Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 52(1), 2025, Pages 272–287, DOI: 10.52846/ami.v52i1.2118
ISSN: 1223-6934

TV Production Application Based on Artificial Intelligence Specific Algorithms

Daniel-Gheorghe Gagiu, Dorin-Gheorghe Sendrescu, Florina-Luminita Petcu, Stefan-Irinel Cismaru, and Razvan-George Dumitrascu

Abstract. The paper proposes an application that integrates an artificial intelligence model and is intended for the television production environment. This uses a pre-trained model, capable of quickly and accurately identifying relevant video files in a vast archive. This process involves the automatic analysis of metadata and audio-visual content to detect the information sought in each file, resulting in a list of files that match the selection criteria made with the help of implemented YOLO-type algorithms. The application is powered by an artificial intelligence model that can be customized and continuously created through additional training, using input data provided by users. Thus, an efficient and adaptable solution is created, which optimizes archive management and improves the production flow. The proposed application aims to eliminate the time spent identifying audio-video material necessary for editing.

2020 *Mathematics Subject Classification.* Primary 68T01; Secondary 68T40.
Key words and phrases. Artificial intelligence, media, YOLO, dynamic management, tv archive.

1. Introduction

Artificial intelligence (AI) is a vast field of computer science dedicated to developing systems capable of performing tasks that require human intelligence, such as visual recognition, natural language processing and decision-making. In this field there are artificial neural networks, which are able to mimic the structure and functioning of the human brain, being composed of interconnected nodes that process information in parallel. Among these, convolutional neural networks (CNNs) stand out for their ability to process and analyse visual data ([21], [25], [5] and [17]), being widely used in image recognition and classification.

A notable example of the application of CNNs is the YOLO (You Only Look Once) algorithm, which integrates the principle of detecting and localizing objects in real time in images or video streams. This synergy between AI, neural networks and advanced algorithms like YOLO highlights the strong interconnections between concepts and applications, opening new horizons in visual data analysis as shown in ([18], [8] and [23]). Neural networks, especially convolutional ones, have proven their usefulness in a wide range of fields, from technology and medicine to industry and entertainment. In the medical field, CNNs are used for automated imaging diagnostics, accurately identifying abnormalities in X-rays, MRIs or CT scans. In the automotive sector, they play a central role in the development of autonomous vehicles, analysing

Received February 25, 2025. Accepted April 9, 2025.

the environment to detect pedestrians, traffic signs and other vehicles. In the retail industry, AI is used for facial recognition and consumer behaviour analysis, thus optimizing the customer experience and sales processes.

In the production and television environment, CNN applications, which integrate YOLO algorithm, are revolutionizing the management of video archives, automating the identification and indexing of visual content. Their applicability can also be extended to the security field, where these technologies are integrated into surveillance systems for real-time detection of potential threats or suspicious behaviour.

This article deepens these relationships, focusing on the use of convolutional neural networks in television specific applications, such as automatic analysis and identification of video content. It will be analysed how a pre-trained model can be customized through additional training to optimize visual data management processes. These applications highlight the extraordinary potential of AI and neural networks to fundamentally transform various industries, from healthcare and transportation to entertainment and security.

Facial recognition is one of the most explored topics in the field of computer vision and biometrics. Currently, video-based facial recognition has become an intensive research topic due to its diverse applications, such as visual surveillance, access control, and video content analysis. Unlike facial recognition based on still images, the video-based approach poses additional challenges, driven by the large amount of data to be processed and the significant intra- and inter-class variations caused by factors such as motion blur, poor video quality, occlusions, frequent scenario changes, and uncontrolled capture conditions.

