SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
Volume 8 No 4
ISSN 2054-7390

Survey on Object Tracking using Deep Learning Paradigms

Saad A. Alhuzami
Department of Computer Science, College of Computer and Information Science, King Saud University, Riyadh, Saudi Arabia;
438106369@student.ksu.edu.sa

ABSTRACT

The field of object tracking extends across different domains. It is a major key player in the field of image processing and pattern recognition. Object tracking is the process of tracking an object over a continuous sequence of image frames to determine over time the relative movements or changes. With the massive advancements in the field of deep learning, the use of deep neural networks has risen due to their impressive accomplishments in object detection and tracking. In this Survey, the objective is to give a comprehensive overview of the recent attempts in the field of object tracking with a focus on the use of deep learning techniques and algorithms. The paper is divided into four sections; at first, we will give an overview of the recent work to highlight the techniques and methods which have been used in object tracking using deep learning. The second section focus is on the object tracking that uses convolutional networks techniques. The third section focuses on some of the recurrent neural networks to tack objects. The final section is concentrated on auto-encoders object tracking.

Keywords: Object tracking, Deep learning, CNN, RNN, Auto-encoders.

1 Introduction

Object tracking is the process of tracking an object over a continuous sequence of video frames to determine the relative movements or changes over time. This means that once an object is located in the first frame, the attributes of this object in the following frames will be tracked and monitored to detect any changes or disappearance. Object tracking can be perceived as an overlap between many fields such as image processing, pattern recognition, and artificial intelligence. The applications of Object tracking vary from video games to military applications. Object tracking adopts the algorithms of object detection, extraction, and recognition, in addition to decision making. On the other hand, Deep Learning (DL) is without a doubt one of the most successful research trends in machine learning. Since its proposal, it made a massive breakthrough in data processing in all of it forms, structured and unstructured, including voice, text, image, and video [37]. The advantage of deep learning consists in its powerful ability to extract and process feature in an unsupervised manner.

Deep learning performs multi-level learning, mapping and processing. In particular, the features extracted using DL proved to be highly discriminative. In opposite to typical hand-crafted feature extraction, DL does not require the expert supervision. With such improvements, DL exploits the big data produced by the modern technologies. In recent years, most of the successful object detection and image/video classification algorithms rely on DL [X]. Object tracking however joined that race later and related

DOI: 10.14738/tmlai.84.8485
Publication Date: 04th July, 2020
URL: http://dx.doi.org/10.14738/tmlai.84.8485

researches witnessed slower growth. Currently there are no recent surveys on object tracking using DL. This represents the main motivation for this research.

The aim of this paper is to give a comprehensive overview on the recent attempts in the field of object tracking based on deep learning in recent years.

As illustrated in Figure 1, the rest of this manuscript is as follows: Section 2 introduces the main DL paradigms. In section 3, we report and discuss existing object tracking works based on Convolutional Neural Networks (CNN), Recurrent Neural networks (RNN) and Auto-encoder respectively. Finally, section 4 concludes the survey.

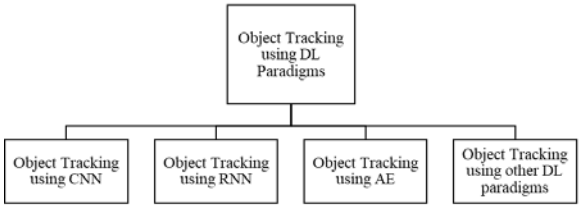


Figure 3: The structure of this survey.

2 Deep Learning Paradigms

Deep Learning (DL) can be introduced as a machine learning technique that relies on artificial neural networks that imitate the human neural system in processing data and recognizing patterns for future use in decision making. Also known as deep neural learning or deep neural networks, DL can be categorized into different paradigms.

2.1 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. Typically, the CNN layers consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers.

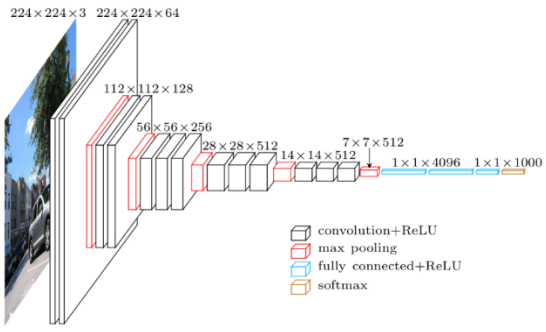


Figure 4: An example of CNN: The VGG network, introduced in 2014.

