ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

Object Detection Using CNN

[1]N. Ramesh, [2] P. Rakshitha Reddy, [3] K. Mahesh, [4] L. Vishnu, [5] Dr. Lakshmaiah,
[1,2,3,4] U.G.Scholor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.
[5]Research Guide, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

***Abstract*:-This paper reviews recent advances in Deep Learning-based Visual Object Tracking Approaches from various algorithms, highlighting key issues with existing research and suggesting future potential directions. It provides a comprehensive review of recent advancements in this complex area of computer vision, highlighting the importance of real-world application scenarios in this field.This paper reviews recent advancements in Deep Learning-based Visual Object Tracking Approaches, analyzing various algorithms. It highlights key issues in existing research and suggests future research directions, concluding with a comprehensive review of recent developments.**

Keywords:MachineLearning,Pre-training, Onlinelearning,Visual Tracking,DeepLearning, NeuralNetworks,CNN,

INTRODUCTION

Tracking is the process of inferring an object's motion from images. Visual object tracking, a field of computer vision, has applications in various fields like driver assistance, medical, military, video surveillance, traffic control, navigation, robotics, augmented reality, and sports. Visual object tracking, extensively studied in computer vision, is utilized in real-world applications such as Advance Driver Assistance Solutions, Medical, Military, Video Surveillance, Traffic Control, Navigation, Robotics, Augmented Reality, and sports.

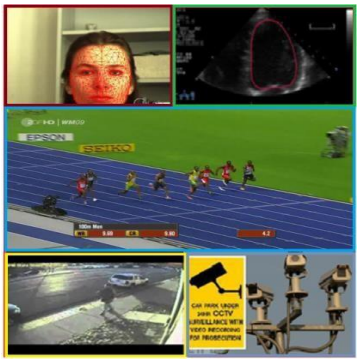

Fig1:SomeVisualTrackingApplications

Basictraditionalvisualtrackingmethodsutilizevariousframeworks like Discriminative Correlation Filters (DCF),silhouettetracking,Kerneltracking, point tracking , and so forth – these methodswerenotabletoprovidesatisfactoryresultsinunconstrained environments.

After careful analysis from the wide variety of algorithmscitedintheliterature,thispaperfocusesonsomepotential

and well performed Deep Learning based visual trackingmethods. The trackers include from diverse classificationmethods like, Network Architecture, Network Exploitation,Training,NetworkObjective,NetworkOutputandCorrelation Filter Exploitation.

AlthoughConvolutionalNeuralNetworks(CNN)havebeenusedinDeepLearning(DL)based methodsfromtheresearchliterature,someofothernetworkarchitectureswerealsoproposedtoimprovetheefficiencyand robustnessofvisualtrackersinrecentyears.The CNN-based visual trackers are classified underthreecategories, robusttargetrepresentation,Balancingtrainingdataand Computationalcomplexityproblem.Someoftheadvantagesof CNNbased methodsutilizedareparametersharing,sparseinteractions,and dominantrepresentations.

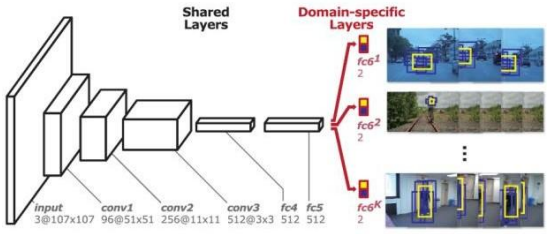

Fig3:ThearchitectureofMulti-DomainNetwork(MDNet)

Fig. 3 consists of shared layers and K branches of domain-specific layers. Yellow and blue bounding boxes denote thepositiveandnegativesamplesineachdomain,respectively.

Toovercomethelimitationsofpre-traineddeep CNNsandtakefulladvantageofend-to-endlearningforreal-time applications, Siamese Neural Network (SNN) hasevolved, someoftheSNNbasedapproachesinrecentyearswere proposedtoachievereal-timespeed.

Siamese Neural Network (SNN) Evolution for Real-Time Applications

- Overcomes limitations of pre-trained deep CNNs.
- Utilizes discriminative target representation, adapting target appearance variation.
- Balances training data for real-time speed.

