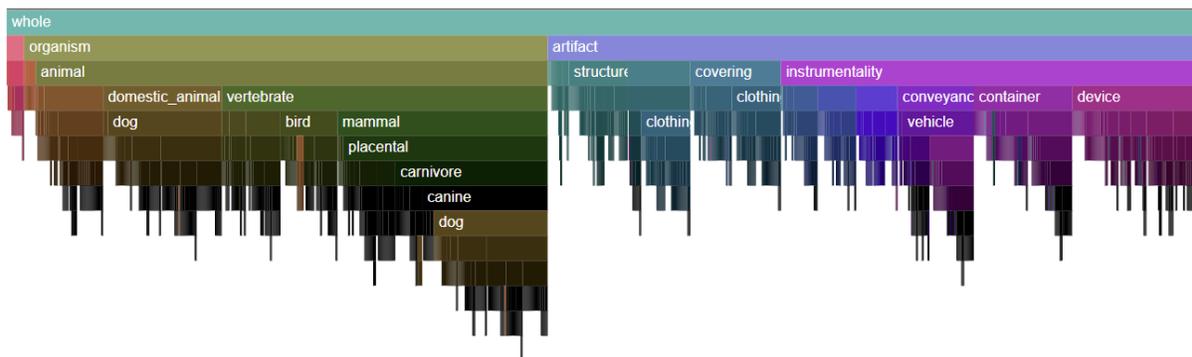


Figure B.1: Concept view of “whole”.

For comparison with the state-of-the-art, we show here a regular grid of images in Figure B.3 from the synset “whole”. Each grid cell contains one random image located within the grid cell on the underlying similarity map. For illustration purposes, we overlaid the splatter boundaries rooted at concept “whole” (violet = “artifact”, green = “organism”, pink = “natural object”). From this view, it is possible to roughly characterize the variability of the chosen root concept, but without concept splatter overlays, it is not possible to determine how this observed variability is reflected in the concept space. In addition, rare categories, such as images of animals or humans labeled as artifacts (see Figure B.2), are hard to find or may not even show up.



Figure B.2: Artifacts inside the organism splatter (a), as well as organisms and artifacts in a region of low inter-class variability between these two synsets.



Figure B.3: Grid view of images. Concept splatter outlines are overlaid for better comparison.

For another comparison, consider the similarity map in Figure B.4. Note that it was necessary to apply the hierarchical concept space here as well as color-coding of 1,000 ground truth classes is not possible. Using this similarity map without concept splatter outlines and annotations, the densities are visible very well, but dense regions cannot be characterized, and rare categories (specially those overlapping with other concepts, such as persons, cray / crabfish shown as food, and teddy / muzzle as shown in Figure B.2) are hardly visible and cannot be identified without splatter annotations. Concept splatters extend similarity maps by adding annotated splatter outlines, which enable interactive exploration.

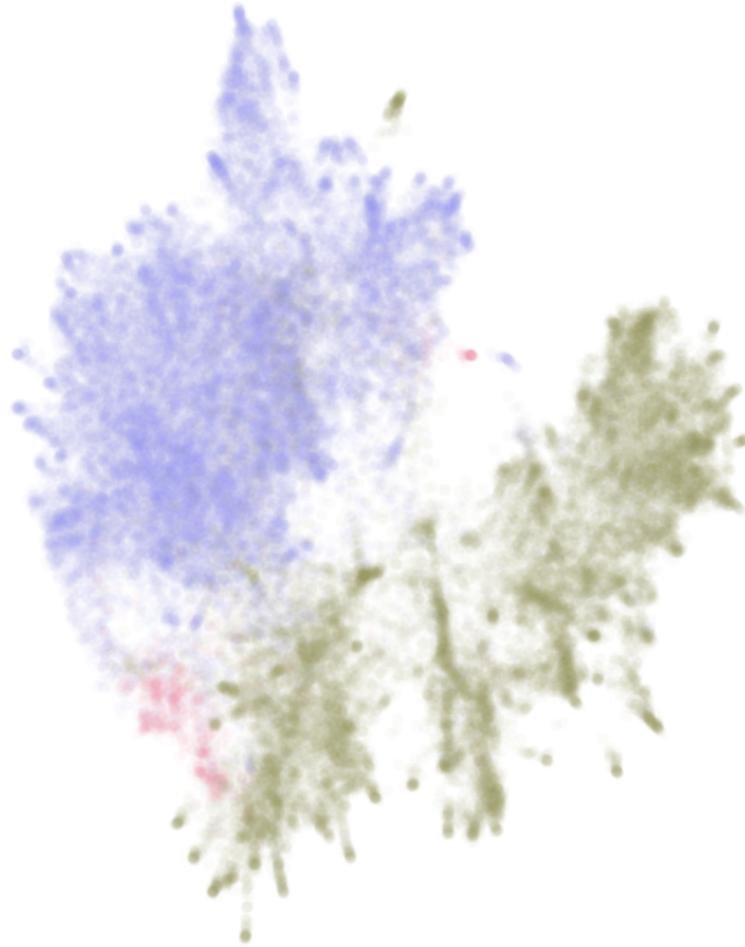


Figure B.4: Similarity map showing descendant instances of “whole”: violet = “artifact”, green = “organism”, pink = “natural object”.

B.2 Rare Category: Tench

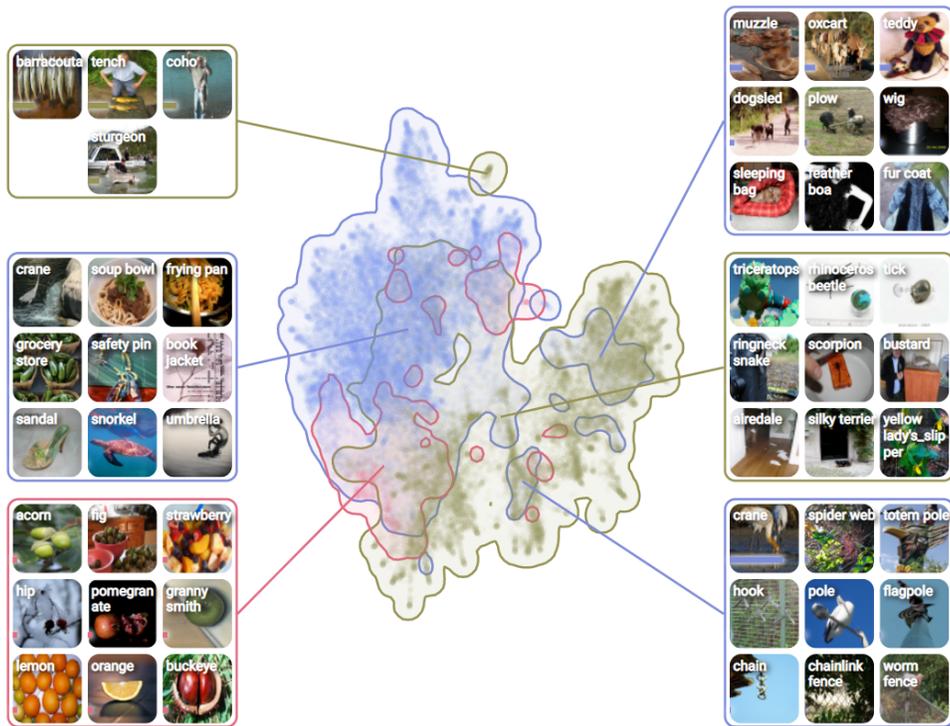
One particularly interesting synset of ImageNet is “tench”. Brendel and Bethge [1] as well as Hohman et al. [2] report that CNNs often identify the tench synset based on hands or fingers instead of visual features of the fish itself. The reason is that many images of tench are taken while persons hold the fish like a trophy.

It can be expected that visual features of human hands differ from visual features of a fish. Therefore, we can expect that “tench” is a rare category of fish that is clearly visually separate from other types of fish. Using concept splatters, we can find such trophy fish surprisingly easily. We first switch to a low detail level to be able to view rare categories in the similarity map. Already on a very high level of abstraction, namely “whole”, we can see trophy fish as one splatter of the child concept “organism” (see Figure 5(a) top left). In the detail view (Figure 5(b)), we can confirm that this splatter only contains fish – but not all fish from the data set. On most images, the fish are indeed held in the hands of a person. However, only 36 of the 118 images in this splatter are tench. Others are barracouta (38 images), coho (23 images), or sturgeon (21 images). Indeed, using the online demo of *Summit* [2], we could confirm that “barracouta” has similar activations as “tench”.