2.2 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is a class of artificial neural network where the connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic

Copyright © Society for Science and Education United Kingdom

behavior. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs.

2.3 Autoencoder

Autoencoder is an unsupervised backpropagation-based learning approach. The back-propagation algorithm is typically a gradient descend method with inherently slow convergence speed and local minima susceptibility. Auto Encoders (AE) utilizes the backpropagation, this allows them to obtain the optimal weights and overcome the limitations of backpropagation.

3 Object Tracking Using Deep Learning

In this section, we provide an overview on object tracking using state-of-the-art deep learning paradigms. The objective is to highlight the breakthroughs and advancements of the fields with more details.

3.1 Object tracking using Convolutional Neural Networks (CNN)

In 2017, Qiao and Xiaohuan [4] used deep convolutional neural networks for thermal Infrared Object Tracking. Specifically, they adopted the pre-trained VGG-Net, which was introduced in 2015 by Simonyan et al. [35], for thermal infrared object tracking. They coupled a correlation filter with convolutional multi-layers to extract the relevant features. Namely, they used the Kernelized Correlation Filter (KCF). Multiple weak trackers were used, each weak tracker gives a response map of the target's location. To improve the accuracy of the tracking, the authors used a simple scale estimation strategy. Their experimental results showed that their method achieved promising performance. One of their findings was that for thermal infrared tracking, the convolution layer is more effective than a fully-connected layer. Schulter et al. [9], in 2017, demonstrated an end-to-end formulation for association problems with the example of network flows for multi object tracking. They outlined a framework for tracking-by detection.

Let d represent the candidate detections in every frame where t is the frame index in a given video sequence. Each detection consists of a bounding box b(d) that describes the spatial location, a detection probability p(d) and a frame number t(d). The tracking algorithm associate each detection with an object trajectory Tk or reject it. A trajectory is defined as a set of detections belonging to the same object, i.e.,

$$Tk = \{d1\ k,\ldots,dNk\ k\ \}$$

where Nk defines the size of the trajectory. Only bounding boxes from different frames can belong to the same trajectory. The number of trajectories |T| is unknown and needs to be inferred as well. The proposed work focuses on network flow formulation from Zhang et. al [32] to solve the association problem.

A hierarchical convolutional feature tracker (HCFT) was proposed by Mozhdehi et al. [5] in 2017. Their novel framework for visual tracking was based on the integration of a deep CNN and a particle filter. the framework has a motion model that predicts the position of the object at each frame and the HCFT-CNN-based tracker refines the particle positions and generating the corresponding weights. The evaluation of the framework performance showed an improvement in HCFT performance in conditions such as deformation, illumination, in-plane and out-of-plane rotations. The motion model in conjunction with the particle filter's ability to sample several image patches allow it to overcome temporary target losses caused by dramatic temporary appearance changes or occlusions. Sequential Monte Carlo methods have

been used to incorporate temporal information and hence develop robust object tracking algorithms that build on features that range from simple color histograms to support vector machines.

