JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AUTOMATIC VISUAL SURVEILLANCE SYSTEM TO DETECT DROWNING INCIDENTS IN SWIMMING POOLS

[1]Mrs.Nita Meshram , [2]Chinmay N.M., [3]Divya Lakshmi J.H., [4]Eshwar Sai Chandra, [5]Gautham Narkodu

[1]Associate professor, [2]student, [3]student,[4]student,,[5]student
Department of computer science and Engineering
K.S. School of engineering and management - Bengaluru

Abstract : The issue of drowning incidents poses a critical challenge to water safety, underscoring the imperative for advanced technologies to swiftly and accurately detect such occurrences. This research introduces a real-time drowning detection system leveraging computer vision and deep learning methodologies. The system integrates the You Only Look Once (YOLO) object detection algorithm to identify objects within video frames, prioritizing potential drowning scenarios. Additionally, a customized Convolutional Neural Network (CNN) model is utilized to categorize individuals' swimming status as either "drowning" or "normal" based on their actions captured in the video footage. Functioning in real-time, the system continuously analyzes video frames from a specified source, delivering visual feedback through the depiction of bounding boxes around detected objects and the annotation of their predicted swimming status.

Index Terms **- You Only Look Once (YOLO), Convolutional Neural Network (CNN).**

I. Introduction

the rising incidence of drowning incidents has underscored the critical demand for cutting-edge technological remedies to address the issue bolster water safety. This project introduces an innovative approach to drowning detection, leveraging the capabilities of the deep learning models, often swift and silent, necessitates a proactive and technologically advanced system that can swiftly identify distress signals and initiate timely responses.

II. Population and Sample

The population of interest in our study comprises individuals who engage in swimming activities in public swimming pools. Given the practical constraints of observing every swimmer, a sample was selected to represent this population effectively.To ensure a comprehensive understanding of drowning risks across diverse demographics, our sample was stratified based on age groups (children, adults, and seniors), swimming proficiency levels, and frequency of pool usage. Additionally, we considered geographic diversity by including public swimming pools from urban, suburban.

Sampling was conducted using a multi-stage approach. First, a random selection of public swimming pools was made from a list of facilities provided by municipal authorities. Within each selected pool, systematic sampling was employed to recruit participants based on their presence during designated observation periods. Before participating in the study, individuals were briefed about its nature and voluntarily provided their consent.

III. Data and Sources of Data

The data collection procedures incorporated both quantitative and qualitative approaches to comprehensively capture drowning risk factors. Quantitative data, such as participant demographics, swimming behavior, and environmental conditions, were gathered through structured surveys and automated sensor systems installed within pool premises. Trained researchers conducted direct observations to document swimming behavior, lifeguard interventions, and potential drowning incidents. Furthermore, in-depth interviews were conducted with lifeguards, pool staff, and eyewitnesses to gain insights into drowning events and response procedures. Additionally, In this study, secondary data was sourced from the Robosoft website, consisting of 957 training images and 89 test images.

IV. Theoretical framework

Comprehending the intricacies of drowning incidents and devising effective detection systems hinges upon a robust theoretical framework. This section elucidates the theoretical frameworks that underpin our study.

4.1 Ecological Systems Theory

Bronfenbrenner's Ecological Systems Theory offers a comprehensive framework for comprehending the myriad factors contributing to drowning incidents. According to this theory, individuals are impacted by various environmental systems, spanning from the immediate microsystem (e.g., personal traits) to the broader macrosystem (e.g., societal norms and policies). By examining these interconnected systems, our goal is to delve into the multifaceted aspects of drowning risk and devise interventions across different levels of influence.

4.2 Human Factors Framework

The Human Factors Framework underscores the cognitive and behavioral aspects of human performance within complex systems, including aquatic environments. This framework enables us to grasp the influence of human error, situational awareness, and decision-making processes on safety outcomes.Incorporating human factors principles into our study enables the development of drowning detection systems that account for users' cognitive limitations and improve system usability.