TheseproposedSNNbasedmethodsutilizecombinationofdiscriminativetargetrepresentation,adaptingtargetappearancevariationandbalancingtrainingdata.

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

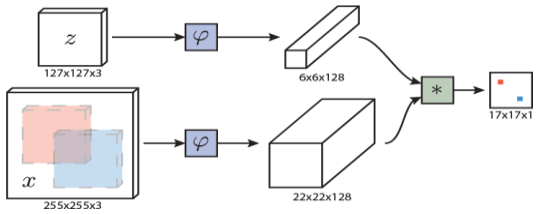

Fig4:Fully-convolutionalSiamesearchitecture

TheSiamesearchitectureisfullyconvolutionalwithrespect tothesearchimagex.Theoutputisascalar-valued score mapwhosedimensiondependsonthesizeofthe search image.Thisenables thesimilarityfunctionto be computed for all translated sub-windows within thesearch image in oneevaluation.Inthisexample,theredandbluepixelsinthescore mapcontainthesimilaritiesforthecorrespondingsub-windows[21].

VisualObjectTrackinginvolvesbothspatialandtemporalinfor mation of video frames, over therecentyears.RecurrentNeuralNetwork(RNN)architecture basedmethodsareproposedtoconsidermotionormovementsi multaneously.Because oftedioustraininganda numerous number ofparameters, the number of RNN-based methods is limitedcomparativelyintheavailableliterature.

Couple of recent methods in literature utilized GenerativeAdversarialNetwork (GAN) architectureto addressthe imbalance distribution of training samples and also todealself-learningproblemofvisualtracking.

Thetrendindescribingcustomnetworksisalsoseen,whichisaco mbinationofmultipleabovementionednetwork architectures like CNN, SNN, RNN and GAN tomainlytacklethelimitationsofordinarymethodsbyexploitin gtheadvantagesofothernetworkstructures.

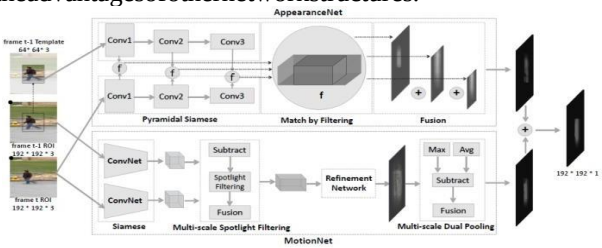

Fig5:Deepcollaborativetrackingnetwork[19].

Themajorityofthebasicdeeplearningbasedvisualtracking methodsexploits end-to-end learning with train orre-trains aDNN by applying gradient based optimizationalgorithms.

The main contributions of this review are summarized asbelow:

- IndepthstudyandanalysisofDeepLearningbasedtrac kersrecentadvancesinvariousaspects.
- Summarizeexperimental resultsoftrackers citedin researchliteraturepublications.
- Describe present research issues of visual trackingresearchandpresentsomeinterestingpotenti alfutureresearchdirections.

Therest ofthis paper isorganizedas follows:Atfirst,various deep visual tracking methods from the availableliteraturearepresentedandthendescribedanddiscuss edsome experimental comparisons of the presentedvisualtrackingmethods.Finally,someoftheresearch issuesarepresentedwithfuturedirectionsattheconclusion.

SOMERECENTMETHODSINTHELITERATURE

Their experiments showthat deep embedding's provide a naturally rich source offeatures for online trackers, and enable simplistic test-timestrategies toperformwell.Theydescribedanalgorithmwith a novel fully-convolutional Siamese network trainedend-to-endontheILSVRC15datasetforobjectdetectioninvideo.

[H. Nam and B. Han. 2016] [4]proposed a multi-domainlearningframeworkbasedonCNNs,whichseparatesdo main-independentinformationfromdomain-specificone,tocapturesharedrepresentationseffectively.They havesuccessfullyimplementedaframeworktolearndomain-specificinformationadaptively.