Using CNN with particle filtering also, Qian et al. [6] proposed a novel object tracking algorithm based on a pre-trained representation. Their algorithm pre-trains a simplified CNN using a huge set of videos. The particle filtering is combined to the fully-connected layer in the pre-trained CNN. Deep representations and hand-crafted features help to model tracking. To optimize the particles' distribution, the velocity and acceleration information aids to calculate dynamic model. Mean while, our algorithm updates the tracking model in a lazy manner to avoid shift and expensive computation. As compared to previous methods, our results demonstrate superior performances in existing tracking benchmarks. In 2018, a deep learning based multiple objects Visual tracking on embedded system for IoT and mobile edge computing applications was proposed in [8]. The results of the present work focus on the frame rate and power consumption of the algorithm, which are hardware dependent. In their work, in order to make state-of-the-art networks useful at IoT end-nodes, more attention must be paid to their power consumption and compute and memory demands. Most of the CNNs energy consumption is related to data movement rather than computation itself. Hence, it is necessary to make a great effort not only at network design but also at hardware implementation in order to exploit CNNs potential in the computer vision field for low-power real time mobile systems and IoT applications. Huang et al. [10] addressed the limitation of existing object tracking methods, where CNN employ either a scalar score or a confidence map as CNN's output. As they stated, these methods suffer the infeasibility of estimating the object's accurate scale and rotation angle. Their work proposes to take a binary mask as output of CNN for tracking. Specifically, they used a state-of-the-art semantic segmentation model named DeepLab [38] as their base model and transfer it to object tracking. The resulting tracker achieved high performance among 20 state-of-the-art real-time trackers on all the three datasets including OTB-2015 [39], VOT2016 [40] and TempleColor [41]. The model works as follow; Given the initial frame, they tune the base model by augmented samples. Then, they pass a region of Interest through the network. After that, they get a rough mask, which is the base to estimate a rectangle as the object's bounding box. To derive a more accurate bounding box, they proposed to use a two-dimensional binary mask as the output of CNN. They adopted one of the state-of-the-arts as a base model and transfer it to tracking. Note that since they supervise the training using a rough label, the model's output is imprecise. Thus, pixel-level segmentation is not reachable, and it is not a necessity since we pursue box-level tracking. In semantic segmentation, the spatial resolution is of crucial importance. However, the resolution is significantly reduced when CNN are employed. To address the resolution reducing challenge, they used Deeplab model which utilizes Atrous convolution [42] to substitute for the down-sampling operators, e.g., max-pooling (Atrous convolution is a technique that inserts holes between nonzero filter taps). To make DeepLab work for their task, they randomize all the fully convolutional layers and leave the output only for two classes, the background and the tracked object. J. Son et al. [15] proposed a quadruplet network that deals with multi-level rank orders by generalizing Siamese and triplet network. They were motivated by the fact that, for data association in multi-object tracking, the embedding should consider not only class labels of detected objects but also their detection timestamps since the object appearances id changing over time. Quadruplet association relies on accurate object localization, they employed bounding-box regression to learn the network. The Quad-CNN minimizes a multi-task loss on each patch in a mini-batch corresponding to both data association and bounding box regression.

Copyright © Society for Science and Education United Kingdom

3.2 Object Tracking using Recurrent Neural Networks

In [14], the authors claimed that it is difficult to use Long Short-Term Memory LSTMs [43] to model object appearance over long sequences. They investigated the issues in utilizing LSTM for tracking. Leveraging intuition from classical recursive least squares regression, they proposed a new type of LSTM that is suitable for learning sequential appearance models. Whereas in a conventional LSTM, the memory and the input have a linear relationship, they proposed a Bilinear LSTM, the LSTM memory serves as the building blocks of a predictor, which leads to the output being based on a multiplicative relationship between the memory and the input appearance. Based on this novel LSTM formulation, they were able to build a RNN for scoring object tracks that combines long-term appearance and motion information. This new track scorer is then utilized in conjunction with an established near-online multi-object tracking approach. Their approach combines the benefits of deep feature learning with the practical utility of a near-online tracker. As part of their contribution, they proposed a training methodology for generating sequential training examples from multi-object tracking datasets that accounts for the cases where detections could be noisy for many frames. They developed systematic data augmentation methods that allow their near-online approach to take advantage of long training sequences and survive scenarios with detection noise and dozens of frames of consecutive missing detections. The results obtained using near-online approaches in the MOT challenge [14] were outstanding. In 2017, A. Milan. et al. [17] were inspired by the success of RNN and its application to language modeling (Vinyals et al. 2015). However, it is not straight forward to apply the same strategies to the problem of multitarget tracking for numerous reasons. Therefore, they introduced a method to address these issues and prove that RNNs can be used as an end-to-end learning model for online tracking of multiple targets in realistic scenarios. One of the advantages of their approach is that it is completely model-free. In other words, it does not require any prior knowledge about the target dynamics and/or the clutter distributions. It can therefore capture the linear, the non-linear, and the higher-order dependencies. Another contribution was to further show, that a model for the challenging combinatorial problem of data association including birth and death of targets can be learned entirely from data. Furthermore, they presented a method to generate random amounts of training data by sampling from a generative model. In [21], the researchers introduced a combination of an LSTM and a residual network (RLSTM) for object tracking. The difference between their RLSTM and the typical LSTM is that the proposed RLSTM captures the spatio-temporal attention based on the residual framework. This spatio-temporal attention is constructed more robustly using a rule-based RLSTM learning with the similarity score and similarity score variation. The obtained results showed that RLSTM tracker outperforms the existing trackers including the Siamese trackers [45], the attention trackers [44], and the correlation trackers [46]. Moreover, they proved that the resulting tracking speed is comparable to the state-of-the-art deep trackers. As part of their future plan, they planned to focus on the robustness of RLSTM tracker using reinforcement learning.

