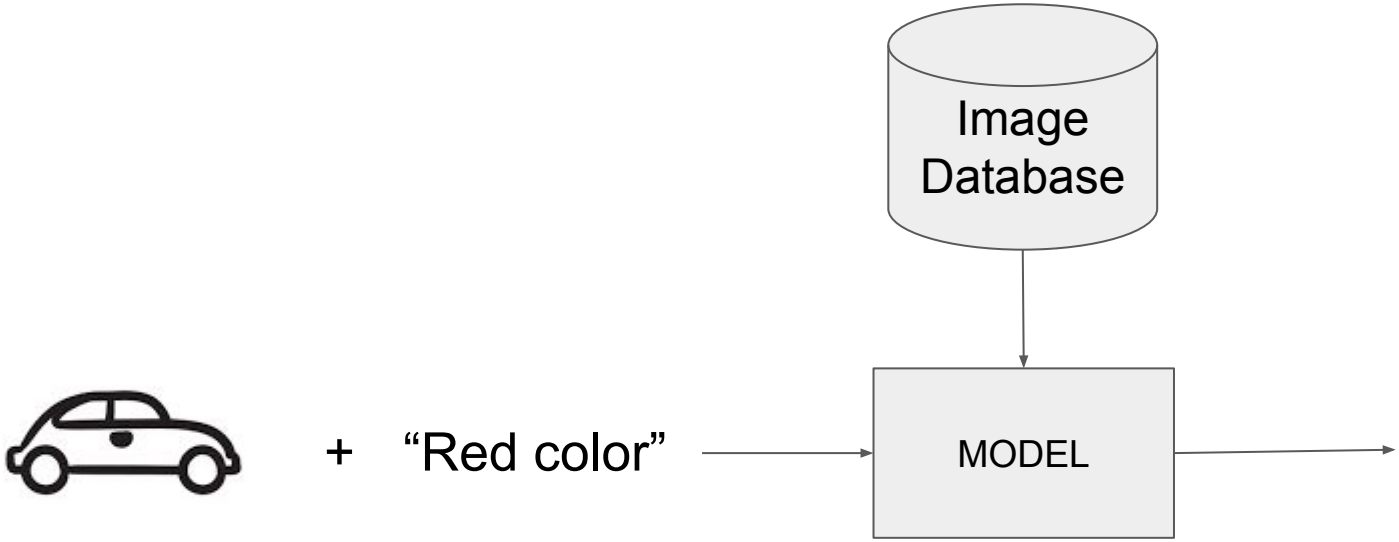
Image Retrieval using Sketch+Text description

Revant Teotia (rt2819)

Problem Statement



Query = Sketch + Text description

Retrieved images

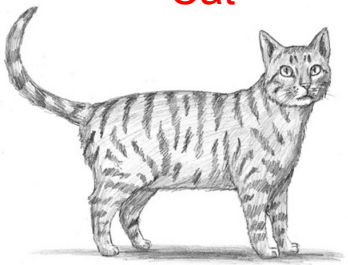
Task : Given a query consisting of a sketch and a text-description, retrieve images from an image database.

Dataset Creation: Method

Visual Genome annotations
Object: “Cat”
Attributes: “laying”, “striped”, “gray”

Query Generation

imagenet-sketch
“Cat”



+ laying striped gray

Query: sketch + text

Target Generation

Target: cropped image

Dataset Creation: Stats

- 20 kinds of object categories: car, cat, dog, umbrella, bench, bottle, …
  - 50 sketches per object category from Imagenet-Sketch
- Used 100k+ VG images and their attribute annotations for dataset creation
- Generated 90k query-target samples

85k training samples + 5k validation samples

Network Architecture

Image encoder (ResNet18)

Sketch feature

Positive image

Image encoder ResNet18

Pos-Image feature

Composition Module (simple Concat)

Comp. feature

Contrastive Learning (soft triplet loss)

"Red color"

Text Encoder (LSTM)

Text feature

Neg-Image feature

Negative image

Image encoder ResNet18

Results: Quantitative

Model	Recall@1	Recall@5	Recall@10	Recall@50	Recall@100
Concat	63.7%	93.8%	98.7%	100%	100%

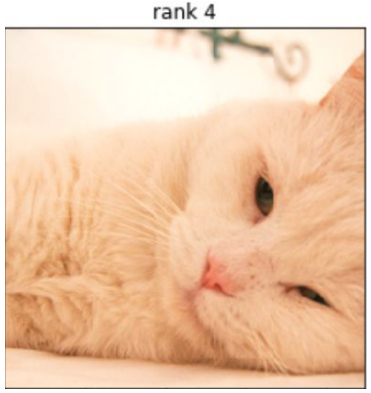
1K gallery

Model	Recall@1	Recall@5	Recall@10	Recall@50	Recall@100
Concat	35.52%	70.74%	82.66%	99.26%	99.98%

