March-April 2020
ISSN: 0193-4120 Page No. 2168 - 2176

Challenges in Scene Interpretation for Video Surveillance

H. N. Lokhande
Department of Electronics and Telecommunication Engineering Sinhgad College of Engineering Pune, India
lokhande.harshad@gmail.com

S. R. Ganorkar
Department of Electronics and Telecommunication Engineering Sinhgad College of Engineering Pune, India
ganorkarsanjay@gmail.com

Article Info
Volume 83
Page Number: 2168 - 2176
Publication Issue:
March - April 2020

Article History
Article Received: *24 July 2019*
Revised: *12 September 2019*
Accepted: *15 February 2020*
Publication: *18 March 2020*

Abstract

The surveillance of local places, shopping malls and traffic signals is essential for safety and security purpose. Though the CCTV's are available, the manual retrieval of footage frame by frame is time intensive and critical task. Hence with video surveillance, the task like object detection, person identification and tracking of suspicious movements will improve the result. But there are several factors affecting the scene like environmental variations, light illumination, camera variations and occlusion. The existing systems are able to identify the object from limited categories under certain conditions with deep learning. But data training and computing system cost are very high. So, different challenges and aspects for scene interpretation around the object are discussed in this paper. The scope will highlight various methods with implementation and their results. Though the deep learning is more powerful tool for video analytics, there are still challenges that must need to improve for real-time scene interpretation.

Keywords – *video surveillance, scene interpretation, video analytics*

I. Introduction

This The surveillance of local places, shopping malls and traffic signals is essential for safety and security purpose. Though the CCTV's are available, the manual retrieval of footage frame by frame is time intensive and critical task. There are several factors affecting like environmental variations, light illumination, camera variations and occlusion. The situation on railway stations, shopping malls and in crowded areas is complicated for structuration. Hence, in order to identify the object from such clustered scene, its most important to analyze and classify the scene in respective domain. And then the corresponding methods are required to apply for identifying the object from such clustered scenes. There are several well-known mechanisms available for detecting the object from the scenes. But these are limited to certain classification groups and under some standard set of rules to define the object.

The classic image processing-based algorithms can give results to identify the object from the scene, but the implementation of method is time consuming task and the results may vary according to user. In this classic algorithm, the humans are defining the key features of an object. These features are unique features for that unique object. And hence even if the model has been trained for identifying that object, the accuracy of identifying from new set of data, is poor because of small variation in object may lead to make it unrecognizable.

The machine learning algorithms may increase the scope to classify the object from scene. The machine learning techniques like SVM, HMM, and other classification techniques, proven better accuracy for detecting object. Though it has potential to detect the objects, but due to small variation in scene and background, the results may get affected. The existing methods are providing object detection in limited scope under fix set of conditions. Hence even the machine learning may classify the object, but there are many challenges to affect the result.

The algorithms of deep learning are improving day by day with the accuracy in object identification. But there are various dependencies to generate the pretrained-model to deploy for final result interpretation. The CNN, ANN and Fast-RCNN are useful for finding the object and classify. But the dataset required for training the method should be extensively large and an inference model should be strong enough to deploy such systems. So, the process is time consuming and hardware dependent. Now once the object is identified, the sematic analysation of scene is key point of this research. The scene around the object has to interpret the details.

Thus, considering above factors, there is need to design a new system, where the appropriate features will be extracted by manually and the deep learning algorithms are used to train them for scene interpretation purpose. This work explores various aspects and challenges in scene interpretation with object identification.

Section -II covers the need of scene understanding, the different methods for scene analysis and the issues to detect object under specific scene conditions.

Section-III is elaborating different object detection techniques like SIFT, KLT and background-foreground subtraction methods.

Section-IV is about existing methods for object identification in scene specific conditions. The various machine learning implementation were compared with state of art results for scene specific identification and detection techniques.

Section-V, goes with improvements and challenges in deep learning for object identification. The scope of advanced methods in deep learning are identified and compared with standard results.

Section-VI, is elaborating the object interpretation techniques. Once the object is identified from the scene, now to analyze the scene and interpret the scene along with the object. The NLP mechanisms are discussed and enlisted some useful methods.

