**报告正文**

**（一）立项依据与研究内容**

**1．项目的立项依据**（研究意义、国内外研究现状及发展动态分析，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）

1.1. **研究意义**

略。

**1.2. 国内外研究现状**

### 2) 视觉目标跟踪

视觉目标跟踪问题[22][52]以输入的图像序列为研究对象，确定其中特定目标在图像平面的运动轨迹。常见的视觉跟踪问题由于面向诸多不同应用场景而演化出了多个分支问题，并产生了很多具体跟踪算法。比如，针对跟踪任意运动目标的通用跟踪问题而开展的模型非固定式在线视觉跟踪算法；针对监控场景下广泛存在的特定运动目标(如行人、车辆等)而开展的基于特定模型检测的离线/在线多目标跟踪算法、针对室内场景基于RGBD(Red, Green, Blue, Depth)摄像机的基于深度信息的RGBD 跟踪算法以及针对机器人自主导航应用中的视觉里程计开展的基于关键特征点匹配的相关跟踪算法。本项目主要针对模型非固定式在线视觉跟踪展开研究。这类跟踪问题只能依赖于待跟踪目标初始位置标注信息及其视觉表观，在没有任何其它关于该目标先验信息的情况下，通过在线学习并更新表观模型来自适应于目标表观变化从而完成跟踪。但是，目标的视觉表观多种多样且具有较大的差异，设计一个泛化的目标跟踪算法仍然具有很大的难度。

模型非固定式在线视觉跟踪既要利用底层的图像分割、运动检测和物体识别等信息来对跟踪进行初始化，同时也可以为上层的运动模式学习以及目标行为理解和描述提供基础，因此具有非常重要的理论研究价值。最近几年，随着视觉跟踪测试库[17][18]和视觉跟踪竞赛[19][20][21]的提出，针对模型非固定式在线视觉跟踪的研究有了更系统、更公平的评价体系和评价指标。考虑到模型非固定式在线视觉跟踪问题的挑战性，特别是针对模型非固定式在线视觉跟踪需要解决由于光照变化、遮挡和姿态形变等剧烈的目标表观变化导致的跟踪漂移问题，这类跟踪问题会涉及到与诸多复杂的机器学习算法的结合，这对于其它计算机视觉问题的解决也有一定的借鉴意义。当前基于深度学习的模型非固定式在线视觉跟踪器主要可分为：基于生成对抗网络的跟踪器、基于图卷积网络的跟踪器、基于循环神经网络的跟踪器、基于孪生网络的跟踪器等。

A) 基于生成对抗网络的跟踪器：生成对抗网络（GAN）可以通过CNN从随机噪声生成逼真的图像。 生成对抗网络包含两个子网，一个充当生成器，另一个充当判别器。 生成器旨在合成图像以欺骗判别器，而判别器则试图正确区分真实图像和生成器合成的图像。通过相互竞争来同时训练生成器和鉴别器。对抗学习的优势在于，所训练的生成器可以生成与训练样本相似的图像统计信息，从而使判别器无法区分。生成对抗网络的进步吸引了包括目标跟踪在内的各种计算机视觉应用的关注。在[119]中，作者利用生成对抗网络产生的样本辅助跟踪器的学习。文章指出，由于以下问题，现有视觉跟踪器的性能可能会受到限制：i）采用密集采样策略生成的正样本会降低样本的多样性；ii）即使收集到大规模的训练数据集，具有挑战性的训练数据也是有限的。作者提出了VITAL算法来通过对抗学习解决这两个问题。为了增加正样本，作者使用一个生成网络随机生成模板，这些模板用于自适应过滤输入特征以捕获各种表观变化。通过对抗学习获得的模板，可以提供最鲁棒的目标特征。此外，为了解决类别不平衡的问题，作者提出了一个高阶成本敏感损失，从而有助于训练分类网络。在[120]中，作者通过对抗生成学习产生难例正样本进行跟踪。具体来说，作者假设目标都位于流形上，因此，引入正样本生成网络（PSGN），通过遍历已构建的目标流形来采样大量训练数据。生成的各种目标图像可以丰富训练数据集并增强目标跟踪器的鲁棒性。为了使跟踪器对遮挡更加鲁棒，作者提出了一个变换网络，该网络可以生成用于跟踪算法的难例样本。