[D. Held, S. Thrun, and S. Savarese 2016] [6] at a highlevel, they feed frames of a video into a neural network andthe network successively outputs the location of the trackedobjectineachframe.Theytrainedthetrackerentirelyoffl inewithvideosequencesandimages.Theirtrackerlearnsoffline agenericrelationshipbetweenanobject'sappearance and its motion, allowing network to track novelobjectsatreal-time (100fps)speeds.

A trackerwhich is controlled by action-decision network (ADNet),pursues the target object by sequential actions iterativelytrained by deep reinforcement learning. They proposed atracker which is designed to achieve a light computation aswell as satisfactory tracking accuracy in both location andscale. In this method the deep network to control actions ispre-trained using various training sequences and fine-tunedduringtrackingforonlineadaptationtotargetandbackgro undchanges.Thistrackerachievedareal-timespeed (15fps).

Described a factorized convolution operatorthat dramatically reduces the number of parameters in theDCF model. Designed a compact generative model of thetrainingsamplespacethateffectivelyreducesthenumberofs amplesinthelearning,whilemaintainingtheirdiversity.

Comprehensive experiments clearly demonstratethatthisapproachconcurrentlyimprovesbothtrac kingperformanceandspeed. Siamese Neural Network (SNN) has evolved to overcome pre-trained deep CNN limitations and optimize end-to-end learning for real-time applications. Recent approaches use discriminative target representation, adapting target appearance variation, and balancing training data to achieve real-time speed. deepanalysisofSiamesetrackersandprovedthatwhenusingdee

networksthedecreaseinaccuracycomesfromthedestroying of the strict translation invariance. Proposed asampling strategy to break the spatial invariance restrictionwhichsuccessfullytrainsSiamesetrackerdrivenbya ResNetarchitecture.

[Y.Song,C.Ma.,2018][2]proposedtouseageneративeadversarialnetwork(GAN)toaugmentpositivesamplesinthefeaturespacetocaptureavarietyofappearancechangesoveratemporalspan.Demonstratedtousehigher-ordercostsensitivelosstominehardnegativesamplestohandleclassimbalance.Conductedextensivevalidationofmethod on benchmark datasets with large-scale sequences.[H.FanandH.Ling2018][5]presentedamultistagetrackingframework,theSiameseCascadedRPN(CRPN),to solve the problem of class imbalance by performing hardnegativesampling.Theydesignedanovelfeaturetransferblock(FTB),enablestofusethehigh-levelfeaturesintolow-levelRPN,whichfurtherimprovesitsdiscriminativepowertodealwithcomplexbackground,resultinginbetterperformanceof C-RPN.Theyhaveconductedextensiveexperimentsonsixbenchmarks

[C. Sun, D. Wang., 2018] [7]in this model, they proposedanddevelopedaspatial-awareKRRmodelbyintroducinga cross-patch similarity kernel. They have implemented amodelwithbothregressioncoefficientsandpatchreliability,whichenablesamodeltoberobusttotheunreliablepatches.Theregressioncoefficientandsimilarity weight vectors are simultaneously optimized viaan end-to-endneural network.

[J. Choi, H. J. Chang., 2018] [8] proposed a visual trackingframeworkbasedoncontext-awaredeepfeaturecompression using multiple auto-encoders. In this modeltheyintroducedacontext-awareschemewhichincludesexpertautoencodersspecializing inonecontext,andacontext-awarenetworkwhichisabletoselectthebestexpertauto-encoderforaspecifictrackingtarget.Conducted experiments lead to the compelling finding thatthis framework achieved a high-speed tracking ability ofover 100fps.

[G. Bhat, J. Johnander., 2018] [9] analyzed the influentialcharacteristicsofdeepandshallowfeaturesforvisual tracking.Theysystematicallystudiedtheimpactofavarietyofdataaugmentationtechniques.Deepinvestigationoftheaccuracy-robustnesstrade-offinthediscriminative learning of the target model has done. Theyproposed a fusion strategy to combine the deep and shallowappearancemodelsleveragingtheircomplementarycharacteristics.Experimentsareperformedonfourchallengingdatasets.