Related work

The work of Hassaballah et al. [6] presents a thorough explanation of face recognition and its corresponding issues, achievements, and future directions. Although it explains improvement in face recognition, it mainly focuses on constrained environments and discusses several issues, such as illumination, pose, occlusion and facial expression. The issues provide obstacles in the field of face recognition, while it is also summarizing the results achieved, particularly in controlled conditions. Future research directions for overcoming existing challenges are also given; however, it does not provide much detail on testing or evaluation of competing approaches. Other studies [13] that are focused on aspects of face recognition cover home security where a side face image is extracted from video footage to recognize the identity of an individual. This mentioned study investigates the use of side face images. In fact, side faces are more difficult and less common in comparison to front faces, and shows automated face recognition approach through video side image. The unique selling point of this article is its focus on a precise application domain along with experiments to validate its assertions; however, it is confined that side face recognition for home security and it may not be applicable to other applications or to general issues in face recognition. It is already clear from [6] that there are many techniques of face identification, but the most versatile among them is said to be the one using non-round networks, as the results obtained by it match greatly even under low-light or partial images. In contrast to [13], which uses a photo database, the approach presented by the authors in this paper involves using an artificial intelligence model to identify more accurately the faces of people in a TV archive.

Video-based facial recognition in uncontrolled environments is a technology that focuses on identifying people using images extracted from video sequences captured under varying conditions. Unlike static facial recognition, which uses individual images, this approach involves processing a large volume of video frames, each containing faces affected by factors such as changes in position, lighting, expression, or even partial occlusions. The systems used in this context are complex and integrate several interdependent modules to achieve robust and accurate results [24].

The process of facial recognition in videos is the detection of faces in each video frame. Specialized algorithms locate faces in images using detectors based on convolutional neural networks (CNN's), which can quickly identify facial contours and other distinctive features. Advanced detectors are able to handle faces of different sizes, positioned at varying angles, or even blurred due to poor video quality. After detection, faces are aligned using fiducial points, such as the eyes, nose, or corners of the mouth. This alignment ensures that all faces are standardized before being analysed in depth, reducing variations caused by different head positions ([12], [7] and [10]).

Another aspect of deep learning-based facial recognition is the matching and classification process. Feature vectors extracted from videos are compared to a database of known identities, using metrics such as cosine distance or Euclidean distance. Advanced classifiers, such as fully connected networks or metric learning methods, are integrated to improve accuracy and reduce errors. A major advantage of deep learning approaches is their ability to learn directly from raw data, without the need for manual feature extraction steps. This feature makes the systems more flexible and able to generalize to complex scenarios. Furthermore, the use of large training datasets, such as MS-Celeb-1M or VGGFace2, allows the models to be robust to extreme variations in faces and capture conditions ([11], [16] and [19]).

Collectively, all these works highlight the astonishing advances made in Convolutional neural networks (CNN) research. Besides, they also identify some challenges yet facing this area, such as data efficiency, interpretability, and the possibility of applying these networks into accounts of complex, real world scenarios. This paper takes these efforts and pushes forward to discover the changing face of CNNs and their applications with a contribution towards overcoming existing limitations and moving forward the state of the art.

2. Materials and Methods

Convolutional neural networks (CNNs) are recognized as the most efficient for object detection and segmentation in images, as well as for scene recognition and classification. They have three main advantages: they eliminate the need for manual feature extraction, as they are automatically learned from training data; they offer remarkable performance comparable to or even superior to human performance; and they allow adaptation to new tasks by reusing pre-trained parts using transfer learning.

CNNs have evolved significantly, becoming the foundation of many artificial intelligence applications, especially in the field of computer vision. They have been essential in the development of object detection algorithms, such as the YOLO (You Only Look Once) series, which allows for the identification and localization of objects in real time [20].

These models are preferred due to their ability to quickly and efficiently detect people in images and video sequences, which makes them ideal not only for applications such as video surveillance but also for identifying video files from television archives.

Recent iterations such as YOLOv7, YOLOv8, and YOLOv10 ([3], [15] and [9]) have brought obvious improvements in terms of speed and accuracy optimization, as can be seen in Figure 1.

