


# A Novel Convolutional Neural Network Based Localization System for Monocular Images

Chen Sun, Tsinghua University, Beijing, China

Chunping Li, Tsinghua University, Beijing, China

Yan Zhu, Southwest Jiaotong University, Chengdu, China

 <https://orcid.org/0000-0002-5804-8926>

## ABSTRACT

The authors present a robust and extendable localization system for monocular images. To have both robustness toward noise factors and extendibility to unfamiliar scenes simultaneously, our system combines traditional content-based image retrieval structure with CNN feature extraction model to localize monocular images. The core model of the system is a deep CNN feature extraction model. The feature extraction model can map an image to a d-dimension space where image pairs in the real world have smaller Euclidean distances. The feature extraction model is achieved using a deep Convnet modified from GoogLeNet. A special way to train the feature extraction model is proposed in the article using localization results from Cambridge Landmarks dataset. Through experiments, it is shown that the system is robust to noise factors supported by high level CNN features. Furthermore, the authors show that the system has a powerful extendibility to other unfamiliar scenes supported by a feature extract model's generic property and structure.

## KEYWORDS

Algorithm, Classification, Content Based Recognition, Convolutional Neural Network, Feature Extraction, Image Based Localization, Image Retrieval, Loss Function

## 1. INTRODUCTION

Localization is crucial for people's life and many applications like navigation, robotics, augmented reality, etc. Though the global positioning system (GPS) can solve the problem in the most of situations, there are still some cases that GPS cannot handle well. Many image-based localization methods are proposed to deal with these cases. This paper proposes a novel localization system named Dis-Retrieval to estimate position from a monocular RGB image.

Our proposed system takes a monocular RGB image as input and outputs a position where this image is taken. The core of our system is a deep convolutional neural network (CNN) model, which can map an image to a d-dimension space where feature pairs of distance-closer image pairs in real

DOI: 10.4018/IJSSCI.2019040103

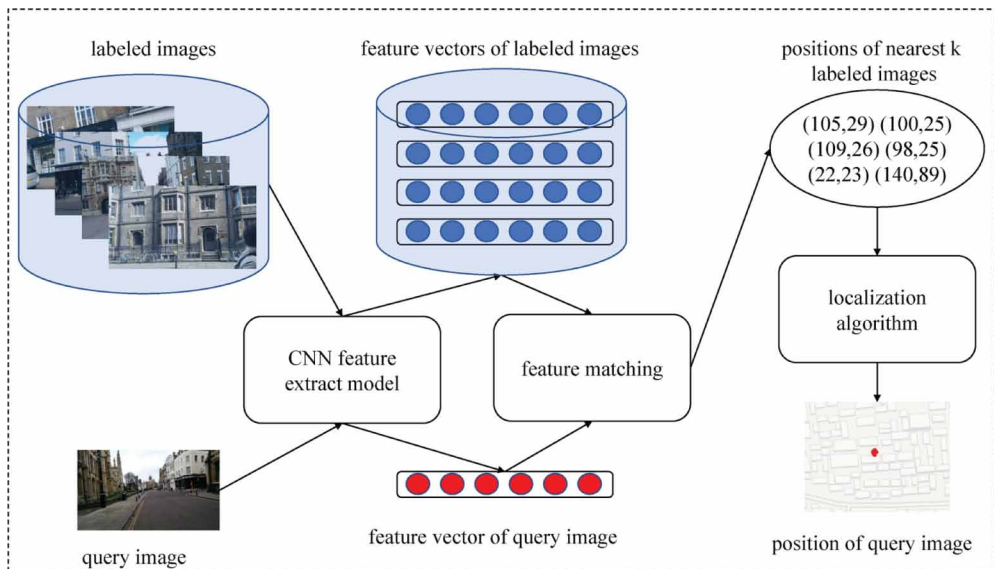
Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

word have smaller Euclidean distances. As Figure 1 illustrates, the structure of our system is similar with content-based image retrieval system (Zhang, Zhao, and Han, 2009). When a query image is coming, our system firstly uses CNN feature extraction model to extract a feature vector from query image, then uses it to match features of labeled images for k nearest images, finally uses k images' position information to estimate query image's position. Similar with content-based image retrieval system, it can operate in real time speeded by hash technology (Gionis, Indyk, and Motwani, 2000).