We further explore the fish synset to see if there are other visual attributes upon which the machine differentiates images of fish. We recompute the dimensionality reduction for the synset “fish” to do that. Figure B.6 shows the corresponding concept space. Figure B.7 shows the resulting grid view and similarity map of the synset “fish” for comparison. In the grid view (Figure 7(a)), it can be seen that there is a region containing images with fish held in hands (left). In the similarity map, we can see that there are two sub-concepts overlapping in this region, namely “food fish” (red) and “teleost fish” (blue).

Figure B.8 shows the resulting concept splatters. We can clearly see the overlapping splatters of “food fish” (red) and “teleost fish” (blue) on the left side, which contain images with humans. Note, however, that “coho” is a descendant of both, “food fish” and “teleost fish”.

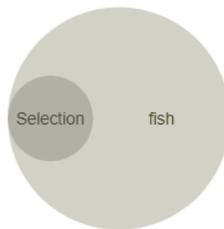
What is not visible in the grid view or similarity map is that there are two splatters of the synset “tench” when we select “cyprinid” as root concept – both with more than a dozen of images each. Figure B.9 shows the Detail Views of these two splatters. The sample images reveal that the machine distinguishes between tench held in hands and tench lying on the ground. If we make a spatial selection around the second tench splatter and step back up the hierarchy, we can observe that no other fish synset is overlapping with this splatter. It seems that there are no other pictures of fish lying on the ground in this data set.



(a) Latent Space View

whole ► organism ► ... ► ► vertebrate ► fish

fish



teleost_fish - 67.80%



food_fish - 51.69%

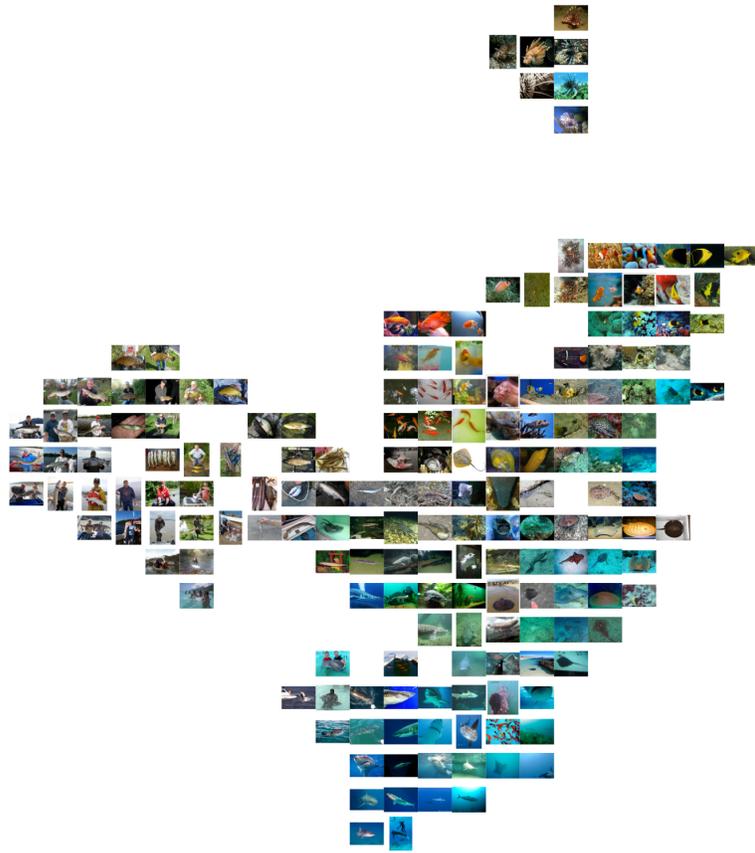


(b) Detail View of top left splatter

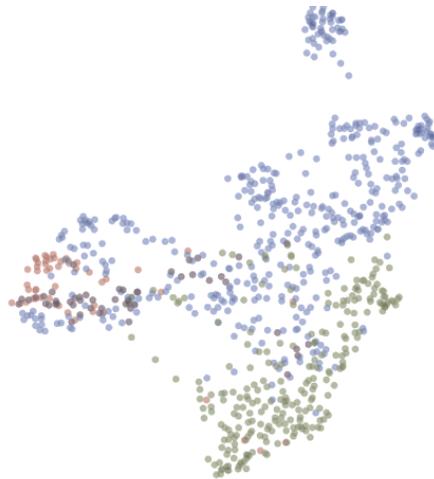
Figure B.5: Trophy fish already show up as rare category splatter of the synset "organism" (top left). This splatter contains only fish, but not only tench. Most fish are held in hands like a trophy.



Figure B.6: Concept space of "fish".



(a) Fish Grid View



(b) Fish Similarity Map

Figure B.7: Synset "fish" shown as grid of images and similarity map.

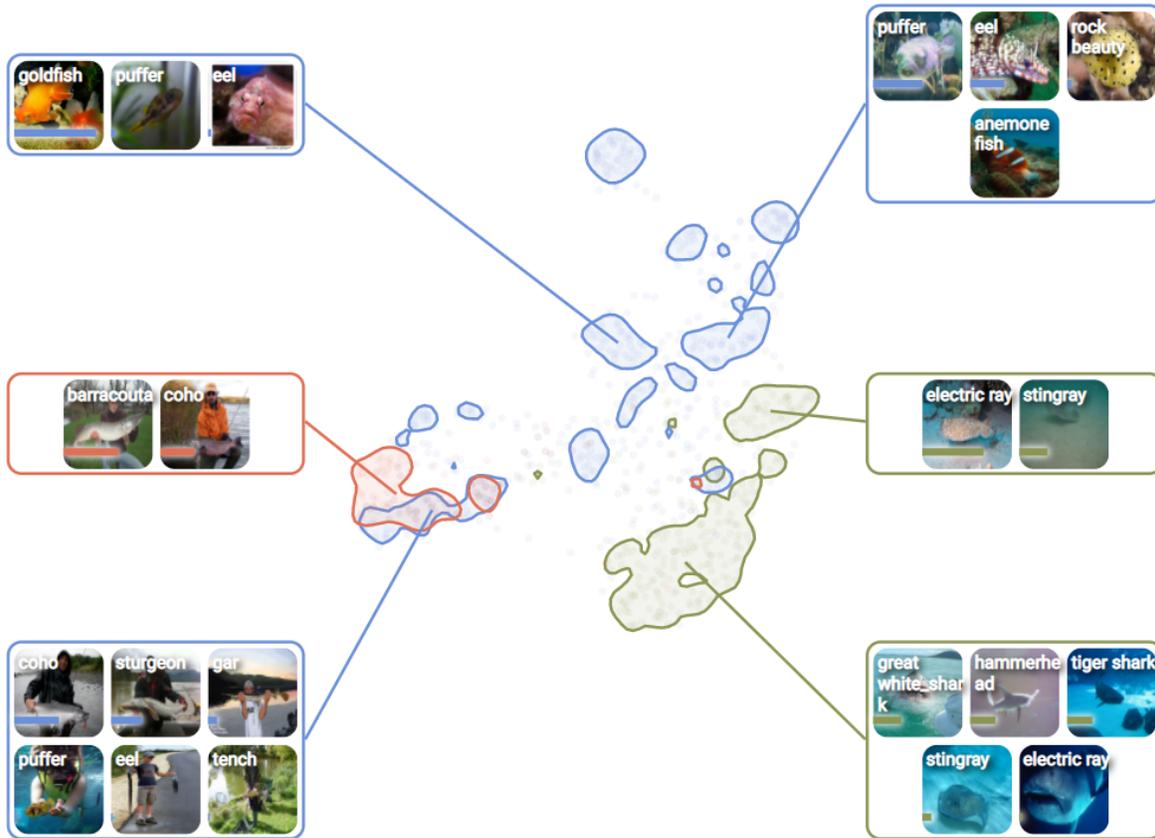


Figure B.8: Concept splatters of the synset fish with descendants “food fish” (red), “elasmobranch” (green), and “teleost fish” (blue). On the left hand side, trophy fish are clearly separate from various other types of fish. Teleost fish have a high intra-class variability.

