LITTER THROW DETECTION SYSTEM USING HISTOGRAM ORIENTED GRADIENTS (HOG)

S. DHARANI[1], P. KANAKA[2], P. TAMILARASI[3], Mr. J. JEBA STANLY[4]

[1] COMPUTER SCIENCE AND ENGINEERING & N.S.N. COLLEGE OF ENGINEERING AND TECHNOLOGY
[2] COMPUTER SCIENCE AND ENGINEERING & N.S.N. COLLEGE OF ENGINEERING AND TECHNOLOGY
[3] COMPUTER SCIENCE AND ENGINEERING & N.S.N. COLLEGE OF ENGINEERING AND TECHNOLOGY
[4] COMPUTER SCIENCE AND ENGINEERING & N.S.N. COLLEGE OF ENGINEERING AND TECHNOLOGY

Abstract -The system will be useful going to detect humans who throw the trash in continental places. Nowadays the waste management system is the major problem in the world. In this society, one of the major problems is that land pollution. People haven't much aware of this issue. Day by day it will be increasing in one particular place and so it can produce some disease from that wastages. The normal places will be converted to dirty places. To overcome this problem, a solution is provided makes the awareness of that people. At the same time avoid making dirty places. Take one particular place like a temple, park, church, college, factories, and so on. first of all, get the input of the photo and phone number of a person and the person will be allowed to get inside. The human position will be detected using the histogram-oriented gradients (HOG) method. The trash will be detected using the tracking algorithm. If the person throws any wastage in the continent place, the person's picture will be captured and the captured image will send to the In Charge. The person's face will be recognized using face recognition and generate the penalty which is sent to the person via SMS. The person should be paid the fine amount. The online payment method is also provided. If the person will not pay the penalty amount inside the place, already the picture is sent to the In Charge. So, the In Charge catches the person when the person returns to the entrance. It can be used in any particular place, private and government places. The cleaning process and land pollution will be reduced due to this project.

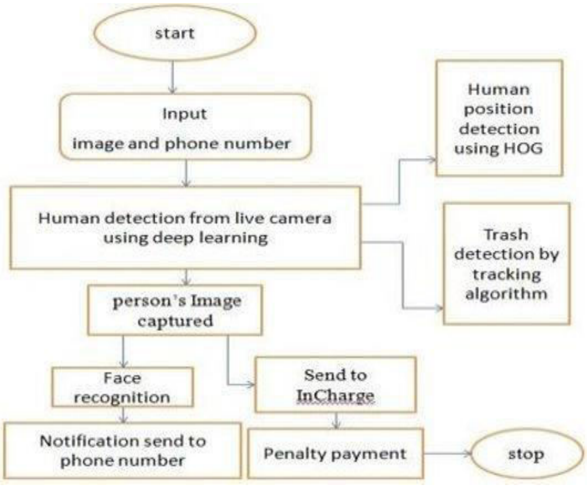
Key Words: convolutional neural network, histogram oriented gradients,

1. INTRODUCTION

Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. First of all, collect the data of people who entered that particular place, and those details are stored in a database using python Tkinter and SQLite. In this section, the photo and phone number of the people data should be collected. All those details are stored in a database and the domain of this project is image processing with deep learning. Using image processing the person's image will be collected and trained. The data are to be trained to check face recognition and finally the data to be checked in the first module. A single image is not convenient for face recognition. So, in this section, the video will be taken and this video is converted to multiple images. Therefore, the face recognition is easily achieved even if the person throws the trash in any direction and angle. The data are trained by the Deep Convolutional Neural Network (DCNN) and at last, the trained model will be checked.