B) 基于图卷积网络的跟踪器：图卷积神经网络（Graph Convolutional Network）是一种能对图数据进行深度学习的方法。在目标跟踪中，图卷积网络用于捕获目标样本的结构特征。在[121]中，作者指出时空信息可以用于增强目标表示，并且上下文信息对于目标的定位很重要。为了全面利用历史目标样本的时空结构并从上下文信息中受益，作者提出了一种用于高性能视觉跟踪的新型图卷积跟踪（GCT）方法。具体而言，GCT将两种类型的图卷积网络（GCN）合并到用于目标表观建模的孪生框架中。作者采用时空GCN来建模历史目标样本的结构化表示。而上下文GCN被设计为利用当前帧的上下文来学习用于目标定位的自适应特征。在[122]中，作者同样使用GCN模块来学习目标跟踪的结构特征。首先，作者利用双路径网络提取异构特征。然后，作者采用GCN模块来构建具有结构化信息的要素。

C) 基于循环神经网络的跟踪器：循环神经网络（Recurrent Neural Network, RNN）是一类以序列（sequence）数据为输入，在序列的演进方向进行递归（recursion）且所有节点（循环单元）按链式连接的神经网络。RNN在建模序列数据方面引起了越来越多的关注。 这些应用程序涵盖了多语言机器翻译，动作识别，场景标记，语音识别等。在目标跟踪中，可利用RNN建模目标的复杂远程依赖关系。RTT[123]尝试识别并利用那些对整个跟踪过程有益的可靠部分。为了解决遮挡并发现可靠的组件，RTT中使用了多方向递归神经网络（RNN），通过从多个方向遍历候选空间区域来捕获远程上下文线索。从RNN生成的置信度图用于抑制背景噪声，同时充分利用来自可靠部分的信息，来自适应地区分判别相关滤波器的学习。在[124]中，作者提出了一种能够将时间信息整合到模型中的实时目标跟踪器。该跟踪器不是专注于有限的一组目标或在测试时训练一个模型来跟踪特定的实例，而是在大量不同的目标上预先训练一个通用跟踪器，并进行实时的在线更新。

D) 基于孪生网络的跟踪器：孪生网络是一种用于度量学习的有监督模型。孪生网络具有两个参数共享的子网络，可以学习两幅输入图像之间的特征相似性。由于优越的性能，基于孪生网络的跟踪器已经成为当前目标跟踪领域的主流。在SiamFC[125]中证明了使用孪生网络解决跟踪问题的有效性。具体来说，作者训练了一个孪生网络以在较大的搜索图像中定位模板图像。利用互相关操作以滑动窗口的方式获得目标位置的响应图，从而对目标进行实时定位。在SiamRPN[126]中，跟踪器由用于特征提取的孪生网络和包括分类分支和回归分支的region proposal子网络组成。受益于跟踪器的改进，传统的多尺度测试和在线微调可以被丢弃。

