Crowd Counting using fine tuning and density maps

Pierpaolo D'Odorico
pierpaolo.dodorico@studenti.unipd.it

Massimiliano Conte
massimiliano.conte@studenti.unipd.it

Abstract

This paper aims to investigate which is the best approach for crowd counting in a simple situation, among different computer vision tecniques. The counting is performed on images captured from a security camera placed in a shopping mall. In this simple but useful task the goal is to estimate the number of people that are present in the image. We can for example count the percentage of visitors who buyed from a specific shop or report to the mall staff that covid-19 social distancing laws are not respected. ttt

1. Introduction

Crowd counting has been a challenge in recent years. The goal of this task is directly connected with crowd control and public safety. The affluence of people in different areas can be also useful for planning spaces and services. For example in our specific task we need to count people from seurity camera images in a mall, and this people counting can be useful for several business and safety applications. There were developed a lot of different tecniques for solving this computer vision task. Our idea is to compare two different tecniques and finding the best one for counting from single security camera images. The first approach is a convolutional neural network fine tuning on VGG16 network. This very deep network was proposed by K. Simonyan and A. Zisserman from the University of Oxford and it was submitted to ImageNet challenge in 2014. The second approach deals with density maps. Those are useful in real life applications since the same number of people could have completely different crowd distributions. (as shown in Fig. 1). After implementing the two approaches on our problem we found that...

2. Related work

We based our project on informations contained in different papers about computer vision tasks and crowd counting. The first one is related to VGG16, the network we fine tuned for dealing with our people counting problem [2]. In this paper the authors investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. In this convolutional neural networks they use very small (3x3) convolution filters, which shows that a significant improvement on the prior state of the art configurations. They also pushed the depth to 16-19 weight layers (Fig.1). Those small convolutional filters can be useful in our application for finding people in images. Fine tuning the final layers and adding a regression one at the final part of the network, we hope that VGG16 model will achieve a good performance.

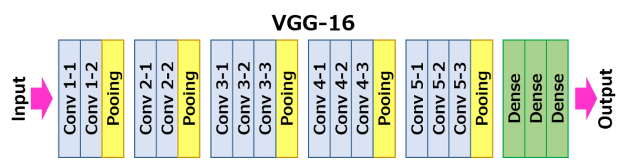

Figure 1. VGG16 deep CNN structure

We relate also to two papers about density map crowd counting approach. In the first one [3] authors introduced a new large-scale crowd dataset named Shanghaitech. It contains nearly 1,200 images with around 330,000 accurately labeled heads. They also used Multi-column CNN for density map estimation (Fig.2). We also relate to a more recent paper that simplifies the convolutional neural network of the previous work and achieves better results. This second paper about density map crowd couting approach [1] uses a CSRNet architecture for capturing high-level features with larger receptive fields and generating high-quality density maps without expanding network complexity.

Figure 2. Density map approach

3. Dataset

References

[1] Yuhong Li, Xiaofan Zhang, and Deming Chen. Csrnet: Dilated convolutional neural networks for understanding the

highly congested scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1091–1100, 2018.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[3] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 589–597, 2016.