Supplement
DeepPatent: Large scale patent drawing recognition and retrieval

Michal Kucer, Diane Oyen, Juan Castorena
Los Alamos National Laboratory
Los Alamos, NM
{michal, doyen, jcastorena}@lanl.gov

Jian Wu
Old Dominion University
Norfolk, VA
jwu@cs.odu.edu

S1. Introduction: Drawing retrieval

S1.1. Patent drawing examples sorted

Figure S1 contains the same drawings from design patents as Figure 1, but in Figure S1 the drawings are grouped together if they come from the same patent.

S1.2. Commercial image retrieval tools

Example web-retrieval results using search-by-image in Google, Baidu, and Yandex demonstrate that the retrieval of drawings is an open problem in computer vision. Google retrieves unrelated cartoonish drawings (Figure S2). Baidu retrieves technical drawings that appear to be from patents (similar style to the query), but not of the same *cactus* subject (Figure S3). Yandex retrieves drawings of similarly-shaped cacti (similar subject) but a simpler style of drawing than patent drawings (Figure S4). In this simple example, Yandex arguably performs the best at content-based drawing retrieval. It is surprising how different the results are among these commercial tools; and an indication that content-based drawing retrieval is an open challenge.

S2. DeepPatent dataset details

S2.1. Ethics statement

Works created for the purpose of USPTO patent application are generally not subject to copyright [6]. Therefore, it is reasonable to assume that the artists and inventors creating these drawings and metadata did so under the expectation that the work would become public domain with no restrictions on re-distribution. These artists and inventors could benefit from computer vision research focused on patent drawing retrieval for tracking how widely their drawings are redistributed (which may be useful for personal and professional purposes, although the artists have waived legal rights of attribution). Being able to search through patent drawings could aid several tasks; such as speeding up the patent approval process by building a hierarchy of prior art; aiding in design ideation; and data-driven design.

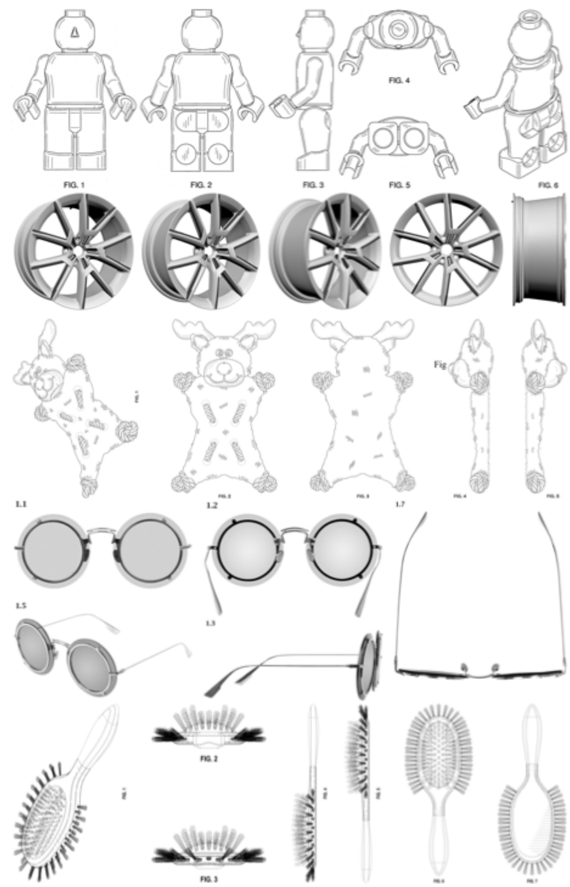


Figure S1: Example drawings from five patents: *toy figure, toy wheel, moose-shaped animal toy, spectacles, hairbrush* organized by row. Each patent contains multiple drawings of a single object from various views.

Our ultimate goal in providing this dataset is to make it easier to find technical information, particularly when contained in images, which should not in and of itself carry any

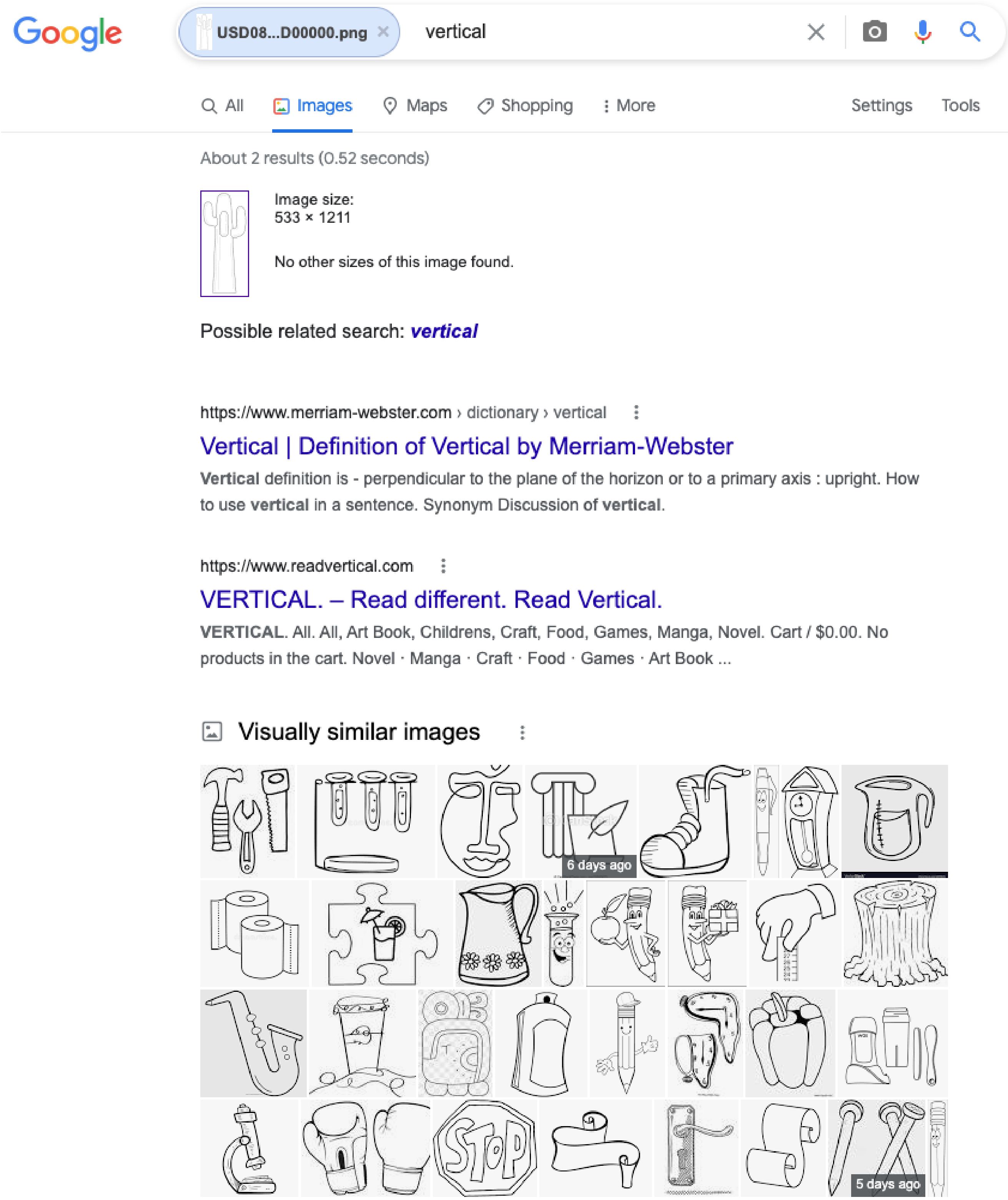


Figure S2: Google search-by-image results from April 6, 2021 for a patent drawing of a cactus.