**参考文献**

1. Y. Ma, and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. *ACM International Conference on Multimedia* **(ACM MM)**, pp. 374-381, 2003.
2. X. Hou, and L. Zhang. Saliency detection: A spectral residual approach. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1-8, 2007.
3. P. Federico, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp.733-740, 2012.
4. J. Zhang, and S. Sclaroff. Saliency detection: A boolean map approach. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 153-160, 2013.
5. T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1-8, 2007.
6. J. Yang, and M. Yang. Top-down visual saliency via joint CRF and dictionary learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 2296-2303, 2012.
7. L. Mai, Y. Niu, and F. Liu. Saliency aggregation: A data-driven approach. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1131-1138, 2013.
8. P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **(PAMI)**, vol. 32, no.9, pp. 1627-1645, 2010.
9. T. Malisiewicz, A. Gupta, and A. Efros. Ensemble of exemplar-svms for object detection and beyond. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 89-96, 2011.
10. X. Ren, and D. Ramanan. Histograms of sparse codes for object detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 3246-3253, 2013.
11. G. Ross, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 580-587, 2014.
12. K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *European Conference on Computer Vision* (**ECCV**), pp. 346-361, 2014.
13. G. Ross. "Fast R-CNN." *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 1140-1448, 2015.
14. S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks." *Advances in Neural Information Processing Systems* **(NIPS)**, pp. 91-99, 2015.
15. R. Joseph, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 779-788, 2016.
16. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A.C. Berg. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision, vol. 1, pp. 21-37, 2016.
17. Y. Wu, J. Lim, and M. Yang. Online object tracking: a benchmark. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**)*,* pp. 2411-2418, 2013.
18. Y. Wu, J. Lim, and M. Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 37(9): 1834-1848, 2015.
19. M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, et al. The visual object tracking VOT2014 challenge results. *European Conference on Computer Vision Workshops***(ECCVW)**, pp. 1-23, 2014.
20. M. Kristan, J. Matas, A. Leonardis, and M. Felsberg, et al. The visual object tracking VOT2015 challenge results. *IEEE International Conference on Computer Vision Workshops* **(ICCVW)**, pp. 1-23, 2015.
21. M. Kristan, A. Leonardis, and J. Matas, et al. The Visual Object Tracking VOT2016 Challenge Results. *European Conference on Computer Vision***(ECCV)**, pp. 191-217, 2016.
22. D. Ross, J. Lim, R. Lin, and M. Yang. Incremental learning for robust visual tracking. *International Journal on Computer Vision* **(IJCV)**, 77(1-3):125–141, 2008.
23. W. Hu, X. Li, W. Luo, X. Zhang, S. Maybank, and Z. Zhang. Single and multiple object tracking using Log-Euclidean Riemannian subspace and block-division appearance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 34(12):2420–2440, 2012.
24. D. Ross, J. Lim, R. Lin, and M. Yang. Incremental learning for robust visual tracking. *International Journal on Computer Vision* **(IJCV)**, 77(1-3):125–141, 2008.
25. W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, and Z. Zhang. Incremental tensor subspace learning and its applications to foreground segmentation and tracking. *International Journal on Computer Vision* (**IJCV**), 91(3):303-327, 2011.
26. J. Wen, X. Li, X. Gao, and D. Tao. Incremental learning of weighted tensor subspace for visual tracking. *IEEE International Conference on Systems, Man and Cybernetics* **(ICSMC)**, pp. 3688-3693, 2009
27. C. Bao, Y. Wu, H. Ling , and H. Ji. Real time robust L1 tracker using accelerated proximal gradient approach. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1830-1837, 2012.
28. T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via multi-task sparse learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 2042-2049, 2012.
29. X. Mei and H. Ling. Robust visual tracking and vehicle classification via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 33(11):2259–2272, 2011.
30. X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai. Minimum error bounded efficient ℓ1 tracker with occlusion detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1257-1264, 2011.
31. T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Low-rank sparse learning for robust visual tracking. *European Conference on Computer Vision* **(ECCV)**, pp. 470-484, 2012.
32. N. Wang, J. Wang, and D. Yeung. Online robust nonnegative dictionary learning for visual tracking. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 657-664, 2013.
33. X. Lan, A. Ma, and P. Yuen. Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1194-1201, 2014.
34. B. Liu, J. Huang, L. Yang, and C. Kulikowsk. Robust tracking using local sparse appearance model and K-selection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1313-1320, 2011.
35. Z. Zhang and K. Wong. Pyramid-based visual tracking using sparsity represented mean transform. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp.1226-1233, 2014.
36. X. Li, C. Shen, Q. Shi, A. Dick, and A. Hengel. Non-sparse linear representations for visual tracking with online reservoir metric learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1760-1767, 2012.
37. D. Wang, H. Lu, and M. Yang. Least soft-threshold squares tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 2371-2378, 2013.
38. J. Kwon and K. Lee. Visual tracking decomposition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1269-1276, 2010.
39. J. Kwon and K. Lee. Tracking by sampling trackers.*IEEE International Conference on Computer Vision* **(ICCV)**, pp. 1195-1202, 2011.
40. J. Kwon and K. Lee. Minimum uncertainty gap for robust visual tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp.2355-2362, 2013.
41. J. Kwon and K. Lee. Interval tracker: Tracking by interval analysis. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 3494-3501, 2014.
42. D. Lee, J. Sim, and C. Kim. Multihypothesis Trajectory Analysis for Robust Visual Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 5088-5096, 2015.
43. Q. Bai, Z. Wu, S. Sclaroff, M. Betke, and C. Monnier. Randomized ensemble tracking. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 2040-2047, 2013.
44. Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 34(7):1409–1422, 2012.
45. J. Zhang, S. Ma, and S. Sclaroff. MEEM: Robust tracking via multiple experts using entropy minimization. *European Conference on Computer Vision* **(ECCV)**, pp. 188-203, 2014.