2. METHODOLOGY

The proposed method detects dumping activities by determining the change in relationship between a person and a hand-held object. This approach can handle various dumping actions and discarded objects but it depends on the performance of joint estimation, assuming that the garbage object is visible. If a pedestrian is accurately tracked even in an occluded situation, dumping action can be deduced by determining the person who carried an object and did not have the object after a specific period. In addition, although learning-based methods such as ST-GCN and Multi-CNN do not show satisfactory performance owing to the challenges in this problem, their performances can be improved by combining them with this method through more specific action modeling. Future work will focus on improving the performance through prior knowledge and learning-based reasoning.


start
Input
image and phone number
Human position detection using HOG
Human detection from live camera using deep learning
Trash detection by tracking algorithm
person's Image captured
Face recognition
Send to InCharge
Notification send to phone number
Penalty payment
stop

Data Collection

This module going to collect the photo and phone numbers of the people. these details will be stored in the database.Data collection is the process of gathering and measuring data, information or any variables of interest in a standardized and established manner that enables the collector to answer or test hypothesis and evaluate outcomes of the particular collection. This is an integral, usually initial, component of any research done in any field of study such as the physical and social sciences, business, humanities and others.

Train the data by DCNN

Convolutional neural network, also known as convnets or CNN, is a well-known method in computer vision applications. This type of architecture is dominant to recognize objects from a picture or video.A convolutional neural network is not very difficult to understand. An input image is processed during the convolution phase and later attributed a label.

A typical convnet architecture can be summarized in the picture below. First of all, an image is pushed to the network; this is called the input image. Then, the input image goes through an infinite number of steps; this is the convolutional part of the network. Finally, the neural network can predict the digit on the image.

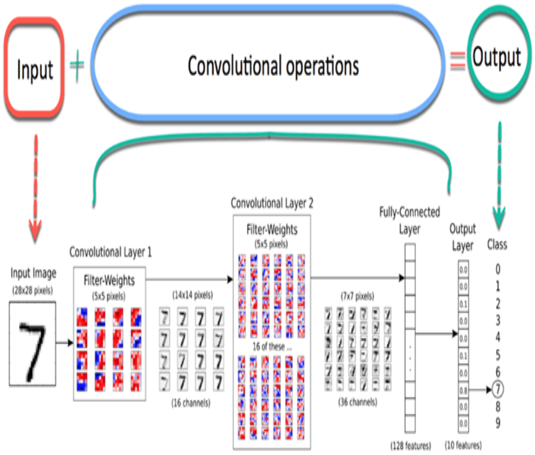
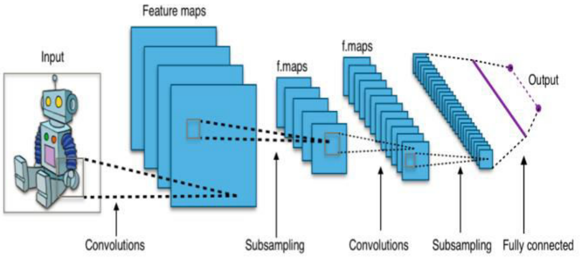


Image recognition with CNNs trained by gradient descent

A system to recognize hand-written ZIP Code numbersinvolved convolutions in which the kernel coefficients had been laboriously hand designed. Yann LeCun used back-propagation to learn the convolution kernel coefficients directly from images of hand-written numbers. Learning was thus fully automatic, performed better than manual coefficient design, and was suited to a broader range of image recognition problems and image types. This approach became a foundation of modern computer vision.

Deep convolutional neural network

A **deep convolutional neural network (DCNN)** consists of many neural network layers. Two different types of layers, convolutional and pooling, are typically alternated. The depth of each filter increases from left to right in the network. The last stage is typically made of one or more fully connected layers:



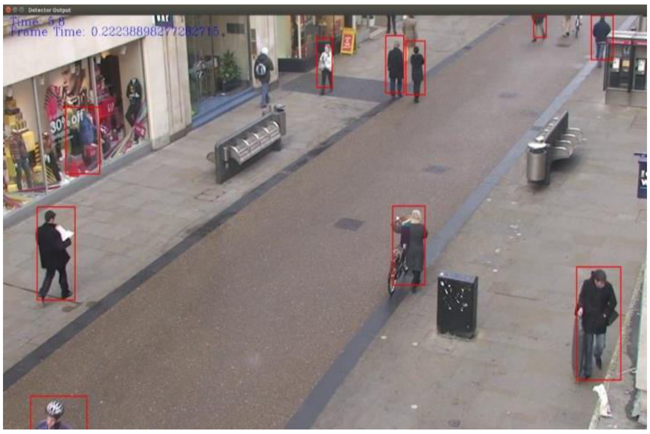
Human Position detection Using Histogram Oriented Gradients (HOG)

Human Detection is a branch of Object Detection. Human presence detection or human sensing refers to technology used to determine if a person is present in a particular environment.This type of technology may be used in a range of applications, whether it's for security or safety, or to enable a device like a smart-home hub to perform key functions. Challenges with Human Sensing Technology.