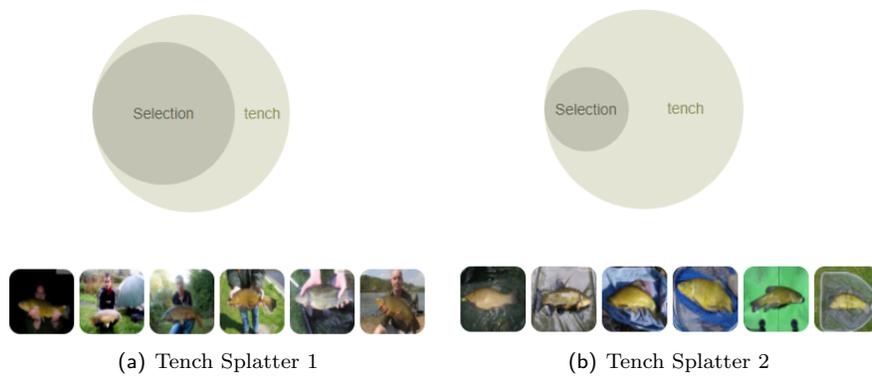


Figure B.9: Detail views of the two “tench” splatters: the larger splatter contains images of persons holding a tench, the smaller one contains images of tench on the ground.

B.3 Inter-Class Variability: Screen – Monitor and Tusker – Elephant

Recht et al. [6] conducted a study to analyze how well models trained on CIFAR-10 and ImageNet generalize to new data. The authors note that one of the most critical aspects is the human annotation step, especially for classes with unclear definition in ImageNet. Recht et al. [6] mention three pairs of classes with ambiguous boundaries in ImageNet: projectile – missile, screen – monitor, as well as tusker – elephant. We will show the latter two here.

For each of these pairs, we recompute the dimensionality reduction for the lowest common hypernym, namely mammal and instrumentality.

Within the WordNet structure, “monitor” and “screen” are rather distant. Their lowest common hypernym is “instrumentality”; “monitor” is a descendant of “equipment” → “electronic equipment”, and “screen” is a descendant of “device” → “electronic device”. We start by looking into “electronic equipment” and make a spatial selection around the single splatter of “monitor”. This splatter is well separated from other electronic equipment, such as “modem” or “CD player”. We then switch to “electronic device”, which has two descendants: “mouse” and “screen”. Figure 10(a) shows that the spatial selection overlaps considerably with the splatter “screen”. Indeed, a majority of “screen” images lie within this spatial selection. Unexpectedly, we can also see a considerable overlap with “mouse”. After inspecting the inset and the detail view (Figure 10(b)), we see that all sample images of “mouse” within this spatial selection also contain a monitor.

“Tusker” is a direct descendant of “mammal”, while “elephant” is a descendant of “placental”. Figure 11(a) shows that tusker can be separated well from all other mammals, except for one small mammal splatter. As expected, this splatter contains the majority of “elephant” images (see detail view in Figure 11(b)).

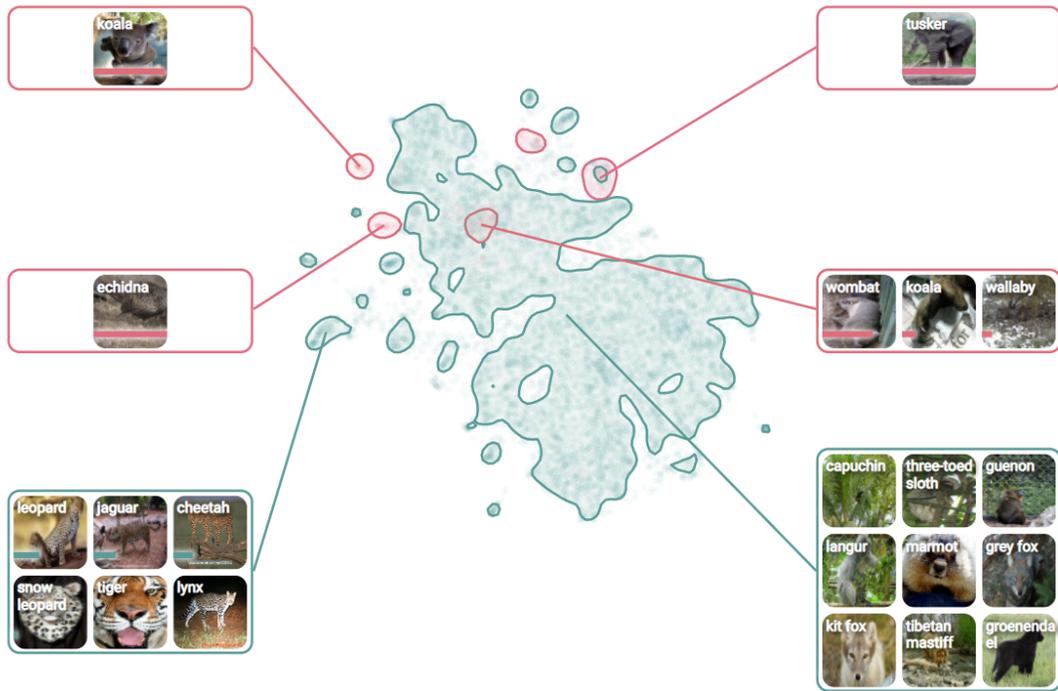


(a) Latent View



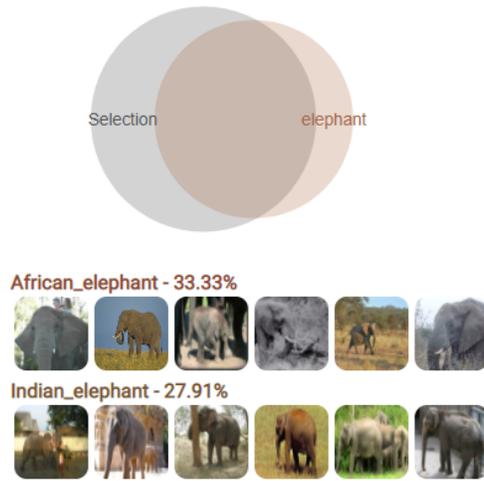
(b) Detail View

Figure B.10: Spatial selection around the electronic equipment “monitor” viewed for electronic devices “screen” (green) and “mouse” (violet) in the latent view. The majority of “screen” images lie within this splatter, but there are also a lot of “mouse” images containing a monitor.



(a) Mammals

elephant



(b) Detail View (Spatial selection around "tusker")

Figure B.11: Concept splatters of "mammals" (a). The detail view of a spatial selection around "tusker" (top right splatter) confirms that the majority of "elephant" images overlap with "tusker" (b).

B.4 Intra-Class Variability: Edible Fruit

Russakovsky and Fei-Fei [7] argue that WordNet’s noun hierarchy is “far from visual”. To demonstrate this, they show example images of direct descendants of the synset “edible fruit”. They argue that *“the high variability within [the] synsets makes classification on this dataset very challenging.”*

Here, we try to visually confirm that descendants of “edible fruit” are visually very heterogeneous. We first recompute UMAP for the “edible fruit” synset. For comparison with the state-of-the-art, we generated a grid view and a similarity map of the synset (Figure B.12): The similarity map indicates that the sub-concepts are not clearly separated. The grid view shows a separation of images into dedicated colors, with green apples on the top, citrus fruits mainly on the right, strawberries on the bottom, and fruits within trees on the left.

Using concept splatters, we can quickly confirm that edible fruit concepts are indeed scattered into multiple dense regions (see Figure B.13(a)). We then select the synset citrus, which is further separated into orange and lemon in the concept space (Figure B.13(a)). It is clearly visible that lemon and orange have a strong overlap in terms of their visual appearance. Using a low bandwidth value, we can see that both citrus fruits are scattered into three distinct regions that can be characterized by their way how the fruit is represented in the image – namely whether it shows the fruits in a tree, a cut fruit, or the fruit as a whole.

We investigate if this separation generalizes to other edible fruit categories by making larger spatial selections around the three “citrus” splatters in Figure B.13(a). Indeed, we can observe the same three visual themes for other fruit categories as well, as shown in Figure B.13(b): On the right, images show mainly cut citrus fruits, but also cut figs, strawberries, jackfruit, and pineapples. The spatial selection on the top contains either single or multiple fruits as a whole (citrus, granny smith, banana, fig, etc.). Finally, on the left, images show jackfruits, citrus fruits, pomegranates, figs, and other types of fruit growing in trees or bushes.