II. Scene Understanding

A scene is an insight of environment with multiple surfaces and objects in an particularizing way. The visual insides are shown with many features like colour, contours or luminance or in the form of parts, shapes and textures or a sematic contextual way. The aim of scene understanding is to make machines think like humans, to have a complete depth and understanding of visual scenes. Scene analysis is inclined by perceptive vision with an association of major fields like software engineering, computer vision and cognitive engineering.

To understand the object, it’s important to understand the scene first. [1] demonstrated an algorithm to classify the objects in 15 different classes based on global geometric correspondence. The image was divided into sub-regions and histogram is computed to obtain the local features to categorize them into weak & strong sets. This bag of features method was only restricted up to 15 different categories. To understand the basic scene definitions, [2] has implemented new set of database for image classification with Turing test.

Due to unclear representation of object in complex scenes, the detection was not up to the mark. [3] introduced mechanism to define the features manually and use to model objects using annotations in real applications of spatial relations and geometric features. The metric spatial relations (MSR) implemented on extracted features to train the dataset. As this method was limited only to table top based objects, it fails for other domains. Though the generation of database is mostly confirmed by labelling images separately, [4] has demonstrated to prepare the dataset based on scene understanding by experienced humans.

So, considering most probable scene with appropriate objects as first input and same scene with dissimilar objects to categorize it in probable and improbable sections. It was useful to structure and eliminate the extreme variations from scene understanding features. But though it depends on manual input from user, the results may get biased. So, need identify methods which will provide an outline for prescribing the scene factors.

III. Object Detection

For scene surveillance, the object detection and tracking are most difficult tasks. A shaped a method with graph representation of moving objects which helped to derive and maintain a dynamic template of moving object to match with still frame [5]. As the method depends on matching, the response was slow and not affected due to partial occlusion. To overcome on this, [6] stated novel four stage wavelet method to detect the change and canny edge method find edges from frame.

Through the current frame's moving edge can be detected and tracked by grouping with the current frame edge, background edge map and past frame's moving edge. But these methods were very prune to noise introduced in frames and not suitable for real-time object detection. [7] Subtly combined multiple low-level features and uniqueness of target object for visual object detection in dynamic scenes.[8] Using approach of combining top-down recognition with bottom up image segmentation, an hypothesis generation step is implemented. With improved shape context feature and verification step with false positive pruning object detection rate was improved. But failed to identify similar category objects and poor performance under occlusions.

For general object category above methods are useful, but for traffic scenes, where the moving car detection is required, these are not much suitable. The pixel-based background – foreground subtraction approach submitted by [9] with thresholding, binary operation and histogram equalization is used for multiple object detection in traffic scenes. The moving object from dynamic background was detected with SIFT algorithm for adaptive detection [10]. In this method, the three images captured and using SIFT, the features were matched with surrounding points on object template of three images. Thus, the speed and rate matching was improved, but still issue with occlusion. In [11], Background subtraction with Mean shift and Kanade-Lucas Tomasi (KLT) applied and the good features were tracked to improve the object detection under occlusion conditions. The tracking of object was achieved by Optical Flow method with KLT.

The challenges like illumination changes in light conditions, rapid variation on target appearance and similar non-target objects and occlusion were damped by [12] video compression techniques like 2D DCT, statistical correlation matching feature points and maximum likelihood method. For object detection under surveillance are, contourlet transform is implemented by [13] with frame separation. However, its limited to foreground and background image classification.

IV. Object Identification

After Object detection is the task where out of all appearing objects from the scene, the specific object has to identify. And the identification stands for, from among all detected objects, identify the specific object. So, an object identification needs extensive set of unique features to learn the model. On railway stations if it has to detect an unattended object, then [14] suggested method with dual tie background subtraction and approximate median model. The system was useful for object identification, but the changes in background affected the results.

Alternative technique [15][16] on unattended object identification improved with background subtraction model on Mixture of Gaussian Techniques. The occlusion issue rectified up to 80% using framework of blob-based HOG and SVM. The above methods are beneficial for still camera and the visible frames. But in real-time the video is not clear and the frame rate is very slow, hence the object identification is critical for long-term period. Hence [17], captured footage from still frames of IPTV video and implemented with sequence of dual background differences. Which was gained by computing the intensity transformation between the current and previous reference background models within a specific time period.

