15-769 Project Checkpoint
User-guided approach to content-based image retrieval

Thomas Kim (andrew ID tskim), Conglong Li (andrew ID conglonl)
$\{thomas.kim, conglonl\}$ $@cs.cmu.edu$

1 Introduction

We propose a system to allow users to provide multiple negative and positive image queries in order to retrieve images with similar content. One important application of our system is to allow users to search through their personal photo collections with content-based queries, and as such we will target datasets consisting of a reasonable number of images for a personal collection (1,000,000 images/ 1tb). We plan to apply modern machine learning techniques in the context of computer vision to compactly express the content and features of the images, then leverage this representation to provide efficient querying of the data set.

Evaluation of content-based image retrieval is qualitative, making it difficult to evaluate the success of this system. As such, the evaluation will be closely tied to qualitative analysis through user studies. Performance will be evaluated against other state of the art content-based image retrieval systems.

Content-based image retrieval allows retrieval of multimedia data based on the content of the image as opposed to metadata such as text annotations. In many cases, such annotations are absent or incomplete, necessitating this content-based approach to improve the accuracy and completeness of search results.

2 Related Work

2.1 Content-Based Image Retrieval

Content-based image retrieval aims at searching for similar images through the analysis of image content. As DNNs learn rich mid-level image descriptors, ImageNet uses the feature vectors from the 7th layer in image retrieval and demonstrated outstanding performance [4]. Babenko *et al.* proposed to compress the DNN features using PAC and discriminative dimensionality reduction to improve the efficiency [1].

Due to the recent growth of visual contents, rapid search in a large database becomes an emerging need. In stead of linear search which has high computational cost, a practical strategy is to use the technique of Approximate Nearest Neighbor (ANN) or hashing based method for speed up [3, 9, 5, 8, 7, 10]. These methods project the high-dimensional features to a lower dimensional space, and then generate the compact binary codes. Benefiting from the produced binary codes, fast image search can be carried out via binary pattern matching or Hamming distance measurement. However, these methods require to use similarity matrix to describe the relationship of the image pairs, which is no practical for a large-scale dataset. Recently Lin *et al.* propose an effective deep learning framework to generate binary hash codes and image representations in a point-wised manner, making it suitable for large-scale datasets [6].

2.2 Image Analysis Applications

AverageExplorer is an interactive framework that allows an user to rapidly explore and visualize a large image collection using the medium of average images [11]. This real-time system provides a way to summarize large amounts of visual data by weighted averages of an image collection, with the weights reflecting user-indicated importance. Zhu *et al.* propose a way to help user-controlled realistic image manipulation by learning the manifold of natural images and defining the image editing operations with constraints lie on the learned manifold at all times. Thus the model automatically adjusts the output keeping all edits as realistic as possible [12].

3 Starter code

We are using starter code for the frontend forked from https://github.com/fancyspeed/py-cbir. We are using tensorflow for the machine learning portions of this project. All of this starter code was acquired and is running and has been modified.

4 Datasets

We are using the ImageNet dataset (140gb) [2]. This dataset has been downloaded to our local machine. ImageNet is a good dataset to use because it is organized and labeled, allowing for fast automatic evaluation of our system. It has about 1 million images, which is a reasonable upper limit for typical personal photograph collections. For the first two goals, we will use a relatively small subset of the ImageNet data set in order to better evaluate the quality of the algorithmic side of this project, and for the last goal we will be using a more substantial subset of ImageNet. We are using the Inception v3 neural network pre-trained on the imagenet dataset (2012).

5 Evaluation

Our goals have somewhat changed, and we are currently working on step 2 of exploring how different methods of calculating similarity between images yield better results. As a result, we are mostly looking at qualitative data rather than performance numbers. We built a simple automatic evaluation framework to score our content-based image retrieval system. Additionally, we inspect the returned images manually. Our final evaluation criteria will probably be some mix of the two.

6 Evaluation Plan and Goals

Our first goal is to take the work done by Lin et al. on representing images as binary strings based on neural network activations, and build a system that takes a simple brute force approach to evaluating similarity by L1 or L2 euclidean distance between the query and the dataset images. **We have already completed this goal.**

Our next step is to explore different methods of calculating similarity between images. For example, we hope to use a 2-3 layer CNN consisting of a ReLU layer sandwiched between fully connected layers, training briefly on the set of input images then classifying images in the database using this CNN. We hope that this approach yields higher accuracy. Running a neural network will be a fairly expensive operation compared to a simple L1/L2 euclidean distance calculation, so it is likely we will need to modify our approach to making this efficient. At the moment it is not clear what this will entail, but our hope is that it will become apparent after analyzing the performance characteristics of the system. **We are working on this goal, and plan to complete it by the week ending Dec 2nd**

Our final stretch goal is to scale this system to larger data sets on the order of hundreds of gigabytes. modifying it to be performant on these larger data sets. To achieve this goal, we will run our system on a larger dataset and work to determine and alleviate performance bottlenecks. **This goal is somewhat in parallel with the second goal, so we hope to complete it by the week ending Dec 2nd or 9th.**

7 Deliverables/Outcomes

We hope to show that machine-learning based approaches to content based image retrieval for multiple query images can be fast and yield qualitatively good results. The reason why this project is challenging is because there is no clear immediate solution to the multi-image query problem, and performing queries on large datasets efficiently is difficult. We will demonstrate success/failure by measuring the end-to-end latency between a user request and a returned result, and by measuring the qualitative "goodness" of the returned result. We will show a demo.

References

[1] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky. Neural codes for image retrieval. In *European Conference on Computer Vision*, pages 584–599. Springer, 2014.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[3] A. Gionis, P. Indyk, R. Motwani, et al. Similarity search in high dimensions via hashing. In *VLDB*, volume 99, pages 518–529, 1999.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] B. Kulis and T. Darrell. Learning to hash with binary reconstructive embeddings. In *Advances in neural information processing systems*, pages 1042–1050, 2009.

[6] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen. Deep learning of binary hash codes for fast image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 27–35, 2015.

[7] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang. Supervised hashing with kernels. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2074–2081. IEEE, 2012.

[8] M. Norouzi and D. M. Blei. Minimal loss hashing for compact binary codes. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 353–360, 2011.

[9] Y. Weiss, A. Torralba, and R. Fergus. Spectral hashing. In *Advances in neural information processing systems*, pages 1753–1760, 2009.

[10] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan. Supervised hashing for image retrieval via image representation learning. In *AAAI*, volume 1, page 2, 2014.

[11] J.-Y. Zhu, Y. J. Lee, and A. A. Efros. Averageexplorer: Interactive exploration and alignment of visual data collections. *ACM Transactions on Graphics (TOG)*, 33(4):160, 2014.

[12] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*, pages 597–613. Springer, 2016.