[S. Pu, Y. Song., 2018] [10] Described and implemented areciprocate learning algorithm to exploit visual attentionwithin the tracking by detection framework. In this method,for temporal robust features they introduced attention mapsas regularization terms coupled with the classification lossto train deep classifiers. They also conducted experimentson benchmarkdatasets.

[Y. Zhang, L. Wang., 2018] [11]proposed a local patterndetectionscheme,whichcanautomaticallyidentifydiscriminativelocalpartsoftargetobjects.Toachievemoreaccuratetrackingresults,theyimplementedthe

message passing process via differentiable operations, andreformulateitthroughaneuralnetworkmodule.

[Z. Zhu, Q. Wang., 2018] [12] Distractor-aware SiameseRegionProposalNetworks(DaSiamRPN)method,Analyzedthefeaturesusedinconventional Siamesetrackers in detail and they found that the imbalance of thenon-semanticbackgroundandsemanticdistractorinthetraining data is the main obstacle for the learning. Proposeda framework to learn distractor-aware features in the off-line training, and explicitly suppress distractors during theinferenceof online tracking.

[C. Sun, D. Wang., 2018] [13]described a model whichincludesbothdiscriminationandreliabilityinformation usingthecorrelationfilterframework.Theyhaveintroduced local response consistency constraint to ensurethat different sub-regions of the base filter have similarimportance. In this model to depict the importance of eachsub-regioninthefilter(i.e.reliabilitylearning)thereliabilityweight mapisexploited.Thistrackerisinsusceptible to the non-uniform distributions of the featuremap,andcanbettersuppressthebackgroundregions.ExtensiveexperimentshaveconductedtoshowthesuperiorityofthealgorithmcomparedandthistrackerachievesremarkabletrackingperformanceontheOTB-2013,OTB-2015andVOT-2016 benchmarks.

[F. Li, C. Tian., 2018] [14] presented a spatial-temporalregularizedcorrelationfilters(STRCF)modelbyincorporating both spatial and temporal regularization intothe DCF framework. The proposed STRCF serves as anapproximation of SRDCF with multiple training samples.TheyhaveimplementedADMMalgorithmforsolvingSTRCF efficiently, where each sub-problem has the closedformsolution.ThisSTRCFModelwithhand-craftedfeaturecanruninreal-time,achievesnotableimprovementsoverSRDCFbytracking accuracy.

Deep CollaborateTrackingNetwork(DCTN)aunifiedframeworkthatjointly encodes both appearance and motion informationforgenericobjecttracking.Theydescribedestablishingaunified tracking framework of a two-stream network thatcanfullycapturecomplementarymotionandappearanceinformation with an end-to-endlearningarchitecture;Described a motion net (MotionNet) to fulfillend-to-end trainable motion detectionand an appearancenet(AppearanceNet)formulti-scaleappearancematchingto achieve object localization.

SiamMask, asimple approach that enables fully-convolutional Siamesetrackerstoproduceclass-agnosticbinarysegmentationmasks of the target object. They proposed two variants ofSiamMaskareinitializedwithasimpleboundingbox,operateonline,runinreal-timeanddonot requireanyadaptation tothe testsequence.

RegularizedCorrelation Filters (ASRCF) model tosimultaneouslyoptimizethefiltercoefficientsandthespatial regularizationweight. Their tracker effectively andefficientlyestimatesbothlocationandscalewithtwoCorrelation Filter models: one exploitscomplicated featuresforaccuratelocalization;andtheotherexploitsshallow featuresforfastscaleestimation.Conductedextensiveexperimentsonfiverecentbenchmarksshowthatthistrackerperformsfavorablyagainstmanystate-of-the-artalgorithms,withreal-timeperformanceof28fps.

This paper presented a systematicstudyonthefactorsofbackbonenetworksthataffecttrackingaccuracy,andprovidesarchitecturaldesignguidelines for the Siamese tracking framework. They havefound and described receptive field size, network paddingand stride are crucial factors. Based on no padding, residualunits they have designed new deeper and wider networkarchitecturesforSiamesetrackers.Conductedmultipleexperimentson fivebenchmarksbaseline datasets.

