IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analyzing Digital Image Processing and Neural Network Models for a Pedestrian Crossing the Road

JAYA RELLI [#1], NARAYANAM.R.S.LAKSHMI PRASANTHI [#2]

[#1, #2]Assistant Professor, Department of CSE,
Vignan's Institute of Information Technology(Autonomous),Visakhaptnam,
.

ABSTRACT

In a surveillance system, pedestrian detection and monitoring are crucial for a variety of utility functions, including the identification of unexpected events, analysis of human gait, appraisal of congested or congested locations, gender classification, and the detection of elderly people falling. The main objective of researchers is to create surveillance systems that can function in a dynamic environment, yet creating such systems involves significant problems and difficulties. The three stages of pedestrian detection—video gathering, person detection, and its tracking—present these difficulties. The difficulties in filming include shifting lighting, sudden movements, a complicated background, shadows, object distortion, etc. Different stances, occlusion, crowd density, area monitoring, etc. provide problems to human identification and tracking. Lower recognition rates are the effect of these. In this proposed study, a brief overview of surveillance systems is provided, along with comparisons of pedestrian recognition and tracking methods in video surveillance. Here, we attempt to evaluate the effectiveness of our suggested work, to explore various datasets, and to contrast many well-known CNN models, including F2DNet, Pedestron, and EGCL. We can finally determine which model will provide greater accuracy and efficiency for pedestrians crossing roads by comparing these two.

KEY WORDS:

Pedestrian Detection, Video Surveillance,F2dnet,Pedestron,EGCL,Pedestrian Detection

1. INTRODUCTION

Shopping malls, ATM machines, public spaces, businesses, banks, educational institutions, hospitals, traffic signals, and other places have surveillance cameras. The four essential elements of a scene are a static background, moving items, the appearance of static and dynamic elements, and moving objects. The field of object tracking in video surveillance is flourishing thanks to the development of video coding technology. It takes a long time to monitor video with a person. Therefore, a machine must examine the video and extract the required data for later use. Numerous soft computing techniques are capable of autonomously identifying a variety of items, including people, vehicles, animals, equipment, etc. Video surveillance is the process of analysing or watching a certain location for commercial or safety reasons. The purpose of this method is to make it safer for elderly persons and those with physical disabilities to cross at stop signs. Multiple industries, such as missile tracking, security needs, medical laparoscopy, moving robot design during building crashes, road and forest accident avoidance, etc., can benefit from video surveillance. At the moment, surveillance systems are being extensively researched and used successfully in a variety of applications, including (a) transportation systems (railway stations, airports, urban and ruler motorway road networks), (b) government organisations (military base camps, prisons, strategic infrastructures, radar centres, laboratories, and hospitals), and (c) industrial settings, automated teller machines (ATM), banks, shopping malls, and public buildings, among others. The majority of surveillance systems used in both public and private settings rely on a human operator observer to spot any questionable pedestrian activity[5] in a scene of video [2],[3]. A person who is running or walking down the street is referred to as a pedestrian. A wheelchair user may also be seen as a pedestrian in some areas.

Finding and following suspicious pedestrian activity is the most difficult challenge for autonomous video monitoring. Because it is challenging to gain prior information about every object in a real-time dynamic environment, learning-based methods did not offer an adequate answer for real-time scene analysis[6]. Nevertheless, learning-based approaches are used because of their reliability and precision. Numerous researchers have found that deep learning (DL) based models are more effective than conventional approaches like perceptron models, probabilistic neural networks (PNN), radial basis neural networks (RBN), etc. for classification purposes in video surveillance. Artificial neural networks (ANN), support vector machines (SVM), AdaBoost, and other learning-based methods abound. Such characteristics as the histogram of oriented gradients (HOG), the sped-up robust features (SURF). To categorise the kind of item, use local binary patterns (LBP), scale and invariant feature transforms (SIFT), etc. The deep belief networks (DBN), recurrent neural networks (RNN), generative adversarial networks (GANs), convolutional neural networks (CNN), restricted Boltzmann machine (RBM), AlphaGo, AlphaZero, capsule networks bidirectional encoder representations for transformers (BERT), etc., are specific deep learning algorithm versions that represent these features[10].

The French prefix sur, which means "over," and the verb veiller, which means "to watch," make up the word surveillance. The term "sousveillance" is first used by Steve Mann in [1] in contrast to surveillance. Contrary to the word "over," "sous" means "under," denoting that the camera is physically near a person (ex. camera mounting on head). Both surveillance and sousveillance are employed to keep a close eye on a suspect, a prisoner, a person, a group, or an ongoing activity in order to gather information. Government and business entities have become more confident in using surveillance technology to enhance traditional security measures. In this proposed work we try to analyze the effectiveness on several datasets like: Caltech, CityPersons, LLVIP and so on[8]. These are well known datasets which are available in Kaggle website and all the information is pre-processed and verified by the certified vendors and then try to apply the well-known

CNN models, including F2DNet, Pedestron, and EGCL. We can finally determine which model will provide greater accuracy and efficiency for pedestrians crossing roads by comparing these two [9]-[15].

