석 사 학 위 논 문
Master's Thesis

포함 관계와 순위 기반 투표를 활용한 정밀 이미지 검색

Accurate Image Search
using Rank-based Voting with Inclusion Relationship

2017

조 재 형 (趙 在 亨 Cho, Jaehyeong)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

포함 관계와 순위 기반 투표를 활용한 정밀 이미지 검색

2017

조 재 형

한 국 과 학 기 술 원

전산학부

포함 관계와 순위 기반 투표를 활용한 정밀 이미지 검색

조 재 형

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2016년 12월 19일

심사위원장 윤 성 의 (인)

심 사 위 원 김 민 혁 (인)

심 사 위 원 박 종 철 (인)

Accurate Image Search using Rank-based Voting with Inclusion Relationship

Jaehyeong Cho

Advisor: Sung-eui Yoon

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Computer Science

Daejeon, Korea
December 19, 2016

Approved by

Sung-eui Yoon
Professor of School of Computing

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MCS
20153607

조재형. 포함 관계와 순위 기반 투표를 활용한 정밀 이미지 검색. 전산학부 . 2017년. 23+ii 쪽. 지도교수: 윤성의. (영문 논문)
Jaehyeong Cho. Accurate Image Search using Rank-based Voting with Inclusion Relationship. School of Computing . 2017. 23+ii pages. Advisor: Sung-eui Yoon. (Text in English)

초 록

본 논문에서는 포함 관계와 순위 기반 투표를 활용하여 높은 품질의 이미지 검색 정확도를 달성하는 기술을 선보인다. 이미지는 하나 이상의 관심 영역을 가질 수 있기 때문에, 이미지 검색 문제에 적합하도록 조정한 최신 물체 검출 기술을 이용하여 이미지로부터 대표적인 물체 영역들을 검출한다. 그 후, 검출한 대표 영역들로부터 지역적으로 CNN 기술자를 추출하며 그와 동시에 영역들 사이의 포함 관계를 확인한다. 질의 이미지와 비슷한 이미지들을 찾기 위해, 본 논문에서는 대표 영역들과 포함 관계에 기반한 새로운 유사도 측정 방식을 제안한다. 본 논문에서 제안하는 유사도 측정 방식은 유사한 물체 영역이 유사한 공간적 관계까지 가질 경우, 높은 유사도를 부여한다. 제안하는 방식의 검증을 위하여 세 종류의 이미지 검색 표준 데이터셋을 활용하며, 검색 정확도를 CNN을 이용하는 다른 최신 이미지 검색 기술들과 비교한다. 실험을 통해 본 기술의 효용성과 견고성을 입증하며, 또한 다른 최신 기술들과 함께 사용될 시 더 높은 성능을 보일 잠재성을 제시한다.

핵 심 낱 말 이미지 검색, 물체 검출, CNN, 이미지 유사도, 공간 관계

Abstract

We present a rank-based voting technique utilizing inclusion relationship for high quality image search. Since images can have multiple regions of interest, we extract representative object regions using a state-of-the-art region proposal method tailored for our search problem. We then extract CNN features locally from those representative regions and identify inclusion relationship between those regions. To identify similar images given a query, we propose a novel similarity measure based on representative regions and their inclusion relationship. Our similarity measure gives a high score to a pair of images that contain similar object regions with similar spatial arrangement. To verify benefits of our method, we test our method in three standard benchmarks and compare it against the state-of-the-art image search methods using CNN features. Our experiment results demonstrate effectiveness and robustness of proposed algorithm, and also show the potential to further improve by cooperating with other state-of-the-art methods.

Keywords Image search, object detection, CNN, image similarity, spatial relationship

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Image search is a task to identify visually similar images from an image database given a query image, by identifying matching features. This is one of the most fundamental tools in many image processing applications, since the performance of many high-level problems often depends on the quality of image search.

Many interesting and novel data-driven applications have been proposed thanks to the advance of image search and other related techniques. Some of well-known applications include photo tourism [39] and scene completion [14]. Furthermore, the data-driven approach keeps to widen its scope into many other directions [38, 27]. The final quality of these data-driven approaches are affected significantly by the search accuracy, since they commonly assume that identified images share similar patches and objects. As a result, we are interested mainly in improving the accuracy of image search in this paper.

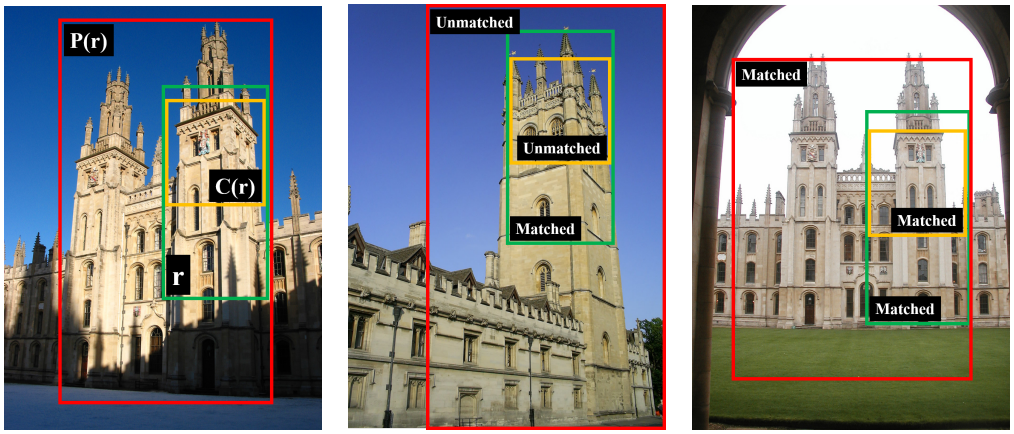
Image search starts from representing images with feature vectors, which have been investigated extensively to develop powerful representation methods. Recently, deep convolutional neural networks (CNNs) [28, 41, 37, 17] demonstrate outstanding classification performance, and the CNN features obtained from a few hidden layers present great generalizability to many other domains or tasks [3, 35, 12, 47]. Such encouraging results are available because the CNN features trained on a large-scale image dataset are sufficiently discriminative and representative in many existing problems.

Recent image search techniques often rely on CNN features to improve performance. A naïve integration of global CNN features demonstrated better performance [3] than the algorithms based on existing hand-crafted image descriptors such as VLAD, Fisher vectors, and triangulation embedding [22, 34, 24]. Recent approaches improve accuracy further by incorporating spatial information of images. Babenko *et al.* [2] regarded CNN features extracted from a convolutional layer as local descriptors, and aggregated those features to encode spatial information in a similar manner of VLAD [22]. On the other hand, Razavian *et al.*[35] extracted CNN features from sub-patches distributed uniformly in an image, in order to consider the spatial information of an image.

