Visual Search at Pinterest

Dmitry Kislyuk, David Liu, Andrew Zhai, Jiajing Xu, Yunsong Guo, Jeff Donahue, Yushi Jing

Visual Discovery, Pinterest

{dkislyuk,dliu,andrew,jiajing,yunsong,jdonahue,jing}@pinterest.com

Introduction

We present our visual search system developed and deployed at Pinterest, and share early results on user engagement from two applications, Related Pins and Similar Looks.

Our scalable and cost-effective visual search system indexes billions of Pinterest Pins. We extract visual features from Pin images using fine-tuned *AlexNet* and *VGG* models, using intermediate layer activations as feature representations for retrieval. Layer activation features have been shown to generalize well to a wide variety of computer vision tasks.

In our applications, we further augment these visual features with a rich set of **Pinterest metadata**, including user categorization of Pins (Travel, Food & Drink, Women's Fashion, Home Decor, DIY, ...), text annotations extracted by mining user captions, pin popularity, performance in search results, and the human-curated boards on which Pins are collected.

Peach System Overview

Peach is a large-scale distributed visual search system that provides real-time k-nearest neighbor lookup on visual features and reranks results using Pinterest metadata. It is built upon many open source tools and widely available platforms, such as Caffe, OpenCV, FLANN, Zookeeper, Thrift, and Amazon EC2.

The indexed image features and metadata are sharded onto many *Peach slices*. Each query gets sent to all the slices, and the collator collects the top results.

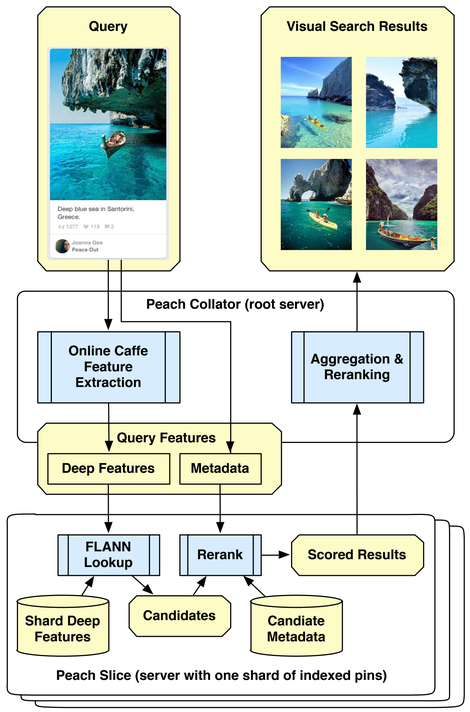

Figure 1: Illustration of how a query is processed in *Peach*.

Application 1: Related Pins

The Related Pins feature recommends Pins similar to the current Pin. Previously, Pinterest relied primarily on user-curated boards (collections of Pins) to generate Related Pins, using only metadata and not the visual features of the image. As a result, we did not have recommendations for 6% of Pins, typically those that were newly created or unpopular.

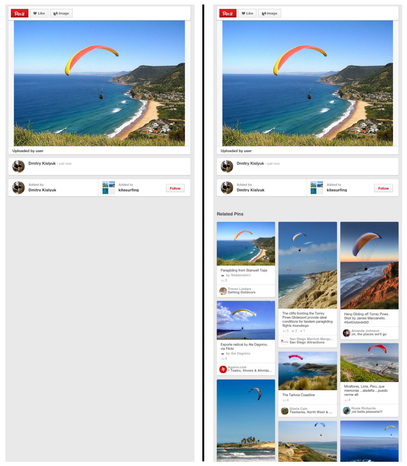

Figure 2: Incorporating Live Related Pins from Peach on a brand-new uploaded photo.

We used Peach to generate these missing recommendations in real time. The Peach index was populated with the 100 million most popular unique Pin images. By displaying visually similar pins for just these 6% of queries that would otherwise have empty recommendations, we observed a **2% increase** in *total* re-pin actions on Related Pins.

Furthermore, when we reranked all recommendations using deep CNN feature similarity, we achieved a **10% increase** in re-pin and click-through engagement on Related Pins in production. These increases were confirmed in internal evaluation by computing the NDCG score for the reranked related pins candidates and comparing against rankings derived from position-normalized user click data.

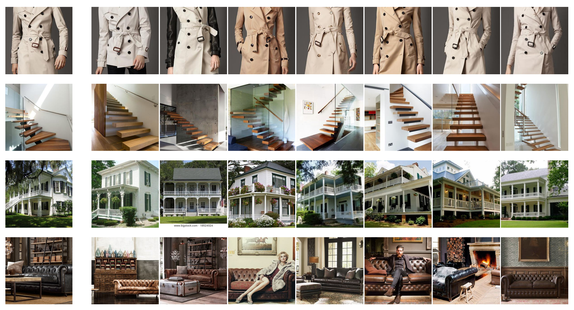

Figure 3: More visual search results used for Related Pins. In each row, the query image is shown, followed by the top results.