2. LITERATURE SURVEY

This section will mostly focus on the background research that has been done to demonstrate the effectiveness of our suggested Method. The most crucial stage of the software development process is the literature review. This step is extremely important for the creation of any programme or application since it affects a number of variables, including time, cost, effort, the number of lines of code, and the strength of the firm. Once each of these many requirements has been met, we must choose the operating system and programming language that will be utilised to create the application. When the programmers begin creating the application, they will first look at the pre-defined inventions that have been made using the same concept before attempting to innovate the task.

MOTIVATION

SI NO.	TITLE OF THE PAPER	AUTHOR NAME AND YEAR OF PUBLICATION	TECHNIQUE USED	MERITS	DEMERITS
1	F2DNet: Fast Focal Detection Network for Pedestrian Detection	Abdul Hannan Khan,2022	F2DNet, a novel two-stage detection architecture	A novel two-stage detection architecture which eliminates redundancy of current two-stage detectors by replacing the region proposal network with our focal detection network and bounding box head with our fast suppression head	F2DNet have significantly lesser inference time compared to the current state-of-the-art.
2.	*Generalizable Pedestrian Detection: The Elephant In The Room*	Irtiza Hasan,, 2021	The authors try to conduct a comprehensive study in this paper, using a general principle of direct cross-dataset evaluation.	Under direct cross-dataset evaluation, surprisingly, we find that a general purpose object detector, without pedestrian-tailored adaptation in	more emphasis should be put on cross-dataset evaluation for the future design of generalizable pedestrian detectors

				design, generalizes much better compared to existing state-of-the-art pedestrian detectors.	
3.	Pedestrian Detection by Exemplar-Guided Contrastive Learning	Zebin Lin,, 2021	Exemplar-Guided Contrastive Learning (EGCL) Model	we propose to perform contrastive learning to guide the feature learning in such a way that the semantic distance between pedestrians with different appearances in the learned feature space is minimized to eliminate the appearance diversities, whilst the distance between pedestrians and background is maximized	Only daytime pedestrian detection is possible and night time detection is not accurate and further left as future work.
4.	Focal Loss for Dense Object Detection	Tsung-Yi Lin,2017	RetinaNet : A simple dense detector	Focal Loss focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training	RetinaNet is not able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors.

3. EXISTING METHODOLOGY

In the existing system there was no proper method to classify the images which are captured from CCTV cameras on road networks and identify where the pedestrians are crossing and whether they are crossing correct signal or not. All the existing approaches are manual approaches, hence following are the main limitations in the existing system.

LIMITATION OF EXISTING SYSTEM

1. More Time Delay in finding the noise in those images which are collected from CCTV cameras.
2. There is no technique available to identify pedestrian crossing road or not.
3. There is no technique which can classify based on image processing.
4. All the existing methods failed to use several CNN models and compare which model gives best accuracy.

4. PROPOSEDSYSTEM

In this proposed study, a brief overview of surveillance systems is provided, along with comparisons of pedestrian recognition and tracking methods in video surveillance. Here, we attempt to evaluate the effectiveness of our suggested work, to explore various datasets, and to contrast many well-known CNN models, including F2DNet, Pedestron, and EGCL. We can finally determine which model will provide greater accuracy and efficiency for pedestrians crossing roads by comparing these two.

ADVANTAGES OF PROPOSED SYSTEM

The following are the advantages of proposed system. They are as follows:

1) By using proposed CNN model it takes less time for the identification of pedestrian who is crossing the roads.
2) In this proposed system we can able to get accurate results by eliminating noise on those images which are collected from video surveillance cameras.
3) Result from using neural networks is nearly 100 % in this paper which can classify images accurately.
4)By using this proposed neural network models ,we can check the best model in order to predict the pedestrian who is crossing the roads.
5) The proposed image processing technique is more efficient in order to segment the image despite of the noise which is captured from video surveillance camera.

5. PROPOSED MODEL ARCHITECTURE

The proposed model gives a brief overview of surveillance systems is provided, along with comparisons of pedestrian recognition and tracking methods in video surveillance. Here, we attempt to evaluate the effectiveness of our suggested work, to explore various datasets, and to contrast many well-known CNN models, including F2DNet, Pedestron, and EGCL as best models. Now let us discuss about these models by taking some sample datasets.

