IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

# Enhancing Live Video Streams with Real-Time Object Detection using Deep Learning

**Kushal C[1], Suma N R[2]**

MCA Student, Department of Computer Application, Bangalore Institute of Technology, Bangalore, India[1]

Assistant Professor, Department of Computer Application, Bangalore Institute of Technology, Bangalore, India[2]

**ABSTRACT:** Object detection in complex scenes, crucial across various fields, can achieve higher accuracy using deep learning. This paper introduces a Convolutional Neural Network (CNN) to detect objects in video frames by decomposing video into image frames, preprocessing, and extracting features through edge detection. Using convolutional layers, activation functions like ReLU, and gradient descent optimization, the system tested on ImageNet shows promising results.

## I. INTRODUCTION

Data Science encompasses a broad domain incorporating Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). AI refers to the development of systems capable of performing tasks that normally require human intelligence, making data-driven decisions. ML allows machines to process data and generate insights by integrating applied statistics and optimization theory. DL, a subset of ML, focuses on constructing, training, and deploying large-scale neural networks, leveraging parallel data processing.

The data science workflow can be envisioned as a journey: data science represents the overarching pathway, ML serves as the toolkit, and AI stands as the ultimate objective. Analytics, the combination of discernment and analysis derived from data, can be categorized into three types:

1. **Descriptive Analytics:** This involves understanding historical data and changes in system states post-algorithm application, often presented via dashboards and underpinned by databases.
2. **Diagnostic Analytics:** Root cause analysis forms the core of this type, utilizing data mining techniques and ETL processes to uncover hidden patterns and facts within large data sets.
3. **Predictive Analytics:** This uses historical data to forecast future events through mathematical models integrating AI and ML.

Prescriptive analytics, the final category, focuses on decision-making processes, answering the question "What should I do next?" using knowledge representation techniques and computational toolkits.

Deep learning addresses limitations in accuracy and performance of existing models by enhancing feature representation, notably through Convolutional Neural Networks (CNNs). CNNs, derived from image processing, use filters (e.g., Canny filter) to detect edges and boundaries across multiple layers, starting from raw data input to final output classification.

Classification in neural networks can be binary (single output neuron) or multiclass (multiple class labels). Neural networks process input values with associated weights using activation functions such as Sigmoid, Tanh, and ReLU for hidden layers, and Softmax for multiclass classification.

Key components of a Multilayer Perceptron (MLP) include:

- **Cost Function:** Measures the error in predictions.
- **Backpropagation:** Optimizes weights by minimizing the cost function.
- **Learning Rate (α):** Controls the update step size.
- **Optimizer:** Algorithms like Gradient Descent enhance learning.
- **Regularization Parameters:** Mitigate overfitting.

Training involves epochs, where one epoch equates to a full forward and backward pass of all training samples. Overfitting, characterized by a gap between training and validation accuracy, can be mitigated using dropout (randomly removing connections in hidden layers) and data augmentation (increasing data volume through transformations).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

Underfitting, where the model fails to capture the data patterns, necessitates model adjustments. Early stopping of iterations during training fine-tunes the model, preventing overfitting.
By employing these techniques, deep learning models enhance their generalization capability and accuracy in complex data scenarios.

**Literature Survey**

Lex et al. introduced a dilated kernel-based Convolutional Neural Network (CNN) algorithm for classifying earth surfaces in satellite imagery. This method replaces traditional kernels with dilated kernels to enhance classification accuracy and uses a hybrid dilated CNN model to address information loss issues [1]. Lin Yuan et al. proposed a CNN model for extracting image features from MRI brain scans to classify tumor cells, employing multiple convolutional layers to extract gradient information and combine spatial features [2]. An ensemble-based CNN approach is applied to medical image classification, utilizing datasets like GoogLeNet and AlexNet to evaluate performance [3]. Hou et al. proposed an object detection method in panchromatic images using a spatial template matching approach, incorporating a single shot multibox detector (SSD) to extract multi-scale features for image classification and localization [4]. Tang et al. proposed an object detection mechanism for linking objects within the same frame, using a cuboid network proposal method to extract spatio-temporal features [5]. Fang et al. introduced a Faster R-CNN algorithm to improve object detection, addressing low object recall in bounding boxes encountered with R-CNN [6]. Haijun et al. compared the performance of state-of-the-art algorithms for object detection in UAV satellite images, successfully detecting larger objects like buildings and vehicles but struggling with smaller objects [7]. Vipul et al. conducted a comprehensive survey on object detection using deep learning algorithms, focusing on image datasets and not extending their applicability to video data [8]. Yong Tian et al. introduced a metal object detection system using a Support Vector Machine (SVM) algorithm for wireless charging technology in electric vehicles [9]. A review of object detection in UAV images discussed several deep learning algorithms, including Faster R-CNN, Cascade R-CNN, R-FCN, YOLO, and its variants, noting their application to static images rather than video data [10].