Figure S3: Baidu search-by-image results from April 6, 2021 for a patent drawing of a cactus.

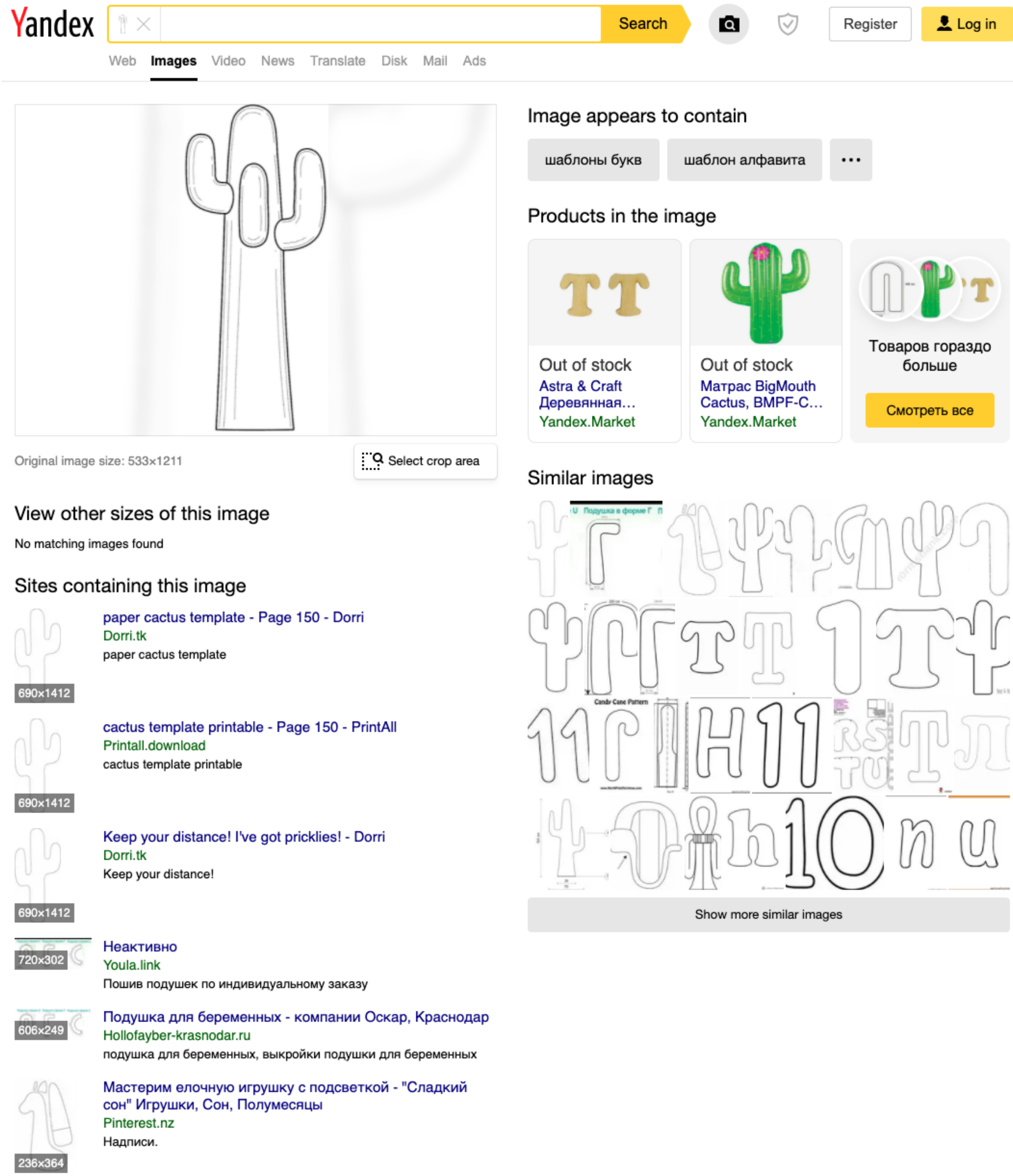


Figure S4: Yandex search-by-image results from April 6, 2021 for a patent drawing of a cactus.

Dataset	Pixel ratio	Comps
DeepPatent	0.030 ± 0.009	705
DeepPatent $\sigma = 3$	0.013 ± 0.001	305
Sketchy sketches	0.040 ± 0.007	2
ImageNet-Sketch	0.163 ± 0.059	244
Sketchy photos $\sigma = 1$	0.138 ± 0.014	265
Sketchy photos $\sigma = 3$	0.038 ± 0.004	34
ImageNet photos $\sigma = 1$	0.131 ± 0.014	673
ImageNet photos $\sigma = 3$	0.034 ± 0.004	84

Table S1: Comparison of statistics of visual features where "Pixel ratio" is the mean $\pm$ standard deviation of the ratio of foreground pixels to all pixels in the binarized image and "Comps" is the median number of connected components. When Canny edge detector is used to create edge maps, the width of the Gaussian filter, σ, is noted.

negative societal impacts. We only caution against over-reliance on models trained on this data if exhaustive prior-art searches are expected in patent review. For example, if a model trained on this data fails to retrieve similar prior-art for a particular design patent, that does not guarantee that the prior-art does not exist; it simply means that the current retrieval method does not identify it.

S2.2. Comparison with other datasets

Table S1 shows that the median number of connected components for DeepPatent is similar to that of edge maps generated from photos of objects in ImageNet. The very large difference in component count between DeepPatent and Sketchy can be attributed to the use of small lines for shading, detail, and for highlighting (through the selective use of dashed lines) that are prevalent in patent drawings but nearly non-existent in quick free-hand sketches. We find that the edge maps with minimal smoothing ($\sigma = 1$) contain more than 10% foreground pixels — an order-of-magnitude higher ratio than is present in DeepPatent or Sketchy (meaning that the edges make up a much greater portion of the image). Therefore, we reduce the number of edges by increasing the value of the σ parameter to $\sigma = 3$ in the Canny edge detector, so that the ratio of foreground pixels is on par with the drawings and sketches.

