PREVENTING POTENTIAL ROBBERY CRIMES USING DEEP LEARNING ALGORITHM OF DATA PROCESSING

Roman Prodan, M.Sc. Student, Denys Shutka, MSc Student, Vasyl Tataryn, PhD, As.-Prof.

Lviv Polytechnic National University, Ukraine;

e-mail: roman.prodan.mknuo.2022@lpnu.ua

Abstract. Recently, deep learning technologies, namely Neural Networks [1], are attracting more and more attention from businesses and the scientific community, as they help optimize processes and find real solutions to problems much more efficiently and economically than many other approaches. In particular, Neural Networks are well suited for situations when you need to detect objects or look for similar patterns in videos and images, making them relevant in the field of information and measurement technologies in mechatronics and robotics. With the increasing number of robbed apartments and houses every year, addressing this issue has become one of the highest priorities in today's society. By leveraging deep learning techniques, such as Neural Networks, in mechatronics and robotics, innovative solutions can be developed to enhance security systems, enabling more effective detection and prevention of apartment crimes.

To evaluate the performance of our trained network, we conducted extensive experiments on a separate test dataset that was distinct from the training data. We meticulously labeled this dataset to obtain accurate ground truth annotations for comparison. By measuring precision scores, we determined the effectiveness of our model in detecting potential crimes.

Our experiments yielded an accuracy rate of 97% in the detection of potential crimes. This achievement demonstrates the capability of YOLO and the effectiveness of our trained network in accurately identifying criminal activities. The high accuracy rate indicates that our system can effectively assist in property protection efforts, providing a valuable tool for security personnel and law enforcement agencies.

Key words: Crimes, Robbery, Neural Networks, Object detection, Machine Learning.

1. Introduction

Today, one of the most priority tasks to solve is the problem of protecting our homes and property. The history of home security systems dates back to when mankind started building their first homes, although they were very different from the security systems we have today. But humans have always had an instinct to keep their homes, families, and possessions safe. Perhaps we have developed this instinct through a history of warfare or an innate responsibility to protect one another. This need opened the door to the development of both simple and complex means of protecting our homes. Security systems have taken many forms, even before the discovery of electricity and its application in modern technology. As technology advanced, people began to use these architectures and designs to create more advanced systems. But every year the number of robbed apartments only increases.

According to statistics, more than 40,000 cases of apartment robbery were recorded in Ukraine last year, and this number is increasing every year. Professional security companies can place different types of sensors to monitor your home. Sensors can track things like movement in your home, doors being opened, windows being broken, smoke, carbon monoxide levels, and even the temperature of your home. Companies that provide security services must constantly monitor video surveillance cameras to capture the unusual behavior of suspicious persons. A large number of people use surveillance cameras, which are only capable of recording video, but do not provide specific information about potential preparations for a crime. The systems and approaches currently on the market do not provide a sufficient level of security and reliability for modern challenges.

Due to the rapid growth in popularity of using Neural Networks for object recognition in videos, there is a growing need to optimize the analysis of human behavior, and Neural Networks are best suited for this. They can process video streams in real time, and analyze and report on the threat of robbery of the user. Also, this approach is very useful for security companies, which can significantly optimize their work. We need to embrace the concept of computer vision in homes rather than shy away from the idea of exchanging personal data to achieve new levels of protection, safety, comfort, and entertainment. Computer vision combined with ML enables computers/systems via digital images or video to understand what they see.

2. Drawbacks

The limitations of deep learning approaches, such as insufficient data and data quality issues, pose significant challenges in achieving optimal performance and reliability of the system. In the domain of property protection, it becomes imperative to acquire the best possible input data that encompasses various situations related to potential robbery.

Collecting and annotating data for training a deep learning model is a manual and time-consuming task that requires meticulous effort. It involves sourcing relevant images and videos depicting potential robbery scenarios from diverse sources, such as surveillance footage, online platforms, and real-life incidents. Ensuring that the collected data adequately represents the range of possible criminal activities and environmental conditions is crucial for training a robust and accurate model.

3. Goal

The goal of the current article is to study and demonstrate the use of deep learning for preventing apartment crimes, showcasing its effectiveness and potential impact on enhancing security measures.