While recent CNN-based image search methods utilize spatial information, they still rely only on weak information compared to techniques developed for object detection and localization. For example, object detection techniques typically employ region proposal methods to identify candidate regions and improve detection accuracy [12]. This paper discusses how to utilize such candidate regions for image search. One of technical challenges is that most of those candidate regions may not contain any meaningful objects since many region proposal techniques have been designed to achieve a high recall. These false-positive candidate regions deteriorate accuracy of image search.

Main contributions. We propose a novel, rank-based voting image search technique, which identifies representative object regions and retrieves similar images based on those regions. Our method identifies such representative regions by utilizing a state-of-the-art region proposal method, while adjusting the objectness measure to be invariant to object sizes (Sec. 3.1). To achieve a high matching quality between regions of images, we introduce a novel similarity measurement method based on a voting scheme, which considers spatial relationship between regions, especially, inclusion relationship (Sec. 3.3). Using the spatial relationship, we can cull out incorrect matching results, resulting in substantial accuracy improvement as we utilize more representative regions in images (Fig. 1.1).

To verify the effectiveness of our approach, we evaluate our method over three standard benchmark



(a) A query image (b) A similar, but incorrect image (c) A correct tested image

Figure 1.1: Our method extracts representative regions from images and identify similar images by matching those regions between the query image and database images. We also improve matching accuracy by considering the inclusion relationship between regions. (a) shows related regions given a query region r shown in the green box. Yellow and red boxes represent child $C(r)$ and parent $P(r)$ regions of the query region, respectively; i.e., $C(r) \subseteq r \subseteq P(r)$. (b) The region r of the query image is matched with a region of this test image. Nonetheless, their related regions do not match. As a result, we can identify that this image is different from the query image. (c) Most of related regions of the query image match with those of this tested image, and our relationship score gives a high weight for this case, resulting in matching with the query image.

datasets [19, 33, 32] and compare it against state-of-the-art image search methods also utilizing CNNs. Overall, our method achieves higher search accuracy across all the three benchmarks compared to the tested methods with same network. Such successful results are attributed to the accurate identification of representative regions and reliable similarity measurement between regions. This paper was submitted to Eurographics 2017 and it is on the review process now.

Chapter 2. Related Works

In this section, we discuss prior methods that are directly related to our work.

2.1 Convolutional Neural Networks and Image Search

Deep convolutional neural networks (CNNs) have achieved remarkable performance improvement in various recognition tasks. Popular applications include image classification [28, 41, 37, 17], object detection and localization [12, 13, 36], semantic segmentation [30, 31, 5], and many others. The main reason to make CNNs successful is their representation power; CNNs capture high-level semantic information in images much more accurately [3, 35, 12] than traditional hand-crafted feature descriptors such as SIFT [29], HOG [8], and so on. Moreover, the CNNs pre-trained on a large-scale image datasets are often directly applicable to other domains [7] or tasks [42], and their performance is even boosted by fine-tuning existing models with additional data.

There are several existing works in image retrieval based on CNN representations. Babenko *et al.* [3] used activations of top three fully connected layers as global descriptors of the input image. They compared the characteristics of the representations depending on their levels of abstraction in image retrieval context. Also, they showed that the features can become more discriminative for particular datasets by retraining the network with specific purpose. Some recent papers proposed a feature aggregation approach to generate improved global descriptors [2, 26, 43, 11]. These ideas are similar to methods designed for dense-SIFTs, *i.e.,* VLAD [22], Fisher vectors [34], and triangular embedding [24]. In an aggregation step, SPoC [2] gives larger weights to features near the center as a simple weighting heuristic, and CroW [26] gives different weights according to the location and channel of each feature. Similarly, R-MAC [43] aggregates region feature vectors, which are generated by collecting the maximum activation of each region.

Gordo *et al.* [11] computes CNN features by leveraging a three-stream Siamese network with a triplet ranking loss, while many existing methods employ CNN features with image classification objectives. This method also suggests an aggregated feature among multiples features by taking the maximum activations.

Departing from this aggregation approach, our method extracts features from multiple regions and utilizes them for achieving high search accuracy.

2.2 Region Detection

It has been widely accepted to extract and use discriminative regions from images for achieving a high search accuracy. Since it is not feasible to employ the naïve sliding window strategy based on CNNs due to a high computational cost, object proposal detection techniques [6, 44, 25] have been studied to identify the regions of interest very efficiently.

One of famous object proposal detection methods is selective search [44], which constructs object proposals from the hierarchical bottom-up image segmentation. Krähenbühl *et al.* [25] introduced geodesic object proposals, where the seeds for objects are selected by minimizing the geodesic distance with all the objects in order to use only a small number of reliable seeds. Cheng *et al.* [6] observed that

various sizes of objects have the correlation with norms of gradients, when their patches are resized to the 8 by 8 fixed resolution. To utilize this observation, a linear model of 64 dimension is trained for objects using SVM, and used to estimate the objectness of each region.

Recently, neural networks considering regions have been proposed too. Faster R-CNN applies the regression to predict box coordinates, while considering multiple anchors [36]. Also, inside-outside net [4] considered context information for regions. Our proposed method can be combined with these recent region network for achieving higher search accuracy.

These object proposal methods are useful to provide object candidates for detection and localization algorithms, since they are optimized to maximize recalls, and this property is also important for image retrieval problems although they need to improve precisions at the same time. Our proposed method uses a region proposal method and improve precision by considering the spatial relationship among regions.

2.3 Similarity Measure

It is also important to have a similarity measure that gives a small distance to a pair of similar images. For this purpose, it is common to use simple vector-to-vector distances such as *L1*, *L2*, and *cosine* metrics.

Some techniques extracted multiple features like our approach and thus considered similarity measures for these features. Razavian *et al.* [35] extracted features from uniformly generated sub-patches, and measured similarity by averaging the minimum *L2* distance between sub-patches of each image. Recently, Xie *et al.* [46] extracted multiple features from their manually defined objects, and applied Naive-Bayes Nearest Neighbor (NBNN) search in order to utilize semantic category information of images. The NBNN search requires categorized feature sets in its process, so this approach also needs additional categorized image dataset for its similarity measure.

Our method utilizes rank-based voting scheme which directly considers the spatial relationship between regions, and thus improves the accuracy of search method.

Chapter 3. Our Approach

In this section, we describe our method that extracts multiple regions from an image and utilize the relationship among those regions. At a high level, our approach is divided into four parts: representative region selection, feature extraction, similarity measurement, and compression/indexing parts. We first explain how to extract regions from an image.

3.1 Representative region selection

The state-of-the-art object localization methods show impressive recalls for detecting objects by employing region proposal techniques [6, 44, 25]. Although these approaches are useful to improve recalls in object detection, it is inappropriate to directly use such region proposals to our problem of image search. While localization methods demonstrate high recalls, there are still a large number of unimportant regions among the candidate proposals as shown in Table 3.1. It is thus unlikely to achieve high search accuracy based on representations involving such unimportant regions.