For DeepPatent, we simply count the number of connected components in each image, calculate the median component count for each week of drawings; then average the medians across the 26 weeks of the first half of 2019 (average is weighted by the number of drawings per week), to get the median estimate for the full DeepPatent dataset at 705. Noting that the shading in a binary image may show up as many separate components, we also generate edge maps with smoothing to focus on counting stroke-like features of these drawings and find that the median number of connected components drops in half to 305 with this smoothing approach. For Sketchy, we use the 256x256-pixel rendering of the sketches provided by the original authors. For Sketchy photos, we use the 256x256-pixel photos (12,500 photos across 125 classes) provided by the original authors and use two different smoothing settings on the Canny edge detector ($\sigma = 1$ and $\sigma = 3$). For ImageNet, we use the validation set of images (50,000 photos across 1000 classes) and use two different smoothing settings on the Canny edge detector ($\sigma = 1$ and $\sigma = 3$).

It is interesting to note that the shape-complexity of photos used for Sketchy is significantly smaller than that of the validation set of ImageNet photos. The Sketchy photos are selected from the ImageNet dataset (though not necessarily from the validation set), with preference given to photos that are judged to be easy to sketch [5].

S3. Comparison method details

S3.1. Classification pre-training with patents

We found that after 100 epochs of classification fine-tuning an ImageNet [2] pre-trained ResNet50, the retrieval performance difference between using the largest 2500 patents was no different from using the all 33000 patents. Therefore, for PatentNet and other comparison models, classification training was done using the 2500 largest patents (i.e. one that have the larges number of images), resulting of approximately 40000 images across 2500 patents.

S3.2. AHDH

As the authors recommend, we use 10 levels of hierarchy and begin the histograms at level 3 for a total image descriptor vector of length 140 [7].

S3.3. Histogram of oriented gradients (HOG)

To obtain a fixed-length descriptor for an image, first it is resized to 256×256 pixels. Then a set of HOG descriptors are computed for a set of equally spaced local regions. These descriptors are then concatenated to generate a single image descriptor.

S3.4. VisHash

We use the settings as described by the authors for a fixed image descriptor of length 144 [4].

S3.5. Local binary patterns (LBP)

For performance evaluation of the local binary pattern (LBP) features in the retrieval task we employ the same query/database selection procedure described in Section 3 with images resized to 256×256. Experiments with different LBP parameter settings were conducted and found that a radius of 16 with a total number of 512 points sampled around this radius and a 'uniform' pattern yielded best performance. Here, the radius parameter and the number of points controls the pixel locations around a pixel center where binary patterns are obtained from. The 'uniform' pattern refers to those patterns that tend to be the most likely (in a statistical sense) as studied in [3] and which are also known to benefit from rotation invariance.

To compute the BP, each sampled pixel value in the radius is subtracted from the center's pixel value and assigned a 0-bit value if the result is less than 0 or a 1-bit value otherwise. After comparisons between all pixels in a patch to its center are completed, the binary pattern is associated to a number between $[1-512]$ and assigned as the pixel's center value. This number indicates the number of 0-1 and 1-0 transitions in the pattern but truncated by prioritizing the number of transitions that are statistically more likely to occur. After completion, of this process throughout all pixels in the image a histogram distribution is computed. Finally, pair-wise similarities between images are obtained by comparing them using the Kullback-Leibler (KL) divergence between the normalized LBP histogram distributions.

S3.6. Fisher vectors (FV)

We follow the best practices outlined by Csurka [1], and generate the Fisher Vector encoding for an image by computing a 128-dimensional SIFT descriptors over a set of local patches densely extracted from the image, where each local patch is of size 48×48. These descriptors are then reduced to 48 dimensions through PCA. These smaller descriptors are then used to compute the FV by computing the gradient of the log-likelihood of the data on the model, in this case a Gaussian mixture model fit over the descriptors extracted from the training set. Similarly to deep descriptors, the FV of an image is unit-normalized before being compared to other images.

S4. Qualitative results

Figure S5 shows the t-SNE embedding of the PatentNet features for all of the test set images. Figures S6 through S10 show additional qualitative comparison of retrieved examples for various benchmarked models.

References

[1] Gabriela Csurka. Document image classification, with a specific view on applications of patent images. In *Current Challenges in Patent Information Retrieval*, 2016. 6

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition*, pages 248–255, 2009. 5

[3] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 6

[4] D. Oyen, M. Kucer, and B. Wohlberg. VisHash: Visual similarity preserving image hashing for diagram retrieval. In *Applications of Machine Learning*, volume 11843. International Society for Optics and Photonics, 2021. 5

[5] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database: Learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, 2016. 5

[6] United States Patent and Trademark Office. Terms of use for USPTO websites. https://www.uspto.gov/terms-use-uspto-websites, August 2021. 1

[7] Stefanos Vrochidis, Symeon Papadopoulos, Anastasia Moumtzidou, Panagiotis Sidiropoulos, Emanuelle Pianta, and Ioannis Kompatsiaris. Towards content-based patent image retrieval: A framework perspective. *World Patent Information*, 32(2):94 – 106, 2010. 5

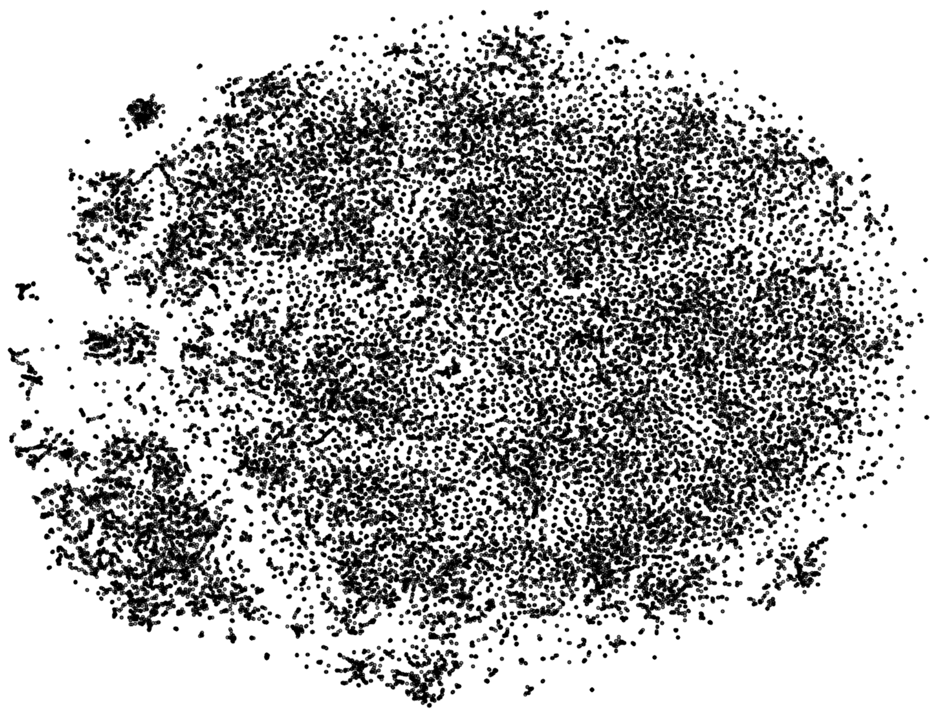
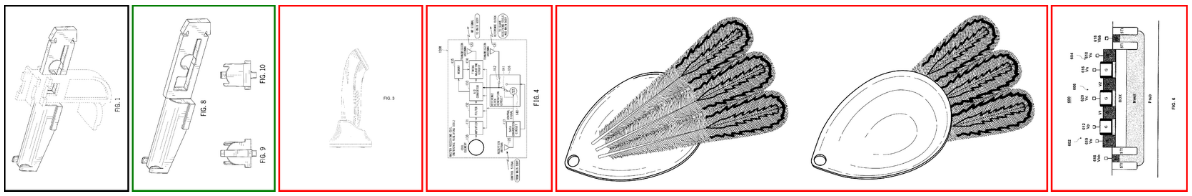
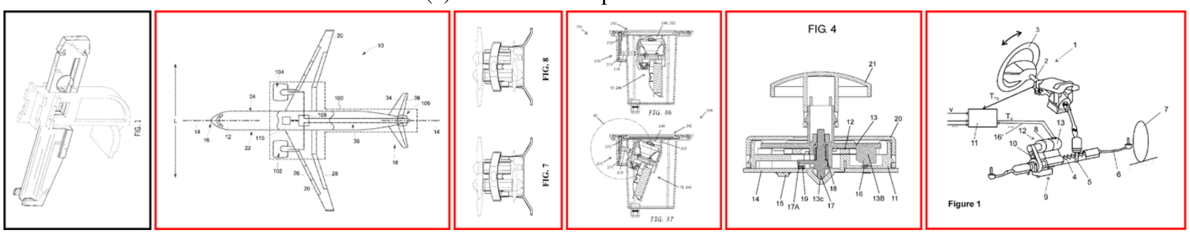