1) F2DNet Model
2) Pedestron Model
3) EGCL Model

Initially the dataset is collected from goggle and the dataset may contain several images both noisy and normal images. Once the images are loaded into the application, some images may contain distinct size, properties, length and width. Hence we need to apply normalization on those input images. In this step we try to re-size all the images into single size and we apply pre-processing technique on sample images. In this data pre-processing we can divide the dataset into two phases: One is test and another is train. Now we try to apply appropraoiate CNN model using digital image processing to classify the images captured from road side cameras and then check whether the pedestrian is crossing the road correctly or not.

1) F2DNet Model

The proposed F2DNet Model contains some main properties such as:

A) Feature Extraction

B) Focal Detection Network

C) Fast Suppression Head

D) Pedestrian Detection

These all can be represented clearly in the below figure 1,which is clear explanation of how Pedestrian crossing is identified and how those detections can be done accurately for static and dynamic objects who try to wait and cross the road.

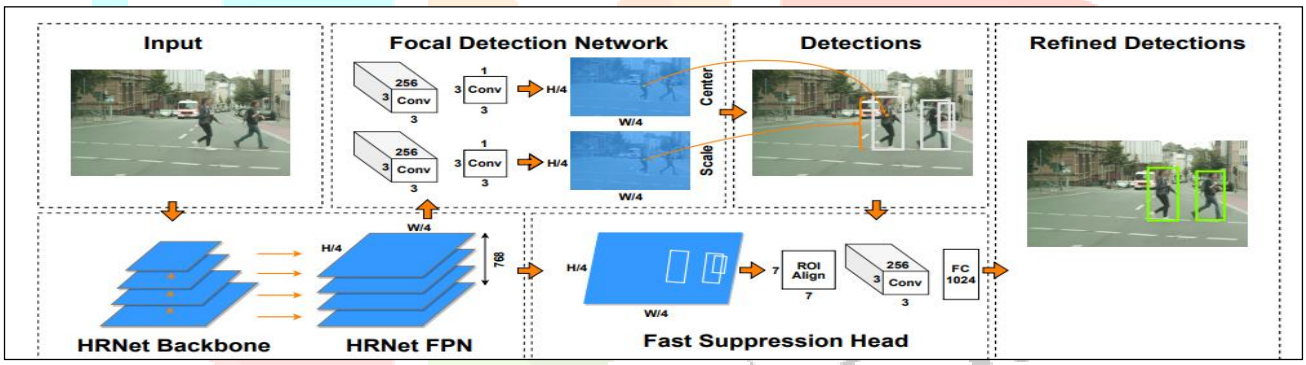

Figure 1. Represent the F2DNet Model for Pedestrian Detection while crossing the roads

From the figure 1,we can clearly identify the proposed model is tested on Caltech Pedestrian dataset which contains nearly 42,782 images with photo resolution of 640 * 480.In this dataset all the images are almost captured from live video sequence collected from different web cameras which are present inside the road.

2) Pedestron Model

In this model we try to gather the Caltech dataset which has around 13K persons extracted from 10 hours of video recorded by a vehicle in Los Angeles, USA. All experiments on Caltech [12] are conducted using new annotations which are recently launched by the several driving datasets.

Evaluation protocol. Following the widely accepted protocol of Caltech [12], the detection performance is evaluated using log average miss rate over False Positive Per Image (FPPI) over range [10−2 , 100] denoted by (MR−2). We evaluate and compare all methods using similar evaluation

settings. We report numbers for different occlusion levels namely, Reasonable, Small, Heavy, Heavy*2 and All unless stated otherwise, definition of each split in table 1.

Setting	Height	Visibility
Reasonable	[50, inf]	[0.65, inf]
Small	[50, 75]	[0.65, inf]
Heavy	[50, inf]	[0.2, 0.65]
Heavy*	[50, inf]	[0.0, 0.65]
All	[20, inf]	[0.2, inf]

Table 1: Represent the Experimental Settings

From the above table 1,we can clearly identify the setting, height and visibility of road images which are captured from video sequence and we apply the Pedestron Model to detect the accuracy of our model.

3) EGCL Model

Our goal is to maximize the distance between pedestrians and background while minimizing the distance between people with different appearances. This is how we want to optimize feature learning for pedestrian recognition. To achieve this, we suggest contrastive learning, where pedestrian detection is seen as a binary classification issue (pedestrian or backdrop). We extract an exemplar dictionary that includes typical pedestrian appearances as prior knowledge to help with contrastive learning in order to increase efficiency and effectiveness. Additionally, the unremarkable exemplar dictionary is used to improve the confidence ratings of hypotheses during inference.