To address the aforementioned problems, we decide to leverage only top-k confident regions. Many effective confidence measuring methods make it possible to determine the ranking of region proposals (e.g., generic objectness measure considering different image cues [1] and aggregating pixel-saliency [45]). Thankfully, among the state-of-the-art object localization methods, BING [6] proposes candidate regions by estimating their objectness in an efficient manner. We therefore decide to utilize this objectness metric for selecting top-k representative regions among many candidates. We tune the objectness score to be more suitable for image search, and select top-150 regions based on the score. Details of the tuning is described in our supplementary material, and the analysis for k, the number of selected regions, is available in Sec. 4.6.

3.2 Feature extraction

After selecting k representative regions from an image, we extract a feature from an image box of each region. We can use any CNN feature as our image descriptor for each region. Additionally, we store inclusion relations of regions together, which are used to improve the confidence of matching (Sec. 3.3).

The advantage of extracting features only from representative regions and maintaining them separately is that the feature extracted from a region has clearer meaning compared to a feature extracted from the whole image space. As an example, Fig. 3.1 shows that while two images contain the same

Table 3.1: Percentage of proposals matched with the ground truth bounding boxes. The experiment is performed with 1 k images from the Pascal VOC 2012 datasets [9]. Both selective search (SS) [44] and BING [6] show high recalls (96.7% and 96.3%, respectively), but have many unmatched proposals.

	Selective Search [44]	BING [6]
Matched proposals	7,422	3,034
Unmatched proposals	251,461	96,966
Unmatched ratio(%)	**97.1**	**97.0**

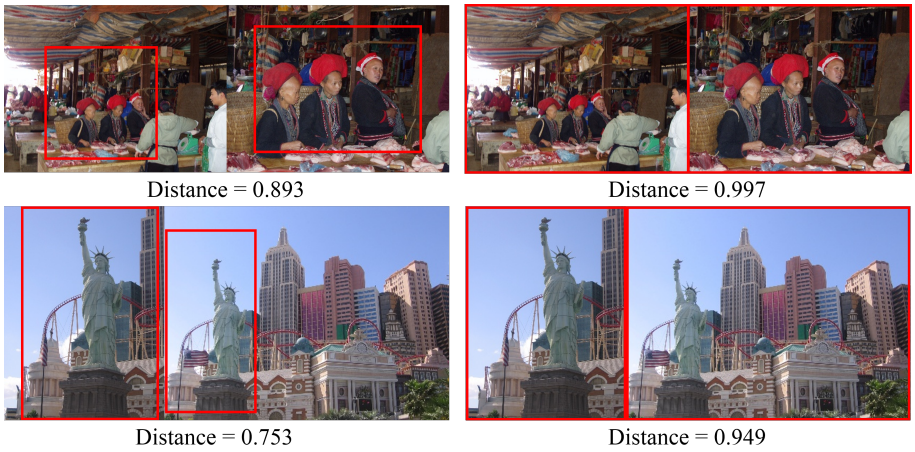


Figure 3.1: Comparison of the Euclidean distances between features extracted only from chosen regions (red boxes in left) or extracted from the whole images (right). Features extracted from chosen regions represent the contents of an image clearly, and enable similarity matching to be more accurate.

object, their Euclidean distance between features extracted from the whole images is far. On the other hand, by extracting features only from representative regions, we achieve a short distance, resulting in better image search. In Sec. 4.2, we show that our method increases recall of image retrieval, compared with using features extracted from the whole images.

During the feature extraction, we also store inclusion relations of regions. There exist many kinds of regions such as regions representing some parts of objects, indicating whole objects, and including even backgrounds. We utilize such inclusion relationships between two regions for accurately identifying similar images as shown in Fig. 1.1(a).

3.3 Similarity measurement

Most existing image search techniques adopted a type of aggregation that computes a single feature from many local features [40, 22, 2]. One of well-known aggregation techniques includes the bag-of-feature model [40].

These techniques then employ a distance metric, *e.g.,* Euclidean distance or cosine measure, between those aggregated features to compute a similarity between two images. We can also use such a similarity metric by aggregating features extracted from regions into a single feature vector. Unfortunately, we found that the traditional distance metrics produce suboptimal results to our method, because the aggregated features dilute the content information extracted from multiple image regions. Instead, we propose a voting-based similarity measure, which is well suited for our features extracted from multiple image regions.

Let $q_i \in \mathcal{R}(I_q)$ be an i-th selected region from the query image I_q, where $\mathcal{R}(I_q)$ denotes a set of regions proposals in the image I_q. Since the order of regions does not matter in our method, the feature set of a query image is then given by $\mathcal{F}(I_q) = \left\{\mathbf{f}(q_i) \mid i = 1, ..., k\right\}$, where $\mathbf{f}(q_i)$ denotes the feature descriptor for q_i and k is the number of representative regions extracted from the image. For each region q_i, we retrieve images from the database that contain a region close to q_i based on the Euclidean distance in the feature space. Suppose that an ordered set of retrieved regions using the query region q_i is denoted by $\mathcal{D}_i = \left\{d_{ij} \mid j = 1, ..., v\right\}$, where j is the rank index in terms of the Euclidean distance and v is the

number of retrieved regions.

We identify similar images using a new similarity measure, which is based on a voting scheme defined on region proposals. The proposed voting algorithm provides a score for a region of an image in the database, where the score is composed of two factors:

$$\text{VotingS}(q_i, d_{ij}) = \text{RankS}(q_i, d_{ij}) \cdot \text{RelS}(q_i, d_{ij}), \tag{3.1}$$

where $\text{RankS}(\cdot,\cdot)$ and $\text{RelS}(\cdot,\cdot)$ denote ranking score and relation score between two regions, respectively. Note that $\text{RankS}(\cdot,\cdot)$ computes rank-based region similarity and $\text{RelS}(\cdot,\cdot)$ considers the inclusion relationship with related regions of input proposals. The final similarity measure considered with all regions between the query image I_q and an arbitrary image I in the database is given by:

$$\text{Similarity}(I_q, I) = \sum_{q_i \in \mathcal{R}(I_q)} \max_{d_{ij} \in \mathcal{D}_i \cap \mathcal{R}(I)} \text{VotingS}(q_i, d_{ij}), \tag{3.2}$$

where $\mathcal{D}_i \cap \mathcal{R}(I)$ indicates a set of retrieved regions for q_i among regions from the image I. This similarity metric simply sums voting scores, each of which is computed by the best match from a region q_i of the query to another region d_{ij} in the test image I.

We first define our rank-based region similarity, $\text{RankS}(\cdot,\cdot)$, which utilizes a rank of a retrieved region from a query region. Simply, the ranking score of d_{ij} for q_i is determined as:

$$\text{RankS}(q_i, d_{ij}) = 1/j. \tag{3.3}$$

Intuitively, as we get a bigger rank order, we give a lower similarity score to the region d_{ij}.