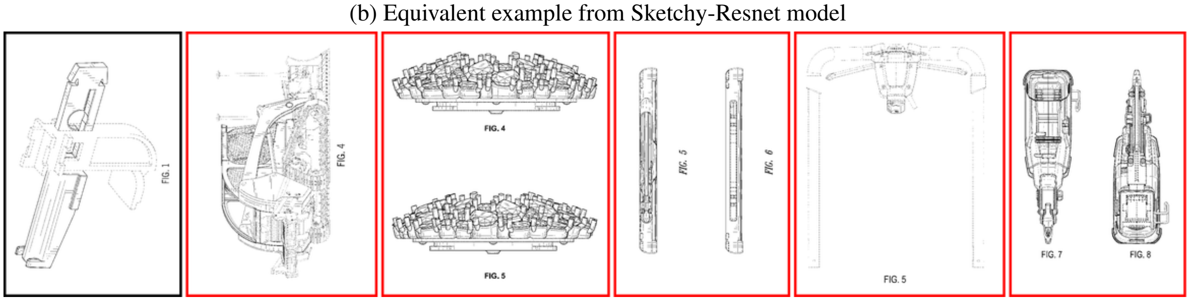
Figure S5: t-SNE embedding of PatentNet features for the test set images.



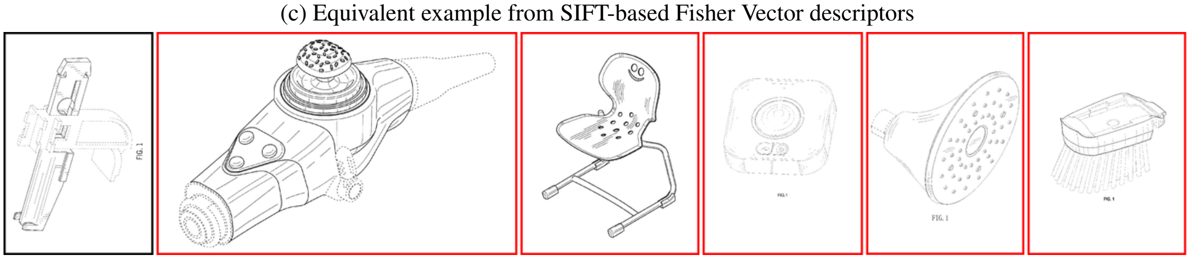
(a) Retrieval example from VisHash



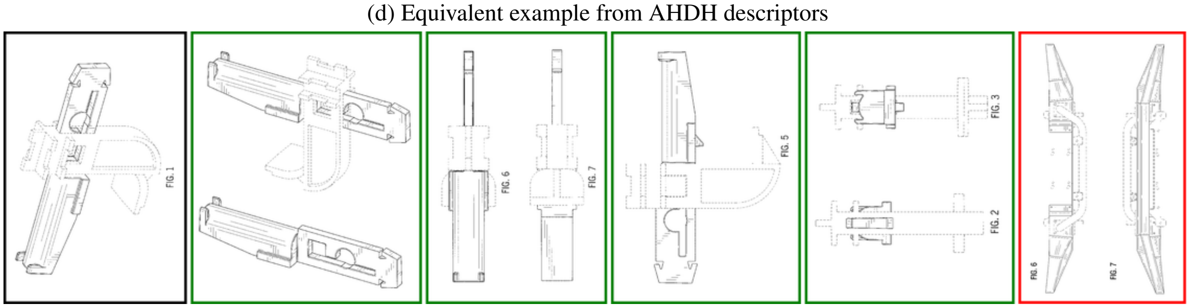
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors

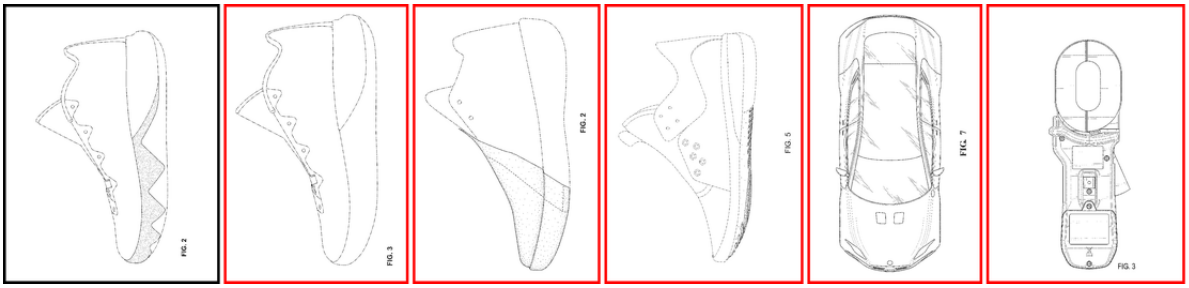


(d) Equivalent example from AHDH descriptors

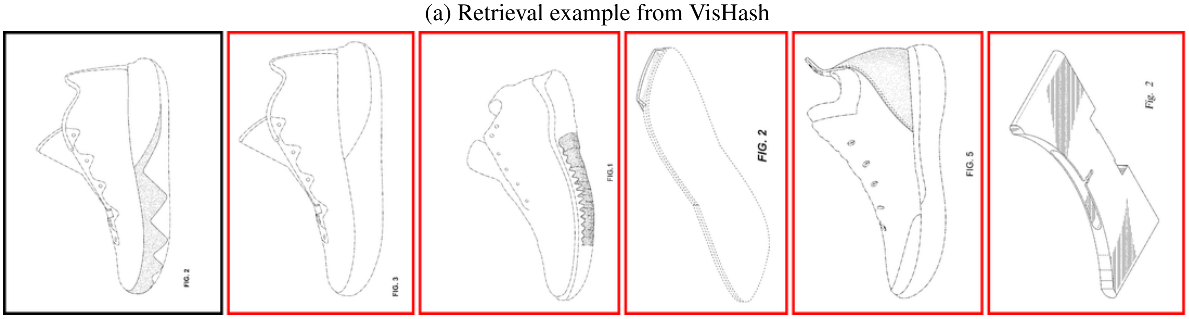


(e) Equivalent example from PatentNet

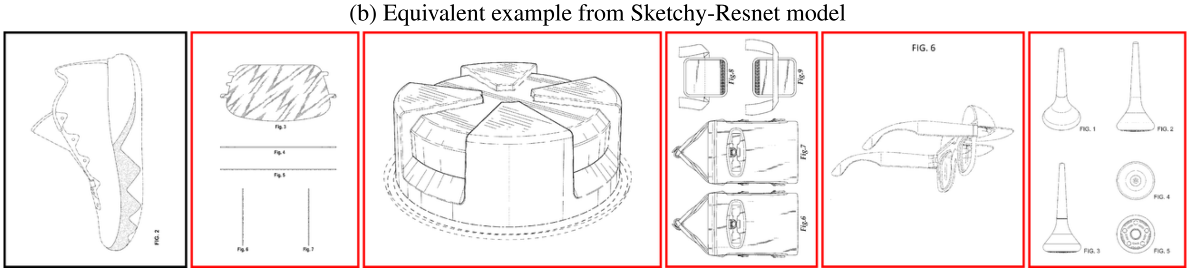
Figure S6: Qualitative comparison of examples from different retrieval models. Complex drawings, such as this query, are particularly challenging for all methods except PatentNet.



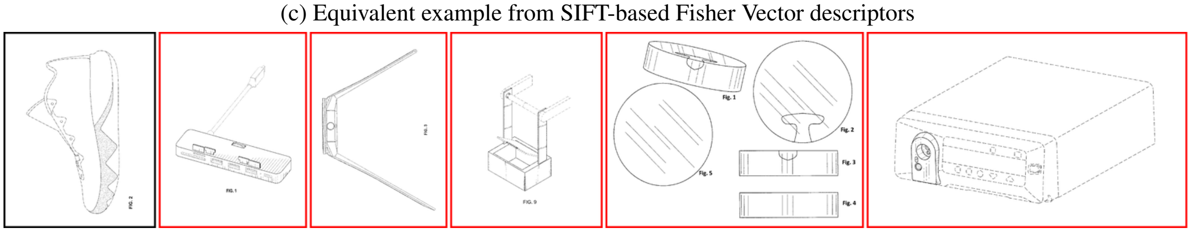
(a) Retrieval example from VisHash



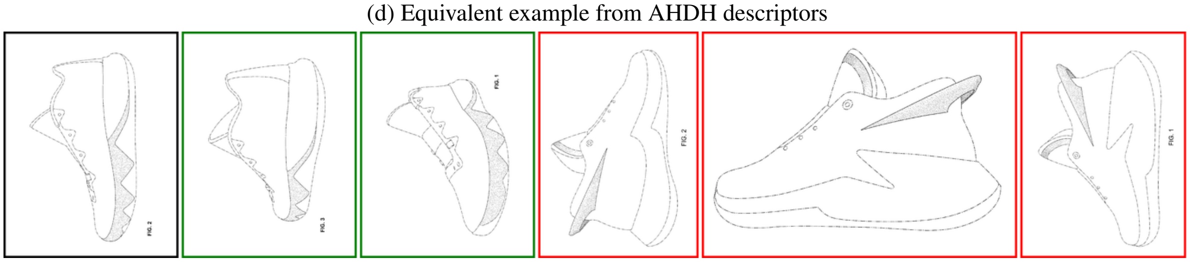
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors

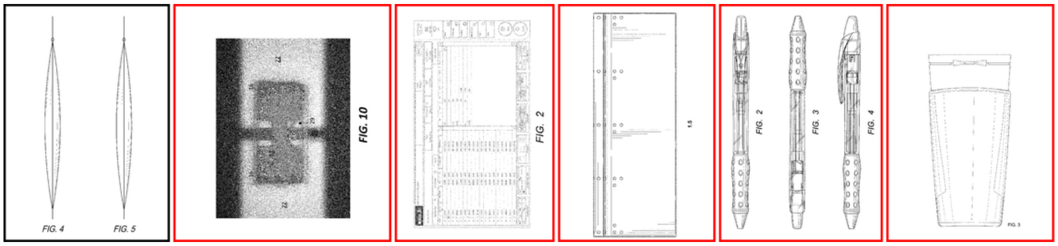


(d) Equivalent example from AHDH descriptors

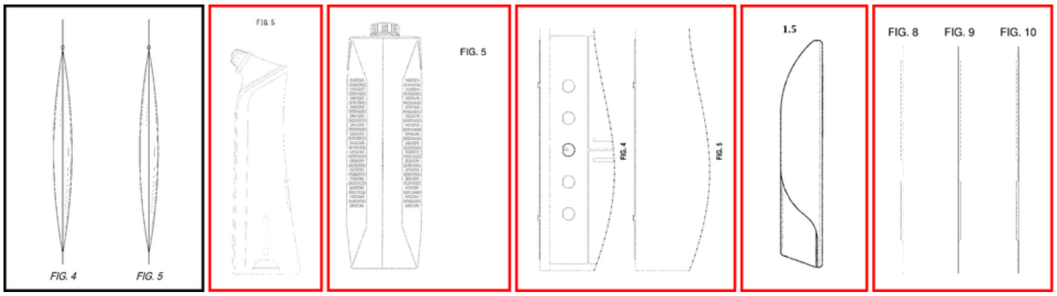


(e) Equivalent example from PatentNet

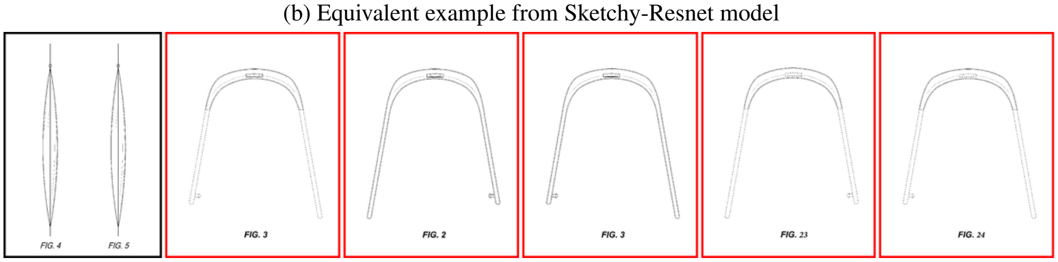
Figure S7: Qualitative comparison of examples from different retrieval models. Even though some of the methods (VisHash and Sketchy-Resnet) succeed in retrieving shoes from different patents, we can see that only the Patent net as able to retrieve different viewpoints from the same patent.