3.3 Object Tracking using Auto-encoder

In 2016, Wua et al. [20] proposed a method based on a stacked autoencoders with Factorial Hidden Markov Model (FHMM) [47] for object tracking. In their method, the observation likelihood is based on sub observation models. Each model is implemented by the Stacked Auto-Encoder architecture and the motion model is extended to FHMM. The proposed method include three main aspects: (i) stable factor, (ii) inference stage, and (iii) training stage. The stable factor was used to characterize the reliability of each

sub-region model. In fact, the original HMM was extended to a FHMM by integrating the stable factor as an additional latent variable. For the inference stage, under the FHMM framework, the inference process for the final tracking result is converted to an estimation problem, where the probability likelihood for each latent state particle is judged by the response values of multiple observation models. Regarding the training stage, the feature learning mechanism is crucial for the robustness of the proposed tracker, and the observation model must be trained carefully during the whole tracking process. There are two parts of training stage for the observation model (i.e. SAE): offline training and online update scheme. The former provides the generic object feature basis, while the latter refines the distinctive feature of the tracking object. Hua et al. [31], inspired by previous reports on successful multiple-level feature representation using deep learning and spare dictionary coding, outlined an object feature learning approach based on Stacked Denoising Autoencoder [48]. Particularly, they proposed a tracking algorithm based on SDAE associated with a genetic algorithm optimization. Each layer of their SDAE architecture represents a different feature space. Thus, in the encoding process, the traditional hand-crafted feature representation is replaced by training a stacked denoising autoencoder on top of a dual-layer sparse-coded representation of the input image.

Recently, J. Xin et al. [7] designed a K-Sparse Stacked Denoising Auto-Encoder (kSSDAE) based tracker for robust outdoor vehicle tracking under particle filter framework. The tracking process was divided into three stages; (i) The offline training of SDAE, (ii) The construction of the classification neural network and (iii) The estimation of the object state. In the offline training stage, they have adopted a pre-trained SDAE model which was trained using a greedy layer-wise [49]. The model was trained on 1 million images as training data. This collection was sampled randomly from the Tiny Images dataset [50]. Their main contribution lies in the classification neural network construction stage. In fact, they proposed kSSDAE model to adapt the object appearance change in the tracking process and train a classification neural network to compute the confidence of each particle. This was the key in the attainment of the high tracking performance. Without kSSDAE, the input cannot be guaranteed to have a sparse representation to extract more effective features. Additionally, they proposed a module called estimation of object state where the object state is estimated using the particle with the maximum confidence calculated by the classification neural network.

3.4 Object Tracking using other Deep learning paradigms

In this section we outline the tracking approaches which cannot be presented as CNN, RNN or AE based approaches. Lately, the authors in [1] proposed a self-adaptive deep neural network suitable for object labeling/tracking. The heart of the proposed self-adaptive deep neural network is a multilayer deep neural network structure, composing of the network pre-training and the classification phase. Network pre-training is responsible for providing a compressed form of the input data, increasing therefore the classification performance. The output of this module is a p-dimensional vector which the elements represent the membership of each image pixel to belong to one of the p pre-defined classes. The Discrete Cosine Transform (DCT) on three color components is first applied on the input data for blocks of 32×32-pixel size. Additionally, DCT is an efficient alternative to capture the spatio-temporal properties of the image data and therefore, very suitable for object classification/tracking. The second stage of the proposed self-adaptive deep neural network structure is a decision-making module. This module includes spatio-temporal coherency metrics which are intended to evaluate the classification performance of the network. The purpose of this module is to determine whether network re-training should be activated or

Copyright © Society for Science and Education United Kingdom

not. In case that the spatial temporal coherency metrics detect no significant changes, network parameters remain intact. On the contrary, in case of a large change, the weights of the network are automatically updated. The few labelled data samples used for network re-training are estimated in this paper by a motion-based tracking algorithm. The purpose of this module is to select a set of representative samples of the background and foreground objects expressing the current environmental conditions. The deep neural network module provides the final and precise classification output. The proposed adaptable learning paradigm exploits both the discriminative and generative constraints. The discriminative constraints model the current visual conditions, while the generative constraints the previous network knowledge.