We now define our relationship score $\text{RelS}(\cdot,\cdot)$. Our main idea of using the inclusion relationship is that while regions of a database image can match with query regions, their related regions may or may not be matched (Fig. 1.1(b)). Depending on whether related regions match well or not, our confidence level of the matching between the query image and the tested database image can vary significantly. To realize this intuition, we introduce the relationship score $\text{RelS}(q_i, d_{ij})$.

For an arbitrary region r, we let $P(r)$ be a set of parent regions including the box of r, and $C(r)$ be another set of child regions included in the box of r. We first define a set of matched parent regions between q_i and d_{ij} as follows:

$$\text{MatchedP}(q_i, d_{ij}) = \left\{ d_p \mid d_p \in \mathcal{D}_x \cap P(d_{ij}) \; for \; q_x \in P(q_i) \right\}, \tag{3.4}$$

where $\mathcal{D}_x$ is a retrieved region set of q_x, a parent region of q_i. Briefly speaking, $\text{MatchedP}(q_i, d_{ij})$ contains the parent regions of d_{ij} which are also retrieved for parent regions of q_i. A set of matched child regions $\text{MatchedC}(\cdot,\cdot)$ is defined in the same manner. We then define a set of matched related regions between q_i and d_{ij} as follows:

$$\text{MatchedRels}(q_i, d_{ij}) = \text{MatchedP}(q_i, d_{ij}) \bigcup \text{MatchedC}(q_i, d_{ij}). \tag{3.5}$$

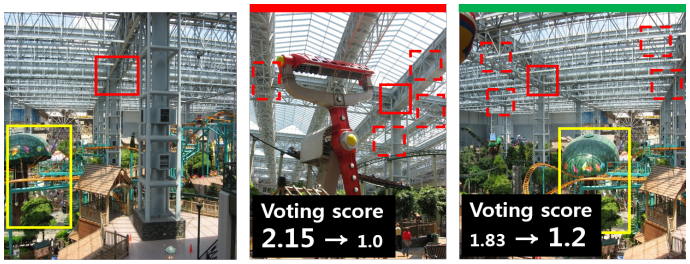
We assume that better matched images are likely to have bigger sets of $\text{MatchedRels}(\cdot,\cdot)$ for the query region q_i. The relationship score $\text{RelS}(q_i, d_{ij})$ is finally defined based on the matched related regions:

$$\text{RelS}(q_i, d_{ij}) = 1 + |\text{MatchedRels}(q_i, d_{ij})|, \tag{3.6}$$

where 1 represents the matching of q_i and d_{ij} themselves, and the cardinality of $\text{MatchedRels}(q_i, d_{ij})$ indicates the number of matching within their related regions.

(a) Ignore dissimilar regions (red boxes) with voting

(b) Deemphasize frequently appearing objects (red boxes) by the rank



(c) Handle burstiness of local regions by having a maximum single vote; dotted boxes are not considered, so the impact of other matches relatively increases (yellow boxes).

Figure 3.2: Left images are queries, and right images are test images. Green and red bars on top of test images indicate correct and incorrect matched results, respectively. Our rank-based voting method handles these issues commonly raised by using local features. Yellow boxes represent similar regions, while red ones indicate less important matches.

We now show visually how our rank-based voting methods achieve high matching quality in Fig. 3.2. Fig. 3.2(a) shows matched regions in yellow boxes and unmatched regions in red boxes between a query and its ground-truth test image. In case of using cross-matching [35] or NBNN [46] for the similarity measurement between multiple regions, all the regions in two images contribute to their similarity. In this approach, dissimilar regions are also matched and lower down the quality of measuring the similarity between two images. On the other hand, our method measures the similarity only with matched regions by utilizing the ranking set of $\mathcal{D}_i$, which is top-v close regions to the query region q_i.

Fig. 3.2(b) points out a problem caused by frequently appearing objects. Objects in the red boxes are visually very similar, so they can have a much higher similarity score than other matches, overwhelming important matches between the object shown in the yellow box. As a simple, yet effective way to normalize the impact of each region in the query, we use a ranking score instead of the direct $L2$ similarity for the final search result as shown in Eq. 3.3. For frequently appearing objects, there are many images containing such objects and thus a ranking of a test image can be very low, as demonstrated in Fig. 3.2(b);

note that this is a similar concept to the term of inverse document frequency (IDF), a common technique deemphasizing frequently appearing words in the text search.

Fig. 3.2(c) demonstrates the burstiness problem, which occurs frequently for local image descriptors [20]. Our method extracting features from multiple local regions can also share the problem. In order to prevent the burstiness, each region in our method votes only once for a test image, by utilizing the maximum score as shown in Eq. 3.2. Finally, the $\mathrm{RelS}(\cdot,\cdot)$ term considered spatial relationship among regions and improve the matching confidence (Fig. 1.1). Effects of this term is elaborated in Sec. 4.4.

3.4 Scalable encoding and indexing

Our method aims to achieve high search accuracy by representing an image with multiple regions. As a result, it may require a large amount of memory for encoding all the images for large image datasets. In order to increase both memory and computational efficiencies, we adopt Principle Component Analysis (PCA) with whitening [18] and Product Quantization (PQ) [21] for our features.

PCA is well-known method for dimensionality reduction, and Jégou *et al.* [18] showed that it could even improve the quality of features by applying whitening. We also apply PCA and whitening to the features first, and convert them into compact binary codes using PQ. In the search time, our method first retrieves a shortlist with size M using the binary codes, and then performs our voting method on the shortlist using the PCA features. Details of our implementation can be found in Sec. 4.7.

We utilize the IVFADC [21] to retrieve a shortlist for our voting method. The IVFADC is built upon the inverted index defined with a coarse quantizer such as K-means clustering. Each feature is indexed based on the inverted index and stored as a compact code compressed by PQ. At the searching stage, we first collect nearest-neighbor candidates by accessing the inverted index, and re-rank them according to the estimated distance between a query feature and a compact code. Since we perform the search by using the inverted index and accessing compact codes before computing the shortlist, it drastically improves the memory footprint required and access time. More details on the IVFADC and shortlist computation can be found in [21, 15].

Chapter 4. Experiment

We now present various experimental results on three standard benchmarks. We also compare the search accuracy of our image search algorithms against the state-of-the-art techniques using CNN-based representations.

4.1 Datasets

Our experiments are performed on the following three benchmark datasets:

INRIA Holidays dataset [19] (Holidays). This dataset consists of 500 groups of photographs, and each group represents the same scene or object taken in vacation. The total number of photographs is 1,491, and the number of queries is 500. The performance is measured by the mean average precision (mAP) over the queries. For a fair comparison, we also use the manually rotated version of the dataset as adopted by previous works [3, 2, 26, 11].