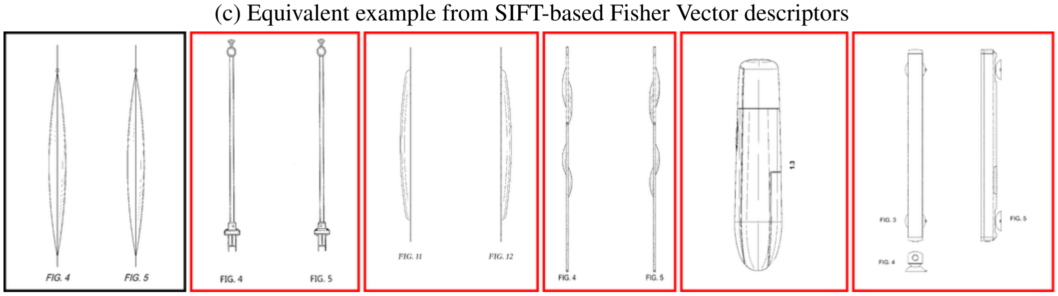
(a) Retrieval example from VisHash



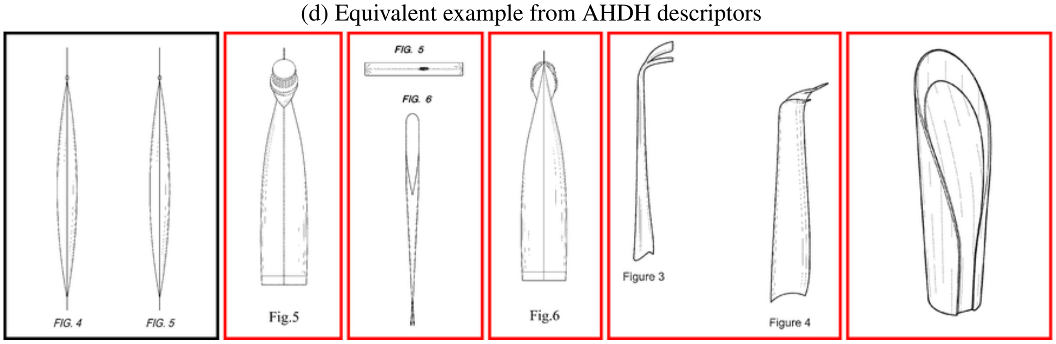
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors

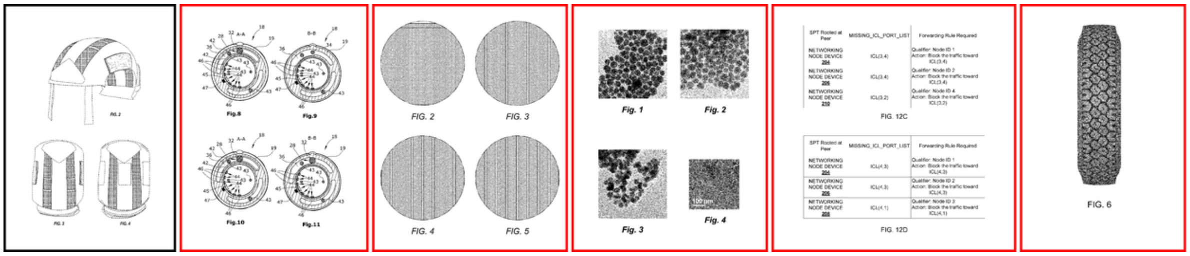


(d) Equivalent example from AHDH descriptors

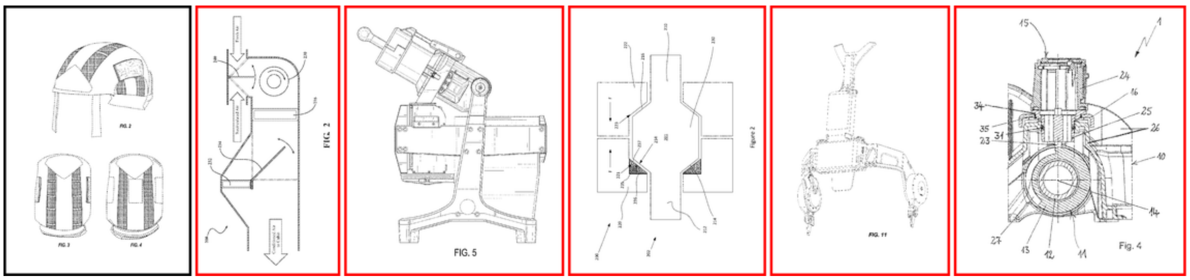


(e) Equivalent example from PatentNet

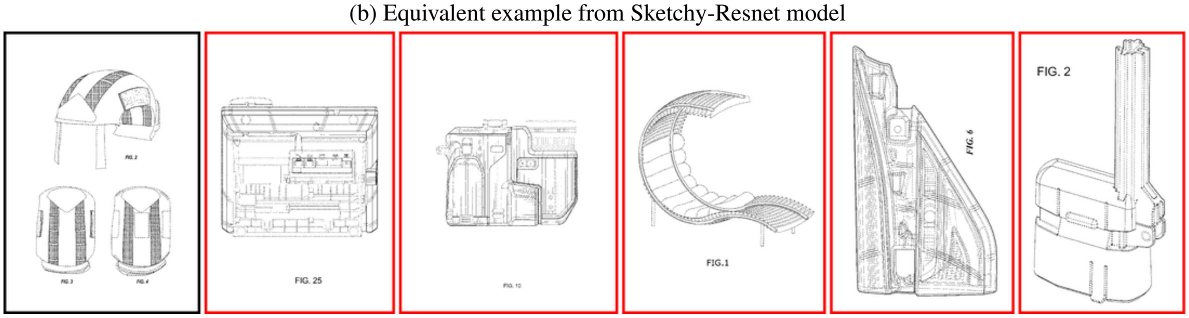
Figure S8: Qualitative comparison of examples from different retrieval models. This figure illustrates a failure case for all of the retrieval methods. Although we can see some interesting patterns in retrievals for PatentNet and AHDH. For AHDH, 4 out of the 5 top retrievals contain two objects similar to the query. In case of PatentNet, we see that the retrieved examples mimic the shape of the objects in the query.



