Deep Learning Tracker via SVM Ranking Vector

Fangfang Li

June 24, 2018

1. The improvements of the network

Wang's DLT tracker applied deep neural networks to visual tracking and achieved excellent performance. However, the frame rate is kept at 13 frames per second on average, which is not enough for practical applications [2]. In the article, the network was redesigned based on some new ideas proposed in recent years to adapt to visual tracking: a large number of pre-training data is limited, the accuracy of the test data set is reduced, and overfitting may occur. In the article, the author adopted a narrow network by reducing some of the model redundancy, which not only alleviates overfitting, but also greatly improves the update speed of the network.Hanton, the pioneer of deep learning, put forward the theory of dropping out of school. He said that half of the feature detectors are not used, which is capable. It is showed prevent the negative influence of different features in Figure 1 to describe the differences between the networks of bringing in the dropout and the original one [1].

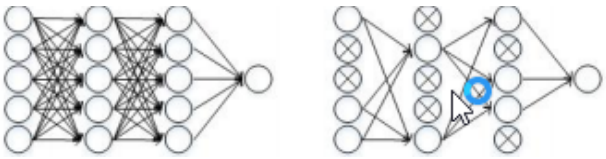

Figure 1. The theory of dropout.

The searching strategy of our algorithm is establishing a particle filter based on the theory of Brownian motion. The position coordinate is defined as an affine transformation:

$$\vec{x} = (x_t : y_t : s_t) \tag{1}$$

Where x_t and y_t denote the current offset describe the scaling change. The state parameters of candidate regions of the target follow the Gaussian distribution and they are mutual independence. In the article, the author design a ranking vector based on SVM to evaluate the particle confidence, which is between 0 and 1. The author selected the highest one as the result. The ranking vector come from the decision function, which is described as:

$$f(x) = sgn(w^* x + b^*) \tag{2}$$

2. Deep learning tarcker via svm ranking vector

Due to the fact that the network of our extractor is deep and the number of the parameter is large, the model is pre-trained under a large scale dataset to achieve an expressive power and construct a more robust visual tracking system [3]. We chose cifar-10 as the pre-training dataset. It contains 50000 training images (RGB, 32×32) and 10000 testing images (RGB, 32×32). The network is pre-trained through unsupervised learning such 50000 non-label images. The author warped the training images from 32×32 to 16×16 and transformed them to grey. Then they put them into the network. Due to the fact that the overall parameter tuning may interfere the completed ones, the pre-training of autoencoder adopt the way of "from the bottom up" and being fixed step by step. The author applied the theory of dropout to add some sparse restraints to the model and achieved a robust performance.

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *CVPR*, pages 3908–3916, 2015. 1

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

[3] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. In *ICCV*, pages 2056–2063, 2014. 1