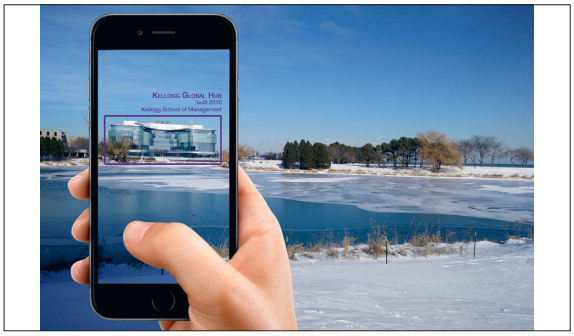
Hannes
Badertscher

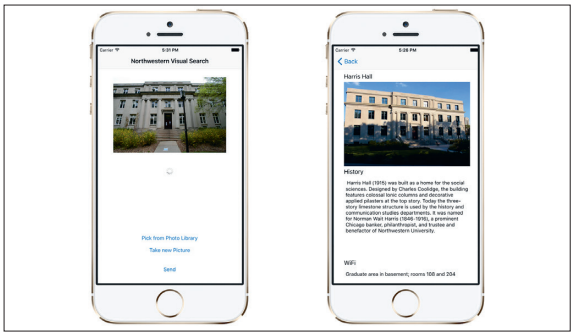
Graduate Candidate	Hannes Badertscher
Examiner	Prof. Dr. Guido Schuster
Co-Examiner	Prof. Dr. Aggelos K. Katsaggelos, Northwestern University, Evanston, IL
Subject Area	Sensor, Actuator and Communication Systems
Project Partner	Image and Video Processing Laboratory (IVPL) of Northwestern University, Evanston, IL

Northwestern Visual Search

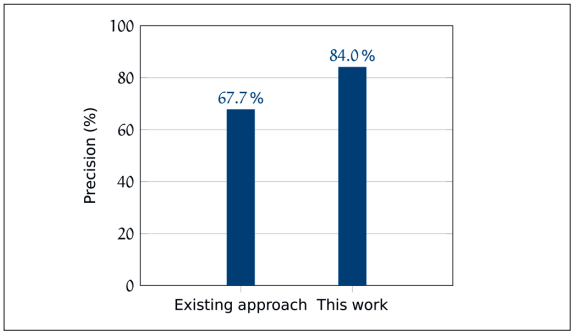
Using Deep Convolutional Networks for Recognizing Northwestern University Buildings



Visualization of how augmented reality (AR) on mobile phones can give useful information on the user's environment.



Screenshots showing the iPhone app developed for recognizing buildings at Northwestern University.



Comparison between the precision of the existing implementation and the newly developed system using deep convolutional neural networks.

Introduction: High-resolution cameras and fast and readily available internet connection make today's smartphones well suited to augmented reality applications. For example, when viewing a building through the smartphone's camera, additional information on that building, such as its name and history, can be useful to the user. The task of finding information based on an image is denoted by image retrieval or visual search. The Image and Video Processing Laboratory (IVPL) at Northwestern University has been working on developing compact descriptors for visual search on mobile devices for the last years. The IVPL has created a database of images of buildings at Northwestern University, and has developed a website and an Android app for recognizing these buildings. This existing approach uses a hand-crafted feature detector to describe the contents of a query image and match it to the database of images.

Approach/Technologies: Over the past years, deep learning has outperformed and replaced hand-crafted features in many subfields of computer vision, especially in image classification. In this thesis, neural networks are applied to the task of visual search. Deep convolutional neural networks (CNN) are trained to create a vector that describes the content of an image. By comparing the descriptor vector of a query image to the descriptor vectors of all database images, the building shown in the query image can be recognized. CNNs are usually trained using millions of training images. As only a few thousand images of Northwestern buildings are available, a CNN that was trained on a different database is reused and applied to the task of recognizing Northwestern buildings.

Result: The existing database of buildings and images at Northwestern was extended from 952 images of 56 buildings to 2,133 images of 134 buildings. The described method for visual search using deep convolutional networks has been implemented and tested using the Caffe deep learning framework. The neural network approach outperforms the existing method, increasing the precision of the visual search from 67,7% to 84,0%. In addition to the existing website and Android app, an iPhone app was developed in this thesis, to make the visual search application available to a broader range of users.