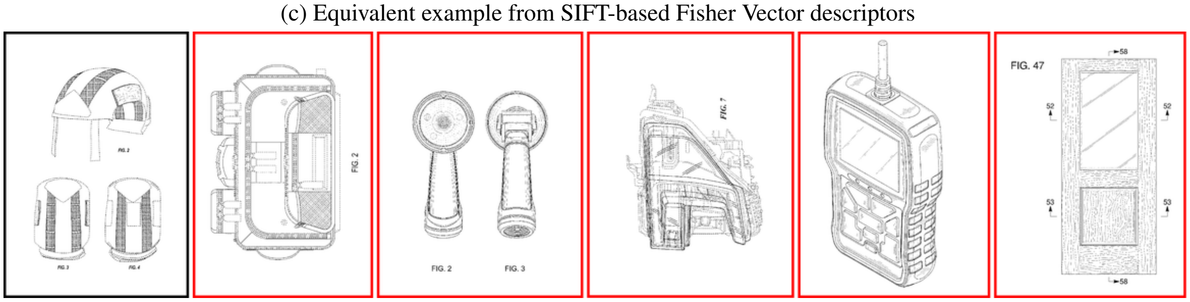
(a) Retrieval example from VisHash



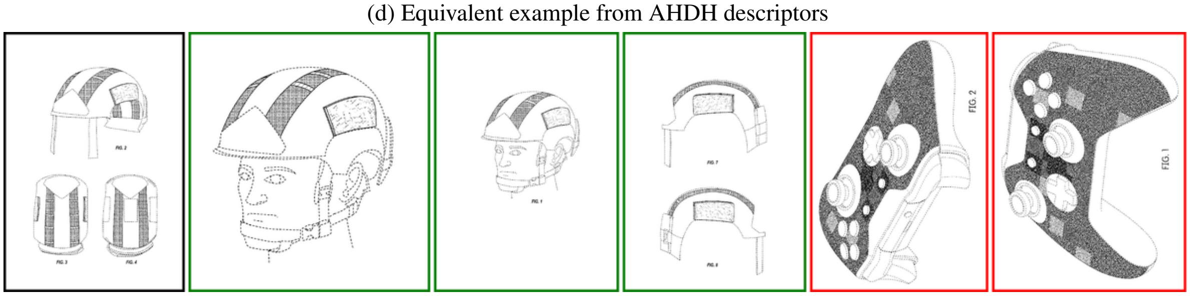
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors

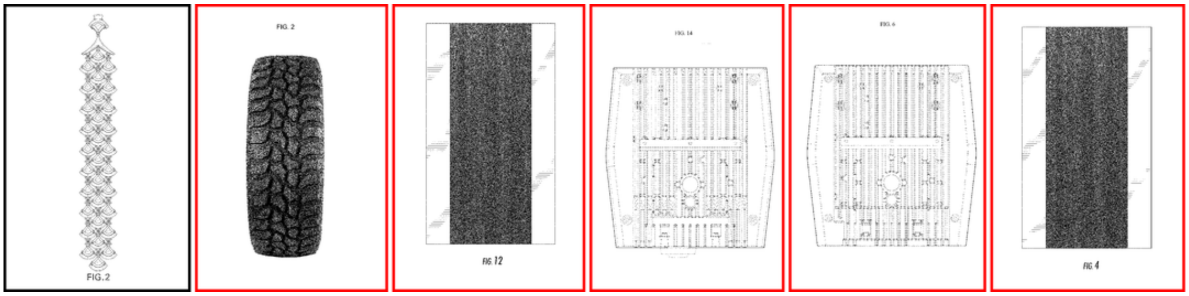


(d) Equivalent example from AHDH descriptors

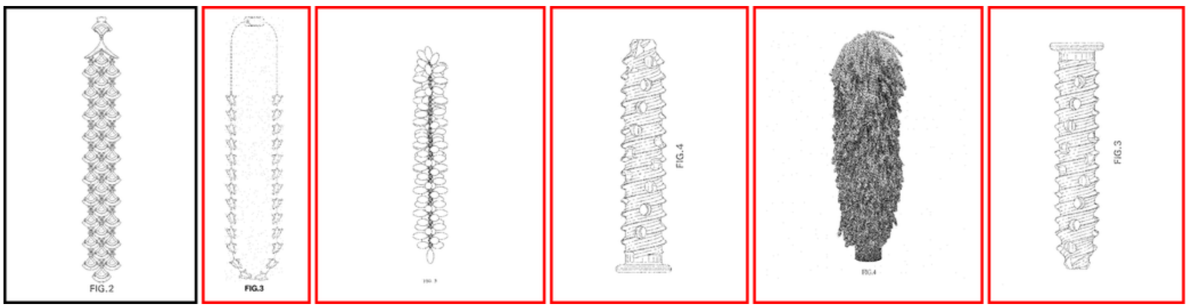


(e) Equivalent example from PatentNet

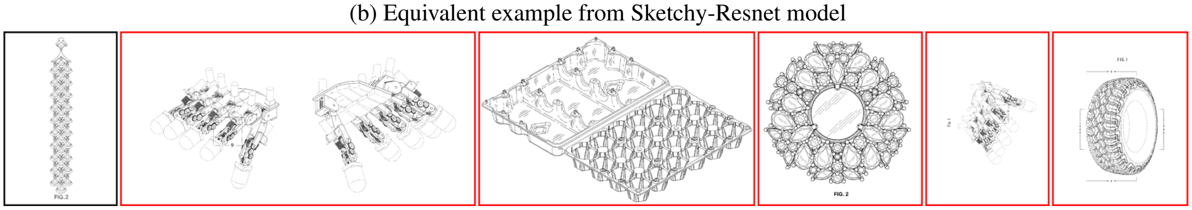
Figure S9: Qualitative comparison of examples from different retrieval models. In this example, we see that PatentNet is able to successfully retrieve drawings from the patent despite there being 3 different figures of a helmet within the query and the top two queries showing a person wearing a helmet. Interestingly, despite VisHash failing to retrieve the correct figures, we see it retrieving figures that have multiple objects in it.



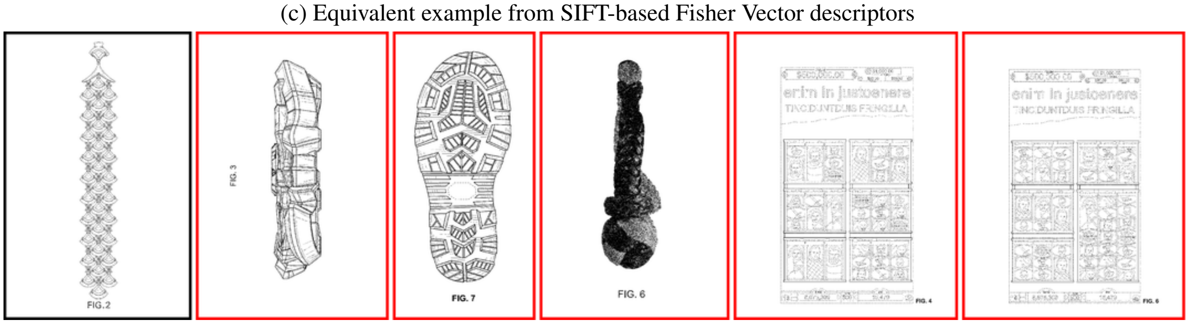
(a) Retrieval example from VisHash



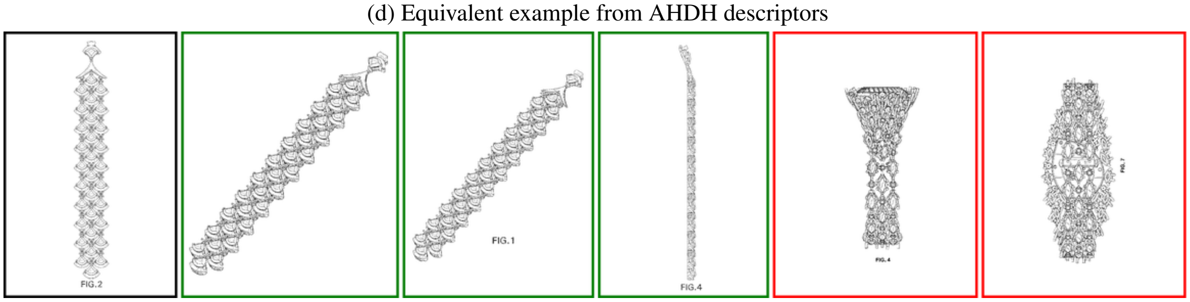
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors

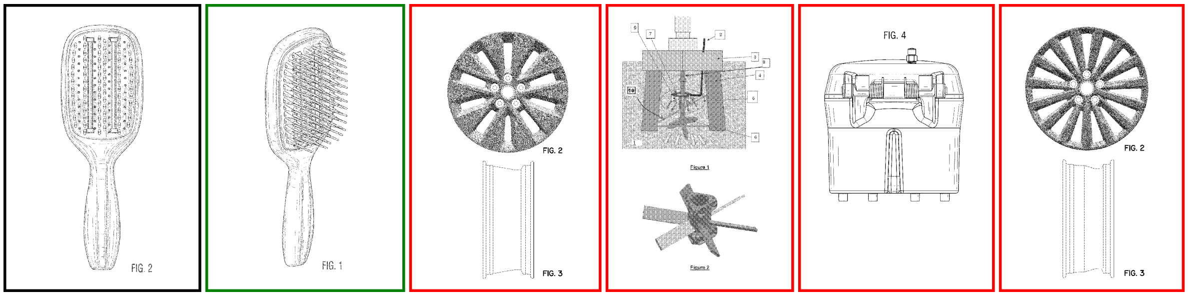


(d) Equivalent example from AHDH descriptors

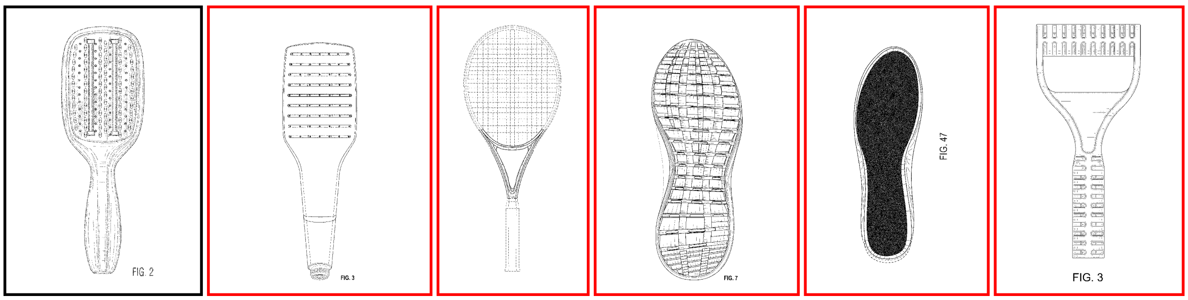


(e) Equivalent example from PatentNet

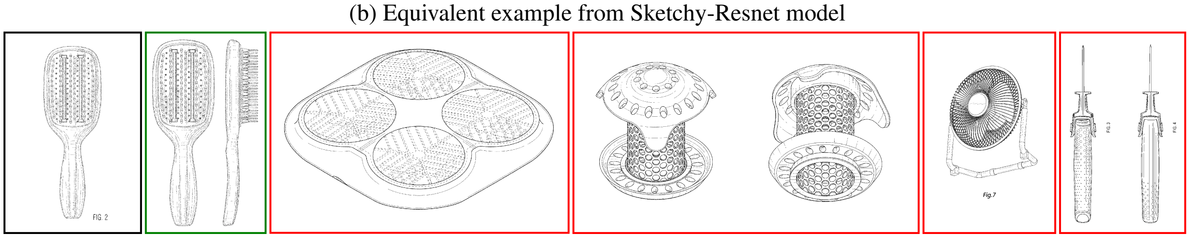
Figure S10: Qualitative comparison of examples from different retrieval models. Even though Sketchy-Resnet fails to retrieved the correct object drawings from the same patent, we see that objects retrieved are all a single elongated object.



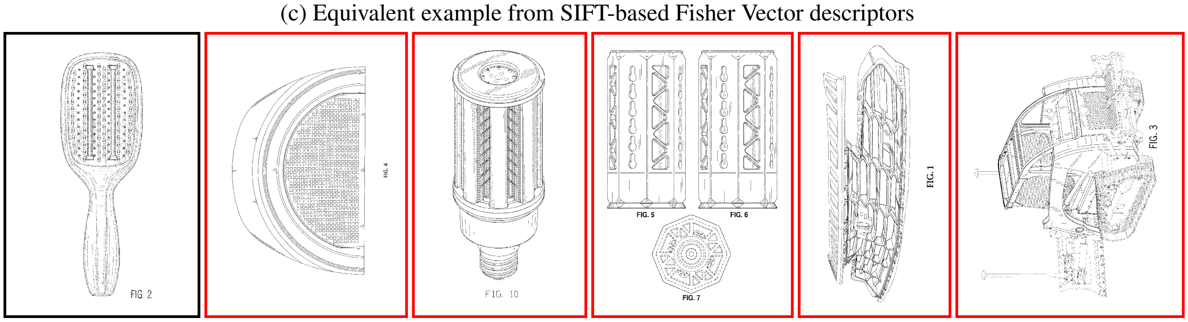
(a) Retrieval example from VisHash



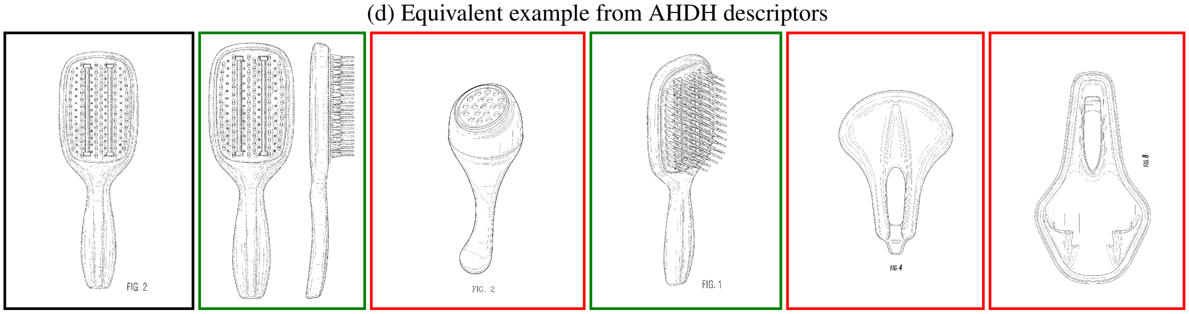
(b) Equivalent example from Sketchy-Resnet model



(c) Equivalent example from SIFT-based Fisher Vector descriptors



(d) Equivalent example from AHDH descriptors



(e) Equivalent example from PatentNet

Figure S11: Qualitative comparison of examples from different retrieval models.