4. Computer vision for video analyzing

Deep learning, with its ability to process large volumes of data and identify complex patterns, offers a promising approach to tackling the challenges associated with preventing apartment crimes. The study aims to showcase the power of advanced deep learning algorithms which can analyze various data sources, surveillance footage, and sensor data, to detect suspicious activities and potential threats in real time. This technology enables proactive measures to be taken, such as alerting security personnel or triggering automated response systems, thereby reducing response times and preventing crimes before they occur.

History. In the 2010s, there was a boom in Neural Networks. In the early 2010s, studies showed that it is best to use the ReLU activation function and the Adam optimization algorithm for training Neural Networks [2]. These two components helped to train Neural Networks effectively.

In 2012, A. Krizhevsky, I. Sutskever, and G. Hinton made a breakthrough in image classification by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with a deep CNN architecture known as AlexNet. AlexNet utilized multiple convolutional layers, nonlinear activation functions, and techniques such as dropout regularization to improve generalization. This victory significantly boosted the popularity and research interest in CNNs.

4.1. Convolution Neural Networks

Nowadays, Convolutional Neural Networks and the algorithm of backpropagation of errors are the main methods for recognizing objects in a picture.

Convolutional Neural Networks are used as the basis of the Neural Network for object detection. Convolution Neural Network (ConvNet/CNN) [3] - In machine learning, it is the main tool for recognizing and classifying objects, faces, symbols, etc. in pictures and videos. Convolutional Neural Networks are based on multilayer perceptrons and are designed to use minimal preprocessing of images and videos. A Convolutional Neural Network uses a convolution operation during training, which at the same time allows the use of these operations to reduce the amount of information used in memory and work better with images or videos of higher resolution. A convolution operation when training a Neural Network learns to extract reference features in images and videos, such as "edges", "contours" or "faces". At the next layer, from these edges and faces of the neuron, the network tries to recognize repeating fragments of textures, which are then composed into a fragment of the image. Each layer of the Neural Network has its convolution operation. If at the first layers the network learns to recognize patterns of edges and faces, then deeper into the layers of the neuron the network tries textures and parts of objects. As a result of such training, we can correctly classify the picture or highlight the object we need at the final stage.

The main components of Convolutional Neural Networks [4]:

1. Input layer. As described above, Convolutional Neural Networks are trained on an image. CNN distinguishes between images in three color representations: red, green, and blue - this is the so-called RGB representation. The first input to any type of Neural Network is input.

2. Convolution layer [5,12]. The convolutional layer is the most important part of CNN. The convolution layer works as follows: Let us have an image D = 300x150x3 where 300 is the height, 150 is the width and 3 is the number of RGB channels. The kernel is the weight matrix used to calculate the convolution operation. As shown in Figure 2.2, the kernel moves from the right corner with a specified stride and performs element-by-element multiplication of pixels, and after that, the result is summed. The result of the matrix depends on several parameters:

- Stride
- Kernel

The main purpose of this operation is to study the low-level pattern of the image:

- lines,
- corners,
- points,
- contours and faces.

In the deeper layers of the neuron, the network uses the patterns it has already learned and tries to understand high-level fragments of the image. The result of the convolution operation is obtaining a new image matrix, which is used in the following layers of Neural Networks.

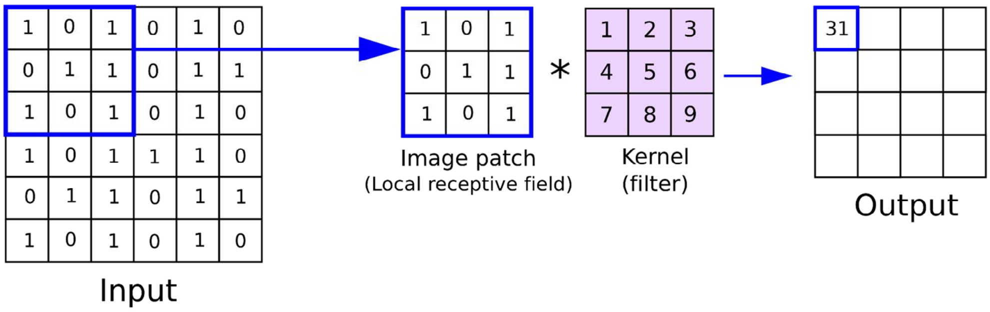

Fig. 1. Convolution layer

3. Pooling layer [6]. The main purpose of this layer is to reduce the dimensionality of the image. Thus, it reduces the number of parameters and the amount of computation performed on the network. Further operations are performed on the reduced image