Before introducing our main contribution, we first simply talk about motivations of this paper. By now, there are mainly two types of methods to solve the image-based localization problem. Methods of first type are traditional methods which estimate position by using traditional image features (like SIFT (Lowe, 2004)) to match images. Methods of this type are easily influenced by noise factors in images such as light, camera's angle, pedestrians, cars, etc. Methods of second type are CNN-feature-based methods, which use CNN to solve the problem. Methods of this type can easily handle the influence of light, camera's angle, pedestrians and cars through data-driven learning. But they are less extendable, that is one trained model can just process one scene or one place. For example, we train a model using labeled images of A University: if we want to have a localization model of B University, we have to train another model using labeled images of B University; even when we want to add some labeled images of A University, we also have to fine tune old model of A University. Maybe you can force one model to process two universities, but what if there are 1000 universities? The model's parameters are limited, so it is hard to extend methods of this type to many places. Our proposed system overcomes these two difficulties through combining traditional content-based image retrieval structure with CNN feature extraction model.

Our main contribution is the deep CNN feature extraction model. The feature extraction model can map an image to a dimension space where feature pairs of distance-closer image pairs in real word have smaller Euclidean distances. The feature extraction model has two good properties. First, the feature extraction model is more robust to light, camera's angle, pedestrians and cars than traditional image feature extraction methods such as SIFT (Lowe, 2004), HOG (Michal, 1995), LBP (Ojala, Pietikainen, and Harwood, 1994), etc. Second, the feature extraction model has a competitive performance for untrained places, that is to say if you use a place's labeled images to train a feature extraction model, the model also works well to other untrained places. We achieve these two properties

Figure 1. Structure of proposed localization system for monocular image



11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/a-novel-convolutional-neural-network-based-localization-system-for-monocular-images/233522](http://www.igi-global.com/article/a-novel-convolutional-neural-network-based-localization-system-for-monocular-images/233522)

## Related Content

---

### Clustering Finger Motion Data from Virtual Reality-Based Training to Analyze Patients with Mild Cognitive Impairment

Niken Prasasti Martono, Takehiko Yamaguchi, Takuya Maeta, Hibiki Fujino, Yuki Kubota, Hayato Ohwada and Tania Giovannetti (2016). *International Journal of Software Science and Computational Intelligence* (pp. 29-42).

[www.irma-international.org/article/clustering-finger-motion-data-from-virtual-reality-based-training-to-analyze-patients-with-mild-cognitive-impairment/174447](http://www.irma-international.org/article/clustering-finger-motion-data-from-virtual-reality-based-training-to-analyze-patients-with-mild-cognitive-impairment/174447)

### Efficiency and Scalability Methods in Cancer Detection Problems

Inna Stainvas and Alexandra Manevitch (2013). *Efficiency and Scalability Methods for Computational Intellect* (pp. 75-94).

[www.irma-international.org/chapter/efficiency-scalability-methods-cancer-detection/76470](http://www.irma-international.org/chapter/efficiency-scalability-methods-cancer-detection/76470)

### Role-Based Autonomic Systems

Haibin Zhu (2010). *International Journal of Software Science and Computational Intelligence* (pp. 32-51).

[www.irma-international.org/article/role-based-autonomic-systems/46145](http://www.irma-international.org/article/role-based-autonomic-systems/46145)

### Combining Ontology with Intelligent Agent to Provide Negotiation Service

Qiumei Pu, Yongcun Cao, Xiuqin Pan, Siyao Fu and Zengguang Hou (2010). *International Journal of Software Science and Computational Intelligence* (pp. 52-61).

[www.irma-international.org/article/combining-ontology-intelligent-agent-provide/46146](http://www.irma-international.org/article/combining-ontology-intelligent-agent-provide/46146)

### An Interactive Visualization of Genetic Algorithm on 2-D Graph

Humera Farooq, Nordin Zakaria and Muhammad Tariq Siddique (2012). *International Journal of Software Science and Computational Intelligence* (pp. 34-54).

[www.irma-international.org/article/interactive-visualization-genetic-algorithm-graph/67997](http://www.irma-international.org/article/interactive-visualization-genetic-algorithm-graph/67997)