University of Kentucky benchmark dataset [32] (UKB). This dataset is composed of 10,200 photographs for 2,550 indoor objects, so there are four images taken from different viewpoints for each object. Each image is used as a query, and the performance is measured by the average number of retrieved images representing the same objects within the top-4 over all the queries.

Oxford Buildings dataset [33] (Oxford 5k). The dataset consists of 5,062 Oxford landmark images, which are collected from Flickr by searching for the landmarks. The dataset has 11 query landmarks, each of which has 5 query images. The performance is evaluated with mAP over the queries. For each query, the dataset provides a single bounding box containing the exact landmark for each query; note that the bounding box is provided only for queries, not for test images. Since the bounding box provides accurate region-of-interest, it is desirable to use this additional information, and thus we utilize the bounding box as one of our region proposals. We also use the well-known *Oxford 105k*, which add additional 100k distractor images to *Oxford 5k*. We use the benchmark for scalability tests.

Holidays consists of outdoor scene images, *Oxford 5k* contains building images, and *UKB* consists of indoor object images. Thanks to the distinct characteristics of the benchmarks, we can validate not only effectiveness, but also robustness of our approach.

4.2 Benefit of using multiple regional descriptors

Before showing the effect of our similarity measurement, we first demonstrate the benefit achieved mainly from our image representation. To present the benefit of using multiple features extracted from the representative regions, we compare recalls of two different types of features: the global feature extracted from a full image and our local feature obtained from regions, under a simple similarity measure, the $L2$ distance.

As shown in Table 4.1, our method using local features from the regions provides higher recalls than the method using a single global feature across all the benchmark datasets. Our multiple features

Table 4.1: Comparison of the average recalls of different features. The single global feature is a feature extracted from a whole image, while our method extracts multiple local features from regions.

Feature types	*Holidays* recall@20	*Oxford 5k* recall@200	*UKB* recall@10
Single global	0.831	0.618	0.931
Multiple regional	**0.977**	**0.865**	**0.990**

capture the details of an image effectively that can be missed in the global feature, and this is the main reason for improved recalls. In the next subsection, we discuss additional benefits of using the rank-based similarity measurement tailored for our representation.

4.3 Effects of the proposed method

We now demonstrate the effectiveness of our proposed method using the multiple features and the rank-based voting method as the similarity measurement. We compare its accuracy against that of a general image search scheme that uses the single global feature and L2 distance metric. To test robustness of our method, we perform experiments with four different image descriptors, which are features from CaffeNet and VGG-19 without any post-processing and ones with PCA-whitening.

Table 4.2 shows the accuracy of two different methods combined with each descriptor. In spite of the diverse conditions, our approach consistently improves the accuracy for all the cases over the general image search method. Based on the high recall from our image representation, our rank-based voting effectively matches and scores the regional features and returns accurate search results. On average, our method improves the accuracies relatively by 29% on *Holidays*, 13% on *UKB*, and surprisingly, 105% on

Table 4.2: Comparison of the accuracies between a general image search scheme and our proposed method with different image descriptors. Across all the different cases and benchmarks, our method robustly improves the accuracy over the general image search that uses the single global feature and $L2$ distance metric.

Holidays (mAP)

search scheme	Caffe	CaffePCA	VGG	VGGPCA
Single-$L2$	0.691	0.743	0.642	0.697
Ours	0.879	0.907	0.884	**0.917**

Oxford 5k (mAP)

search scheme	Caffe	CaffePCA	VGG	VGGPCA
Single-$L2$	0.302	0.356	0.319	0.415
Ours	0.678	0.735	0.661	**0.768**

UKB (recall@4)

search scheme	Caffe	CaffePCA	VGG	VGGPCA
Single-$L2$	0.832	0.877	0.806	0.866
Ours	0.943	0.957	0.952	**0.970**

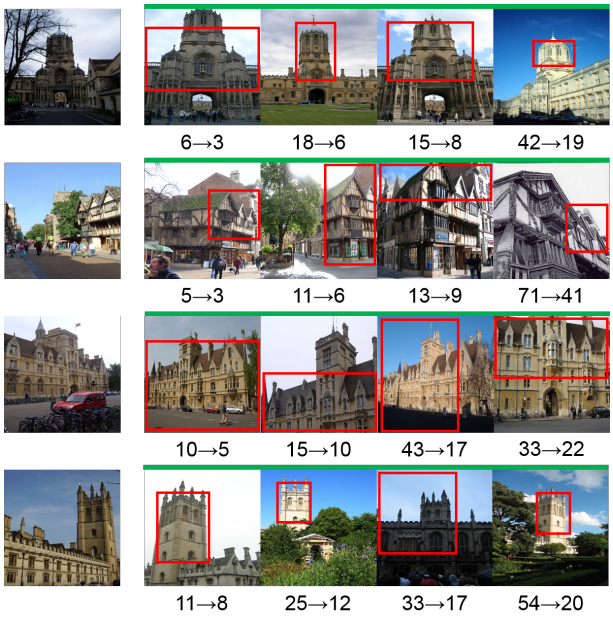


Figure 4.1: This figure shows how ranks of true positive images change by using $RelS(\cdot,\cdot)$.

Oxford 5k. Due to the small visual variances of buildings and frequent viewpoint changes in *Oxford 5k*, multiple local features work much better than a single global feature on this dataset.

4.4 Comparison with the state-of-the-arts

So far, we discussed advantages of our image representation and similarity measure methods. We now compare the search accuracy of our approach with the state-of-the-art image search methods that also utilize CNNs. Before comparing the accuracies with other methods, we categorize the methods according to their main objectives. The first category, the improved descriptors category, includes approaches that propose compact descriptors with improved accuracies [3, 2, 26, 43, 11]. The second category, improved similarity measurements category, consists of approaches proposing novel similarity measurements for utilizing multiple regional features [35, 46].

The main objective of the first category is improving discriminative power of a descriptor, while maintaining the memory efficiency. On the other hand, the second category focuses on improving search results by effectively utilizing larger amount of information from images. Since it is not reasonable to compare the accuracies directly between the memory-efficient methods and accuracy-focused methods, we show results of these two categories separately in Table 4.3.

In Table 4.3, we present results without any post-processing, *i.e.*, query expansion, re-ranking, and costly spatial verification, for fair comparison. We also show the results of our approach using both CaffeNet [23] and VGG-19 [41], to clearly compare the performance with others. Among those methods, neural code and off-the-shelf used the networks with the CaffeNet structure, and other methods used the VGG structure.