A clustering and an object detector are then integrated to identify the unattended objects. Thus, object identification was improved. But still there are so many challenges to define the parameters for multiple object identification. For each object, the scope of parameters and features is changing. The extraction of features with limited set of parameters and aligning them for learning under other special methods like deep learning may improve the chance to identify the object under scene specific conditions.

V. Deep Learning For Object Identification

Deep convolutional neural network (CNN) becomes a mostly used tool for object identification. Various works have achieved excellent performance on object detection benchmarks. But these works best with present generic detectors where their performance will drop rapidly when they are applied to a surveillance scene. In order to improve the object identification rate, deep background subtraction method is used by [18].

A background image is used with input new image, for subtracting & finding object. Subtraction is done on spatial features by CNN(ConvNETS) and trained under RMSProp for optimization in CNN. The model is trained on LE-NET-5 dataset and the results were compared with CDNET2014 dataset results. Here problem occurred in subtraction operation where ConvNet trained with scene specific data. The category specific salient mode detection [19] worked with 3D objects in 2D & the images are trained in different viewing angles. Such images are then classified in novel framework to determine natural orientations and salient views of 3D objects.

But issue was raised in generalizing the salient features and again train by per categories. Based on datamining, [20] has suggested image mining with instigated as Scale Invariant Feature Transform (SIFT) on K-Nearest Neighbours (KNN) as classifier. Results were obtained by matching the two set of key points that are similar by extracting low- level features in the neighbourhood. The available CNN systems are not well known how to effectively adapt the pretrained model into the new task, so [21] recommended recursive neural network structure on features extracted by a deep convolutional neural network pre-trained on a large data set like Washington RGBD image data set.

This was much improved in results and gained positive point over well-known Alex Net feature extraction technique. Using this pay-as-you-go method, no training at the feature extraction phase is required, and can be performed very effectively as the output features are compact and extremely discriminative. The improvement in the speed of training & object detection is observed in [22] with the trained AlexNet using RNN. System was created on AlexNet-RNN to enhance the accuracy of identifying the object. Now though deep learning is providing some good methods to identify the object, once the model is pretrained, it will not check an new object which is not a set of pretrained features.

This is a most dangerous drawback of above mentioned systems. The results were improved by learning new objects in adaptive manner with new set of features. So, [23] put forward an adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection by three methods of Adaptive-CNN, Sequential CNN (S-CNN) , ADCNN. The context was learned from target domain for adaptive scene. The bounding box is used to identify the object from the scene. Though result was improved, but learning from scene, is critical and having limitations in scene data.

The labelling of objects to find under bounding box is critical and time-consuming task. So, [24][25] proposed method to avoid the labour-intensive task of manual annotations, the system proposed with a semi-supervised approach for training deep convolutional networks on partially labelled data. It is only applicable for scene specific human detection. The best method was introduced by [26] named as You Look Only Once (YOLO). He implemented a single neural network that classifies objects using bounding boxes and class probabilities is utilized. He used multi-scale training that varies between sizes and recognizable patterns and applied on 9000 different objects. The result was improved by 80% than other methods in term of speed. The complete method was initiated under the DarkNET for training purpose with extensive use of heavy GPU's.

Hence, in order to train our customize database model, needs much more complicated GPU's. Though the YOLO is giving accurate object identification results, as the view changes or perception of view changes, the YOLO gives negative results [27][28]. Hence in order to find an object on aerial view, Minimum Distance Bag-of-Words Approach is used with Speeded up Robust Feature Transform (SURF) based algorithm. Using this method results are improved by 16% compared to regular bag-of words approach.

So, considering above factors, the object identification under pure deep learning is possible but, the identification under scene specific condition is difficult. Following are the issues affecting the deep learning approach for scene specific object identification:

• The linking of a series of events or actions taken over time and synchronizing them to meaningful information that relate to achieving outcomes [29]

• The hardware required for deep learning is highly complex. Deep learning default works on black box mechanism, hence direct control is not over the training models[30]

• For optimization and regularization in deep learning, there is an issue in nature of the gradient of the model [31]

• Overfitting is an issue in CNN. Due to network depth an accuracy of output gets saturated and then starts to degrade rapidly. CNN's are not keen to Spatial invariance [32].