ANALYTICALREVIEW

According to the in depth analysis, the deep visual objecttrackingmethodsthatareconsistentinperformanceonsomestandardvisualtrackingdatasetsare[2],[4],[10],[3],[1],[5],[11],[12],[9],[7],[14],and[13].These methods performed well in terms of precisionsuccess measures and accuracy robustness on some famousstandard Data Sets.The research results are promising whenconsidered individual visual challenging attributes, but theresults are not encouraging when considered all the visualchallenging attributessimultaneously.

The methods [2], [4], and [10] take advantage from bothOfflineandOnlineTrainingofDeepNeuralNetwork(DNN),Howeverthesemethodslackinginspeed(<=1FramesPerSecond)becauseofhugecomputationalcomplexity, for real time applications these methods willnotbesuitable.

According the analysis, though [2] out performed in coupleof visual challenging scenarios(deformation (DEF),in-planerotation (IPR), out-of-plane rotation (OPR)) but lacking inrobustnesswhensignificantscalevariation(SV)presentinascene.

Basedonsystematicstudyofrevelsthatthismethod failed to follow the abrupt movement of the targetand the proposed actions could not adapt to thesuddenaspectratiochange.

Though [1] performed well in terms of accuracy, Overallsystemspeedis far less comparative to real time. In[7]utilizeskernelizedridgeregression(KRR)toconcentrate on reliable regions of the target, consideration of rotationinformationandonlineadaptationofConvolutionNeuralNetwork (CNN) models.This method responded well to thedeformationandin-planerotationvisual challenges.

The methods [14], [13] and [9] are the DCF based anddescribe on fusing the HOG with deep off the self-featuresto improve the consistency of the results.Though thesemethodsshow the competitive performance but very muchsuffer from the limitations of the computational complexityofappearance variationand deepfeatures.

According to the research results[1], [5], [3], [7],[13],

[9], and [14] on some standard data sets have failed inscenariosthatconsistofsimultaneousmultiplecriticalvisualattributes.

In depth study of[9], [14], [13], [7]describesexploitingondeepoff-the-shelffeaturesandtakethe advantageofDCFframeworktoaddresssomeofthechallengingvisual attributes.

Themethods[3],[1],[5]and[12]arebasedonthefast SiamRPN method and exploit on one-shot detectiontask to solve some ofthe visual trackingproblems.Basedon research results of[4] performs well on deformation,lowresolutionandfastmotionscenarios.

In[6]Whenthereisasizechangeornovariation,thetracker performs slightly worse when using the previousframe.Underalargesizechange,thecorrespondingappearance change is too drastic for network to perform anaccurate comparison between the previous frame and thecurrentframe.Thetrackerisactingasalocalgenericobject detector insucha case.

In [5], [11], and [12] methods exploit the shallow AlexNetas their backbone network. In [11] describes to decrease thesensitivityofSNN-based(SiameseNetwork)methodsspecificallyfornon-rigidappearancechangeandpartialocclusion (POC) attributes, this method detects contextualinformationof local patterns andtheirrelationships andmatchesthem byaSiamese network inreal-timespeed.

Inthemethodpresentthree-brancharchitecturetoestimatethetargetlocationbyarotated BoundingBox,which includes the binary mask of the target.The mostfailure reasons of SiamMask are the motion blur (MB) andout-of-view (OV) attributes that produce erroneous targetmasks.

RESEARCHCHALLENGES

Despite rapid considerable advancements that are emergedinDeepVisualTracking,thementionedtrackersfrom research literature are still unable to handle the real-worldchallengesefficiently.

Studiesonreferencesfromliteraturerevealedthatdeeplearning based methods are still not reliable for real-worldapplicationsastheyarelackingintelligentsituationunderstandingwithrealtime speed.

It is understood that despite decades of research still havetheproblemstosimultaneouslyhandlechallengingscenarios which significantly consists of visual attributessuchasOCC(Occlusion),OV(Out-of-View),DEF(Deformation),SV(ScaleVariation)andFM(FastMotion).

Sometrackingapproachesperformwellinspecificvideo/imagescenariosandStandardDataSets.Whileappliedtootherca es,however,theymaynotproducesatisfying results.