46. X. Zhang, W. Hu, S. Maybank, and X. Li. Graph based discriminative learning for robust and efficient object tracking. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 700-707, 2007.
47. W. Zhong, H. Lu, and M. Yang. Robust object tracking via sparsity-based collaborative model. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1838-1845, 2012.
48. D. Chen, Z. Yuan, G. Hua, Y. Wu, and N. Zheng. Description-discrimination collaborative tracking. *European Conference on Computer Vision* **(ECCV)**, pp. 345-360, 2014.
49. J. Santner, C. Leistner, A. Saffari, T. Pock, and H. Bischof. PROST: Parallel robust online simple tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 723-730, 2010.
50. Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao. Multi-Store Tracker (MUSTer): a cognitive psychology inspired approach to object tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 749-758, 2015.
51. C. Ma, X. Yang, C. Zhang, and M. Yang. Long-term correlation tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 5388-5396, 2015.
52. C. Bailer, A. Pagani, and D. Stricker. A superior tracking approach: building a strong tracker through fusion. *European Conference on Computer Vision***(ECCV)**, pp. 170-185, 2014.
53. N. Wang and D. Yeung. Ensemble-based tracking: Aggregating crowdsourced structured time series data. *International Conference on Machine Learning* **(ICML)**, pp. 1107-1115, 2014.
54. A. Adam, E. Rivlin, and I. Shimshoni. Robust fragments-based tracking using the integral histogram. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 798-805, 2006.
55. S. He, Q. Yang, R. Lau, J. Wang, and M. Yang. Visual tracking via locality sensitive histograms. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 2427-2434, 2013.
56. E. Erdem, S. Dubuisson, and I. Bloch. Fragments based tracking with adaptive cue integration. *Computer Vision and Image Understanding* **(CVIU)**,116(7):827-841,2012.
57. Y. Lu, T. Wu, and S. Zhu. Online object tracking, learning, and parsing with and-or graphs. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 3462-3496, 2014.
58. L. Zhang and L. Maaten. Preserving structure in model-free tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 36(4):756–769, 2014.
59. R. Yao, Q. Shi, C. Shen, Y. Zhang, and A. Hengel. Part-based visual tracking with online latent structural learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp.2363-2370, 2013.
60. X. Jia, H. Lu, and M. Yang. Visual tracking via adaptive structural local sparse appearance model. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1822-1829, 2012.
61. D. Bolme, J. Beveridge, B. Draper, and Y. Lui. Visual object tracking using adaptive correlation filters. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 2544-2550, 2010.
62. D. Martin, H. Gustav, K. Fahad and F. Michael. Accurate scale estimation for robust visual tracking. *British Machine Vision Conference* **(BMVC)**, pp. 1-5, 2014.
63. K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. Yang. Fast visual tracking via dense spatio-temporal context learning. *European Conference on Computer Vision***(ECCV)**, pp. 127-141, 2014.
64. J. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. *European Conference on Computer Vision***(ECCV)**, pp. 702-715, 2012.
65. J. Henriques, R. Caseiro, P. Martins, and J. Batista. High speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 37(3):583–596, 2015.
66. Y. Li and J. Zhu. A scale adaptive kernel correlation filter tracker with feature integration. *European Conference on Computer Vision Worshops***(ECCVW)**, pp. 254-265, 2014.
67. M. Danelljan, F. Khan, M. Felsberg, and J. Weijer. Adaptive color attributes for real-time visual tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1090-1097, 2014.
68. G. Tanisik and E. Gundogdu. Multiple model adaptive visual tracking with correlation filters. *IEEE International Conference on Image Processing* **(ICIP)**, pp. 661-665, 2015.
69. M. Zhang, J. Xing, J. Gao and W. Hu. Robust visual tracking using joint scale-spatial correlation filters. *IEEE International Conference on Image Processing* **(ICIP)**, pp. 1468-1472, 2015.
70. M. Zhang, J. Xing, J. Gao, X. Shi, Q. Wang and W. Hu. Joint scale-spatial correlation tracking with adaptive rotation.*IEEE International Conference on Computer Vision Workshops* **(ICCVW)**, pp. 32-40, 2015.
71. T. Liu, G. Wang, and Q. Yang. Real-time part-based visual tracking via adaptive correlation filters. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 4902-4912, 2015.
72. M. Danelljan, G. Hager, F. Khan, and M. Felsberg. Learning spatially regularized correlation filters for visual tracking. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 4310-4318, 2015.
73. M. Yang, Y. Wu, and G. Hua. Context-aware visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**PAMI**), 31(7):1195 – 1209, 2009.
74. W. Luo, T. Kim, B. Stenger, X. Zhao, and R. Cipolla. Bi-label propagation for generic multiple object tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1290-1297, 2014.
75. M. Mueller, N. Smith and B. Ghanem. Context-aware correlation filter tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2017.
76. H. Grabner, J. Matas, L. Gool, and P. Cattin. Tracking the invisible: Learning where the object might be. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1285-1292, 2010.
77. T. Dinh, N. Vo, and G. Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1177-1184, 2011.
78. L. Wen, Z. Cai, Z. Lei, D. Yi, and S. Li. Online spatio-temporal structural context learning for visual tracking. *European Conference on Computer Vision* **(ECCV)**, pp. 716-729, 2012.
79. S. Hong, T. You, S. Kwak, and B. Han. Online tracking by learning discriminative saliency map with convolutional neural network. *International Conference on Machine Learning* **(ICML)**, pp. 597-606, 2015.
80. L. Wang, W. Ouyang, X. Wang, and H. Lu. Visual tracking with fully convolutional networks. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 3119-3127, 2015
81. H. Nam and B. Han. Learning multi-domain convolutional neural networks for visual tracking. arXiv preprint arXiv:1510.07945, 2015.
82. R. Tao, E. Gavves, and A. W. M. Smeulders. Siamese instance search for tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**) ,pp. 1420-1429, 2016.
83. M. Tian, W. Zhang, and F. Liu. On-line ensemble SVM for Robust Object Tracking. *Asian Conference on Computer Vision* **(ACCV)**, pp. 355-364, 2007.
84. F. Tang, S. Brennan, Q. Zhao, and H. Tao. Co-Tracking using semi-supervised support vector machines. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 1-8, 2007.
85. H. Grabner, M. Grabner, and H. Bischof. Real-time tracking via on-line boosting. *British Machine Vision Conference* ***(BMVC)***, pp. 1-6, 2006.