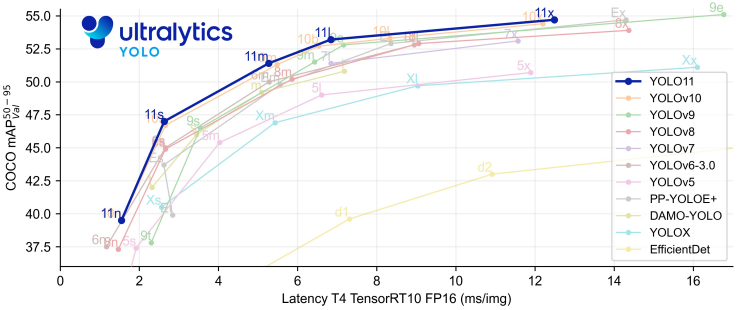

FIGURE 1. Performance of YOLO releases.

The implementation of the application presented in this paper considered testing two convolutional network models, namely YOLOv7 and YOLOv10, both of which have favourable results in terms of correct identification of people. Convolutional neural networks (CNNs) are recognized as the most efficient technologies for detecting, segmenting and classifying objects and scenes in images. They have three essential advantages: they eliminate the need for manual feature extraction, as they are automatically learned from training data; they offer remarkable recognition performance, comparable or superior to the human level; and they allow adaptation to new tasks using transfer learning methods. This flexibility and efficiency make them indispensable in numerous applications in fields such as security, medicine or autonomous vehicles.

Structurally, CNNs consist of two main parts. The first part, known as the convolutional component, consists of an alternation between neural layers with local convolutional connections and aggregation layers that automatically extract relevant features from images by using convolutional and aggregation layers (max pooling/average pooling), transforming the raw image into a feature vector. This vector becomes the input for the final part of the network, a fully connected multilayer perceptron (MLP) or other classical classifier, responsible for identifying or classifying data. The modular structure of CNNs allows them to be versatile, efficient and applicable in various scenarios.

In order to apply these algorithms in image processing, their representation is taken into account, so that digital images are represented as two-dimensional (2D) matrices of integer values, to which is added information about the number of colour channels, also known as colour depth. For example, colour images in RGB format are described by three channels, each corresponding to one of the red, green and blue components. In this context, convolutional filters applied to an RGB image, which represents a three-dimensional volume (dimension $Hi \times Wi \times 3$), must also be defined as three-dimensional volumes ($Hf \times Wf \times 3$). Applying a single such filter generates

a two-dimensional output ($Ho \times Wo$), and using a set of K filters leads to a three-dimensional volume of dimension $Ho \times Wo \times K$, where the additional dimension is determined by the number of filters applied.

For example, in the case of using $K = 5$ filters of size $2 \times 2 \times 3$ applied without padding on an RGB input image with dimensions $4 \times 4 \times 3$ and a translation step $s = 1$, the total number of parameters is determined by the relationship: $(2 \times 2 \times 3 + 1) \times 5 = 65$ and the resulting activation map contains $3 \times 3 \times 5 = 45$ neurons. In an extended scenario, with $K = 50$ filters applied under the same conditions, the total number of parameters becomes 650, and the activation map will have $3 \times 3 \times 50 = 450$ neurons. Figure 2 shows an example of filtering.

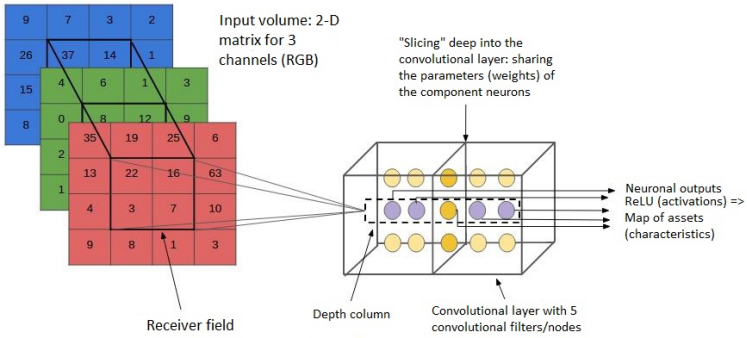

FIGURE 2. Filtering RGB channels of a picture.