Table 1 summarizes the reviewed literature. It shows the published literature, the techniques and the algorithms used to address the object tracking problem, and the publication date.

#	Journal/Conference	Year	Title	DL Technique	Model/Method
1	Springer - Multimedia Tools Applications	2018	Adaptable deep learning structures for object labeling/ tracking under dynamic visual environments	deep multi-layered neural networks	Self-adaptive deep neural network requiring few labelled data for pre-training and low complexity
2	IEEE International Conference on Image Processing (ICIP)	2014	An ensemble of deep neural networks for object tracking	Autoencoder	stacked denoising autoencoder and deep neural network
3	Applied soft computing	2015	CNN Tracker: online discriminative object tracking via deep convolutional neural networks	CNN	Deep CNN to enhance the appearance model
4	Knowledge-Based Systems	2017	Deep Convolutional Neural Networks for Thermal Infrared Object Tracking	CNN	Deep Convolutional Neural Networks
5	IEEE International Conference on Image Processing (ICIP)	2017	Deep Convolutional particle filter for visual tracking	CNN and Particle filters	Hierarchical Convolutional Feature Tracker
6	Signal processing: Image communication	2017	Deep learning assisted robust visual tracking with adaptive particle filtering	CNN and Particle filters	Combining PF to fully-connected layer of a pre-trained CNN
7	Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)	2017	Deep learning for robust outdoor vehicle visual tracking	Autoencoder-CNN	k-sparse stacked denoising auto-encoder connected to CNN
8	IEEE Internet of Things Journal	2018	Deep Learning-Based Multiple Object Visual Tracking on Embedded System for IoT and Mobile Edge Computing Applications	CNN	low-power and real time deep learning-based multiple object visual tracking
9	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2017	Deep Network Flow for Multi-Object Tracking	CNN	Deep convolutional neural networks

10	Springer – Multimedia Tools and Applications	2019	Exploiting weak mask representation with convolutional neural networks for accurate object tracking	CNN	CNN with semantic segmentation
11	Human Interaction Prediction Using Deep Temporal Features (book)	2016	Fully-Convolutional Siamese Networks for Object Tracking	CNN	Fully-Convolutional Siamese Networks
12	IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	2016	Joint Learning of Siamese CNNs and Temporally Constrained Metrics for Tracklet Association	CNN	Convolutional neural networks and temporally constrained metric
13	IEEE: Special section on sequential data modeling and its emerging applications	2018	Learning Deconvolutional Network for Object Tracking	CNN	Deconvolutional neural network
14	15th European Conference	2018	Multi-object Tracking with Neural Gating Using Bilinear LSTM	RNN	Bilinear LSTM to store building blocks of a linear predictor coupled with a multiplicative relation between input and memory
15	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2018	Multi-Object Tracking with Quadruplet Convolutional Neural Networks	CNN	Quadruplet Convolutional Neural Networks
16	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2015	Multiple People Tracking by Lifted Multi-cut and Person Re-identification	CNN	Lifted Multi-cut and Person Re-identification
17	Association for the Advancement of Artificial Intelligence	2017	Online Multi-Target Tracking Using Recurrent Neural Networks	RNN	temporal RNN for state prediction and LSTM for data association
18	Neurocomputing	2017	Online Object tracking based on CNN with spatial-temporal saliency guided sampling	CNN	CNN tracking incorporated with spatial-temporal saliency detection
19	IEEE Winter Conference on Applications of Computer Vision (WACV)	2018	Recurrent Autoregressive Networks for Online Multi-Object Tracking	RNN	Temporal generative model with internal and external memories
20	Neurocomputing	2016	Regional deep learning model for visual tracking	Autoencoder	SAE with extension to Factorial Hidden Markov Model
21	IEEE Signal processing letters	2018	Residual LSTM Attention Network for Object Tracking	RNN	Rule-based Residual LSTM
22	IEEE Transactions on Multimedia	2019	Robust Object Tracking Using Manifold Regularized Convolutional Neural Network	CNN	Manifold regularized CNNs with label propagation
23	Springer – Multimedia tools and applications	2018	Robust tracking based on H-CNN with low-resource sampling and scaling by frame-wise motion localization	CNN	Hierarchical CNN Incorporated with inter-frame motion prediction