Methods in the improved descriptors category show great performance with a single global feature. Neural code [3] improves the descriptor for each benchmark by retraining the networks. SPoC [2], CroW [26], and R-MAC [43] enhance descriptors with their own aggregation methods. They implicitly embed spatial information into the descriptors with centering prior, spatial weighting, and max-pooling,

Table 4.3: Comparison of accuracies with the state-of-the-art search methods. Methods included in the improved descriptors category generate memory-efficient compact features, and methods contained in the improved similarity measurement category propose new similarity measures utilizing multiple regional features for accurate search results. Results marked with an * are achieved with retrained or fine-tuned networks.

Methods	*Holidays* mAP	*Oxford 5k* mAP	*UKB* recall@4
Improved descriptors			
Neural code [3]	0.793*	0.557*	0.890*
SPoC [2]	0.784	0.657	0.915
CroW [26]	0.851	0.708	-
R-MAC [43]	-	0.669	-
DeepIR [11]	**0.891***	**0.831***	-
Improved similarity measurements			
Off-the-shelf [35]	0.843	0.680	0.911
ONE [46]	0.887	-	0.968
Ours with CaffeNet	0.907	0.735	0.957
Ours with VGG-19	**0.917**	**0.768**	**0.970**

respectively. While other methods use the networks originally trained for image classification, DeepIR [11] trained the networks based on image similarities. With this search-specialized network, it shows the best performance among the methods in the first category, and even shows better performance than the methods in the second category for some cases. Since these approaches propose single compact descriptors, all of them are highly memory-efficient.

As the second category utilizes additional information from images, techniques in the category can show better accuracy than that of the first category on the condition of using equally trained networks. Fig. 4.2 shows qualitative comparisons between ours and SPoC. In the second category, Off-the-shelf [35] uses an average $L2$ distance, and ONE [46] utilizes NBNN for similarity measure. Comparing the results with the same network structures, our rank-based voting achieves better accuracy than others across all the benchmarks. Also, compared to the ONE method using 220 regions per image, our method achieves higher accuracy with 68% of the regions, 150 regions per image. Furthermore, our method does not require any additional datasets during the search process, while ONE requires additional categorized feature sets for the similarity measure using NBNN.

Table 4.4: Comparison of search accuracy with and without using $RelS(\cdot,\cdot)$. The $RelS(\cdot,\cdot)$ improves the accuracy significantly for difficult queries that produced frequent mismatches w/o using the term in *Oxford 5k*. Considering the term increases 0.137 AP on average over these four queries.

queries	w/ $RelS(\cdot,\cdot)$	w/o $RelS(\cdot,\cdot)$
christ church 5	**0.461**	0.304
cornmarket 2	**0.782**	0.641
balliol 1	**0.758**	0.625
magdalen 3	**0.506**	0.387

Figure 4.2: Comparison of the qualitative results between SPoC and ours. The leftmost images are queries, and right images are the results in order. In each row, the top images represent the result of SPoC, and the bottom images show the result of our method. We mark the correctness of a result image with a color bar above the image. Green and red bars indicate true positives and false positives, respectively. Blue bar indicates "junk" images of the $Oxford5k$ dataset, which contain the correct building to the query but with high occlusion or distortion. We also mark the regions with a high voting score on our result images. Our method works well for the queries where SPoC finds correct images well (top row). Moreover, our method shows good performance even for the queries where SPoC cannot retrieve correct images well (middle and bottom rows).

Discussions. While we compared different methods in two different categories, they can be combined together, since their goals, improving image descriptors and similarity measurements, are complementing each other. As shown in Table 4.2, our method consistently improves the accuracy over various descriptors by using them as multiple regional features instead of single global features. Also, DeepIR [11] leverages the network trained with triplet ranking loss. Since the ranking loss is evaluated based on image similarity ranks, this network is more appropriate and thus brings improvement to image search than the general networks trained with the classification loss. While our method currently utilizes the networks trained with the classification loss, it can be further improved by leveraging the network trained with the ranking loss.

Effects of relation term. We analyze the effect of relation term in our voting method. In order to study the effect attributed by the relation term $RelS(\cdot,\cdot)$, we compare search accuracy with and without using $RelS(\cdot,\cdot)$ (Table. 4.4). We found that using the term improves accuracy, especially for difficult cases, where incorrect matches occur frequently. Since it gives higher weights to more confident matches,

(a) Query w/ mask (b) Completed image (c) Retrieved image

Figure 4.3: This shows examples of the scene completion application of image search. (a) show queries with red masks for the completion, while (b) are completed images using the retrieved images (c) by our method in *Oxford 5k*.

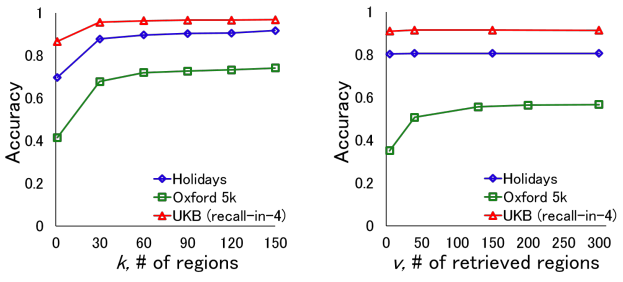
impact of incorrect matches relatively decrease. Therefore, it improves the final ranking of correct images appropriately as shown in Fig. 4.1.

4.5 Scene completion as an application

We apply our image search method to the application of scene completion that utilizes similar images identified from an image database [14]. Although we use a small size of the dataset, our search result effectively finds visually similar images to an input query in the $Oxford5k$ dataset. Fig. 4.3 shows two example images of the application based on our image search method.

4.6 Analysis

Our similarity measurement method uses two parameters k and v, which are the number of representative regions in an image and the number of retrieved regions from a query region, respectively.



(a) Accuracy according to k (b) Accuracy according to v

Figure 4.4: (a) shows the accuracy according to k, the number of representative regions in each image. Accuracy increases as k increases, but its rate becomes smaller as k becomes larger. (b) shows the accuracy as a function of v, the number of retrieved regions from each query region. We also observe the similar trend to that of k.

Fig. 4.4 shows retrieval accuracy as a function of k and v, individually. Interestingly, all data represent a similar tendency, where accuracies rapidly increase in early phases and slowly keep increasing until they reach the stable points. As the results indicate non-decreasing shapes, our method can show its best performance by simply assigning large values to k and v. Nonetheless, larger k and v need more time and memory resources. As a result, we set $k = 150$ and $v = 250$ for our experiments as a practical balance between the time and memory requirements.

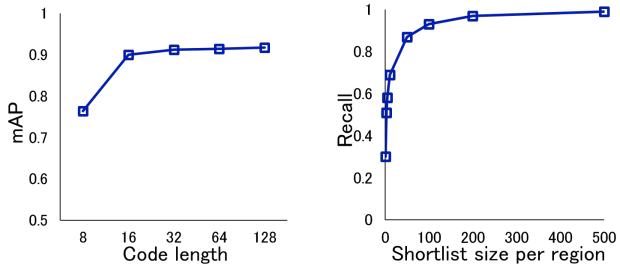
4.7 Implementation details

We use the python implementation of BING objectness estimation [6, 10] for object localization. For CNN feature extraction, we utilize Caffe [23] deep learning framework. We test our method with the features from both CaffeNet [23] and VGG-19 [41] networks for fair comparison with related works. For each region in an image, we extract a 4096-dimensional feature from a fully-connected layer of the networks.