An essential aspect to highlight is that, unlike fully connected neural networks of the MLP (Multi-Layer Perceptron) type, in the case of CNN (Convolutional Neural Networks), the number of parameters is not dependent on the size of the input image. For example, if the size of the input image is expanded by 100 times ($40 \times 40 \times 34$), the convolutional layer with 50 filters, according to the previous example, will still require 650 parameters, keeping the complexity of the model constant. In contrast, in a fully connected neural network, the number of parameters would increase proportionally to the size of the image, reaching 65,000 parameters. This fundamental characteristic of CNN highlights their efficiency in terms of scalability and optimal use of computational resources, contributing to the robustness of image processing models.

In the case of convolutional layers, unlike the fully connected layers in an MLP architecture, each neuron in such a layer is connected only to a subset of neurons in the previous layer, in a limited area called the receptive field. This mechanism is analogous to the biological visual system, where specialized neurons respond to stimuli from a restricted region of the visual field. The operation of a convolutional layer involves the application of a convolution filter/kernel, which is translated over the entire distribution of input data. The elements of this filter are connection weights that determine the relationships between the receptive field of the previous layer and the corresponding neuron in the current layer. These weights are learned during the training process of the network, thus optimizing the detection of essential features in images.

Convolutional neural networks (CNNs) are currently one of the most advanced and frequently used technologies in the field of computer vision, being essential for the detection and segmentation of objects in images, as well as for the recognition and classification of scenes and objects. These networks have multiple advantages, among which the following stand out:

- eliminating the need for manual feature extraction, a process performed automatically by learning directly from training data, thus optimizing visual analysis;
- a high level of performance in recognition tasks, comparable or even superior to human capabilities, thanks to its optimized architecture for processing visual data;
- retraining capacity for new recognition tasks, allowing the reuse and adaptation of pretrained components through transfer learning techniques, which facilitates the rapid development of specialized models for various applications.
- Due to these characteristics, CNNs are fundamental in numerous fields, including facial recognition, medical imaging diagnostics, autonomous driving, and advanced video analytics.

3. Results

As a single-stage object detector, YOLOv7 uses an advanced architecture, called YOLOv7-Net, which includes a ResNet-50-based backbone for feature extraction, a Feature Pyramid Network (FPN)-based neck for feature aggregation at different scales, and a Spatial Pyramid Pooling (SPP)-based head for class and coordinate prediction. The algorithm introduces an innovative loss function, YOLOv7-Loss, which combines classification, coordinate regression, confidence, and model regularization, optimizing performance. The YOLOv7-Train training strategy uses pretrained initializations on ImageNet, advanced techniques such as Mosaic augmentation and optimization via Stochastic Gradient Descent (SGD), achieving top results on the MS COCO dataset, with an average accuracy (AP) of 51.4% at a resolution of 640×640 pixels and a speed of 161 FPS on high-performance GPUs.

The algorithm allows for fine-tuning the balance between speed and accuracy, with each implementation using more advanced backbones such as EfficientNet or DenseNet. YOLOv7 achieves competitive performance, outperforming its predecessors and other popular architectures, including SSD, RetinaNet, and Faster R-CNN, in terms of efficiency and scalability.

The mathematical concepts associated with the YOLOv7 architecture involve extracting relevant features from the raw image through convolutional layers. These reduce the gradient degradation problem using skip connections given by the equation:

$$Y = F(x, \{Wi\}) + x \tag{1}$$

where x represents the input to the block, $F(x, \{Wi\})$ is the output of the layer, and Wi represents the weights.