86. J. Santner, C. Leistner, A. Saffari, T. Pock, and H. Bischof. PROST: Parallel robust online simple tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 723-730, 2010.
87. B. Babenko, M. Yang, and S. Belongie. Visual tracking with online multiple instance learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 983-990, 2009.
88. X. Zhang, W. Hu, S. Maybank, and X. Li. Graph based discriminative learning for robust and efficient object tracking. *IEEE International Conference on Computer Vision* **(ICCV)**, pp. 1-8, 2007.
89. N. Oza. On-line ensemble learning. *PhD Thesis, University of California, Berkeley*, 2001.
90. Grabner, Helmut, C. Leistner, and H. Bischof. Semi-supervised on-line boosting for robust tracking. *European Conference on Computer Vision* **(ECCV)**, pp. 234-247, 2008.
91. P. Viola, J. Platt, and C. Zhang. Multiple instance boosting for object detection. *Advances in Neural Information Processing Systems* **(NIPS)**, 2005.
92. M. Danelljan, A. Robinson, and F. Khan, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking. *European Conference on Computer Vision* **(ECCV)**, pp. 472-488, 2016.
93. A. Krizhevsky, I. Sutskever, and G.E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* **(NIPS)**, pp. 1-8, 2012.
94. J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li. ImageNet: a large-scale hierarchical image database. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 248-255, 2009.
95. J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. S¨ackinger, and R. Shah. Signature verification using a siamese time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* (**JPRAI**), 7(04):669–688, 1993.
96. K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in Neural Informationt Processing Systems* (**NIPS**)*,* pp. 568-576, 2014.
97. L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, Towards good practices for very deep two-stream convnets. *arXiv preprint arXiv:1507.02159*, 2014.
98. C. Feichtenhofer, A. Pinz, and A. Zisserman, “Convolutional two-stream network fusion for video action recognition,” *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (**CVPR**), pp. 1933-1941, 2016.
99. L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L.V. Gool, “Temporal segment networks: towards good practices for deep action recognition,” *In Proceedings of the European Conference on Computer Vision* (**ECCV**), 2016.
100. W. Zhu, J. Hu, G. Sun, X. Cao and Y. Qiao. A key volume mining deep framework for action recognition. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1991-1999, 2016.
101. S. Ji, W. Xu, M. Yang, and K. Yu, 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* **(TPAMI),** 35(1):221–231, 2013.
102. D. Tran, B. Lubomir, F. Rob, and T. Lorenzo. Learning spatiotemporal features with 3D convolutional. IEEE International Conference on Computer Vision, pp. 4489-4497, 2015.
103. D. Jeffrey, A. Lisa, G. Sergio, R. Marcus, V. Subhashini, S. Kate, and D. Trevor. Long-term recurrent convolutional networks for visual recognition and description. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 2625-2634, 2015.
104. B. Nicolas, Y. Li, P. Chris, and C. Aaron, Delving deeper into convolutional networks for learning video representations. *Computer Science* **(CS)**, pp. 1-8, 2015.
105. Y. Joe, H. Matthew, V. Sudheendra, V. Oriol, M. Rajat, and T. George, Beyond short snippets: Deep networks for video classification. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 4694-4702, 2015.
106. W. Yang, and M. Greg, Human action recognition by semilatent topic models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **(PAMI)**, 31(10):1762-1774, 2009.
107. N. Nguyen, Q. Phung, V. Svetha, and B. Hung. Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 955-960, 2005.
108. Z. Dong, D. Gatica, S. Bengio, and I. Mccowan, Modeling individual and group actions in meetings with layered hmms. *IEEE Transactions on Multimedia* **(TMM)**, 8(3):509–520, 2006.
109. P. Dai, H. Di, L. Dong, L. Tao, and G. Xu. Group interaction analysis in dynamic context. *IEEE Transactions on Systems, Man, and Cybernetics* **(TSMC)**, Part B, 39(1):34–42, 2009.
110. Y. Wang, P. Sabzmeydani, and G. Mori. Semi-latent dirichlet allocation: A hierarchical model for human action recognition. *In Proceedings of the 2Nd Conference on Human Motion: Understanding, Modeling, Capture and Animation* **(UMCA)**, pp. 240-254, 2007.
111. D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research* **(JMLR)**, 3:993–1022, 2003.
112. D. Blei, and J. Lafferty. Correlated Topic models. *Advances in Neural Information Processing Systems* **(NIPS)**, pp. 147-154, 2006.
113. Y. Li, W. Li, V. Mahadevan and N. Vasconcelos. VLAD3: Encoding dynamics of deep features for action recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1951-1960, 2016.
114. J. Niebles, and F. Li. A hierarchical model of shape and appearance for human action classification. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1-8, 2007.
115. J. Niebles, H. Wang, and F. Li. Unsupervised learning of human action categories using spatial-temporal words. *International Journal of Computer Vision* **(IJCV)**, 79(3):299-318, 2008.
116. M. Ryoo, and J. Aggarwal. Hierarchical recognition of human activities interacting with objects. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1-8, 2007.
117. K. Andrej, T. George, S. Sanketh. L. Thomas, S. Rahul, and F. Li, Large-scale video classification with convolutional neural networks. *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1725-1732, 2014.
118. G. Otkrist, R. Dan, and R. Ramesh. Deep video gesture recognition using illumination invariants. *arXiv preprint arXiv:1603.06531*, 2016.
119. Song Y, Ma C, Wu X, et al. Vital: Visual tracking via adversarial learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 8990-8999, 2018.
120. Wang X, Li C, Luo B, et al. SINT++: robust visual tracking via adversarial positive instance generation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 4864-4873, 2018.
121. Gao J, Zhang T, Xu C. Graph convolutional tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 4649-4659, 2019.
122. Tu Z, Zhou A, Jiang B, et al. Visual Object Tracking via Graph Convolutional Representation. *IEEE International Conference on Multimedia & Expo Workshops* **(ICMEW)**, pp. 234-239, 2019.
123. Cui Z, Xiao S, Feng J, et al. Recurrently target-attending tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 1449-1458, 2016.
124. Gordon D, Farhadi A, Fox D. Re $^ 3$: Re al-Time Recurrent Regression Networks for Visual Tracking of Generic Objects. *IEEE Robotics and Automation Letters* **(RA-L)**, 3(2): 788-795, 2018.
125. Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking. *Proceedings of the European Conference on Computer Vision* (**ECCV**), pp. 850-865, 2016.
126. Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **(CVPR)**, pp. 8971-8980, 2018.