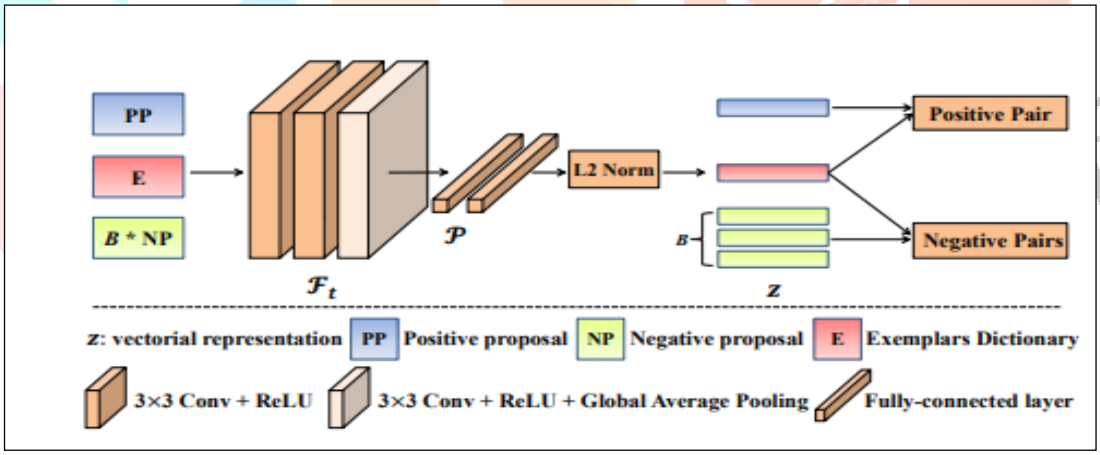

Figure 2. Represent the EGCL Model for Pedestrian Detection while crossing the roads

6. PROPOSED METHODOLOGY

Here in this section we try to discuss about the several models which are used to detect the pedestrian who is crossing the road which is collected from live video sequence.

. The proposed application is mainly divided into 4 modules. They are as follows:

1) DATA GATHERING

Here we try to load the data set from Google and once dataset is downloaded we try to load the dataset to the system for performing the operations[2]-[8]. Here we try to download the data from

https://github.com/AbdulHannanKhan/F2DNet and all the images contained both noise and general images.

2) PRE PROCESSING

Data pre-processing is a technique that is used to convert raw data into a clean dataset. The data is gathered from different sources is in raw format which is not feasible for the analysis. **Training and Test data:** Splitting the Dataset into Training set and Test Set Now the next step is to split our dataset into two. Training set and a Test set. We will train our machine learning models on our training set, i.e our machine learning models will try to understand any correlations in our training set and then we will test the models on our test set to examine how accurately it will predict. A general rule of the thumb is to assign 80% of the dataset to training set and therefore the remaining 20% to test set.

3)APPLY MODELS

Once the data pre-processing is completed now we apply several models on the pre-processed data , in order to check which model is having more accuracy and how effectively this can be applied on several datasets.

4) INTERPRETATION

The data set used for is further spitted into two sets consisting of two third as training set and one third as testing set. Here we apply 3 distinct model to classify the images and find out the best model which can predict the accuracy and efficiency in order to predict the pedestrian who is crossing the road.

7. CONCLUSION

In the proposed work, we collected many datasets from Google and Kaggle and used CNN models like F2DNet, Pedestron, and EGCL to provide a quick summary of surveillance systems. All of the information was gathered using video surveillance's pedestrian tracking and recognition techniques. After running numerous tests on our suggested model, we have finally come to the conclusion that the F2DNet model performs better than the other two models when detecting pedestrians crossing the street using security cameras.

8. REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," vol. 28, 2015.

[2] Y. Pang, J. Xie, M. H. Khan, R. M. Anwer, F. S. Khan, and L. Shao, "Mask-guided attention network for occluded pedestrian detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4967–4975.

[3] Z. Xu, B. Li, Y. Yuan, and A. Dang, "Beta r-cnn: Looking into pedestrian detection from another perspective," Advances in Neural Information Processing Systems, vol. 33, pp. 19 953–19 963, 2020.

[4] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6154–6162.

[5] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 4, pp. 743–761, 2011.

[6] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213– 3223.

[7] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "Eurocity persons: A novel benchmark for person detection in traffic scenes," IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 8, pp. 1844–1861, 2019.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21–37.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss ´ for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.

[11] W. Liu, S. Liao, and W. Hu, "Efficient single-stage pedestrian detector by asymptotic localization fitting and multi-scale context encoding," IEEE transactions on image processing, vol. 29, pp. 1413–1425, 2019.

[12] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional onestage object detection," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 9627–9636.

[13] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 734–750.

[14] W. Liu, S. Liao, W. Ren, W. Hu, and Y. Yu, "High-level semantic feature detection: A new perspective for pedestrian detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5187–5196.

[15] W. Wang, "Adapted center and scale prediction: more stable and more accurate," arXiv preprint arXiv:2002.09053, 2020.