Though some of the methods presented performs well in achallenging scenarios but they are not robust enough tohandlethediversity ofsituations.

Basedonthereviewitisunderstoodthatmaintainingaccuracy innumeroussituations,robustnesstovisualvariances and computational efficiency all at once is anexisting researchchallenge.

Based on thorough analysis computational complexity andmemory usage is the biggest research challenge in DeepLearningBasedVisualTrackingeventoaddresssinglec

hallengingvisual attribute.

All these issues restrict further development of the VisualTracking research and its applications in real-world, real-time systems. Recently, attempts to deal with some of theseissues have been made,for example, the Benchmark newdata sets provides a large set oftesting video sequences,standardbaselineevaluationtools,newmethodsofe valuation etc. This is likely to advancethe further studiesanddevelopments ofVisualObjectTrackingtechniques.

CONCLUSIONANDFUTUREDIRECTIONS

The study reveals that the most challenging attributes for Deep Learning-based visual tracking methods are OCC, DEF, OV, SV, and FM. To improve robustness, multiple complementary features from efficient methods should be exploited. Integrating both offline and online training methods can lead to more robust visual trackers. To achieve success rates, more efficient combination of network architectures and intelligent methods should be employed.

REFERENCES

[1] B.Li,W.Wu,Q.Wang,F.Zhang,J.Xing,andJ.Yan,"SiamRPN++: Evolution of Siamese visual tracking with verydeepnetworks,"2018.[Online].Available:http://arxiv.org/abs/1812.11703

[2] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R.W.Lau, and M. H. Yang, "VITAL: Visual tracking via adversariallearning,"in Proc.IEEE CVPR,2018, pp.8990–8999.

[3] Z. Zhang and H. Peng, "Deeper and wider Siamese networks forreal-timevisualtracking,"2019.[Online].Available:http://arxiv.org/abs/1901.01660

[4] H.NamandB.Han,"Learningmulti-domainconvolutionalneural networks for visual tracking," in Proc. IEEE CVPR, 2016,pp.4293–4302.

[5] H. Fan and H. Ling, "Siamese cascaded region proposal networksforreal-timevisualtracking,"2018.[Online].Available:http://arxiv.org/abs/1812.06148

[6] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPSwith deep regression networks," in Proc. ECCV, 2016, pp. 749–765.

[7] C. Sun, D. Wang, H. Lu, and M. Yang, "Learning spatial-awareregressions for visual tracking," in Proc. IEEE CVPR, 2018, pp.8962–8970.

[8] J. Choi, H. J. Chang, T. Fischer, S. Yun, K. Lee, J. Jeong, Y.Demiris,andJ.Y.Choi,"Context-awaredeepfeaturecompressionforhigh-speedvisualtracking,"inProc.IEEECVPR,2018,pp.479–488.

[9] G. Bhat, J. Johnander, M. Danelljan, F. S. Khan, and M. Felsberg,"Unveilingthepowerofdeeptracking,"inProc.ECCV,2018,pp.493–509.

[10] S.Pu,Y.Song,C.Ma,H.Zhang,andM.H.Yang,"Deepattentivetrackingviareciprocativelearning,"inProc.NIPS,2018,pp.1931–1941.

[11] Y.Zhang,L.Wang,J.Qi,D.Wang,M.Feng,andH.Lu,"Structured Siamese network for real-time visual tracking," inProc.ECCV,2018,pp.355–370.

[12] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, "DistractorawareSiamesenetworksforvisualobjecttracking,"inProc.ECCV,vol.11213 LNCS,2018,pp.103–119.

[13] Li, C. Tian, W. Zuo, L. Zhang, and M. H. Yang, "Learningspatial-temporalregularizedcorrelationfiltersforvisualtracking,"in Proc.IEEE CVPR,2018,pp.4904–4913.

[14] S. Yun, J. J. Y. Choi, Y. Yoo, K. Yun, and J. J. Y. Choi, "Actiondecision networks for visual tracking with deep reinforcementlearning,"inProc.IEEECVPR,2016,pp.2–6.