Retraining and fine-tuning are well-known techniques customize CNN features to a particular dataset. However, our main purpose is to introduce an effective similarity measurement that consistently improves the accuracy across different benchmarks. Therefore, we did not perform any retraining or fine-tuning.

For dimensionality reduction, we perform *L2* normalization, PCA and whitening, and *L2* normalization again to the features in sequence same as other methods [35, 26, 43]. With PCA, we reduce the dimension of features from 4,096 to 512. We utilize 4,096 inverted indices to construct an inverted file and represent each feature by a compact code whose length is 64 bits, which gives high accuracy with a drastic compression ratio (Fig. 4.5 (a)). The memory footprint for storing the feature is reduced by a factor of 178 times. For instance, we only need 74MB for storing 6,413,174 features, while the size of PCA-whitened features is 12GB for *Oxford 105k* with $k = 60$.

When identifying the nearest neighbors per each region, we fix to collect $M = 10,000$ candidates from the inverted file for re-ranking and return the shortlist as the final search result. We chose to return 500 shortlist, since it achieves high accuracy given our PQ based encoding and indexing (Fig. 4.5 (b)). The nearest neighbor search is 4,000 times accelerated compared to the exhaustive scan of the raw features. On average, identifying 500 shortlist per each region takes 0.002 s in *Oxford 105k*. We also



(a) mAP on *Holidays* (b) Recall for 1-NN search

Figure 4.5: Experiments to determine parameters for PQ. (a) shows the mAP on *Holidays* according to the code length. Surprisingly, 64-bit codes do not cause much decrease compared to the result (0.917 mAP) achieved without using PQ. (b) shows the recall for 1-NN search according to the shortlist size. We achieve 0.99 recall with the shortlist of size 500.

found that the approximate nearest neighbor search by using the inverted index degraded the final image search accuracy by only 0.4%, while providing significant performance improvement. In this setting, our method takes 42.7 s for processing 55 queries and achieves 0.675 mAP accuracy on *Oxford 105k*.

Chapter 5. Conclusion and Future Work

We have introduced a novel image search method, which uses representative regions with an associated similarity measure, to improve the search accuracy. Our method employed a state-of-the-art region proposal method and customized it for our problem to define the objectness of representative regions. By extracting CNN features from the regions, our algorithm represents images more effectively, resulting in a high search accuracy. We have also proposed a new similarity measure, which incorporates the spatial relation of regions and improves the confidence of the matching. To show the usefulness and robustness of our method, we have performed evaluation on three different image datasets and have compared the performance with the state-of-the-art search methods. We observed that the proposed algorithm achieves the best accuracy among the methods using equally trained networks.

There are many interesting directions to further improve the performance for image search. Our method showed accuracy improvement by adopting the state-of-the-art region proposal and pre-trained CNN features. Fortunately, our method is well modularized, so there is a high chance to be combined with new cutting-edge object localization methods or new discriminative features. Second, our method achieved higher accuracy by utilizing the inclusion relationship of regions. Embedding a high-level spatial relation into the similarity measure can further increase the search accuracy. Finally, the current approach extracts only features from each region. If we classify the regions and estimates their labels, we can additionally consider the co-occurrence and semantic relations of the regions. Use of such higher-level information can further improve the accuracy of image search.

Bibliography

[1] Alexe B., Deselaers T., Ferrari V., *What is an object?*, In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (2010), IEEE, pp. 73–80.

[2] Babenko A., Lempitsky V., *Aggregating local deep features for image retrieval*, In Proceedings of the IEEE International Conference on Computer Vision (2015), pp. 1269–1277.

[3] Babenko A., Slesarev A., Chigorin A., Lempitsky V., *Neural codes for image retrieval*, In Computer Vision–ECCV 2014. Springer, 2014, pp. 584–599.

[4] Bell S., Zitnick C. L., Bala K., Girshick R., *Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks*, Computer Vision and Pattern Recognition (CVPR) (2016).

[5] Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L., *Semantic image segmentation with deep convolutional nets and fully connected crfs*, arXiv preprint arXiv:1412.7062 (2014).

[6] Cheng M.-M., Zhang Z., Lin W.-Y., Torr P., *Bing: Binarized normed gradients for objectness estimation at 300fps*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014), pp. 3286–3293.

[7] Donahue J., Jia Y., Vinyals O., Hoffman J., Zhang N., Tzeng E., Darrell T., *Decaf: A deep convolutional activation feature for generic visual recognition*, arXiv preprint arXiv:1310.1531 (2013).

[8] Dalal N., Triggs B., *Histograms of oriented gradients for human detection*, In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (2005), vol. 1, IEEE, pp. 886–893.

[9] Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A., *The pascal visual object classes (voc) challenge*, International Journal of Computer Vision 88, 2 (June 2010), 303–338.

[10] Ferrari A., *Python implementation of bing objectness method from "bing: Binarized normed gradients for objectness estimation at 300fps"*, https://github.com/alessandroferrari/BING-Objectness, 2015.

[11] Gordo A., Almazan J., Revaud J., Larlus D., *Deep image retrieval: Learning global representations for image search*, arXiv preprint arXiv:1604.01325 (2016).

[12] Girshick R., Donahue J., Darrell T., Malik J., *Rich feature hierarchies for accurate object detection and semantic segmentation*, In Proceedings of the IEEE conference on computer vision and pattern recognition (2014), pp. 580–587.

[13] Girshick R., *Fast r-cnn*, In Proceedings of the IEEE International Conference on Computer Vision (2015), pp. 1440–1448.

[14] Hays J., Efros A. A., *Scene completion using millions of photographs*, ACM Transactions on Graphics (SIGGRAPH 2007) 26, 3 (2007).

[15] Heo J.-P., Lin Z., Shen X., Brandt J., Yoon S.-E., *Shortlist selection with residual-aware distance estimator for k-nearest neighbor search*, In CVPR (2016).

[16] Hou X., Zhang L., *Saliency detection: A spectral residual approach*, In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (2007), IEEE, pp. 1–8.

[17] He K., Zhang X., Ren S., Sun J., *Deep residual learning for image recognition*, arXiv preprint arXiv:1512.03385 (2015).

[18] Jégou H., Chum O., *Negative evidences and co-occurences in image retrieval: The benefit of pca and whitening*, In Computer Vision–ECCV 2012. Springer, 2012, pp. 774–787.

