Improving Image Search by Augmenting Image Queries

Conrad Burchert
Philipp Neubeck
Hanna Maria Pasula
Filip Pavetić

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Burchert, Conrad; Neubeck, Philipp; Pasula, Hanna Maria; and Pavetić, Filip, "Improving Image Search by Augmenting Image Queries", Technical Disclosure Commons, (March 01, 2021)
https://www.tdcommons.org/dpubs_series/4115

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
Improving Image Search by Augmenting Image Queries

ABSTRACT

Matching image or video content to other content is an important requirement for content hosting platforms. A common mechanism is to construct an index of known content, e.g., that include multi-dimensional embeddings generated from the content, and match new content against the index. The precision and recall of such techniques require a high quality fingerprint, and tradeoffs between recall performance and the cost of filtering out false positives. This disclosure describes improvements to content matching techniques that generate multiple transformations of the input content, look up each transformation in the index, and limit detection of false positives or other downstream analysis to content that has at least a threshold number of matches. Performance improvements in the recall vs. cost tradeoff are obtained due to the shape of the volume in the embedding space is no longer spherical, and instead, including many smaller spheres around the different transformed versions.

KEYWORDS

- Image search
- Image matching
- Video search
- Content match
- Content fingerprinting
- Query rewrite
- Nearest neighbor search
- k-approximate nearest neighbor search (kANN)
- Convolutional Neural Network (CNN)

BACKGROUND

For content platforms that host image or video data, detecting reuploads of known material is an important requirement. Such platforms typically implement technologies to match new content to determine if it matches prior content. Such technologies utilize indexes generated...
based on prior content, e.g., very large numbers of known video frames for a video
hosting/sharing platform. The indexes include multi-dimensional embeddings of the frames, that
are generated using machine learning models through application of techniques such as
convolutional neural networks. Such embeddings are typically invariant to transformations of the
input image, such as rotations, crops and other potentially adversarial transforms.

The indexes are usable to look up frames from newly uploaded content. Embeddings of
the probe image (e.g., newly uploaded content) are looked up in the indexes. Matches from the
index can be further verified using additional techniques such that false positives are filtered out.
The performance of such content matching techniques is measured based on its recall (ratio of
the detected matches to the true matches) and the cost of filtering out false positives (matches
that are incorrect). The radius around the searched embedding (e.g., which acts as a match
threshold) can be selected to optimize the tradeoff between the recall performance and the cost of
filtering out false positives.

Returning results from a larger radius around a looked up frame results in a higher cost of
processing them in downstream systems that filter out false positives, but also has a higher recall,
e.g., is more likely to detect transformed versions of the image. High performing matching
techniques would detect every instance of the probe image, including images with various
transformations applied, while not returning too many results that have a similar embedding, but
are otherwise unrelated (false positives). However, while improvements in the recall-cost trade-
off can be obtained by training a better fingerprint that is more resistant to transformations and
therefore, pulls transformed versions of the same image further together relative to distractors,
such approaches may be expensive and/or difficult to implement (e.g., may require reindexing all
content).
DESCRIPTION

This disclosure describes techniques to improve the performance of matching content detection algorithms. The techniques involve generation of transformations of the looked up image. These transformations are generated prior to creating the embeddings for the image that are looked up in an index.

![Diagram of process flow]

**Fig. 1: Improving content matching by using transformations of input image**

Fig. 1 illustrates an example process flow to improve content matching by using transformations of the input image (looked up image). Multiple transformations of the image that is to be looked up are generated (102). For example, such transformations can include application of crops, rotations, crops and rotations, etc.
Embeddings of the original image and each of the transformed images are created by application of a machine learning technique such as convolutional neural networks (104). Each generated embedding is independently looked up in the index that includes embeddings of existing content (106). After the look up, a union of each stored image that matches the original image or any of the transformations in the embedding space is obtained (108).

From the set of matched images, candidates where the number of successful matches does not meet a threshold are dropped (110). This set is further trimmed by applying filters based on metrics such as similarity and embedding magnitude (112).

The use of transformed versions of the input image in addition to the original image itself can improve recall and can reduce the number of false positives. This is analogous to query rewriting which involves creation of multiple versions of the query in text search. A key reason for the improved performance is that the shape of the volume that is being matched in the embedding space is no longer spherical. Instead, it includes many smaller spheres around the different transformed versions. This helps avoid matching distractors that are close to the original image being looked up but in a direction that does not correspond to a transformation that is being matched. Additionally, the knowledge about which of the transformed images were most similar to a match from the index can be used by downstream systems to further verify the match and weed out false positives.

The described technique involves fingerprinting more image material which adds to the computational cost. Hardware accelerators can be used to at least partially offset this cost. Furthermore, the increased workload for content matching scales with the number of looked up images and not with the index size.
CONCLUSION

This disclosure describes improvements to content matching techniques that generate multiple transformations of the input content, look up each transformation in the index, and limit detection of false positives or other downstream analysis to content that has at least a threshold number of matches. Performance improvements in the recall vs. cost tradeoff are obtained due to the shape of the volume in the embedding space is no longer spherical, and instead, including many smaller spheres around the different transformed versions.