4. Activation function [7]. This is a mathematical function [6] that changes every negative value of the image to zero and amplifies the non-linear properties of the function. Simply put, an activation function is a function that is added to a Neural Network to help the network learn complex patterns in the data. The activation function takes the output signal from the previous neuron and transforms it into a certain form that can be accepted as input to the next neuron.

5. Forward propagation [8]. In Neural Networks, the forward propagation algorithm is of great importance and performs two main functions during training:

- The sum of products. The algorithm goes in one direction and multiplies the weights of the Neural network by the input vector (image) and then sums them.
- Application of the activation function. When passing through each layer of the Neural Network and obtaining the output matrix as a result of multiplying the weights by the input vector, the Forward Propagation Algorithm transfers the sum to the activation function. Namely, the activation function activates neurons that are most suitable for training data and transmits these values to the next layer of the Neural Network.

6. Loss function [9]. The loss function is an important stage of the Neural Network, it evaluates how well and accurately the Neural Network works. The cost function is a measure of how wrong the model is in estimating the relationship between the predicted data and the actual data. The function is applied after each pass of the forward propagation algorithm. The main goal is to provide information for the error backpropagation algorithm by how much the weights need to be updated to get a better result at the next iteration of the network.

7. Backpropagation [10] became one of the methods that revolutionized the presentation and efficiency of Neural Networks. Backpropagation is an iterative and recursive method. After the forward propagation algorithm has run and the cost function is calculated, we know how well our model performs on a given training iteration. According to the value of the cost function, we minimize the error of the Neural Network by differentiating a complex function and calculating gradients. In other words, the backpropagation algorithm goes back through the network, adjusting the weights of the Neural Network so that in the next iteration we get a better result.

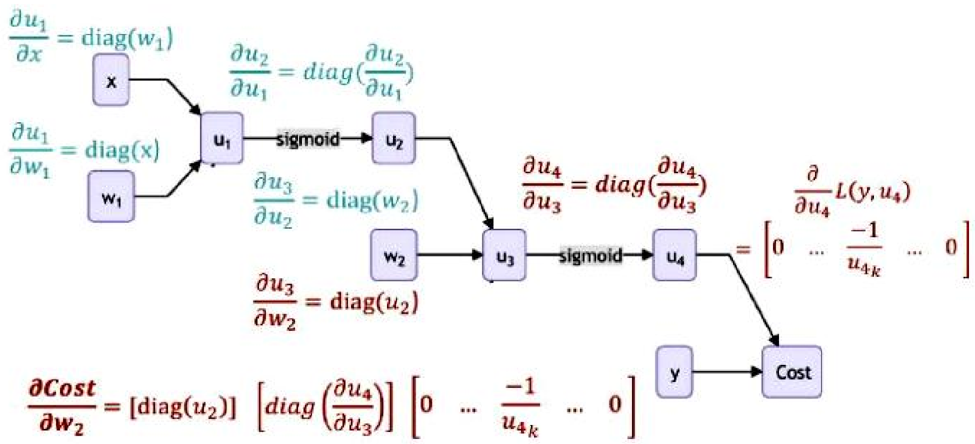

Figure 2. Backpropagation algorithm

8. Detecting objects on images. Neural Networks for detecting objects in the image consist of two parts. The first is an information encoder that uses convolutional networks, and the second is a decoder that predicts bounding boxes and labels for each object.

For our experiments, we used a Neural Network with a Single shot detector [11]. SSD relies on a set of predefined regions. A grid of anchor points is superimposed on the input image, and at each anchor point, objects of different shapes and sizes serve as regions. For each region at each point, the model predicts whether an object exists in that region and changes the location and size of the bounding box to better fit the object. Because each anchor point can have multiple regions and the anchor points can be close together, SSDs create many potential overlapping detections. Post-processing must be applied to the SSD outputs to remove most of these predictions and select the best. The most popular post-processing method is known as non-maximum suppression.