## II. PROPOSED METHODOLOGY

The proposed method utilizes video input to detect objects of interest. The video is decomposed into individual frames, and selected frames are processed for object detection. In the initial step, feature extraction is performed on the images, where not all pixels are fed into the neural network; rather, the shape and structure of the image suffice to identify significant features. Through the feature extraction process, image edges are detected, allowing for the determination of shapes. These shapes are then used to detect higher-level features such as vehicles, people, and signboards.

The first layer in a Convolutional Neural Network (CNN) is the convolutional layer, which comprises four components: the number of filters, filter size, pooling layer, and max pooling layer. The sequence of these components is crucial in the CNN method. The process begins with the convolutional layer, where feature extraction occurs. A low-pass filter is employed to smooth edges, while a high-pass filter identifies edges within the image. Filters are used to train the image, and in this approach, the Sobel operator is utilized for image training. For corner pixel features, zero-padding is applied.
.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

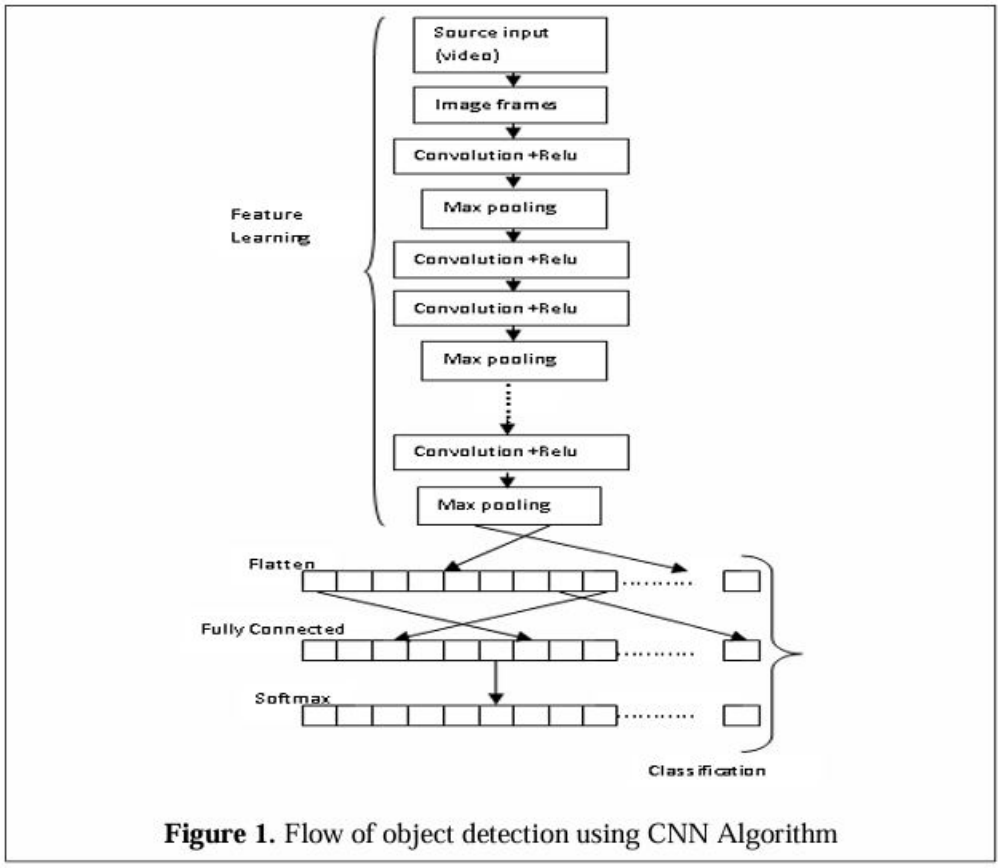


**Figure 1.** Flow of object detection using CNN Algorithm

The kernel value is defined using the formula. If H*W and F*F are defined as 5*5 and 3 * 3, then K can be computed as