Human Detection

The basic idea of human pose estimation is understanding people's movements in videos and images. By defining key points (joints) on a human body like wrists, elbows, knees, and ankles in images or videos, the deep learning-based system recognizes a specific posture in space.

Human Detection is a branch of Object Detection. Object Detection is the task of identifying the presence of predefined types of objects in an image. This task involves both **identification of the presence of the objects** and **identification of the rectangular boundary surrounding each object** (Object Localization). An object detection system which can detect the class "Human" can System work as a Human Detection.



Human Position Detection using HOG

HOG pedestrian detection approach is proposed by **N. Dalal** and **B. Triggs** in their paper "**Histograms of oriented gradients for human detection**" published in 2005.**OpenCV** includes inbuilt functionality to provide HOG based detection. It also includes a pre-trained model for Human Detection.

This Python code snippet shows application of **HOG Human Detection using Open CV 3.4.** It shows a frame time of approximately 150–170 milliseconds per frame (equivalent to 6.25 frames-per-second) in my test bench.

3. MODEL IMPLEMENTATION

Foreground detection

Because the video is composed of consecutive images, useful information can be obtained from the temporal property. Among these images, foreground detection provides information about the region where the change occurs in the video. To obtain the foreground region, we first modeled the background, which is a stable area, without changing the input frame and subtracting the input frame from the background. The latest methods using R-PCA 21, 22 or CNN 23, 24 exhibit good performance in foreground detection but they sacrifice speed and require scene-specific training and future frame information. Therefore, our framework adopts a scene conditional background modeling method that can handle the moving camera 25.

Joint confidence map and joint estimation

The information about the posture and joint of a person is useful abstracted information for many applications such as human-computer interaction and behavior understanding. In particular, the extensive pose data 26, 27 and state-of-the-art deep learning-based methods have enabled robust and fast joint detection 28, 29. Among them, we adopt the algorithm used by 28 called Openpose. This estimates a multiperson 2D pose using the multistage convolutional neural network that learns joint locations and associations. After an input image passes through this network, a joint confidence map that has the same input size and joint coordinates with discrete positions are obtained.

Pedestrian tracking

In the previous section, we obtained the joint confidence map and discrete joint coordinates by the pose estimation method. However, because it is a detection-based method that operates on single frame without temporal information, we cannot accurately determine the information about every person owing to false positives and false negatives (missing). Apart from the problems of missing and false alarms, pose estimation generates an output regardless of the order of existence of multiple people. To ensure that each person has the same ID over time, a multitarget tracking scheme is required. We employed the tracking-by-detection framework 30 based on the Hungarian method 31, which operates online in real time. While the original method used a full-body bounding box from the deformable part model 32, the whole-body boxes often are overlapped and occluded when people walk together. When each person's bounding box overlaps, a complex cost function for matching is required to correctly link it with the tracking process. Additionally, the size of full-body bounding box significantly changes depending on the movement of the hand and foot but the change in the size of head bounding box is relatively smaller. The position of the head bounding box has a tendency to move linearly in the direction that the person is heading to, which assists in linking the detections to the tracking process. Therefore, we used the head bounding boxes as inputs to pedestrian tracking.