Application 2: Similar Looks

Using an offline object detection and localization pipeline, we built a Similar Looks product, which detected 80 million "clickable objects" in the fashion category on Pinterest.

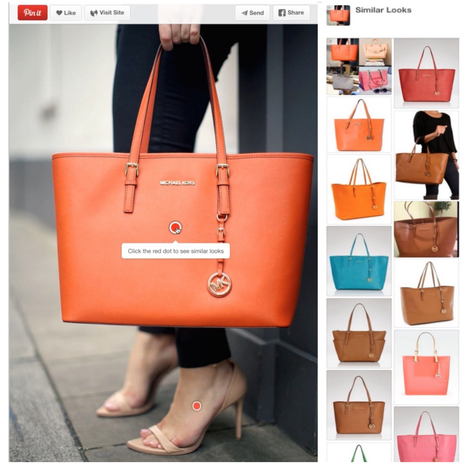

Figure 4: Similar Looks example: "object tags" are placed on each detected object; clicking a tag would show visual search results for that object.

Similar Looks uses a heavily optimized implementation of cascading deformable parts models [1], with text annotation matching to reduce the number of candidate images considered for each detector. This two-step process of text filtering followed by detection was critical: it dramatically reduced computational costs and lowered the false positive rate, as shown below.

	Text Filter		Object Detector		Combined	
Object Class	TP	FP	TP	FP	TP	FP
shoe	79.8	6.0	41.8	3.1	34.4	1.0
dress	75.5	6.2	58.8	12.3	47.0	2.0
glasses	75.2	18.8	63.0	0.4	50.0	0.2
bag	66.2	5.3	59.8	2.9	43.6	0.5
watch	55.6	6.0	66.7	0.5	41.7	0.0
pants	75.9	2.0	60.9	2.2	48.2	0.1
shorts	73.0	10.1	44.9	1.2	31.5	0.2
bikini	71.9	1.0	31.3	0.2	28.1	0.0
earrings	81.5	4.7	18.5	0.0	18.5	0.0
Average	72.7	6.7	49.5	2.5	38.1	0.5

Table 1: Object localization and classification accuracy (%) for Similar Looks evaluation.

In live experiments, the Similar Looks "tags" (red dots) had a click-through rate of 12%, although they decreased overall engagement with Pins. To instead test the relevance of visually similar results (independently of the bias resulting from introducing new UI for the object tags), we designed an experiment to blend Similar Looks results into Related Pins. For images containing detected visual objects, we found that adding Similar Looks results increased re-pin engagement by **5%**. This experiment was launched to production.

Offline Feature Evaluation

To evaluate different visual features for retrieval, we used the following evaluation dataset. We queried Pinterest's text search to retrieve 3,000 results for each of 1,000 top search terms on Pinterest, yielding about 1.6M unique images. We label each image with the queries that produced it, and images are considered relevant if they share a label.

We used features taken from several "Generic" (pre-trained for ILSVRC) and "Fine-Tuned" (on Pinterest annotation classification) models. The precision@k metrics for Peach retrieval on the evaluation dataset is shown below.

Model	p@5	p@10	latency
Generic AlexNet FC6	0.051	0.040	193ms
Pinterest AlexNet FC6	0.234	0.210	234ms
Generic GoogLeNet	0.223	0.202	1207ms
Generic VGG-16	0.302	0.269	642ms

Table 2: Precision of top retrieval results using various visual features.

We observed that VGG-16, even without any fine-tuning, can easily outperform both the GoogLeNet and fine-tuned AlexNet, although AlexNet is significantly faster.

Conclusions and Future Work

We present a visual search system built by a small team that indexes Web-scale image collections. We found it very effective to integrate the rich metadata available on Pinterest for more relevant Related Pins results as well as better object detection.

In recent efforts on our visual search system, we have moved away from manually trained per-class object detectors, and instead took advantage of the near-complete coverage for visually similar results offered by Peach (as well as better performing deep learning features). We are currently exploring generic region proposal algorithms for object localization.

Given the excellent performance of the pre-trained VGG-16 features, we are investigating fine-tuning models based on this architecture with Pinterest-specific data sets.

References

[1] P. F. Felzenszwalb, R. B. Girshick, and D. A. McAllester.
Cascade object detection with deformable part models.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2241–2248, 2010.

[2] M. Muja and D. G. Lowe.
Fast matching of binary features.
In *Proceedings of the Conference on Computer and Robot Vision (CRV)*, 12, pages 404–410, Washington, DC, USA, 2012. IEEE Computer Society.

[3] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman.
Return of the devil in the details: Delving deep into convolutional nets.
In *British Machine Vision Conference*, 2014.