Feature aggregation is the next step that takes place and aims to combine features of different resolution levels, improving the detection of objects with different sizes. Responsible for this stage is the FPN component using combinations of concatenation and up-sampling.

$$P1 = Conv(C1 + Upsample(P1 + 1)) \tag{2}$$

where $C1$ represents the characteristics of level 1, $P1$ the output of the pyramid at level 1, $Upsample$ represents the scaling operation.

Class and coordinate prediction aims to detect objects for each bounding box through the SPP architecture where pooling is applied to regions of variable size. This aspect is represented by

$$z = Conact(Pool_k(f)), k \in \{1, 2, 3, 4\} \tag{3}$$

where $Pool_k(f)$ represents pooling applied with region $k \times k$, f represents the input features, and $Conact$ represents the concatenation of the results.

The fundamental mathematical concepts in YOLOv7 highlight an architecture optimized for fast and accurate object detection in images. Its modular structure, consisting of a backbone based on residual blocks for feature extraction, a neck that aggregates information at multiple scales via FPN, and a head that uses Spatial Pyramid Pooling for spatial context, allows the model to be scalable and robust.

Image data extraction involves the use of object delineation methods to isolate regions of interest. These methods, whether based on squares or polygons, are fundamental in computer vision and image analysis. They allow the location and characterization of relevant features, transforming raw visual data into quantifiable and useful information. Each approach has its own characteristics that influence how the data is processed and analysed, with different applications depending on the needs of the project.

Square-based delineation is a simple and effective method for quickly locating objects. This technique involves drawing a rectangle or square around the region of interest, including the entire object within its boundaries. In many applications, this approach has proven to be robust enough to identify and locate objects in a fast manner. The square allows for a well-defined workflow, in which its coordinates are used to isolate the region and extract information such as the position and size of the object. For example, a square is defined by two opposite coordinates (x_1, y_1) and (x_2, y_2), its area being given by the formula.

$$Area = |x_2 - x_1| \cdot |y_2 - y_1| \tag{4}$$

At the same time, the square is well suited to situations where objects have regular shapes and predictable dimensions. This is the case in facial recognition, where a rectangle can easily enclose a human face, facilitating further processing. However, there are obvious limitations in the case of objects with complex or irregular shapes, as the square often includes irrelevant areas in the background. This can lead to noise and affect the accuracy of algorithms, especially in situations where object details are important.

On the other hand, polygon based delineation is a more advanced and precise method for isolating objects. Polygons are used to trace the exact outline of an object, which significantly reduces the inclusion of background areas and increases the relevance of the extracted data. This methodology is used in applications such as semantic segmentation, where it is necessary that each pixel is correctly assigned to a class. For example, polygons can be defined by a list of coordinates $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, and the area of a simple polygon can be calculated using the formula

$$Perimeter = \sum_{i=1}^{n} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{5}$$

where coordinates (x_{n+1}, y_{n+1}) are equal to (x_1, y_1) for closing the polygon. Therefore, the use of polygons is ideal for objects with complex shapes, such as buildings, vehicles, or biological structures. In medical analysis, for example, polygon based delineation is indispensable for identifying tumours or other anatomical structures that cannot be precisely surrounded by a rectangle. Polygons allow for the capture of every detail, which is crucial for image-based decision-making. However, this approach requires more processing power and resources to generate accurate contours. Also, algorithms using polygons are more complex, requiring advanced neural networks or specialized data processing models. Polygon generation often requires manual intervention or well trained segmentation models. In data labelling applications for machine learning, drawing polygonal contours is a time-consuming process, but indispensable for achieving high accuracy. This method becomes extremely useful in mapping applications or advanced robotics, where objects need to be accurately identified in three-dimensional space. For example, the distances between points of a polygon can be calculated to estimate the perimeter using the relationship:

$$Perimeter = \sum_{i=1}^{n} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{6}$$

where $(x_{n+1}, y_{n+1}) = (x_1, y_1)$.