V. Equations

Commonly used evaluation metrics comprise accuracy, sensitivity, specificity, precision, F1 score, and Matthews correlation coefficient (MCC). The accuracy (Ac) metric gauges the percentage of correctly predicted drowning and swimming instances in the testing dataset. Sensitivity or recall (R) quantifies the proportion of drowning instances that were correctly predicted relative to all drowning cases in the dataset. Precision (Pr) measures the percentage of accurately predicted drowning cases out of all predicted drowning cases. Specificity (Sp) assesses the percentage of accurately predicted non-drowning cases out of all non-drowning cases in the dataset. These metrics can be expressed mathematically as follows:

This is equation 1:

$$Ac = (TP+TN) / (TP+TN+FN+FP)$$

This is equation 2:

$$R = TP / (TP+FN)$$

This is equation 3:

$$Pr = TP / (TP+FP)$$

This is equation 4:

$$Sp = TN / (TP+FP)$$

Where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative instances, respectively, the F-measure (F1) is an evaluation metric that combines precision and recall into a unified value. A statistic that strikes a compromise between prediction sensitivity and specificity is the Matthews correlation coefficient (MCC). MCC is a numeric scale that goes from -1, which denotes an inverse prediction, through 0, which stands for a random classifier, to +1, which denotes a flawless prediction.

This is equation 5:

$$F1 = 2PR / (P+R)$$

This is equation 6:

$$MCC = TP*TN - FP*FN / \sqrt{((TP+FN)(TP+FP)(TN+FP)(TN+FN))}$$

Table 1
Sample classification

Real Value	Predicted Value (Positive)	Predicted Value (Negative)
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

VI. RESEARCH METHODOLOGY

6.1 YOLO Object Detection

Employ YOLO to detect objects within video frames. This involves preprocessing the frames, passing them through the YOLO network, and extracting bounding boxes and confidence scores for detected objects.

6.2 Custom CNN Model Inference

Utilize a custom CNN model to perform inference on preprocessed video frames. This involves preprocessing the frames, passing them through the CNN model, and extracting predicted swimming statuses.

6.3 Integration of Open CV and Torch

Integrate Open CV for video stream processing and Torch for deep learning model inference. This includes handling video stream input/output, preprocessing frames, and feeding them to the CNN model.

6.4 Real-time Processing

Implement real-time processing of video frames by continuously reading frames from the video source, performing object detection and swimming status classification on each frame, and displaying the processed frames with visual feedback.

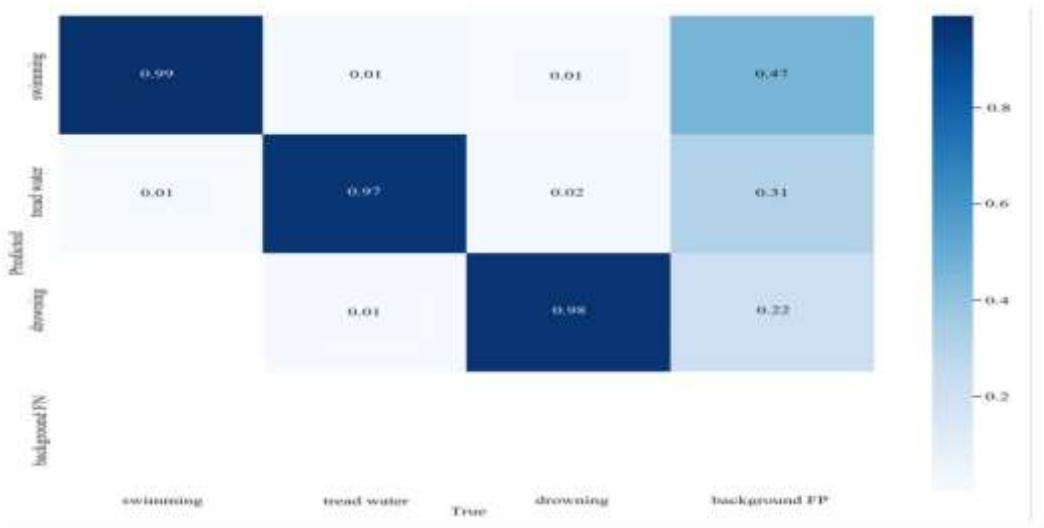

Fig.1 The confusion matrix of the YOLOv3 on the self-made datasets.