4.2. Realization

To implement the detection of robbers, we used the architecture of a Neural Network for detecting objects in real-time - YOLO4. This model is the fourth generation of the YOLO model. Abbreviation for You Only Looks Once, which means that the network predicts one pass through the network.

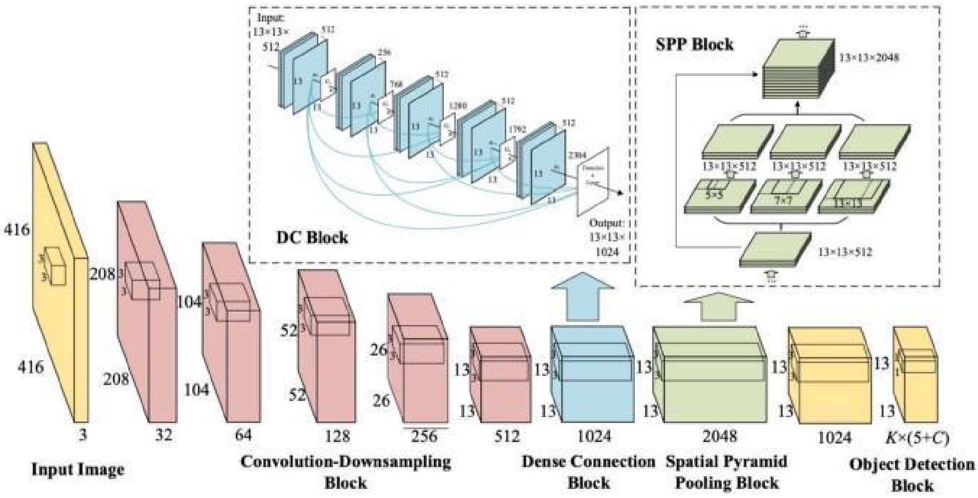

Figure 3. YOLOv4 architecture

Figure 4: Potential robber

YOLO has three main advantages over other types of models:

- Speed: This network has greatly improved the speed of object detection and enables real-time prediction.
- High accuracy: Works very accurately with minimal background errors.
- Training: YOLO architecture is designed to enhance training opportunities.

The YOLO architecture consists of two parts:

- Convolutional Neural Network.

Single Shot detectors, SSDsThe training data set was collected on the Internet. After analyzing several dozens of videos and finding relevant patterns of behavior in robbers. A data dataset of 1000 images was formed, which includes two classes of human behavior:

- potential robber
- normal behavior

After analyzing the data, each image had to be annotated with a corresponding class of behavior using the LabelMe tool.

To train the YOLO Neural Network, we used the Darknet framework, which already has tools for starting the training and debugging the process. To start training, you need to create two TXT files where the dataset will be divided into training and test samples. The dataset was split concerning 80% of the data for training and 20% for testing. The training process is launched using built-in Darknet methods.

5. Results

In our experiments, we apply Mean Average Precision(mAP) [13] to evaluate the network. This is a metric for the evaluation of a model for detecting objects in an image. mAP is based on the following indicators:

- Confusion matrix;
- Precision

mAP - is calculated as the average accuracy for each class and then averaged over all classes.

$$mAP = \frac{1}{n}\sum_{i=1}^{N} AP_i.$$

After the implementation of all components, some experiments were conducted with the training of the Yolo Neural Network, and the results of the metrics for the evaluation of the Neural Network were obtained.

During training, the cost function as the main indicator of the stable training of the Neural Network always appeared as a decreasing graph, which indicates that the Neural Network is well trained.