Copyright © Society for Science and Education United Kingdom

24	IEEE Transactions on image processing	2016	Robust Visual Tracking via Convolutional Networks Without Training	CNN	CNN with a lightweight structure for visual tracking
25	IEEE International Conference on Image Processing, ICIP	2014	Semi-supervised deep learning for object tracking and classification	Adaptive semi supervised learning	SAE to configure the network, Hidden Layer Weight Adaptation (US) Output Layer Weight Adaptation (S)
26	IEEE International Conference on Image Processing, ICIP	2017	Simple Online and Real-time tracking with deep association metric	CNN	Integrating pre-trained CNN with Simple Online and Realtime Tracking (SORT)
27	Springer – Multimedia tools and applications	2018	Spatial and semantic convolutional features for robust visual object tracking	CNN	Translation correlation filters to localize the target with spatial and semantic convolutional features with spatial features reduction for information retainment.
28	Springer – Multimedia tools and applications	2018	Taking full advantage of convolutional network for robust visual tracking	CNN	leverages hierarchical convolution layers and the FC layers to enhance the tracking robustness
29	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2017	Tracking the untrackable: Learning to Track Multiple Cues with Long-Term Dependencies	RNN and CNN	RNN constructed on CNN for appearance model and LSTM for motion
30	IEEE First International Smart Cities Conference (ISC2)	2015	Visual Object Tracking via Deep Neural Network	Autoencoder	Training an SDAE to form a classifier for target training.
31	Springer – Multimedia tools and applications	2017	Visual tracking based on stacked Denoising Autoencoder network with genetic algorithm optimization	Autoencoder	Stacked Denoising Autoencoder with genetic algorithm optimization

Table 2: Published work of Object Tracking using deep learning

4 Conclusion

The objective of this paper is to highlight the recent work in using different and various representative deep learning techniques for object tracking. It was shown that the expansion of deep learning and the vast improvements of its techniques had a massive impact on the field of object tracking. These deep learning object trackers lately often cast the problem of tracking a target as a spatio-temporal detection in a sequence of multiple adjacent frames. However, the wide usage of deep learning techniques in this fields has yielded a new type of challenges, which was presented in the surveyed work. The traditional object tracking is not omitted yet due to the processing power consumption of using deep learning. Most of the observed work are dependent on a pre-trained network, which means that in the near-future, the results of the tracking mechanism will enhance as the pre-trained model are exposes to larger and more comprehensive datasets. Finally, most work currently focuses on the adoption of neural networks to increase implicitly learning the temporal dependencies during tracking directly in a bottom-up fashion from the data itself.

REFERENCES

[1] Nikolaos Doulamis. 2018. Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. Multimedia Tools Appl. 77, 8 (April 2018), 9651-9689. DOI: https://doi.org/10.1007/s11042-017-5349-7

[2] X. Zhou, L. Xie, P. Zhang and Y. Zhang, "An ensemble of deep neural networks for object tracking," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 843-847.

[3] Chen, Yan & Yang, Xiangnan & Zhong, Bineng & Pan, Shengnan & Chen, Duansheng & Zhang, Huizhen. (2015). CNNTracker: Online discriminative object tracking via deep convolutional neural network. Applied Soft Computing. 38. 10.1016/j.asoc.2015.06.048.

[4] Liu, Qiao & Lu, Xiaohuan & He, Zhenyu & Zhang, Chunkai & Chen, Wen-Sheng. (2017). Deep Convolutional Neural Networks for Thermal Infrared Object Tracking. Knowledge-Based Systems. Knowledge-Based Systems Volume 134, 15 October 2017, Pages 189-198 10.1016/j.knosys.2017.07.032.