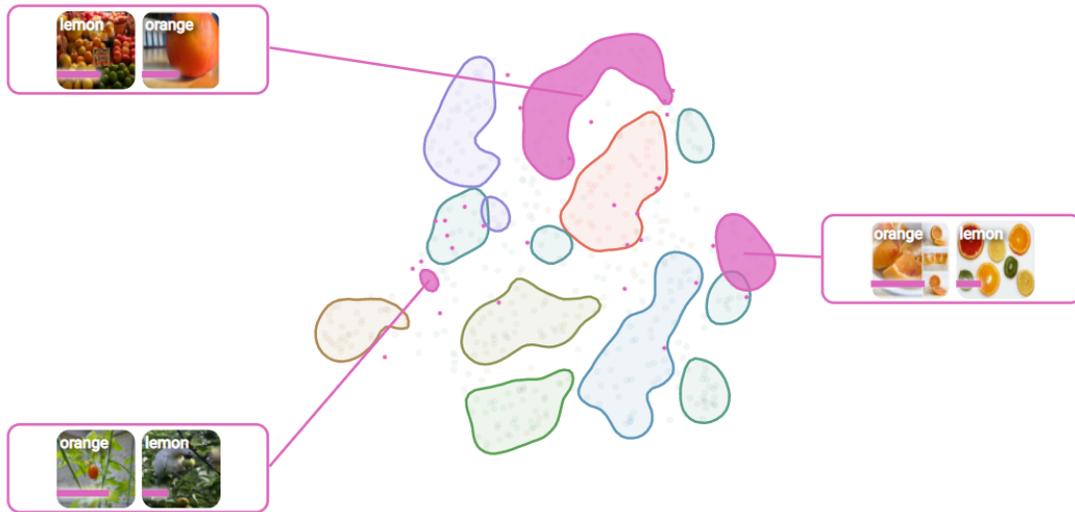


(a) Grid view

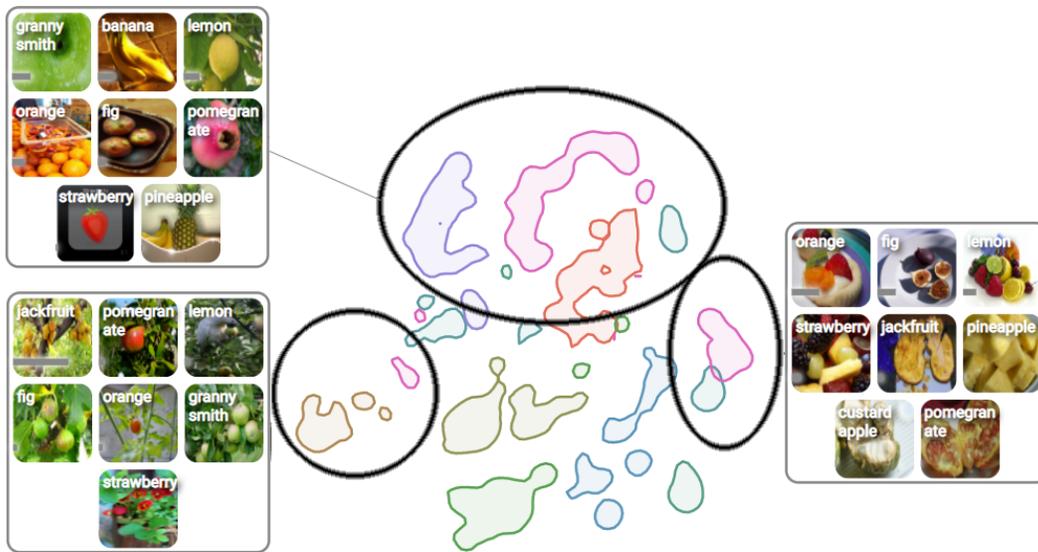


(b) Similarity map

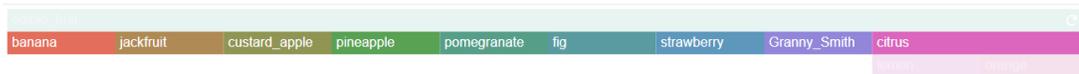
Figure B.12: Grid view and similarity map of synset "edible fruit".



(a) Edible Fruit – Latent Space (Citrus Selected)



(b) Edible Fruit – Latent Space



(c) Edible Fruit – Concept Space

Figure B.13: Citrus fruit splatters of lemons and oranges (a): The two concepts are separated into three distinct splatters that can be characterized as whole fruit (top), cut fruit (right), and fruit in trees (bottom). Spatial selections around the three splatters (b) confirm that these three categories generalize to other types of edible fruit as well (image manually composed of three separate spatial selections for illustration purposes). The concept space is shown in (c).

C Inception-V1 and Oxford Flowers

For the Oxford flowers data set, we apply the botanical taxonomy as concept space, as shown in Figure C.1. Note that not all flower categories are present so that several concepts only have one descendant.

Plantae														c		
Spermatophyta																
Angiospermae																
Monocotyledones							Dicotyledones									
Liliidae							Primales			Dilleniidae		Magnoliidae		Asteridae		
Liliales							Parietales		Ranales		Campanulales					
Iridaceae			Liliaceae				Violaceae			Ranunculaceae		Compositae				
daffodil	crocus	genus_Iris	iris	Convallaria	Tulipa	Fritillaria	Lilium	cowslip	Viola	Ranunculus	anemone	sunflower	Bellis	Tussilago	Taraxacum	
					tulip	fritillary	tiger_lily		pansy	buttercup			daisy	coltsfoot	dandelion	

Figure C.1: Biological taxonomy as concept space.

For comparison to the state-of-the-art, we show again a grid view and scatterplot of the images, overlaid by the “angiospermae” root concept, which is separated into “Monocotyledones” and “Dicotyledones” (Figure C.2). In the grid view, it is clearly visible that daisies and sunflowers are separated from the remaining flowers. In the scatterplot, some overlap between the two plant classes can be inferred.



(a) Flowers Grid View

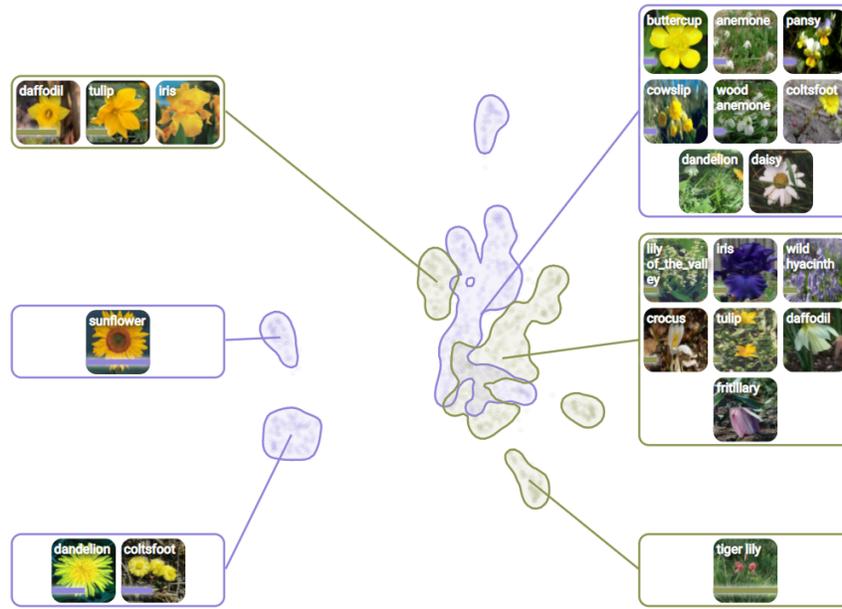


(b) Flowers Similarity Map

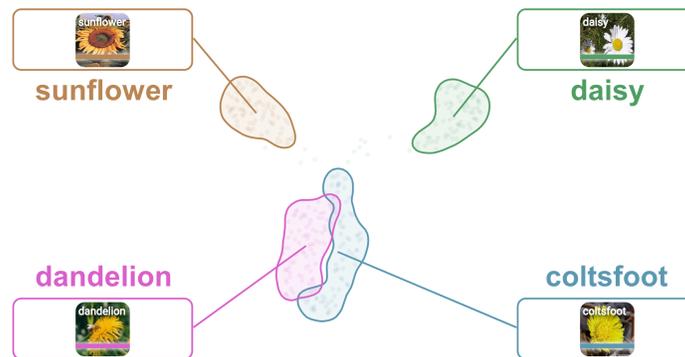
Figure C.2: Grid view overlaid with concept splatter boundaries “Monocotyledones” (green) and “Dicotyledones” (violet) and similarity map of Oxford flowers dataset with Inception-V1.

C.1 Inter-Class Variability: Coltsfoot and Dandelion

Using Concept Splatters, we can see that the two flower classes are quite well separated (Figure C.3(a)). When recomputing UMAP for the respective plant families, we can confirm that the separation works well even for visually very similar families, such as coltsfoot and dandelion (Figure C.3(b)), which are known to have a rather low inter-class variation [5].



(a) Angiospermae



(b) Compositae

Figure C.3: Using Inception-V1 trained on ImageNet for fine-grained flower classification: flower classes monocotyledones (green) and dicotyledones (violet) (a) and recalculated latent space for the family of compositae (b).

