GATE CONNECTED CONVOLUTIONAL NEURAL NETWORK FOR OBJECT TRACKING

T.Kokul[⋆†] *C.Fookes*[⋆] *S.Sridharan*[⋆] *A.Ramanan*[‡] *U.A.J.Pinidiyaarachchi*[◇]

[⋆]Image and Video Lab, SAIVT Program, Queensland University of Technology, Australia
[‡]Dept. of Computer Science, University of Jaffna, Sri Lanka
{[†]PGIS, [◇]Dept. of Statistics and Computer Science}, University of Peradeniya, Sri Lanka

ABSTRACT

Convolutional neural networks (CNNs) have been employed in visual tracking due to their rich levels of feature representation. While the learning capability of a CNN increases with its depth, unfortunately spatial information is diluted in deeper layers which hinders its important ability to localise targets. To successfully manage this trade-off, we propose a novel residual network based gating CNN architecture for object tracking. Our deep model connects the front and bottom convolutional features with a gate layer. This new network learns discriminative features while reducing the spatial information lost. This architecture is pre-trained to learn generic tracking characteristics. In online tracking, an efficient domain adaptation mechanism is used to accurately learn the target appearance with limited samples. Extensive evaluation performed on a publicly available benchmark dataset demonstrates our proposed tracker outperforms state-of-the-art approaches.

Index Terms— object tracking, CNN, domain adaptation

1. INTRODUCTION

Visual object tracking is a fundamental task in computer vision and many other applications [1]. Object tracking has attracted considerable research in the past and much progress has been made recently. However, it is still far from reaching the accuracy of the tracking ability of humans, due to appearance changes, pose variations, occlusions, illumination variations and background clutter.

Many of the appearance-based tracking frameworks [2, 3] depend on low-level hand-crafted features. These features fail to capture semantic information of targets and are not robust to significant appearance changes. Therefore sophisticated learning methods are needed to improve the feature representation in tracking frameworks.

Through a rapid increase in computational power, Deep Neural Networks (DNNs) can directly learn features from raw data without resorting to hand-crafting. DNNs, especially convolutional neural networks (CNNs), have demonstrated state-of-the-art performance in several vision tasks [4, 5, 6]. However, few visual tracking frameworks that have been proposed make use of DNNs. The main reason for this is as DNNs employ an extensive set of parameters, they require enormous amounts of training data, which is not yet available for visual tracking.

Several early deep learning based tracking approaches [7, 8] manage the training data deficiency by transferring offline learned DNN features to online tracking. These approaches obtain generic image features from object classification models. However, they do not learn similar local structure and inner geometric layout information among the targets, which are more important to discriminate a target from distractors.

Recent deep tracking approaches [9, 10, 11] analyse internal properties of CNN features from the perspective of visual tracking and propose tracking algorithms based on that. Even though these approaches showed state-of-the-art-performance, they suffer from over-fitting as they are fine-tuned online with only limited samples.

A robust visual tracking approach should learn different representations of targets and adapt to their appearance changes. In addition, deep learning based trackers should successfully manage the trade-off between over-fitting and the learning capability of the network with limited samples. Based on these requirements, we propose a novel deep tracking approach (called GNET) to learn generic tracking features. Our main contributions are:

- A residual network based gating CNN architecture, which learns generic tracking characteristics by effectively capturing features from front and end convolutional layers.
- An online domain adaptation mechanism, which is used to train the model with limited samples and reduce over-fitting.

The rest of this paper is organised as follows: Section 2 analyses state-of-the-art tracking trends and deep domain adaptation techniques. Section 3 describes the pros and cons of deep learning based tracking and an overview of our approach. Section 4 demonstrates our proposed approach in detail. The experimental setup and testing results are described in Section 5. Finally the paper is concluded in Section 6.

978-1-5090-2175-8/17/$31.00 ©2017 IEEE