$$K = \frac{H - F + 2P}{S} + 1$$

Where s is a stride value. For example H=5, F=3 and P=0, the filter size is odd. This means that the filter size is odd which used to fitting center pixel. Padding can be defined using the formula

$$p = \frac{F - 1}{2}$$

where p=1

Next, we need to define the max pooling layer. This layer is used to reduce the spatial dimensions of the feature maps. There are two types of pooling: max pooling and average pooling. Max pooling reduces the spatial size of the feature map by selecting the maximum pixel value within a defined window. For instance, a 4x4 window size is reduced to a 2x2 window. In our system, the max pooling layer reduces the window size to 2x2, and the stride value ensures non-overlapping movement of the filter. Max pooling generally yields better results in practice.

The convolutional neural network (CNN) is a fundamental algorithm in deep learning. Applications of CNNs include object detection, face recognition, and image classification. Essentially, the computer processes images as arrays of pixels with specific height, width, and depth dimensions. For example, a 6x6x3 array represents an RGB image, while a 6x6x1 array represents a grayscale image. CNNs typically work with training and testing datasets. The process involves applying convolutional filters (kernels), pooling layers, fully connected layers, and finally, the softmax function to classify objects, producing probabilistic values between 0 and 1.

**Convolutional Layer**: This layer extracts features from an input image using a sample window (h*w*d) and a kernel (fh*fw*d). The output is a 2D feature map.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

**Stride**: Stride determines the number of pixel shifts over the input image. A stride of 1 shifts by one pixel, while a stride of 2 shifts by two pixels. Our work uses a stride of 2.

**Padding**: Padding helps fit the kernel in the image window. Zero padding covers the boundary region, ensuring the kernel can process edge pixels.

**ReLU (Rectified Linear Unit)**: The ReLU function, f(x)=max⁡(0,x)f(x) = \max(0, x)f(x)=max(0,x), introduces non-linearity to the network. It provides non-negative linear values, outperforming Tanh and Sigmoid functions.

**Pooling Layer**: This layer reduces the feature map dimensions. Types include:

- **Max Pooling**: Takes the maximum pixel value in a (2*2) grid with a stride of 2.
- **Average Pooling**: Sums all pixel values in the grid and computes the average.
- **Sum Pooling**: Sums all elements in the selected grid.

**Fully Connected Layer**: Flattens the matrix into a vector, which is fed into the neural network. This layer combines features, classifies objects using the softmax function, and optimizes accuracy with the gradient descent algorithm.

## III. EXPERIMENTAL RESULT

The proposed system employs a Convolutional Neural Network (CNN) algorithm for implementation. Video input is processed and decomposed into individual frames, which are then utilized for subsequent processing. The frames undergo preprocessing to remove noise. The Sobel operator is applied for edge and corner pixel detection. Shapes are identified, and based on these shapes, higher-level features are extracted and fed into the CNN layers.

The CNN network consists of several steps, including convolutional filters with activation functions and max-pooling layers.

Figure 2. Sample image for detecting the objects

Figure 2 illustrates an input image of a road view, where the CNN algorithm is applied. This image contains multiple objects such as cars, people, buses, and autos. The system is trained to detect these objects individually. The image is segmented into meaningful regions, and regions of interest (ROIs) are extracted.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