C.2 Intra-Concept Variation: Coltsfoot

By lowering the bandwidth and increasing the density threshold, we can observe previously reported intra-class variability of coltsfoot [5] as small sub-concept splatters differ primarily based on the camera angle as seen in Figure C.4.

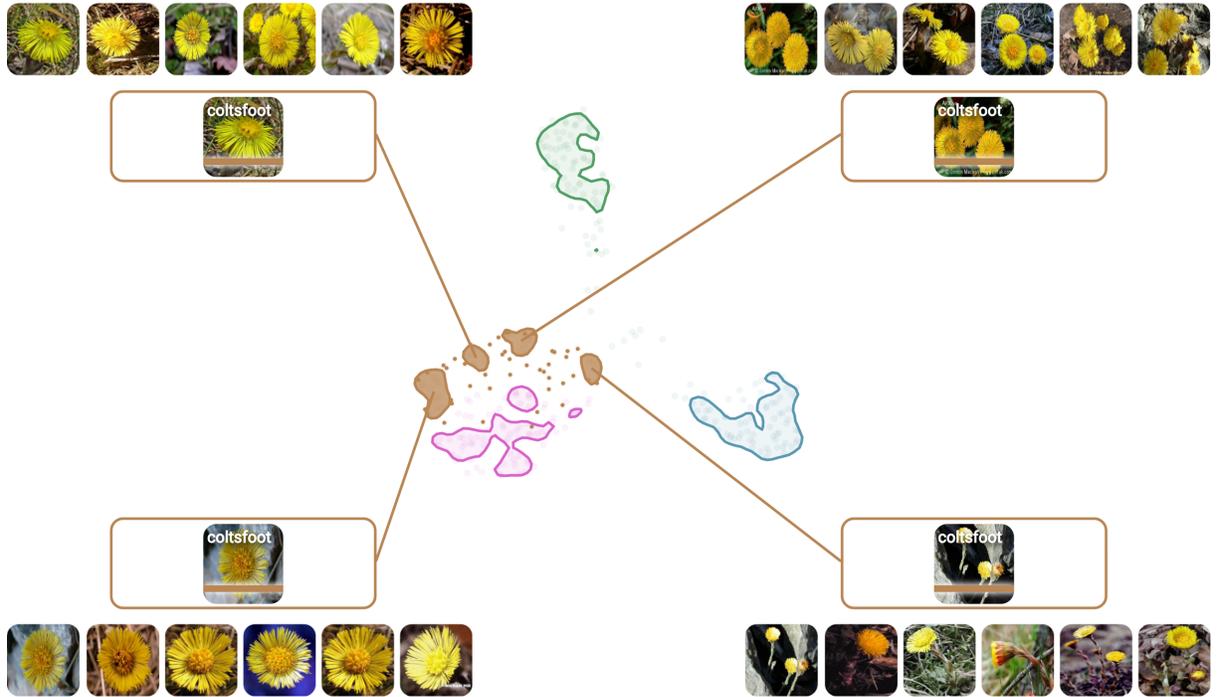


Figure C.4: Concept Splatters of coltsfoot showing separate groups of images with different perspectives.

C.3 Rare Categories: Fritillary

Sometimes, flowers of the same genus have very distinct looks, and uncommon appearances show up as rare category splatters. One example is the genus “fritillary”. Figure C.5 shows that “fritillary” has a small rare category splatter containing only three images. These images differ from the main splatter by their color.

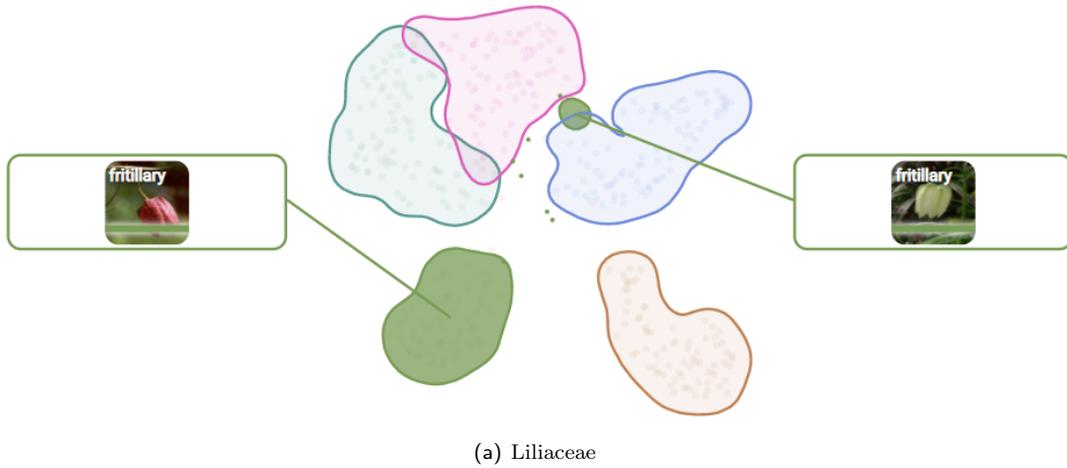


Figure C.5: Concept splatters of 'liliaceae', with the sub-concept “fritillary” selected. The bottom row shows the detail views of the two “fritillary” splatters.

D word2vec (Google News) and WordNet

For comparison with the state-of-the-art, we show a similarity map of all words in Figure D.1, colored by their parts of speech. It is obvious that the parts of speech overlap, and some dense regions (mainly nouns) are also visible.

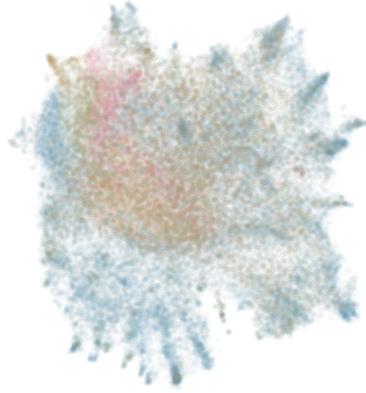


Figure D.1: Scatterplot of words, colored by their parts of speech (noun = blue, adjective = green, verb = brown, adverb = pink).

Figure D.2 shows the similarity map of Figure D.1 augmented by Concept Splatters on the highest level of abstraction. The overlap between the different parts-of-speech, such as verbs and nouns, is expected, since different parts-of-speech naturally co-occur.

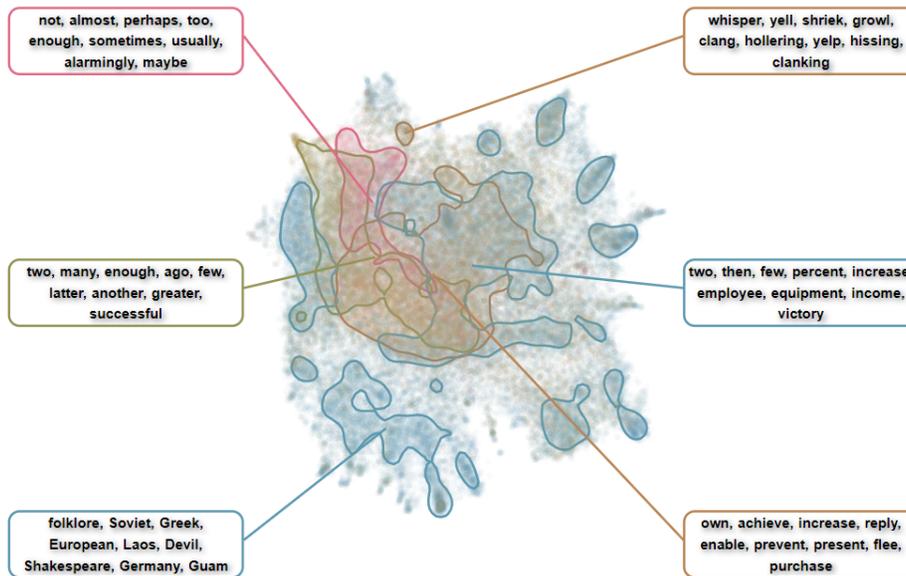


Figure D.2: Parts-of-speech splatters for word2vec: nouns, verbs, adjective, and adverbs have a strong overlap, but there are multiple separated noun splatters.

D.1 Inter-Class Variability: Organisms & Artifacts

We replicate our exploration of the synset “whole” (see Section B.1) for the pre-trained word embedding to test if WordNet is a useful hierarchy to structure large word embeddings. After recomputing UMAP for the synset “whole”, Figure D.3 shows that, indeed, the pre-trained word embedding can separate “artifact” (green) from “living thing” (violet) and “natural object” (pink) fairly well.