Table 2
Architectures of the convolutional neural networks (CNN) used within this study

Classification Model	Architecture	Parameters (1,000,000 s)
Convolutional Neural Network (CNN)	3-layer CNN with Max Pooling.Batch normalization and Dropout Regularization.	~1
Mobile Net	28-layer CNN, Residual Blocks, Max Pooling, Batch Normalization and Dropout Regularization.	~4

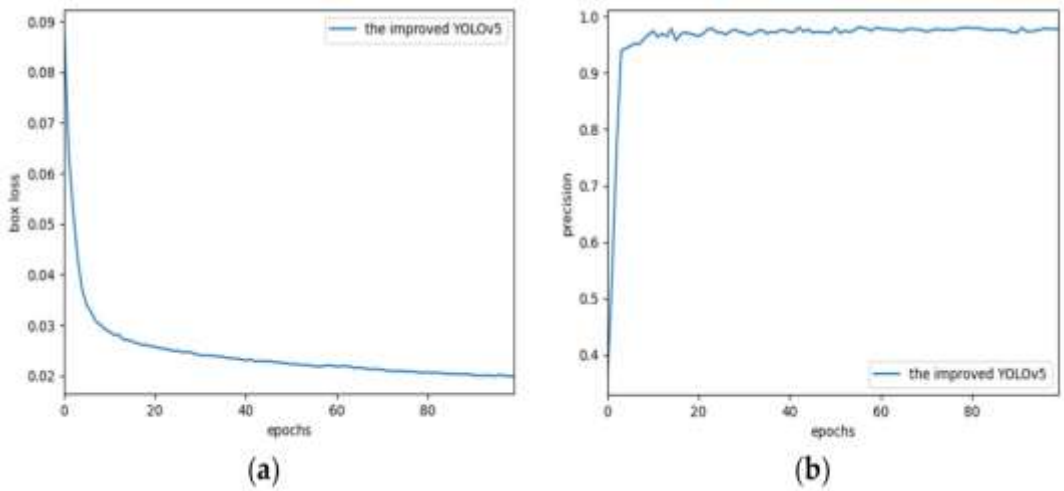

Fig.2 The box loss convergence curve and precision curve on the training of the improved YOLOv3. (**a**) the box loss convergence curve; (**b**) the precision curve.

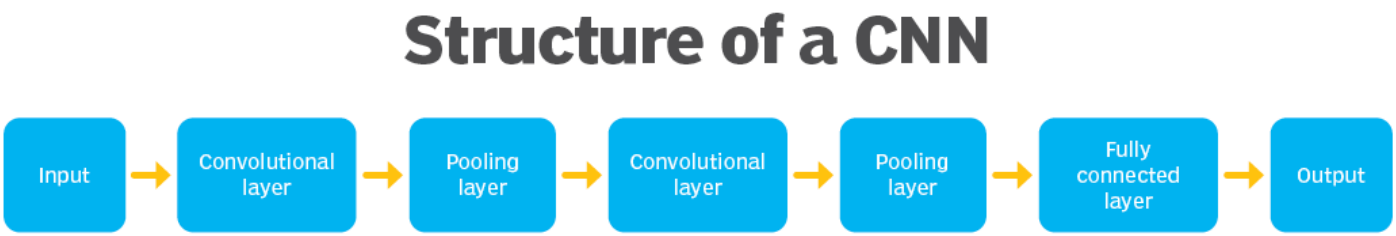

Fig.3 Structure of CNN

A convolutional neural network (CNN) is a category of machine learning models designed specifically for analyzing visual data within deep learning frameworks. Also referred to as convnets, these networks harness principles rooted in linear algebra, particularly convolution operations, to extract features and identify patterns inherent in images. While primarily employed for image processing tasks, CNNs exhibit adaptability in handling other data types, including audio and various signal formats.

Drawn from the structural organization observed in the human brain, particularly in the visual cortex responsible for visual perception and processing, CNN architecture replicates these connectivity patterns.Within a CNN, artificial neurons are strategically arranged to effectively interpret visual stimuli, enabling these models to thoroughly analyze entire images. Leveraging their remarkable ability in object identification, CNNs are extensively applied in computer vision tasks, encompassing image recognition, object detection, and various other domains such as autonomous vehicles, facial recognition systems, and medical image analysis.

In contrast to older neural network architectures, which often necessitated processing visual data in a fragmented manner through segmented or lower-resolution input images, CNNs provide a holistic approach to image recognition. This capability allows them to outperform traditional neural networks in various image-related tasks and, to a lesser extent, in speech and audio processing.