**1.3. 发展动态分析**

**2) 模拟大脑与小脑协作关系的实时在线目标跟踪方法：**由于跟踪场景的复杂性，如物体表观的显著变化、姿态变化、严重遮挡、背景混乱等，进行鲁棒的实时在线目标跟踪具有较大的挑战性。近年来，受大脑神经工作机制启发的基于CNN的跟踪器展示了巨大的潜力。在离线网络的预训练阶段，使用AlexNet和VGGNet等主干CNN架构在外部海量视频数据集ILSVRC2015上学习用于分类或回归的语义嵌入空间。与手工制作的特征不同，在CNN学习的语义嵌入空间中的特征投影包含丰富的高级语义信息，并且对于区分不同类别的目标非常有效。该嵌入特征还具有跨数据集的泛化能力，可确保跟踪的泛化性。在线跟踪阶段，CNN跟踪器仅通过一次前馈网络传递就可以快速估计目标位置，而无需进行任何网络微调。尽管受大脑神经工作机制启发的基于CNN的跟踪器具有良好的性能，但仍有一些局限性。首先，语义嵌入空间中的特征表示通常具有较低的分辨率，并且会损失一些特定于实例的细节和细粒度的定位信息。因此，一方面，基于CNN的跟踪器可能对细节不太敏感，并且在比较具有相似属性或语义的两个目标时会降低判别性。另一方面，基于CNN的跟踪器可能会发生领域偏移问题，特别是当跟踪器遇到未知的目标或目标发生突然变形时。其次，基于CNN的跟踪器通常不执行在线网络更新以提高跟踪速度，这不可避免地影响模型的适应性，从而损害跟踪精度。在人类视觉系统中，大脑和小脑通过协同工作完成各类视觉任务。小脑更多地参与需要精细调节与定位的动作执行，而大脑更侧重于高层的语义感知与理解。在目标跟踪器的设计中，模拟小脑的空间定位能力和大脑的高层认知能力，给目标跟踪提供了一个新的思路，未来目标跟踪将朝着更快、更好、更通用的方向发展。