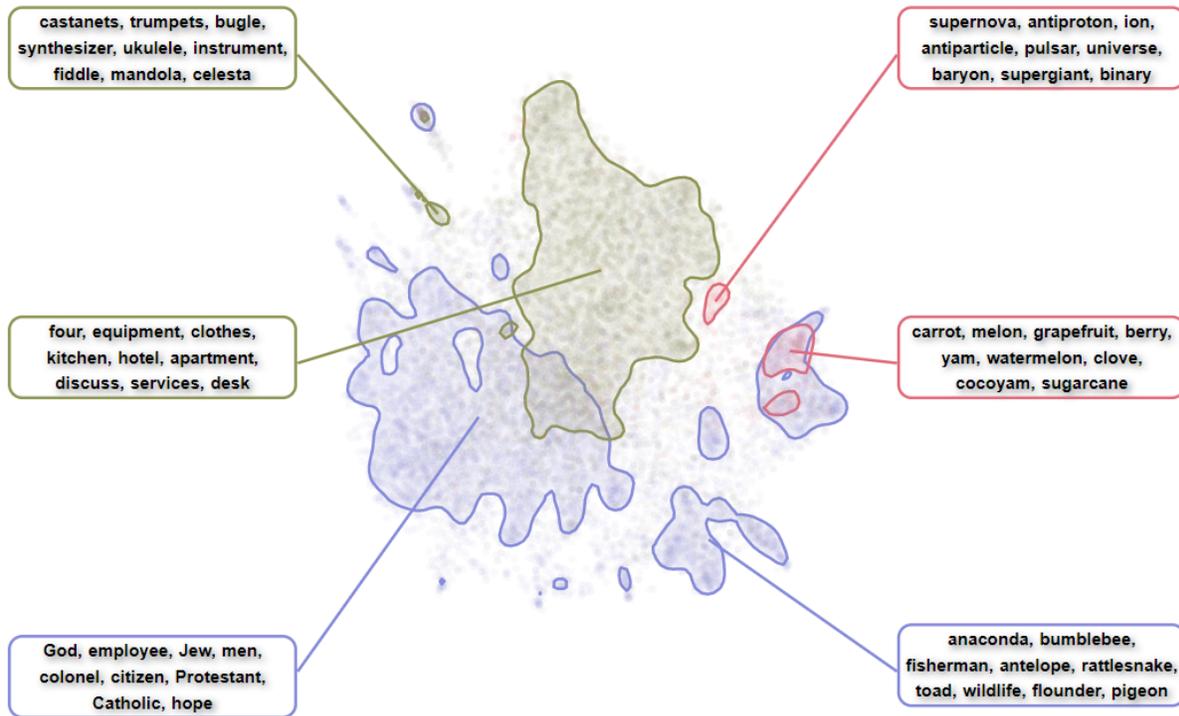
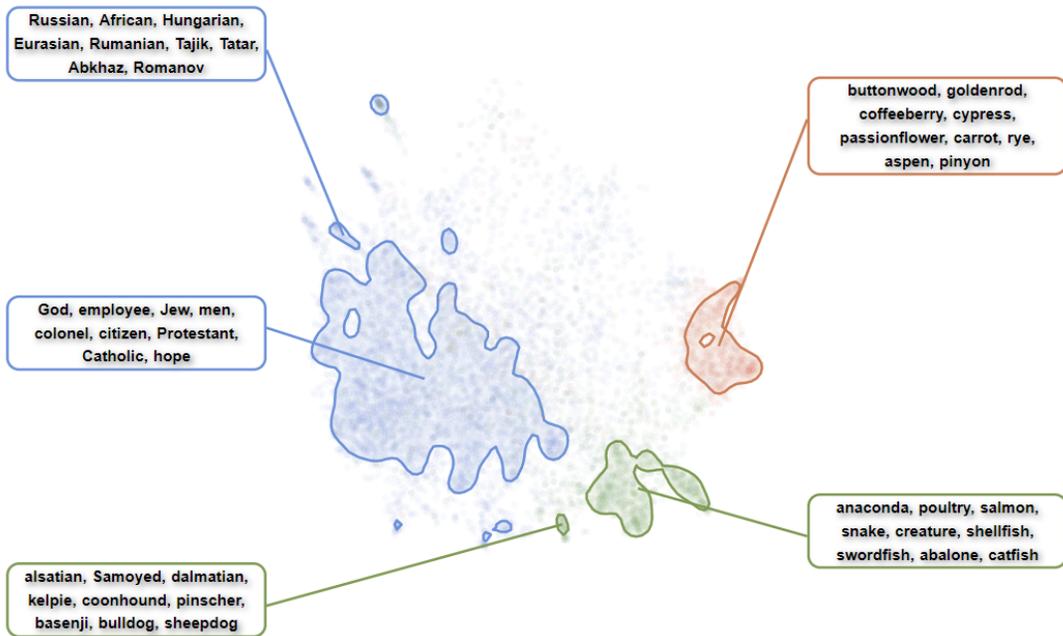
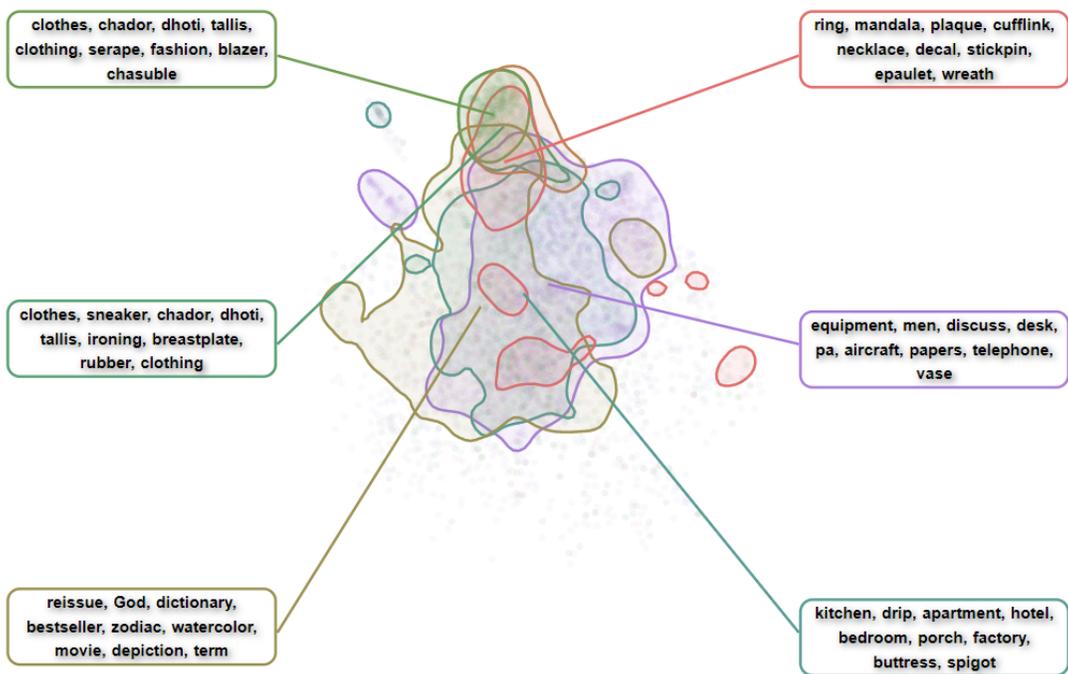


Figure D.3: Concept splatters rooted at synset “whole”. Note that, due to the dynamic color assignment, the color coding differs from the ImageNet figures in the main paper.

This is also true for the sub-concept “organism”, as shown in Figure 4(a): the three sub-concepts “person”, “animal”, and “plant” are clearly distinct. For the sub-concepts of “artifact”, however, this is not the case (see Figure 4(b)). Sub-concepts include “decoration”, “fabric”, “covering”, and “instrumentality”, which has sub-concepts, such as “vehicle” and “instrument”. Especially concepts related to fashion and jewelry (on the very top of Figure 4(b)) have a strong overlap.



(a) Organism



(b) Artifact

Figure D.4: Sub-concept splatters of "organism" and "artifact".

D.2 Intra-Class Variability: “Topics”

In word embeddings, semantically similar words are geometrically close. Therefore, clustering can be used to extract groups of semantically related words [8], i.e., “topics”. To explore latent “topics” in the word embedding, we only consider nouns, and we qualitatively describe the content of noun splatters to characterize these “topics”.

To explore dense noun regions, we select the most general noun sub-concept “entity” as root. “Entity” has two descendants: “abstraction” and “physical entity”. Figure D.5 illustrates that it is possible to derive topical word splatters from this high-level concept. The two central abstraction and physical entity splatter represent broad artifacts, persons, and events, respectively. As illustrated in Figure D.5, the peripheral splatters describe more specific topics, such as food and drinks (top right), medicine (bottom right), and religious and spiritual topics (left).