[5] R. J. Mozhdehi and H. Medeiros, "Deep convolutional particle filter for visual tracking," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 3650-3654.

[6] Xiaoyan Qian, Lei Han, Yuedong Wang, and Meng Ding. 2018. Deep learning assisted robust visual tracking with adaptive particle filtering. Image Communication. 60, C (February 2018), 183-192.

[7] J. Xin, X. Du and J. Zhang, "Deep learning for robust outdoor vehicle visual tracking," 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, 2017, pp. 613-618.

[8] Blanco-Filgueira, Beatriz & Garcia-Lesta, Daniel & Fernandez-Sanjurjo, Mauro & Manuel Brea, Victor & Lopez, Paula. (2019). Deep Learning-Based Multiple Object Visual Tracking on Embedded System for IoT and Mobile Edge Computing Applications. IEEE Internet of Things Journal. PP. 1-1. 10.1109/JIOT.2019.2902141.

[9] Schulter, Samuel et al. “Deep Network Flow for Multi-object Tracking.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 2730-2739.

[10] Huang, J., Zhou, W., Tian, Q. et al. “Exploiting weak mask representation with convolutional neural networks for accurate object tracking” Multimedia Tools Appl (2019). https://doi.org/10.1007/s11042-019-7219-y

[11] Bertinetto, Luca & Valmadre, Jack & Henriques, Joao & Vedaldi, Andrea & H. S. Torr, Philip. (2016). Fully-Convolutional Siamese Networks for Object Tracking. 9914. 850-865. 10.1007/978-3-319-48881-3_56.

[12] Wang, Bing & Wang, Li & Shuai, Bing & Zuo, Zhen & Liu, Ting & Luk Chan, Kap & Wang, Gang. (2016). Joint Learning of Convolutional Neural Networks and Temporally Constrained Metrics for Tracklet Association. 386-393. 10.1109/CVPRW.2016.55.

[13] X. Lu, H. Huo, T. Fang and H. Zhang, "Learning Deconvolutional Network for Object Tracking," in IEEE Access, vol. 6, pp. 18032-18041, 2018.

[14] Kim, Chanho & Li, Fuxin & Rehg, James. (2018). Multi-object Tracking with Neural Gating Using Bilinear LSTM: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VIII. 10.1007/978-3-030-01237-3_13.

Copyright © Society for Science and Education United Kingdom

[15] J. Son, M. Baek, M. Cho and B. Han, "Multi-object Tracking with Quadruplet Convolutional Neural Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3786-3795.

[16] S. Tang, M. Andriluka, B. Andres and B. Schiele, "Multiple People Tracking by Lifted Multicut and Person Re-identification," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3701-3710.

[17] Milan, Anton & Rezatofighi, Hamid & Dick, Anthony & Schindler, Konrad & Reid, Ian. "Online Multi-target Tracking using Recurrent Neural Networks". 2016 Association for the Advancement of Artificial Intelligence.

[18] Zhang, Peng et al. "Online object tracking based on CNN with spatial-temporal saliency guided sampling." Neurocomputing 257 (2017): 115-127.

[19] K. Fang, Y. Xiang, X. Li and S. Savarese, "Recurrent Autoregressive Networks for Online Multi-object Tracking," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 466-475.

[20] G. Wu, W. Lu, G. Gao, C. Zhao, J. Liua. "Regional deep learning model for visual tracking". 2016 Neurocomputing Volume 175, Part A, 29 January 2016, Pages 310-323.

[21] H. Kim and R. Park, "Residual LSTM Attention Network for Object Tracking," in IEEE Signal Processing Letters, vol. 25, no. 7, pp. 1029-1033, July 2018.

[22] H. Hu, B. Ma, J. Shen, H. Sun, L. Shao and F. Porikli, "Robust Object Tracking Using Manifold Regularized Convolutional Neural Networks," in IEEE Transactions on Multimedia, vol. 21, no. 2, pp. 510-521, Feb. 2019.

[23] Zhang, Peng & Zhuo, Tao & Huang, Hanqiao & Chen, Kangli & Zhang, Bo & Kankanhalli, Mohan. (2017). Robust tracking based on H-CNN with low-resource sampling and scaling by frame-wise motion localization. Multimedia Tools and Applications. 77. 10.1007/s11042-017-4493-4.