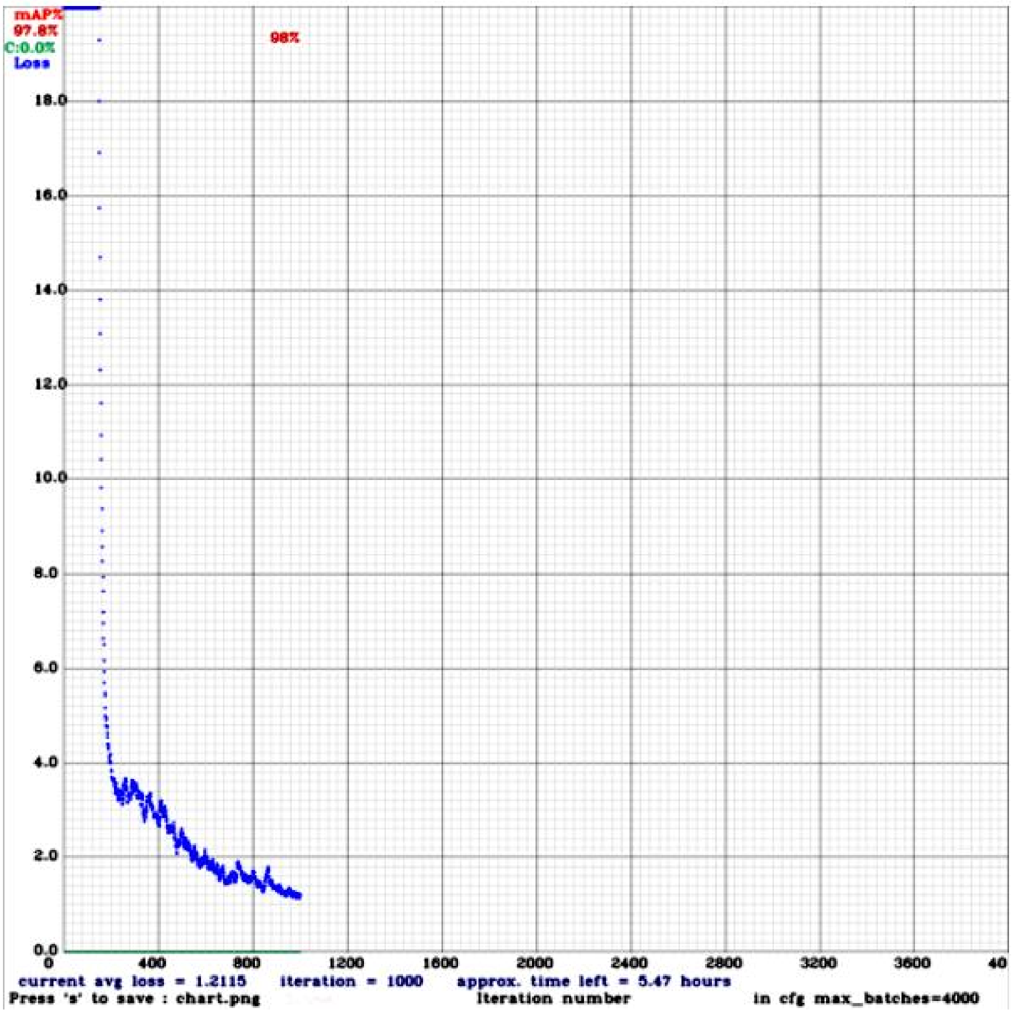

Figure 5: Results of the cost function

```
 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.978022, or 97.80 %
Total Detection Time: 1 Seconds
```