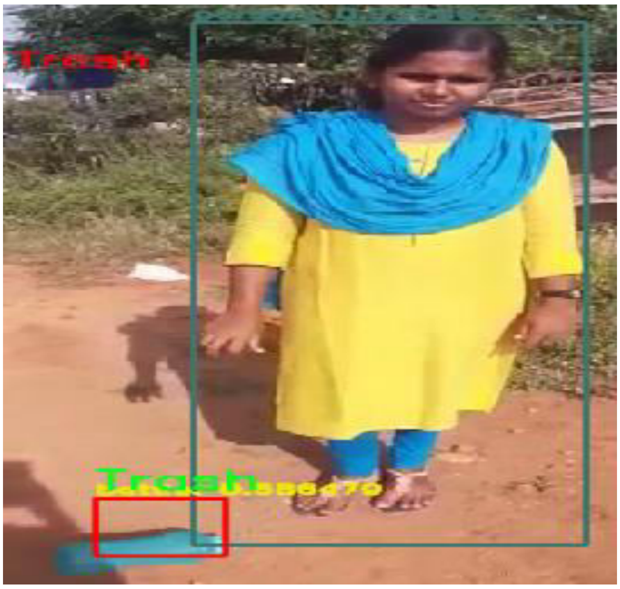
Carrying object detection

The dumping action can be defined as a situation in which a person is separated from a human carrying an object. If a bounding box of a person and an object is given as the ground truth, the situation where two objects are moving away can be easily detected 12. However, objects that people carry are difficult to detect even with state-of-the-art detection algorithms. Figure 4 shows the detection results of Faster R-CNN 3 trained on the COCO dataset 26. Pedestrians are relatively well-detected but human-carried objects are rarely detectable because it is difficult to define the shape characteristics of humans carrying objects. In other words, because the type of garbage is diverse, the performance of garbage detection is not satisfactory owing to the intra-variation problem.

Tracking of carrying object

Although brefine as a human-held object rectangle is obtained, it is not efficiently detected when the confidence of hand joint is low or the foreground has some noise. To compensate for this instability, we combined single-target tracking to maintain temporal consistency in the object region. Among the various single-target tracking algorithms 34-36, we adopted the kernelized correlation filter method (KCF) 37. This correlation filter-based tracking is very efficient as well as provides a comparable performance in comparison with deep learning-based tracking. However, single-target tracking methods are generally assumed when the position of an object is accurately given in the first frame. Instead, in our problem, the position brefine in Section 3.4 is used for the object's initial position. We also implemented a reinitialization scheme when the tracker confidence was smaller than the given threshold, implying that our method discards the error accumulated in the tracker and finds the object using the scheme in Section 3.4, which enables robust tracking like a tracking-by-detection scheme. It brings together the local features from the earlier convolutional layers. Figure below shows 1D vector as the input layer.

Finally, the features of 2 fully connected, Dense layers, which are fundamentally ANN (Artificial Neural Networks) classiers. The net outputs distribution of probability of every class can be found in the final layer, where the net outputs distribution of probability of each class. The diagram above shows 1D vector as the input layer. The Figure below shows 1D vector as the input layer.



4. CONCLUSIONS

A two stage model has been built for classification of a skin lesion into melanoma and non melanoma type. Labeled dataset of dermoscopic images consisting of both melanoma and non melanoma images were used to train the two stages. Results obtained for segmentation stage using UNET is good enough for the targeted application. Good accuracy was obtained from FCRN classification stage, but since the targeted application is for medical diagnosis of false negatives needed to be less. Therefore, increasing recall factor had to be given more priority than increasing the overall accuracy. Using weighted cross entropy as the loss function improved the recall factor by 7% than recall factor obtained from binary cross entropy as the loss function.

REFERENCES

[1] S. Ren "Towards real-time object detection with region proposal networks" , March-2017.

[2] K. Simonyan and A. Zisserman " Two-stream convolutional networks for action recognition in videos" , December-2014.

[3] D. Tran "Learning spatiotemporal features with 3D convolutional networks" , December-2015.

[4] L. Wang " Temporal segment networks: Towards good practices for deep action recognition" , October-2016.

[5] F. Porikli, Y. Ivanov, and T. Haga " Robust abandoned object detection using dual foregrounds" , January-2008.

[6] R.H. Evangelio and T. Sikora "Complementary background models for the detection of static and moving objects in crowded environments" , March-20011.

[7] G. Chunhui "A video dataset of spatio-temporally localized atomic visual actions" , May-2017.

[8] K. Yun, Y. Yoo, and J.Y. Choi "Motion interaction field for detection of abnormal interactions" , April-2017.

[9] R. Csordas, L. Havasi, and T. Sziranyi "Detecting objects thrown over Fence in outdoor scenes" , Jun-2015.

[10] S. Mahankali "Identification of illegal garbage dumping with video analytics" , September-2018.