Deep Learning Tracker via SVM Ranking Vector

Fangfang Li

June 22, 2018

Abstract

The deep learning based trackers can always achieve high tracking precision and strong adaptability in different scenarios. However, due to the fact that the number of the parameter is large and the fine-tuning is challenging, the time complexity is high. In order to improve the efficiency, author proposed a tracker based on fast deep learning through constructing a new network with less redundancy. Based on the theory of deep learning, author proposed a deep neural network to describe essential features of images. Furthermore, fast deep learning can be achieved by restricting the size of network. With the help of GPU, the time complexity of the network training is released to a large extent.

1. Introduction

Object tracking is an important branch of computer vision. It plays a key role in many areas that contain objects Identification, human-computer interaction and video surveillance. A complete visual tracking system should establish observation models Based on some appropriate search methods. Particle filter and Is a widely used search method. We choose The former one, it can handle non-linear and non-Gaussian problem. It is widely used in algorithm research and Engineering practice, which is suitable for static vision Platforms and mobile platforms. A reliable observation model is The key is to improve tracking robustness, including color histograms, textures, contours, and the fusion of these features. based on High-level visual cues work well when the appearance of the target changes Contains translations, rotations, and scaling changes, but it may drift over time. Some occlusion appeared. Based on the primary visual cues Prevents occlusion, but lacks robustness in complex work background. Deep learning [3] based on features combined with upper grades. Leads and juniors express data in a more efficient way. At the same time, it can be used in digital imaging and other fields Processing, speech recognition, in-depth Neural network and big data pretraining.

2. The deep learning based feature extractor

From a pattern recognition point of view, visual tracking is a model The classification problem separates the goal from the background. The Feature selection and extraction limit the execution of these tasks. The author uses a deep learning method as a feature extractor, which is possible. Features are represented by multiple descriptions.

Automatic encoder is an important theoretical basis for deep learning Consists of input, hidden layers, and output. The encoding and decoding way is described as:

$$h = sigm(Wx + b) \tag{1}$$

$$\widehat{x} = sigm(W^T h + c) \tag{2}$$

where $sigm(z) = \frac{1}{1+exp(-z)}$ denotes the sigmoid function, x denotes the vector transformed from the input data. Different from the traditional neural network, the weight matrix W and W^T are transposed. b and c respectively denote the encoding bias and the decoding bias [2]. The autoencoder learning is aimed at reconstructing the input to a maximum limit, which optimizes the target function:

$$min \sum_{i=1}^{N} ||x^j - \widehat{x}^j||_2^2 + \left(||W||_F^2 + ||W^T||_F^2\right) \tag{3}$$

The Denoising Autoencoder (DAE) [1] put some noise into the input data and thex is transformed to $\widehat{x}$, which is defined as:

$$\widehat{x} = x \times B \tag{4}$$

Based on greedy algorithm, superimposed denoising automatic encoder (SDAE). It's built with a few DAEs that have been trained in one step step. The author uses this network to extract functionality. Where elements of the matrix B are sampled from a Bernoulli distribution. The training process works similarly to AE as shown in Figure 1.

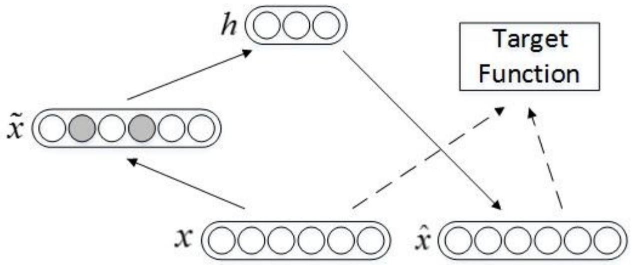

Figure 1. The working principle of DAE.

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3908–3916, 2015. 1

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012. 1

[3] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. In *IEEE International Conference on Computer Vision*, pages 2056–2063, 2014. 1