**2．项目的研究内容、研究目标，以及拟解决的关键科学问题**（此部分为重点阐述内容）

**2.1. 研究内容**

**2) 探索大脑与小脑的协作关系在目标跟踪中的应用，通过模拟小脑的空间定位能力和大脑的高层认知能力，实现相关联滤波学习和深度特征学习相融合的鲁棒的目标跟踪**：深度神经网络自动学习特征的方式模拟了大脑对于目标最具区别性的关键特征信息的提取。在深度学习的训练过程中，判别网络以及特征学习都是针对特定的目标，在训练过程中会使得网络更关注于语义信息，逐步丢弃掉底层的信息，而这些底层信息对于精确的目标跟踪问题尤为关键。同时，CNN所学习到的特征缺乏时空连续性以及对抖动、混杂背景等不确定性因素的表述能力。对于目标跟踪问题而言，需要模拟小脑的空间定位能力和大脑的高层认知能力，实现快速、强大、自适应的目标跟踪。具体地说，利用相关滤波器模拟小脑对底层视觉信号的编码与空间定位，利用卷积反卷积神经网络模拟大脑对视觉信号的编解码，实现相关滤波学习和深度特征学习相融合的实时在线目标跟踪。

**2.3. 拟解决的关键科学问题**

**2)** **如何在目标跟踪中模拟大脑与小脑的协作关系，建立相关滤波学习和深度特征学习相结合的模型以实现开放、不确定环境下鲁棒的目标跟踪是本项目拟解决的关键科学问题**：目前基于模拟大脑语义感知的CNN目标跟踪算法将特征学习和判别模型训练相分离，直接将图片分类、目标检测中学习到的卷积神经网络应用到判别学习的框架中，这些特征具有较强语义性，但由于在网络的设计中缺乏了对小脑精确定位功能的模拟，导致网络缺乏针对平移缩放等不确定因素的判别性，因而无法有效地提升开放、不确定环境下目标跟踪的精度。另一方面，深度学习的判别模型训练过程需要大量的标签样本，而跟踪问题中的目标只能提供一帧的目标信息，这对于判别模型的收敛与泛化性能提出了极大的挑战。对于不同视频序列中的目标，最能刻画它们区别的特征并不相同，通过模拟小脑自适应性强的特点，选择适合的特征来更好的适应开放环境与不确定条件，可以降低模型所需要的规模，提高跟踪速度。如何在目标跟踪中模拟大脑与小脑的协作关系，建立相关滤波学习和深度特征学习相结合的模型以实现开放、不确定环境下鲁棒的目标跟踪是高效的目标跟踪的关键所在。

**3．拟采取的研究方案及可行性分析**（包括研究方法、技术路线、实验手段、关键技术等说明）

**3.2. 技术路线**

**2) 对于基****于模拟大脑与小脑协作关系的实时在线目标跟踪，本项目拟利用相关滤波器模拟小脑对底层视觉信号的编码与空间定位，利用卷积反卷积神经网络模拟大脑对视觉信号的编解码，实现相关滤波学习和深度特征学习相融合的实时在线目标跟踪。**本项目首先离线训练卷积神经网络模型，对特征学习网络进行初始化通过调节感受野和网络深度实现精简的网络模型架构；在此基础上，通过在视频目标检测数据集上进行特征网络参数的迁移学习，这样可以使得深度特征学习中加入更多的动态信息。通过这种监督训练获得的卷积神经网络模型可以利用视频序列中在初始帧标注的待跟踪目标信息以及相邻帧之间时间和空间上的连续性这一共性先验来对离线训练得到的卷积神经网络进行反馈学习和进一步的特征层筛选，从而获得能够有效刻画被跟踪目标的特征表示。在实际跟踪过程中，以初始帧目标为模板，将当前搜索样本的从卷积神经前馈来的深度特征与目标模板深度特征作对比，根据相似性学习度量每个搜索样本和目标的相似概率，通过融合空间位置先验来估计目标位置。在训练过程中随机选取同一段视频内的正负样本进行相似性对比，学习适应于开放、不确定环境的特征表示。跟踪过程中，为了对于目标形变自适应，可以产生候选样本，然后将候选样本与第一帧模板进行对比，在位移空间以外增加对尺度空间的建模，以增强跟踪模型对于尺度变换的鲁棒性。为了进一步提高算法的判别性，可引入半监督学习与迁移学习来在线调整基于深度特征的相似性估计，从而减缓或者矫正开放、不确定环境下跟踪漂移问题。本项目拟通过卷积神经网络抽取具有底层纹理以及运动描述和语义信息的分层特征，将该特征输入到相关滤波网络当中，得到目标位置的概率分布，可以有效结合不同层次特征，针对不同的特征给出相应分辨率的响应，从而模拟大脑与小脑的协作机制，通过多种分辨率的响应进行特征融合。在物体表观清晰的时候，可以直接利用底层特征得到精确的跟踪位置，而当物体发生旋转或部分遮挡时，可以增加高层特征的权重，维持对目标的跟踪以适应表观的剧烈变化。同时，由于判别相关滤波方法利用快速傅里叶变换高效地进行稠密采样，使得训练样本相对于粒子滤波框架得到大幅提升，增加了跟踪算法的鲁棒性和实时性。