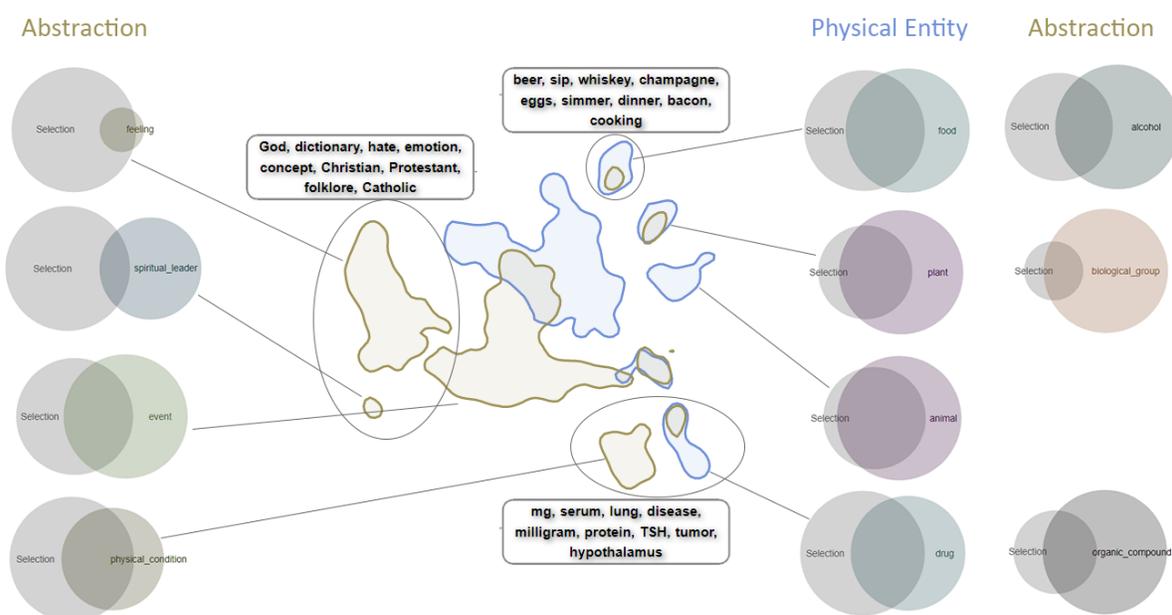


Figure D.5: Sub-concept splatters of “entity”: the peripheral splatters are annotated by the detail view Euler diagrams of their most prominent sub-concepts for illustration purposes. Insets of three spatial selections are also added for illustration purposes. From top right, the splatters represent the following topics, in clockwise order: food and drinks, plants, animals, drugs and chemistry, diseases, events, religion, and feelings.

We can now use Concept Splatters to find other parts of speech related to one of these topics. For instance, we can identify typical medical and chemical adjectives by selecting the respective spatial region for the concept “adjectives”. As illustrated in Figure D.6, the splatter on the bottom right indeed contains adjectives related to medicine, chemistry, and biochemistry.



Figure D.6: Splatters of general adjectives.

E User Discoveries

In addition to the visual confirmation of known network properties, we will here illustrate a selection of findings made by our users during the qualitative evaluation.

Table 1: List of selected discoveries by our users.

User	Discoveries in Concept View	Reference
DS	Overlap between <i>substances</i> and <i>food</i> in <i>matter</i>	Figure E.1
DS	<i>Piano</i> is associated with three parents	Figure E.2
MM	<i>Matter</i> has three descendents, where two of them are overlapping	Figure E.1
MM	<i>Whole</i> is separated into <i>organism</i> , <i>artifact</i> , and a small third group	Figure B.1
MM	ImageNet is highly imbalanced towards dogs	Figure B.1
User	Inter-Class Variability Discoveries in the Latent View	Reference
DS	<i>Food</i> and <i>beverages</i> are separated	Figure E.1
DS	Visual similarity between <i>piano</i> , <i>marimba</i> , and <i>accordion</i>	Figure E.2
SC	Visual similarity between <i>sea lion</i> and <i>megalith</i>	Figure E.3
SC	Visual similarity between <i>airliner</i> and trophy fish	Figure E.4
PS	Visual similarity between <i>dandelion</i> and <i>coltsfoot</i>	Section C.1
PS	Visual similarity between <i>buttercup</i> and <i>cowslip</i>	main paper
MM	Few <i>dogs</i> looking similar to <i>cats</i> and vice versa	Figure E.5
User	Intra-Class Variability Discoveries in the Latent View	Reference
SC	<i>Musical instruments</i> are segmented into many groups	Figure E.2
SC	<i>Reptiles</i> and <i>insects</i> are visually similar and separated from <i>dogs</i> and <i>birds</i>	main paper
PS	Influence of camera setting on clustering	Figure C.4
MM	<i>Organism</i> split into two groups: <i>fish</i> + <i>reptiles</i> and the rest	main paper
User	Rare Categories Discoveries in the Latent View	Reference
SC	<i>Whales</i> as rare category of <i>organism</i>	Figure E.3
SC	Rare category of fish with humans	Section B.2
SC	<i>Teddy</i> or dogs with sunglasses as part of rare category of <i>artifact</i> overlapping with <i>organism</i>	Figure B.2
MM	Multiple <i>artifacts</i> as rare category overlapping with <i>organism</i>	Figure B.2
MM	Mislabeled <i>domestic cats</i>	Figure E.5
MM	<i>Artifact</i> images containing humans	Figure B.2
MM	Snail shell labeled as <i>coil</i> (descendent of <i>artifact</i>)	Figure E.6
MM	<i>Support structures</i> resembling the eyes of insects	Figure E.6



Figure E.1: *Matter* has two identical descendants *substance*, where one contains *beverages* (red), and one contains primarily *food* (violet). Insets are shown for the violet *substance* concept.

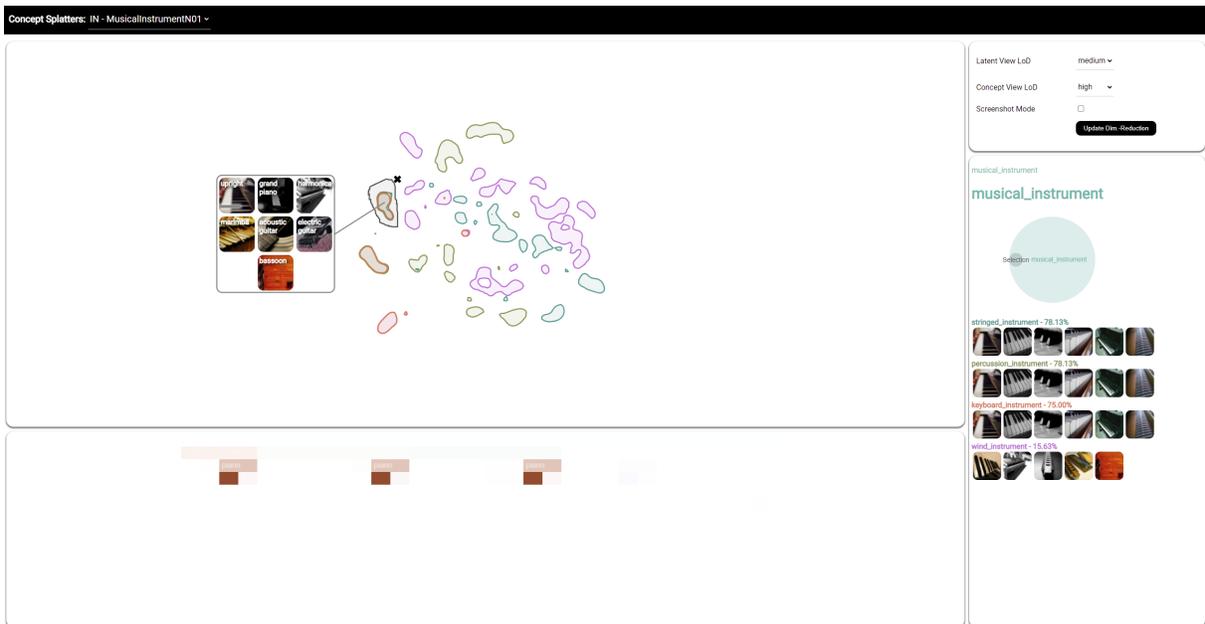


Figure E.2: Musical instruments: the synset “piano” is associated with three parents. Therefore, splatters of these three parent synsets often overlap. Please zoom in for details.