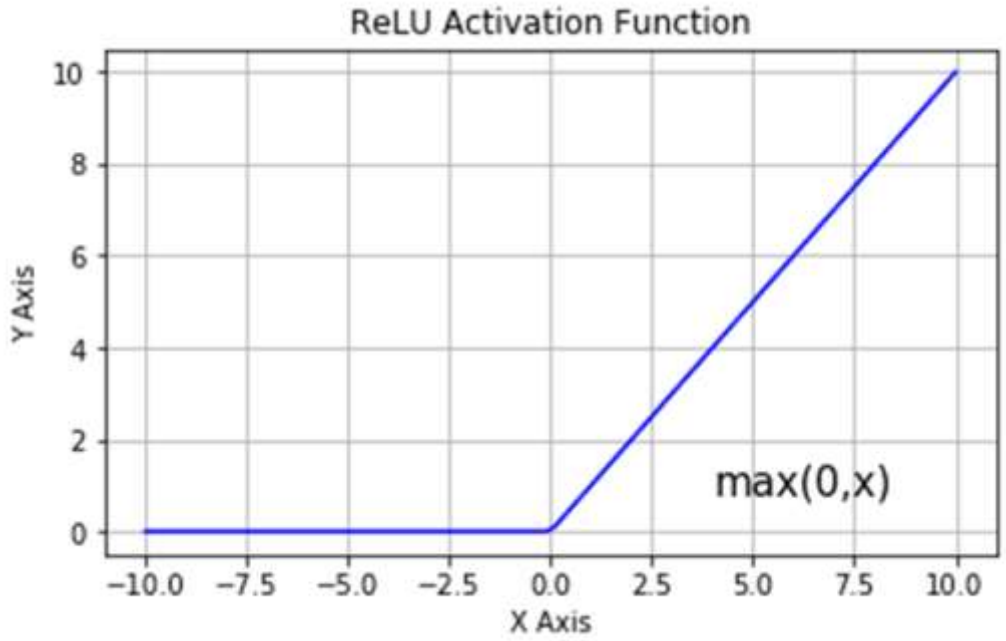

Fig.4 ReLu activation

The Rectified Linear Unit (ReLU) Activation Function is a widely-employed mathematical function in artificial neural networks. Its main role is to introduce non-linearity into the output of a neuron, allowing the network to recognize and understand complex patterns and correlations within the data. The ReLU function outputs the input value if it is positive; however, if the input value is negative, it outputs zero.
Operatively, the ReLU activation function performs a simple mathematical operation on the input value. When the input value is greater than or equal to zero, the output equals the input value. Conversely, if the input value is negative, the output is zero. This function can be succinctly expressed as: \(f(x) = \max(0, x) \).

The ReLU activation function holds significant importance in deep learning and machine learning due to several key attributes:

- Introducing Non-linearity: It is pivotal for enabling the network to capture and understand intricate patterns and relationships in the data.

- Mitigating the Vanishing Gradient Problem: This is a common challenge that can impede the learning process in deep neural networks.
- Computational Efficiency: ReLU has been observed to outperform other activation functions in many cases. such as sigmoid and tanh in terms of computational efficiency.
- Sparse Activation: It offers sparse activation compared to alternative activation functions, potentially enhancing the effectiveness and efficiency of the network.

