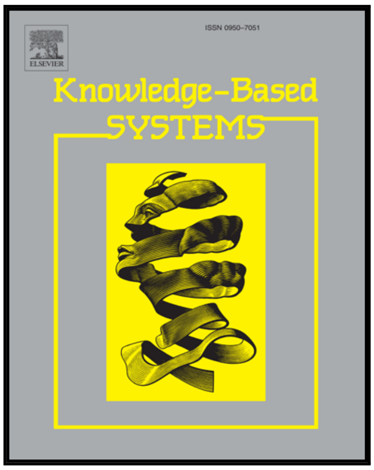
Accepted Manuscript



Improving Deep Ensemble Vehicle Classification by Using Selected Adversarial Samples

Wei Liu, Zhiming Luo, Shaozi Li

PII: S0950-7051(18)30321-6
DOI: 10.1016/j.knosys.2018.06.035
Reference: KNOSYS 4415

To appear in: *Knowledge-Based Systems*

Received date: 3 January 2018
Revised date: 16 June 2018
Accepted date: 19 June 2018

Please cite this article as: Wei Liu, Zhiming Luo, Shaozi Li, Improving Deep Ensemble Vehicle Classification by Using Selected Adversarial Samples, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.06.035

Improving Deep Ensemble Vehicle Classification by Using Selected Adversarial Samples

Wei Liu[a,d], Zhiming Luo[b,e,*], Shaozi Li[c]

[a] *Virtual Reality and Interactive Techniques Institute, East China Jiaotong University, Jiangxi, China.*
[b] *Postdoc Center of Information and Communication Engineering, Xiamen University, China*
[c] *Cognitive Science Department, Xiamen University, Fujian, China.*
[d] *Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, Yi Chang,China.*
[e] *The Key Laboratory of Cognitive Computing and Intelligent Information Processing of Fujian Education Institutions, Wuyishan, China*

Abstract

Most image classification algorithms aim to maximize the percentage of class labels that are predicted correctly. These algorithms often missclassify images from minority categories as into the dominant categories . To overcome the issue of unbalanced data for classifying vehicles from traffic surveillance images, we propose a semi supervised pipeline focused on integrating deep neural networks with data augmentation based on generative adversarial nets (GANs). The proposed approach consists of three main stages. In the first stage, we trained several GANs on the original dataset to generate adversarial samples for the rare classes. In the second stage, an ensemble of CNN models with different architectures are trained on the original imbalanced data set, and then a sample selection step is performed to filter out the low-quality adversarial samples. In the final stage, the aforementioned ensemble model is refined on the augmented dataset by adding the selected adversarial samples. Experiments on the highly imbalanced large benchmark "MIOvision Traffic Camera Dataset (MIO-TCD)" classification challenge dataset demonstrate that the proposed framework is able to increase the mean performance of some categories to some extent, while maintaining a high overall accuracy, compared with the baseline.

Keywords: Imbalanced classification, image classification, generative adversarial nets, ensemble learning
2010 MSC: 00-01, 99-00

1. Introduction

In the last decade, the widespread use of visual traffic systems has led to rapid growth of the available video data that must be processed. However, while parsing surveillance video content is repetitive, it is suitable for computers to perform tedious tasks that require a long attention span. Therefore, automated video surveillance is attracting more attention than manual surveillance. With the increasing number of available images, image processing techniques such as image classification have become a hot topic in the field of artificial intelligence. Although image classification has been widely studied in academia and applied in various fields, it remains an open problem. For example, many practical image classification problems are imbalanced, i.e., at least one of the classes is represented by only a few samples, while other categories make up the majority. It is a difficult task to classify images with multiple labels using only a small number of labeled samples, and this difficulty is compounded by images with an unbalanced distribution.

Moreover, image classification plays an important role in visual intelligent transport systems. It is a prerequisite for the semantic analysis of visual traffic surveillance systems. In the field of traffic surveillance, a visual traffic surveillance system needs to detect vehicles or pedestrians and classify them if possible. In practice, *Pedestrians*, *Bicycles* and *Motorcycles* often constitute a minority of the data set, in contrast with *Cars* and *Buses*. Consequently, to avoid the misclassification of images from rare categories as majority classes, it is not appropriate to assume that the misclassification error costs for all samples are equal. If the misclassification error costs are implicitly assumed to be equal, images from minority categories are prone to be misclassified as from dominant categories. Therefore, to effectively reduce the number of fatalities, it is reasonable to focus on enhancing the mean accuracy of all categories, in the condition of high overall accuracy.

In the last decade, deep neural networks have led to a series of breakthroughs on a variety of machine learning tasks, such as computer vision, text analysis and voice recognition. Large-scale labeled training datasets are becoming increasingly important with the rise in the capacity of deep learning methods. However, such datasets are not always available. In such cases, data augmentation techniques to enlarge training image data sets with given labels become a viable solution. Generative adversarial networks (GANs)[1] have been used to generate synthetic images, owing to the relative sharpness of samples generated by these models compared to other approaches. However, learning from training data enlarged by GANs may not achieve the desired performance, owing to a gap between the synthetic and real image distributions for rare classes.

The contributions of this work are as follows:

*Corresponding author
Email address: zmluo_xmu@qq.com (Zhiming Luo)