VI. Scene Interpretation

From the clustered scene, an identification of object is acute task and only machine learning or deep learning is not sufficient. Now once the object is identified, the sematic analysation of

March-April 2020
ISSN: 0193-4120 Page No. 2168 - 2176

scene is key point of this research. The scene around the object has to interpret the details. So, it will be helpful to take the decisions smartly.

For. E.g. On railway stations, if an unattended bag has been found. So, by manual approach, we may check the CCTV footage and try to find the suspect. But, by applying the sematic analysis, if the bag found as unattended or suspicious, then it will try to interpret the scene as "an blue colour bags found unattended since last 20 minutes".

So, scene interpretation contains various stages as shown in figure 1 below,

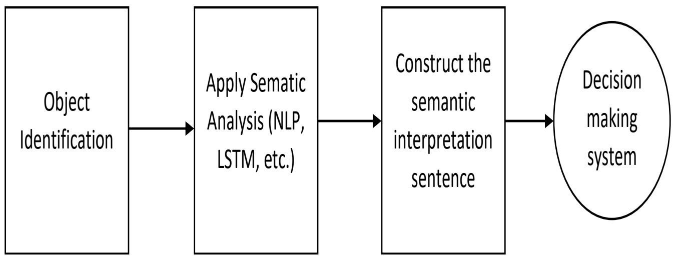

Fig. 1 Stages in scene interpretation

Here, the major objects will be identified from the scene. Then the sematic model is trained using Natural Language Processing techniques like (NLP, LSTM and most advanced like BERT). In last stage, these generated semantics are used to take the automatic decisions after final implementation. Considering the requirement, the extraction of words in sequence with the situation or the association with past and future words to create a sentence, following are state-of-art methods discussed.

In[33], the word2vec model is introduced, where the word embeddings trained on a task where the aim is to predict a word based on its context, mostly using a shallow neural network. But word2vec models using smaller window sizes tends to produce similar embeddings for contrasting words such as "good" and "bad", which is not acceptable for predictive sentence generation. [34], showing some comparison between different NLP methods where, CNN is used. The CNN is useful for learning at multiple output predictions for NLP tasks such as portions, named-entity tags, semantic roles, semantically-similar words and a language model.

It uses look-up table to transform each word into a vector of user-defined dimensions. But the major issue with CNN is long distance dependencies. For semantic analysing machine translation require perseverance of sequential information and long-term dependency. Which is not available in CNN. An improvement on CNN is given by new R-CNN approach with sequence modelling ability to model variable length of text [34]. The deep LSTMs have been presented to generate reasonable task-specific text in tasks such as machine translation and image captioning.

Using LSTM, the semantic analysis was improved with condition to preserve the part and future words predictions in memory. But repeatedly it fails due to attend to the input text and image to form episodes of information improved at each node. To overcome on this, [35] developed a system as SWAG, means Situations With Adversarial Generations. Its implementation was creating a large-scale dataset to support research toward Natural Language Inference (NLI) with natural reasoning.

The model was built with 113K multiple choice questions along with video frames, but still its not with the Adversarial property for defining new dataset by his own. In 2018, [36] google has developed latest technique known as (BERT) Bidirectional Encoder Representations from Transformers. Which helps the model to consider the context from both the left and the right sides of each word. The BERT works for different NLP tasks, including question answering, named entity recognition and other tasks related to general language understanding. They defined the strategy for training a deep bidirectional model with randomly masking few input tokens allowing BERT to better realize relation between sentences.

So, considering above facts, it's important to select the best suitable NLP model for generating the semantics for scene analysis and interpretation.

VII. Conclusion

The smart surveillance needs the object identification with accuracy in clustered areas. From above survey, there ae some methods like global geometric correspondence, bag of features and metric spatial regions (MSR) to label the image dataset. It's important to identify the key factors for outlining the scene factor in accurate semantic. For detection of object, methods alike graph representation of moving objects, four stage wavelet method are tested. But these are very prune to noise in frames, and not suitable for real-time object detection. Thus, by combining top-down recognition with bottom-up scene, improved the issues on noise, but shown poor performance under occlusion conditions. Later pixel-based background-foreground subtraction enhanced the object detection. To track the object an optical flow method is useful with KLT technique.