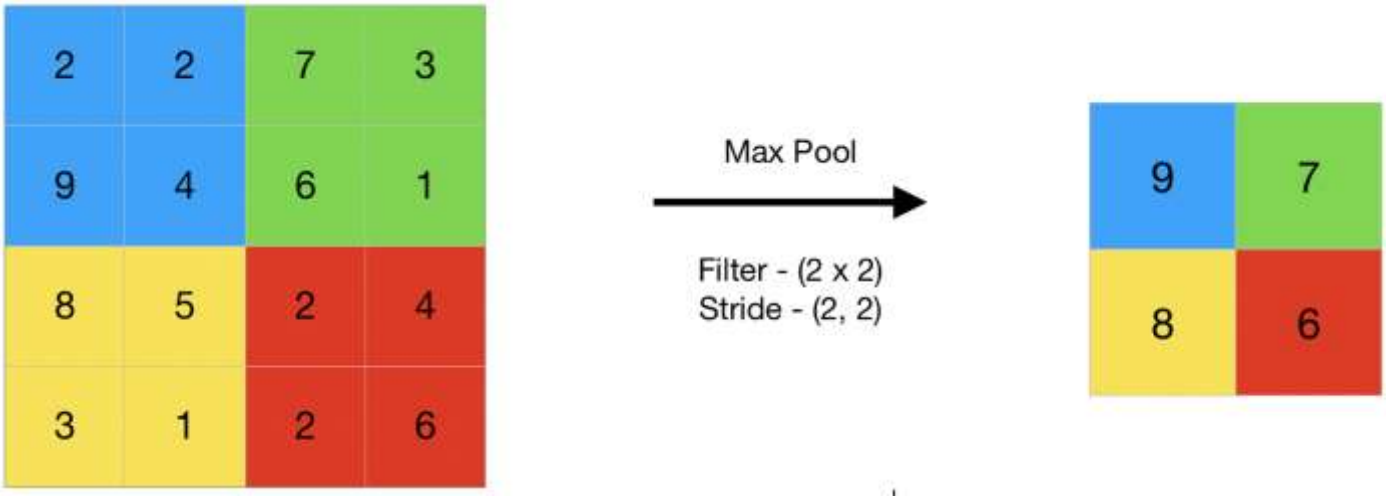

Fig.5 Max pooling

Max pooling is a commonly used downsampling technique in convolutional neural networks (CNNs) to efficiently reduce the spatial dimensions of an input volume. It serves as a form of non-linear downsampling aimed at condensing the representation, thus decreasing the computational load and parameter count within the network. This technique operates independently on each depth slice of the input, modifying its spatial dimensions.

The core objective of max pooling serves two main purposes: firstly, to reduce the volume of information within an image while preserving important features necessary for accurate image recognition. Secondly, it enables the creation of feature detectors that remain unaffected by changes in scale and orientation, thereby improving the network's robustness. Additionally, max pooling helps prevent overfitting by ensuring the model generalizes effectively to unseen data, as it prevents the memorization of irrelevant details.

VII. RESULTS AND DISCUSSION

In conclusion, this paper presents an innovative approach to tackle the critical issue of drowning incidents by employing deep learning techniques. By harnessing computer vision and deep learning algorithms, a real-time drowning detection system has been created, demonstrating the ability to accurately identify potential instances of drowning from video footage. The integration of the You Only Look Once (YOLO) object detection algorithm and a custom Convolutional Neural Network (CNN) model empowers the system to autonomously detect and classify individuals' swimming status as "drowning" or "normal."This approach not only improves the efficiency and accuracy of drowning detection but also facilitates timely intervention to prevent potential tragedies.

Overall, this project showcases the potential of advanced technologies, such as deep learning and computer vision, to greatly enhance water safety through real-time monitoring and detection capabilities.Ongoing research and development in this field hold promise for saving lives and reducing the frequency of drowning incidents in aquatic environments.

VIII. REFERENCES

[1] Awati, R. (2022, September 30). convolutional neural network (CNN). Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural_network Chan, M.(2021,December 14)

[2] Step by Step Implementation: 3D Convolutional Neural Network in Keras. Medium. https://towardsdatascience.com/step-by-step-implementation-3d-convolutionalneural-networkin-keras-12efbdd7b130 Ko, B. (2017, October 16).

[3] Long-term Recurrent Convolutional Network (LRCN).Home. https://kobiso.github.io/research/research-lrcn/ Mahapatra , S. (2018, June 15).

[4]A simple 2D CNN for MNIST digit recognition - Towards Data Science. Medium. https://towardsdatascience.com/a-simple-2d-cnn-for-mnist-digitrecognitiona998dbc1e79a

[5]Video classification with a 3D convolutional neural network | TensorFlow Core. (n.d.). TensorFlow. https://www.tensorflow.org/tutorials/video/video_classification

[6]A Video-Based Drowning Detection System - Alvin H. Kam, Wenmiao Lu, and Wei-Yun Yau https://link.springer.com/content/pdf/10.1007/3-540-47979-1_20.pdf

[7]Long-term Recurrent Convolutional Networks for Visual Recognition and Description https://cseweb.ucsd.edu//classes/wi19/cse291g/student_presentations/Image_Caption_LRCN.pdf