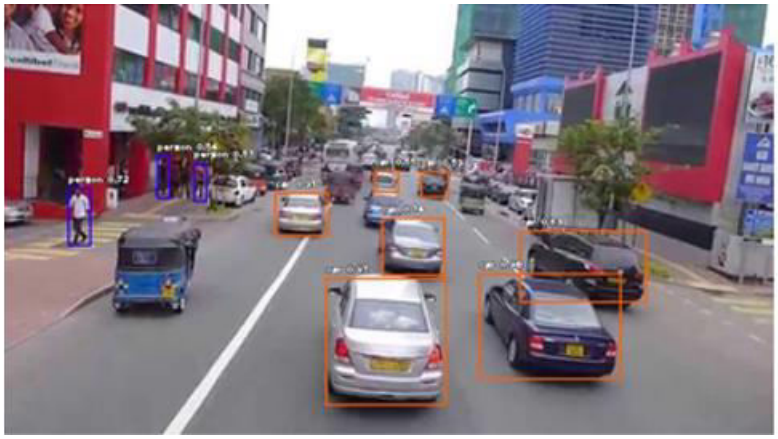

Figure 3. Output image with objects detected using the CNN algorithm

Figure 3 demonstrates object detection in a video frame using the CNN algorithm. Objects such as cars, people, and trees are detected based on the trained data. The output clearly indicates that the algorithm successfully detects these objects.

**Table 1** performance comparison

| Algorithm | Accuracy (%) |
|---|---|
| Random forest | 85 |
| CNN with Gradient descent | 92 |

Table 1 demonstrates the enhancement in detection accuracy achieved through the implementation of the gradient descent algorithm for object detection. Utilizing this algorithm, the accuracy improved from 85% to 92%.

## IV. CONCLUSION

Detecting objects in streaming video is inherently complex due to the substantial amount of information contained within each frame. The continuous video feed is segmented into individual image frames, which are subsequently identified for processing. These frames are divided into training and testing datasets for experimentation. The system successfully processes these images using a Convolutional Neural Network (CNN) optimized with the Gradient Descent algorithm. The integration of the gradient optimization method with the CNN yields highly promising results. This system can be further extended by incorporating the Fast Region-based Convolutional Neural Network (Fast R-CNN) method for analogous datasets.

## REFERENCES

[1] Lei, X., Pan, H., & Huang, X. (2019). A dilated CNN model for image classification. IEEE Access, PP:124087-124094 1–1. doi:10.1109/access. 2019.2927169

[2] Yuan, L., Wei, X., Shen, H., Zeng, L.-L., & Hu, D. (2018). Multi-center Brain Imaging Classification Using A Novel 3D CNN Approach. IEEE Access, PP: 124082-124095, 1 1. doi:10.1109/access.2018.2868813

[3] Kumar, A., Kim, J., Lyndon, D., Fulham, M., & Feng, D. (2017). An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. IEEE Journal of Biomedical and Health Informatics, 21(1), 31–40. doi:10.1109/jbhi.2016.2635663

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207046 |

[4] B. Hou, Z. Ren, W. Zhao, Q. Wu and L. Jiao, "Object Detection in High-Resolution Panchromatic Images Using Deep Models and Spatial Template Matching," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 2, pp. 956-970, Feb. 2020, doi: 10.1109/TGRS.2019.2942103.

[5] P. Tang, C. Wang, X. Wang, W. Liu, W. Zeng and J. Wang, "Object Detection in Videos by High Quality Object Linking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 1272-1278, 1 May 2020, doi: 10.1109/TPAMI.2019.2910529.

[6] F. Fang, L. Li, H. Zhu and J. Lim, "Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection," in IEEE Transactions on Image Processing, vol. 29, pp. 2052 2065, 2020, doi: 10.1109/TIP.2019.2947792.

[7] Haijun Zhang, Mingshan Sun, Qun Li, Linlin Liu, Ming Liu, Yuzhu Ji, "An empirical study of multi-scale object detection in high resolutionUAV images", Neuro Computing 421(20201), PP:173-182

[8] Vipul Sharma Roohie Naaz Mir, A comprehensive and systematic look up into deep learning based object detection techniques: A review, Computer Science Review 38 (2020), 100301

[9] YongTian, ZhengLi, Yawen Lin, LijuanXiang, Xiaoyu Li, Yonghong Shao, JindongTian, Metal object detection for electric vehicle inductive power transfer systems based on hyperspectral imaging, Journal of Measurement 168 (2021) 108493

[10] Payal Mittal, Raman Singh, Akashdeep Sharma, Deep learning-based object detection in low altitude UAV datasets:A survey, Image and Vision Computing, 104 (2020) 104046

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details