Mixture of Gaussian Technique gives better object identification with approximate median model. Issue of occlusion rectified up to 80% with blob based HOG and SVM. In real-time the processing rate of video frames is slow, hence need to map another method with high frame rate for processing.

Deep learning does give access to choose what is happening inside the model on features, hence tough following mentioned deep learning method like CNN, RCNN, YOLO are giving the improved results. Still they are restricted to certain categories and extensively depends on hardware. Hence, for our project, need to identify such deep learning model, which will train the features defined manually.

Once the object is identified, the linking of object with scene and making an interpretation of sentence is a tough challenge. So, the methods like word2vec, C-CNN on words, LSTM are presenting reasonable text interpretation like image captioning, named entity recognition, etc. The advanced methods like BERT developed by google, improved the accuracy by considering both left and right side of the words for predicting future words. But the real-time video scene prediction with object in short time have still have a scope for improvement in latency, object accuracy in prediction and in inferencing systems.

VIII. References

[1] S. Lazebnik, C. Schmid, J. Ponce., "Beyond Bags of Features : Spatial Pyramid Matching for Recognizing Natural Scene Categories," Comput. Vis. pattern recognition, 2006 IEEE Comput. Soc. Conf., vol. 2, pp. 2169–2178, 2010.

[2] J. Xiao et al. , "Basic level scene understanding: from labels to structure and beyond," SIGGRAPH Asia 2012 Tech. Briefs, p. 36:1--36:4, 2012.

[3] A. Thippur et al., "A Comparison of Qualitative and Metric Spatial Relation Models for Scene Understanding," Aaai, no. Section 4, pp. 1632–1640, 2015.

[4] M. R. Greene et al., "What you see is what you expect: rapid scene understanding benefits from prior experience," Attention, Perception, Psychophys., vol. 77, no. 4, pp. 1239–1251, 2015.

[5] I. Cohen and G. Medioni, "Detecting and tracking moving objects for video surveillance," in Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), 1999, pp. 319–325.

[6] X. Y. Zhang and R. C. Zhao, "Automatic video object segmentation using wavelet transform and moving edge detection," Proc. 2006 Int. Conf. Mach. Learn. Cybern., vol. 2006, no. August, pp. 3929–3933, 2006.

[7] X. Liu, B. Dai, and H. He, "Real-time object segmentation for visual object detection in dynamic scenes," Proc. 2011 Int. Conf. Soft Comput. Pattern Recognition, SoCPaR 2011, pp. 423–428, 2011.

[8] L. Wang, J. Shi, G. Song, and I. Shen, "Object Detection Combining Recognition and Segmentation," Comput. Vis. – ACCV 2007, pp. 189–199, 2007.

[9] R. Venkatesan and A. B. Ganesh, "Real time implementation on moving object tracking and