Figure 6: Result of Mean Average Precision

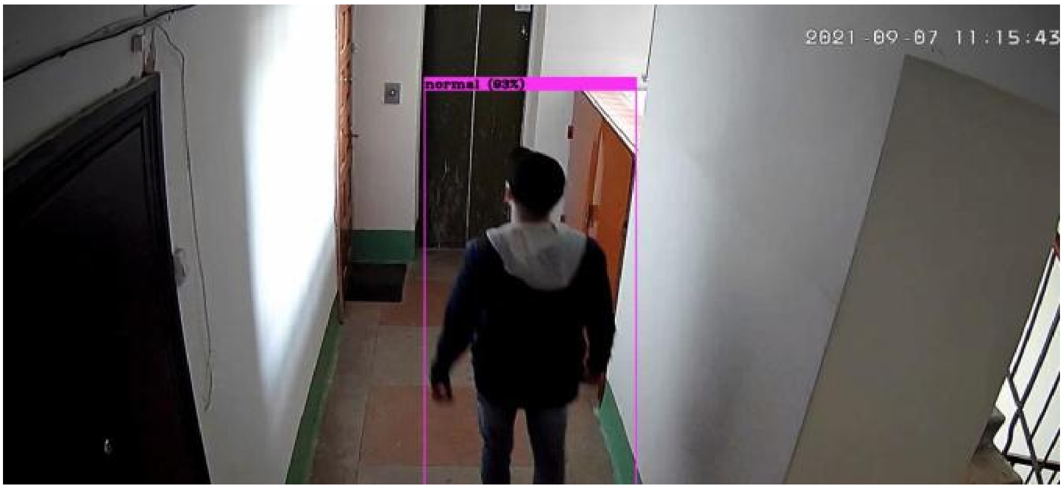

Figure 7: Detecting normal behavior.

Figure 8. Detecting potential robber behavior.

After training, the cost function stopped at a value of 0.93. Also, the Mean Average Precision (mAP) metric showed a good result for evaluating how well the Neural Network distinguishes different human behaviors and can accurately predict them.

The result of Mean Average precision reached 0.97, which is a good result.

The method described by us for detecting potential robbery based on the YOLO Neural Network works quite accurately. The Neural Network recognizes the potential behavior of criminals with an exactness of more than 95%.

6. Conclusions

Using Neural Networks to detect the behavior of criminals and prevent potential crimes by analyzing the behavior patterns of robbers. Based on the Darknet framework and the YOLO Neural network, the necessary detection accuracy of various human behaviors was achieved by network training, and it was tested in real situations. To operate better, a large set of data is needed. It enables the model to reflect exactly different human behaviors. The carried-out experiments permit the evaluation of the prospects of such a method in the field of recognizing the behavior of criminals.

7. Gratitude

The authors express their gratitude to the Staff members of the Department of Information-Measuring Technologies of Lviv Polytechnic National University.

8. Mutual claims of authors

The authors declare the absence of any financial or other potential conflict related to this work.

References

[1] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model for Information Storage And Organization In The Brain". Available - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.335.3398&rep=rep 1&type=pdf

[2] Diederik P. Kingma, Jimmy Ba. (2010). “Adam: A Method for Stochastic Optimization”. Available - https://arxiv.org/abs/1412.6980

[3] Bengio, Y. & Lecun, Yann. (1997). “Convolutional Networks for Images, Speech, and Time-Series.”. Available - https://www.researchgate.net/publication/216792820_Convolutional_Networks_for_Images_Speech_and_Time-Series.

[4] Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016). “Deep learning book”. Available - https://www.deeplearningbook.org/

[5] Keiron O'Shea, Ryan Nash. (2015). “An Introduction to Convolutional Neural Networks”. Available - https://arxiv.org/abs/1511.08458

[6] Hossein Gholamalinezhad1, Hossein Khosravi. “Pooling Methods in Deep Neural Networks” (2009). Available - https://arxiv.org/pdf/2009.07485.pdf

[7] Shiv Ram Dubey, Satish Kumar Singh, Bidyut Baran Chaudhuri. (2021). “Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark”. Available - https://arxiv.org/abs/2109.14545

[8] Geoffrey Hinton (2022). “The Forward-Forward Algorithm: Some Preliminary Investigations”. Available - https://arxiv.org/abs/2212.13345

[9] Katarzyna Janocha, Wojciech Marian Czarnecki. (2017). “On Loss Functions for Deep Neural Networks in Classification”. Available - https://arxiv.org/abs/1702.05659

[10] David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams. (1986). “Learning representations by back-propagating errors”. Available - https://www.nature.com/articles/323533a0

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg (2015) “SSD: Single Shot MultiBox Detector”. Available - https://arxiv.org/abs/1512.02325.

[12] Bernacki, Mariusz; Włodarczyk, Przemysław (2004). "Principles of training Multi-layer Neural Network using backpropagation". Available - http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

[13] Rafael Padilla, Sergio L. Netto, Eduardo A. B. da Silva (2022). “A Survey on Performance Metrics for Object-Detection Algorithms”. Available - https://ieeexplore.ieee.org/document/9145130