[19] Jégou H., Douze M., Schmid C., *Hamming embedding and weak geometric consistency for large scale image search*, In European Conference on Computer Vision (oct 2008), David Forsyth Philip Torr A. Z., (Ed.), vol. I of LNCS, Springer, pp. 304–317.

[20] Jégou H., Douze M., Schmid C., *On the burstiness of visual elements*, In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (2009), IEEE, pp. 1169–1176.

[21] Jégou H., Douze M., Schmid C., *Product quantization for nearest neighbor search*, IEEE transactions on pattern analysis and machine intelligence 33, 1 (2011), 117–128.

[22] Jégou H., Douze M., Schmid C., Pérez P., *Aggregating local descriptors into a compact image representation*, In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (2010), IEEE, pp. 3304–3311.

[23] Jia Y., Shelhamer E., Donahue J., Karayev S., Long J., Girshick R., Guadarrama S., Darrell T., *Caffe: Convolutional architecture for fast feature embedding*, arXiv preprint arXiv:1408.5093 (2014).

[24] Jégou H., Zisserman A., *Triangulation embedding and democratic aggregation for image search*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014), pp. 3310–3317.

[25] Krähenbühl P., Koltun V., *Geodesic object proposals*, In Computer Vision–ECCV 2014. Springer, 2014, pp. 725–739.

[26] Kalantidis Y., Mellina C., Osindero S., *Crossdimensional weighting for aggregated deep convolutional features*, arXiv preprint arXiv:1512.04065 (2015).

[27] Kemelmacher-Shlizerman I., *Transfiguring portraits*, ACM Trans. Graph. 35, 4 (2016).

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

[29] Lowe D. G., *Distinctive image features from scale-invariant keypoints*, International journal of computer vision 60, 2 (2004), 91–110.

[30] Long J., Shelhamer E., Darrell T., *Fully convolutional networks for semantic segmentation*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp. 3431–3440.

[31] Noh H., Hong S., Han B., *Learning deconvolution network for semantic segmentation*, In Computer Vision (ICCV), 2015 IEEE International Conference on (2015).

[32] Nister D., Stewenius H., *Scalable recognition with a vocabulary tree*, In Computer vision and pattern recognition, 2006 IEEE computer society conference on (2006), vol. 2, IEEE, pp. 2161–2168.

[33] Philbin J., Chum O., Isard M., Sivic J., Zisserman A., *Object retrieval with large vocabularies and fast spatial matching*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2007).

[34] Perronnin F., Liu Y., Sánchez J., Poirier H., *Large-scale image retrieval with compressed fisher vectors*, In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (2010), IEEE, pp. 3384–3391.

[35] Razavian A., Azizpour H., Sullivan J., Carlsson S., *Cnn features off-the-shelf: an astounding baseline for recognition*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2014), pp. 806–813.

[36] Ren S., He K., Girshick R., Sun J., *Faster r-cnn: Towards real-time object detection with region proposal networks*, In Advances in Neural Information Processing Systems (2015), pp. 91–99.

[37] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A., *Going deeper with convolutions*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp. 1–9.

[38] Samii A., Měch R., Lin Z., *Data-driven automatic cropping using semantic composition search*, Comput. Graph. Forum 34, 1 (2015), 141–151.

[39] Snavely N., Seitz S. M., Szeliski R., *Photo tourism: Exploring photo collections in 3d*, In SIGGRAPH Conference Proceedings (2006), pp. 835–846.

[40] Sivic J., Zisserman A., *Video google: A text retrieval approach to object matching in videos*, In ICCV (2003).

[41] Simonyan K., Zisserman A., *Very deep convolutional networks for large-scale image recognition*, CoRR abs/1409.1556 (2014).

[42] Tzeng E., Hoffman J., Darrell T., Saenko K., *Simultaneous deep transfer across domains and tasks*, In ICCV (2015).

[43] Tolias G., Sicre R., Jégou H., *Particular object retrieval with integral max-pooling of cnn activations*, arXiv preprint arXiv:1511.05879 (2015).

[44] Van De Sande K. E., Uijlings J. R., Gevers T., Smeulders A. W., *Segmentation as selective search for object recognition*, In Computer Vision (ICCV), 2011 IEEE International Conference on (2011), IEEE, pp. 1879–1886.

[45] Valenti R., Sebe N., Gevers T., *Image saliency by isocentric curvedness and color*, In Computer Vision, 2009 IEEE 12th International Conference on (2009), IEEE, pp. 2185–2192.

[46] Xie L., Hong R., Zhang B., Tian Q., *Image classification and retrieval are one*, In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (2015), ACM, pp. 3–10.

[47] Zhang N., Donahue J., Girshick R., Darrell T., *Part-based r-cnns for fine-grained category detection*, In Computer Vision–ECCV 2014. Springer, 2014, pp. 834–849.

Acknowledgments in Korean

연구 내용을 발전시켜 나갈 수 있도록 열과 성을 다해 이끌어 주시고, 항상 “Aim high” 하여 스스로 발전해 나아가도록 지도해주신 윤성의 교수님께 감사드립니다. 아직도 부족한 점이 많지만 교수님의 가르침이 있었기에 2년 전과 비교하여 실력 뿐만 아니라 정신적으로도 크게 성장할 수 있었습니다.

석사 기간 동안 함께 생활한 모든 연구실 동료분들께도 감사드립니다. 자유롭고 수평적인 연구실 문화를 만들어주신 덕분에 행복하게 대학원 생활을 보냈습니다. 외적인 스트레스 없이 연구에만 집중할 수 있는 환경이었기에 무사히 논문을 제출하고 졸업할 수 있었습니다. 특히 직접적으로 연구에 도움을 주신 재필 형님과 태영이형 덕분에 논문의 완성도를 한층 더 높일 수 있었습니다. 2년간 항상 가장 든든한 지원군이 되어주신 모든 연구실 멤버분들께 이자리를 빌어 다시 한 번 감사의 인사를 드립니다.

마지막으로 항상 저를 믿어주시고 제 생각을 존중해주시는 부모님과 동생 민영이에게도 감사의 말씀을 드립니다. 타지에서 공부하느라 아들, 오빠로서의 역할을 충분히 하지 못한 것 같아 아쉽습니다. 앞으로 더욱 열심히 살면서 크게 효도하겠습니다. 사랑합니다.

Curriculum Vitae in Korean

이 름: 조 재 형

생 년 월 일: 1992년 03월 18일

주 소: 대전 유성구 대학로 291 한국과학기술원 E3-1 3433호

전 자 주 소: dil122001@gmail.com

학 력

2008. 3. – 2010. 2. 대구과학고등학교

2010. 2. – 2015. 2. 한국과학기술원 전산학부 (학사)

경 력

2013. 2. – 2013. 12. 주식회사 마이뮤직테이스트 개발 인턴

2015. 3. – 2016. 8. 한국과학기술원 전산학부 조교