March-April 2020
ISSN: 0193-4120 Page No. 2168 - 2176

recognisation using Matlab," in 2012 International Conference on Computing, Communication and Applications, 2012, no. February, pp. 1–8.
[10] Y. Lu, X. Xu, Y. Dai, and B. Zheng, "An adaptive object detection scope algorithm based on SIFT," Proc. 2012 4th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2012, vol. 2, pp. 100–103, 2012.
[11] B. Karasulu and S. Korukoglu, Performance Evaluation Software. New York, NY: Springer New York, 2013.
[12] K. Kalirajan and M. Sudha, "Moving Object Detection for Video Surveillance," Sci. World J., vol. 2015, pp. 1–10, 2015.
[13] R. Venkatesan and A. B. Ganesh, "Real time implementation on moving object tracking and recognisation using Matlab," in 2012 International Conference on Computing, Communication and Applications, 2012, no. February, pp. 1–8.
[14] L. Xiya, W. Jingling, and Z. Qin, "An Abandoned Object Detection System Based on Dual Background and Motion Analysis," in 2012 International Conference on Computer Science and Service System, 2012, pp. 2293–2296.
[15] M. Chitra, "Occlusion and Abandoned Object Detection for Surveillance Applications," vol. 2, no. 6, pp. 708–713, 2013.
[16] L. H. Jadhav and B. F. Momin, "Detection and Identification of Unattended / Removed Objects in Video Surveillance," pp. 1770–1773, 2016.
[17] Wahyono, A. Filonenko, and K. H. Jo, "Unattended Object Identification for Intelligent Surveillance Systems Using Sequence of Dual Background Difference," IEEE Trans. Ind. Informatics, vol. 12, no. 6, pp. 2247–2255, 2016.
[18] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," 2016 International Conference on Systems, Signals and Image Processing (IWSSIP), Bratislava, 2016, pp. 1-4.
[19] S.-H. Kim, Y.-W. Tai, J.-Y. Lee, J. Park, and I. S. Kweon, "Category-Specific Salient View Selection via Deep Convolutional Neural Networks," Computer Graphics Forum, vol. 36, no. 8, pp. 313–328, 2017.
[20] S. Regina and L. Suganthi, "Image Mining using SIFT Based Object Identification and Tagging," pp. 261–266, 2016.
[21] T. Wu, C. J. Zhang, L. Chen, P. Hui, and S. Liu, "Object identification with Pay-As-You-Go crowdsourcing," in 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 578–585.
[22] H. M. Bui, M. Lech, E. Cheng, K. Neville, and I. S. Burnett, "Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure," IEEE Access, vol. 4, no. c, pp. 10059–10066, 2016.
[23] X. Li et al., "Adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection," IEEE Trans. Circuits Syst. Video Technol., vol. 8215, no. c, pp. 1–1, 2017.
[24] S. Wu, S. et al., "Exploiting Target Data to Learn Deep Convolutional Networks for Scene-Adapted Human Detection," IEEE Trans. Image Process., vol. 27, no. 3, pp. 1418–1432, Mar. 2018.
[25] E. Unlu, E. Zenou, and N. Riviere, "Ordered minimum distance bag-of-words approach for aerial object identification," in 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2017, pp. 1–6.
[26] R. Alimuin, A. Guiron, and E. Dadios, "Surveillance systems integration for real time object identification using weighted bounding single neural network," HNICEM 2017 - 9th Int. Conf. Humanoid, Nanotechnology, Inf. Technol. Commun. Control. Environ. Manag., vol. 2018–Janua, pp. 1–6, 2018.
[27] D. Library et al., "Real time object identification using deep convolutional neural networks," no. 0, pp. 0–2.
[28] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey," IEEE Signal Process. Mag., vol. 35, no. 1, pp. 84–100, 2018.
[29] Dhar, Vasant. "The Scope and Challenges for Deep Learning." Big data 3 3 (2015): 127-9 .
[30] P. Shrivastava, "Challenges in Deep Learning," ParallelDots, 26-Jul-2018. [Online]. Available: https://blog.paralleldots.com/data-science/challenges-in-deep-learning/. [Accessed: 04-Apr-2019].
[31] J. Rodriguez and J. Rodriguez, "The Challenges of Optimizing Deep Learning Models," Medium, 26-Dec-2017. [Online]. Available: https://medium.com/@jrodthoughts/the-challenges-of-optimizing-deep-learning models-1c4562aab312. [Accessed: 04-Apr-2019].
[32] "TUM WikiTUM Wiki," Unresolved issues/ Limiting Factors - Convolutional Neural Networks for Image and Video Processing - TUM Wiki. [Online]. Available:

March-April 2020
ISSN: 0193-4120 Page No. 2168 - 2176

https://wiki.tum.de/pages/viewpage.action?pageId=22577466. [Accessed: 04-Apr-2019].

[33] T. Mikolov et al., “10.1162/Jmlr.2003.3.4-5.951,” CrossRef List. Deleted DOIs, vol. 1, pp. 1–9, 2006.

[34] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent Trends in Deep Learning Based Natural Language Processing,” pp. 1–32.

[35] R. Zellers et al., “SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference,” pp. 93–104, 2018.

[36] Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018.

[37] Parekh et al., " A Survey on Object Detection and Tracking Methods," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, pp. 2970-2978, 2014.

[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. arXivpreprint arXiv:1506.02640, 2015.