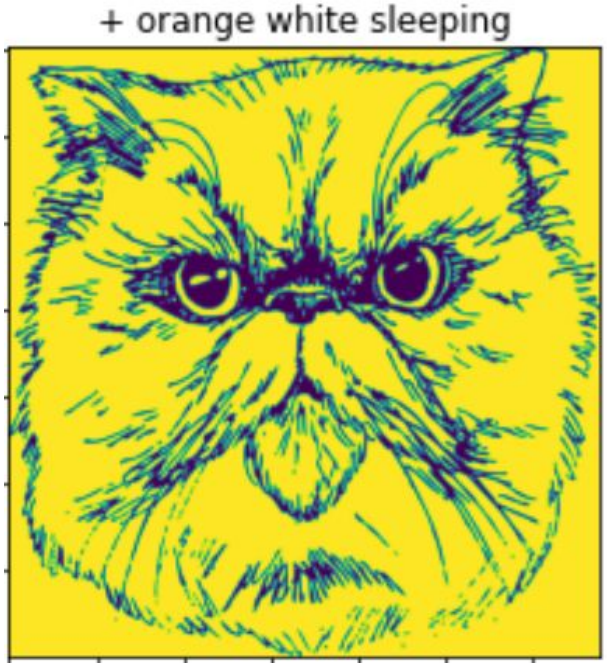
5K gallery

Results: Qualitative

Query



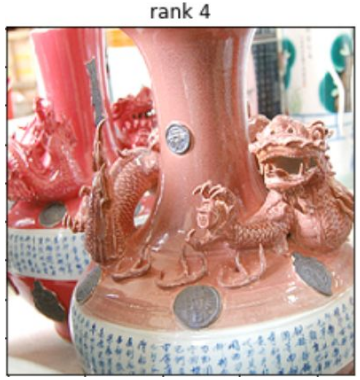
Results: Qualitative

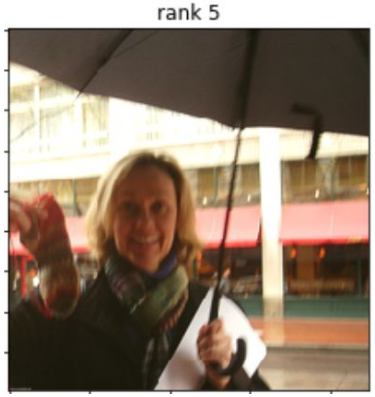


Query

Top retrieved images

Results: Qualitative

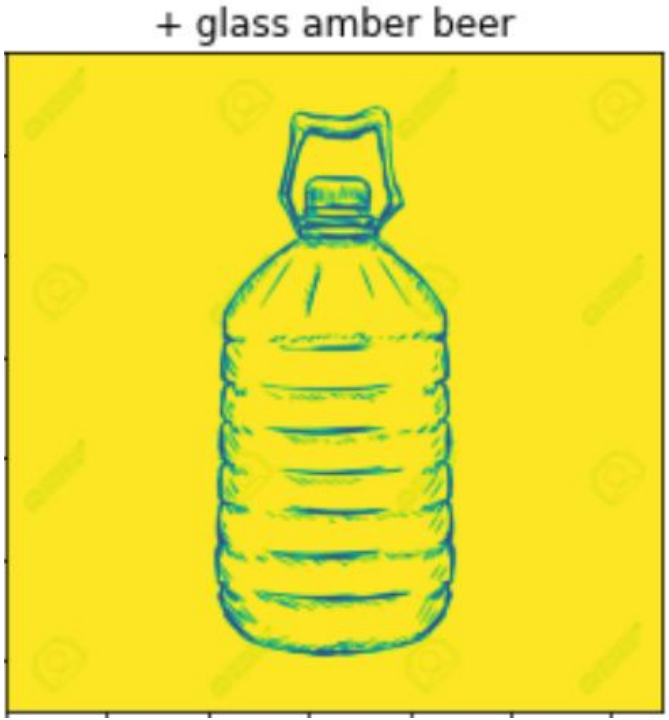




Query

Top retrieved images

Results: Qualitative



Query

Top retrieved images

Possible improvements

- Increasing dataset size/diversity: adding more object categories and more diverse kinds of sketches
- Using complex model: currently using only concat model. Can implement more complex sketch + text composition
- Using a different optimization: like hard triplet mining and SmoothAP loss (which is designed for retrieval tasks)

Thank you!

Code available: https://github.com/revantteotia/sketch-text-image-retrieval