[24] K. Zhang, Q. Liu, Y. Wu and M. Yang, "Robust Visual Tracking via Convolutional Networks Without Training," in IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1779-1792, April 2016.

[25] N. Doulamis and A. Doulamis, "Semi-supervised deep learning for object tracking and classification," *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, 2014, pp. 848-852.

[26] N. Wojke, A. Bewley and D. Paulus, "Simple online and real-time tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 3645-3649.

[27] Zhang, Jianming & Jin, Xiaokang & Sun, Juan & Wang, Jin & Kumar, Arun. (2018). Spatial and semantic convolutional features for robust visual object tracking. Multimedia Tools and Applications. 10.1007/s11042-018-6562-8.

[28] Guan, Hao and Baozhong Cheng. "Taking full advantage of convolutional network for robust visual tracking." Multimedia Tools and Applications (2018): 1-15.

[29] Sadeghian, Amir & Alahi, Alexandre & Savarese, Silvio. (2017). Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies.

[30] T. Xu and X. Wu, "Visual object tracking via deep neural network," 2015 IEEE First International Smart Cities Conference (ISC2), Guadalajara, 2015, pp. 1-6.

[31] Hua, Weixin et al. “Visual tracking based on stacked Denoising Autoencoder network with genetic algorithm optimization.” Multimedia Tools and Applications 77 (2017): 4253-4269.

[32] Li Zhang, Yuan Li and R. Nevatia, "Global data association for multi-object tracking using network flows," 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8.

[33] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[34] S. Tang, B. Andres, M. Andriluka and B. Schiele, "Subgraph decomposition for multi-target tracking," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 5033-5041.

[35] Tang, Siyu & Andres, Bjoern & Andriluka, Mykhaylo & Schiele, Bernt. (2016). Multi-person Tracking by Multicut and Deep Matching. 9914. 100-111. 10.1007/978-3-319-48881-3_8.

[36] Kim, Chanho & Li, Fuxin & Ciptadi, Arridhana & Rehg, James. (2015). Multiple Hypothesis Tracking Revisited. 4696-4704. 10.1109/ICCV.2015.533.

[37] Feng, Xiaoyu & Mei, Wei & Hu, Dashuai. (2016). A Review of Visual Tracking with Deep Learning. 10.2991/aiie-16.2016.54.

[38] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2016) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915

[39] Wu Y, Lim J, Yang MH (2015) Object tracking benchmark. TPAMI 37(9):1834–1848

[40] Kristan M, Leonardis A, Matas J, Felsberg M, Pflugfelder R, Cehovin L, Vojir T, et al (2016) The visual object tracking VOT2016 challenge results, pp 777–823

[41] Liang P, Blasch E, Ling H (2015) Encoding color information for visual tracking: algorithms and benchmark. TIP 24(12):5630–5644

[42] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2016) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915

[43] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9, 1735–80 (12 1997)

[44] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, “Recurrent models of visual attention,” in Proc. Int. Conf. IEEE Neural Inf. Process. Syst., Montreal, QC, Canada, Dec. 2014, pp. 2204–2212.

[45] R. Tao, E. Gravves, and A. W. M. Smeulders, “Siamese instance search for tracking,” in Proc. Int. Conf. IEEE Comput. Vis. Pattern Recog., Las Vegas, NV, USA, Jun. 2016, pp. 1420–1429.

[46] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, “Long-term correlation tracking,” in Proc. Int. Conf. IEEE Comput. Vis. Pattern Recog., Boston, MA, USA, Jun. 2015, pp. 5388–5396.

[47] M.I.J. Zoubin Ghahramani, Factorial hidden Markov models, Mach . learn. 29 (2-3) (1997) 245-273.

Copyright © Society for Science and Education United Kingdom

[48] Vincent P, Larochelle H, Lajoie I et al (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 11:3371–3408

[49] Wang N, Yeung DY, "Learning a deep compact image representation for visual tracking," Advances in Neural Information Processing Systems, pp. 809-817, Jan. 2013.

[50] Torralba A, et al, "80 million tiny images: A large data set for nonparametric object and scene recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1958-1970, Nov. 2008.