Report

Matteo Gambino s287572 | Michele Pierro s287846 | Fabio Grillo s287873

Abstract

In order to predict the location of a query image by retrieving annotated photographs with similar descriptors needs an efficient and reliable generation of those descriptors. In order to accomplish that objective, is fundamental that the network focuses on portion of the various images that contains useful information and at the same time ignore not informative areas like the ones containing elements like cars or pedestrians. For that reason attention layers are fundamental in the proposed network. In addition to that we are comparing state of the art techniques for the visual geolocalization task like GEM [?] and NetVLAD [?].

1. Introduction

2. Related works

3. Methods

4. Experiments

	R@1	R@2	R@10	R@20
lr = 1e-3	81.9	91.8	94.8	96.8
lr = 1e-4	82.2	93.0	**95.4**	**97.1**
lr = 1e-5	**83.5**	**93.1**	94.3	**97.1**

5. Ablation study

6. Conclusions