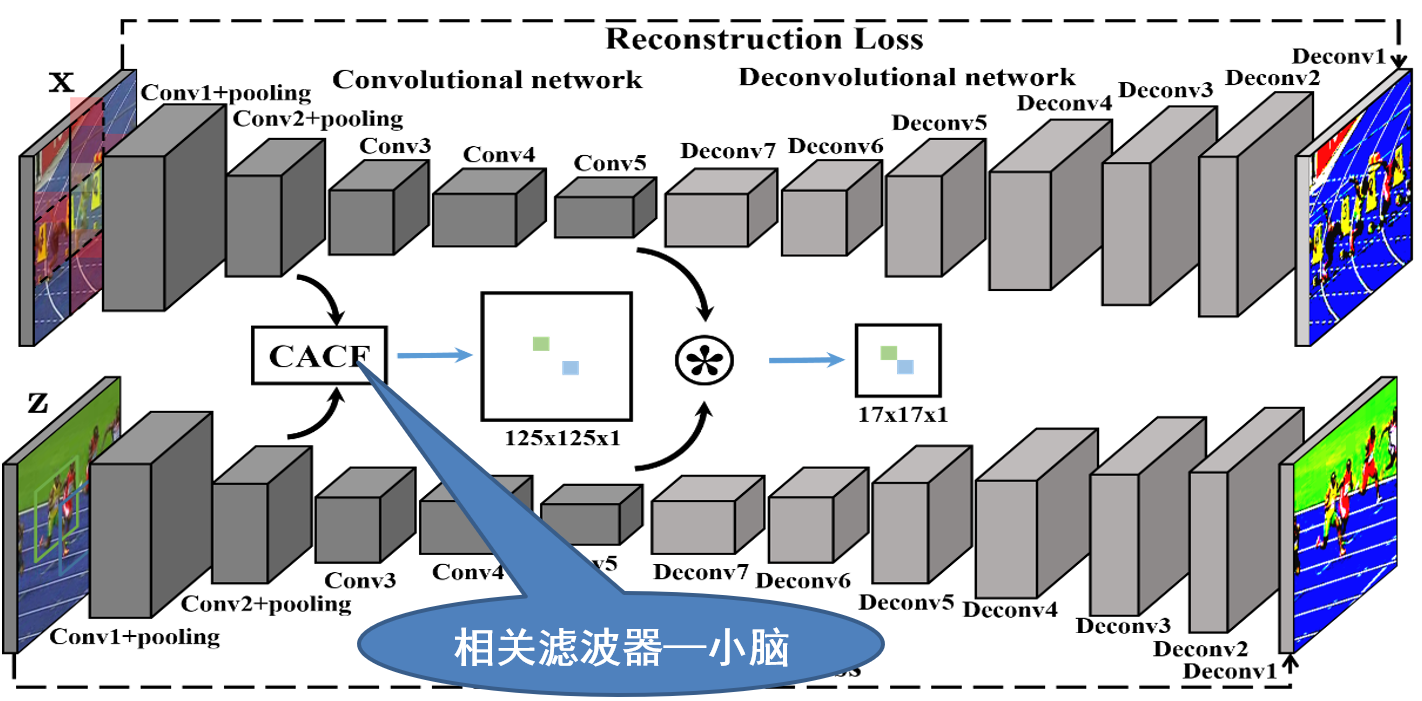


图1. 基于模拟大脑与小脑协作关系的目标跟踪神经网络结构