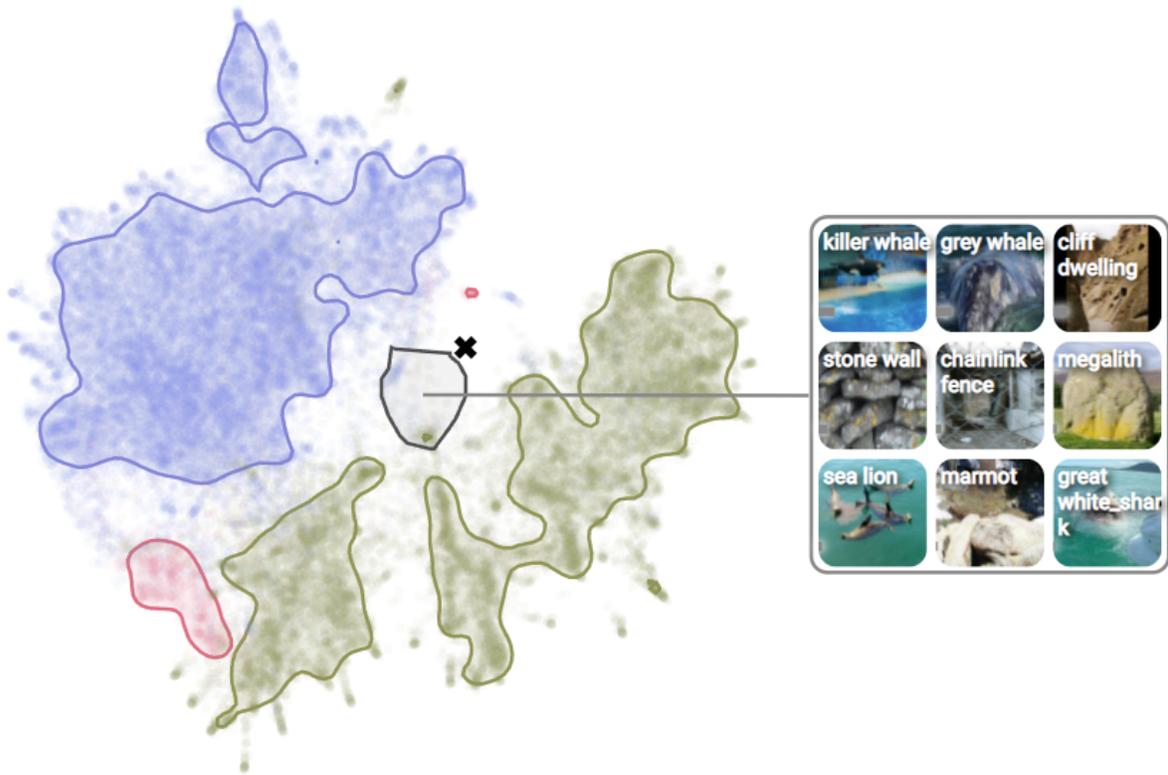


Figure E.3: Overlap between whales, sea lions, megaliths etc.

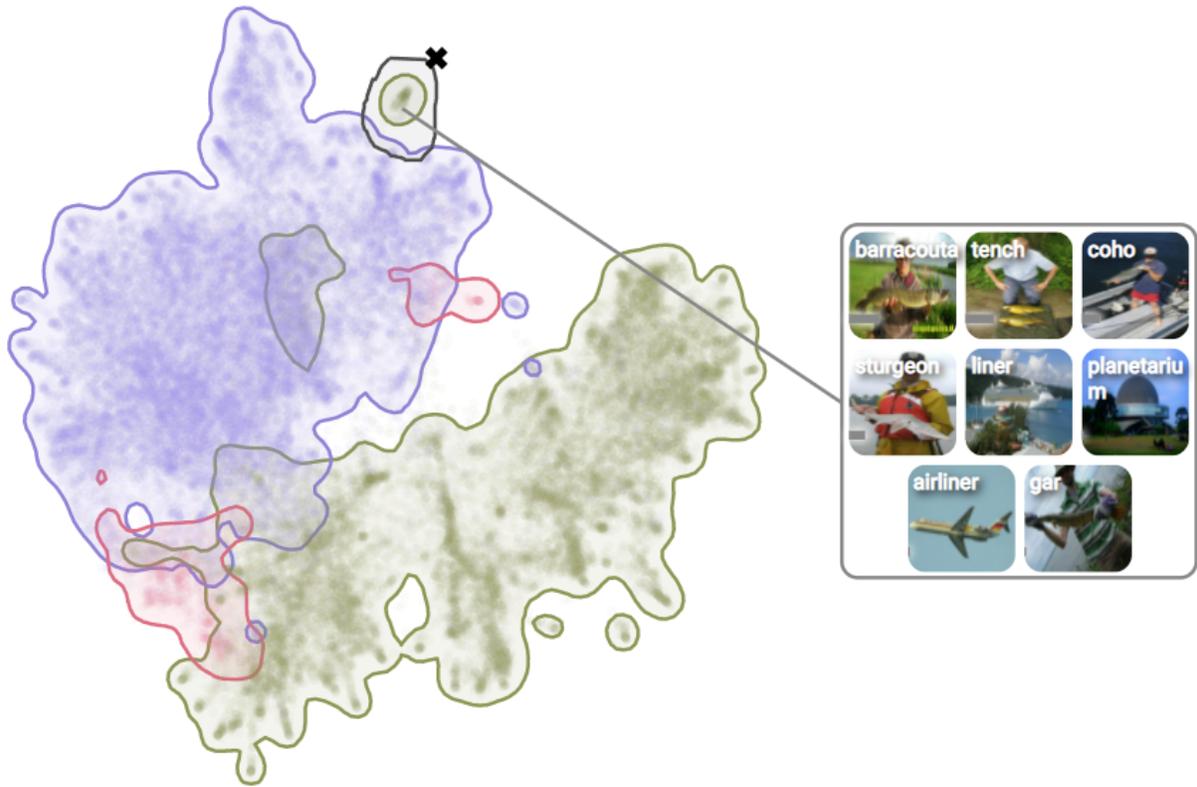
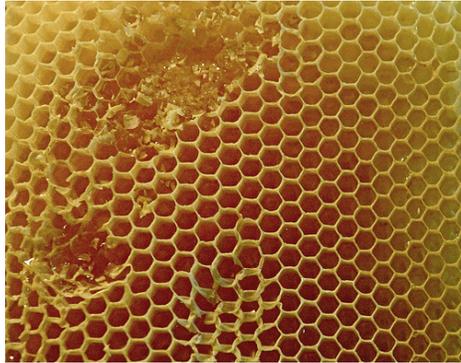


Figure E.4: "Airliner" and other artifacts similar to trophy fish.



Figure E.5: Examples for *domestic cats* not well separable from *domestic dogs*: a cat looking similar to a white dog (a), a mislabeled instance (b), and an unusual view of cats (c).



(a)



(b)



(c)



(d)

Figure E.6: Examples for *structure* artifacts not well separable from *organisms*: *honeycomb* (a), a *fence* with a toy animal (b), and two instances of *coil* (c, d).

References

- [1] Wieland Brendel and Matthias Bethge. “Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet”. In: *International Conference on Learning Representations*. 2018.
- [2] Fred Hohman, Haekyu Park, Caleb Robinson, and Duen Horng Polo Chau. “Summit: Scaling deep learning interpretability by visualizing activation and attribution summarizations”. In: *IEEE transactions on visualization and computer graphics* 26.1 (2019), pp. 1096–1106.
- [3] Mingwei Li, Zhenge Zhao, and Carlos Scheidegger. “Visualizing neural networks with the grand tour”. In: *Distill* 5.3 (2020), e25.
- [4] Dominik Mautz, Claudia Plant, and Christian Böhm. “Deep embedded cluster tree”. In: *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE. 2019, pp. 1258–1263.
- [5] M-E Nilsback and Andrew Zisserman. “A visual vocabulary for flower classification”. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*. Vol. 2. IEEE. 2006, pp. 1447–1454.
- [6] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishal Shankar. “Do imagenet classifiers generalize to imagenet?” In: *arXiv preprint arXiv:1902.10811* (2019).
- [7] Olga Russakovsky and Li Fei-Fei. “Attribute learning in large-scale datasets”. In: *European Conference on Computer Vision*. Springer. 2010, pp. 1–14.
- [8] Peng Wang, Bo Xu, Jiaming Xu, Guanhua Tian, Cheng-Lin Liu, and Hongwei Hao. “Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification”. In: *Neurocomputing* 174 (2016), pp. 806–814.