The integration of polygons and squares into modern workflows also depends on advanced image processing technologies. Deep learning algorithms play a crucial role in creating these delineation methods. Convolutional neural networks, for example, can automatically generate squares or polygons depending on the application requirements. Thus, the process becomes more efficient and accurate, reducing errors and improving data interpretation. In this context, hybrid models that combine the advantages of squares and polygons can be used to meet complex needs, providing balanced solutions for various scenarios. The application developed and presented in this paper involves collecting a significant number of images containing the character.

These images should be representative of the variability that the model should manage. Using specialized tools, the classes that the model should detect (e.g., "John") were defined and labels were also used on each image with the character class. This process often involves drawing a bounding box or polygons around the character and assigning a corresponding label, Figure 3.

Converting the labelled data into a format accepted by the YOLO architecture. This format involves specifying the coordinates of the border and the associated class within the image as shown in Figure 4 and Figure 5.

Once the training images are provided, features are extracted during training epochs. These represent a complete cycle through the entire training data set in the process of learning the artificial intelligence model. Within each epoch, the model goes through all the training examples and adjusts the weights using optimization algorithms. Finally, the epochs have a direct effect on the confusion matrix which has the role of a detailed representation of the results of a model's predictions compared to the true state of the data. Thus, precision and recall are directly influenced and the optimal point of stopping the training is determined.

Both the training images and the test images are analysed using an algorithm that extracts the most relevant textural features. The texture analysis process is the key part of the classification system, having the greatest influence on its performance.

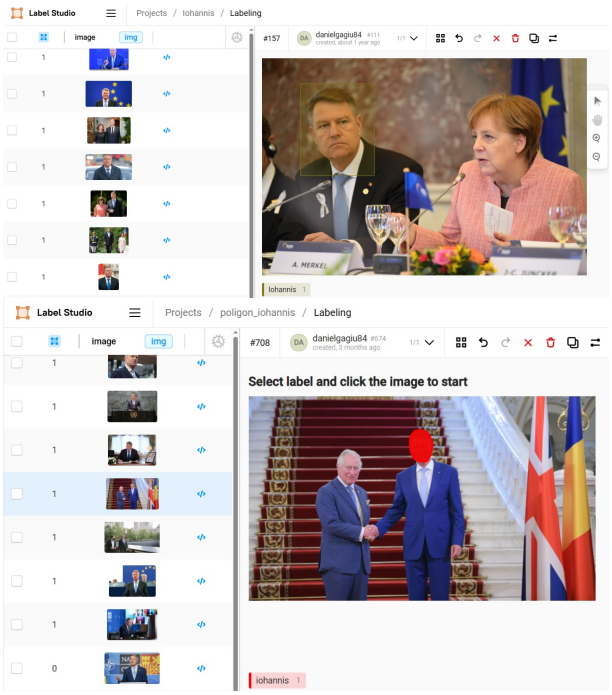

FIGURE 3. Class labelling.

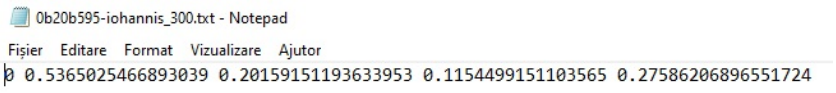

FIGURE 4. Bounding box coordinates.

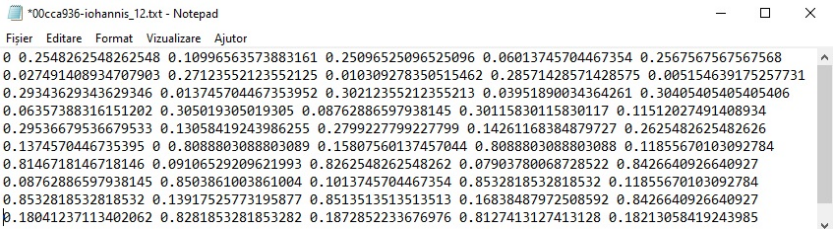

FIGURE 5. Polygon coordinates.

To obtain an effective description, the extracted features should be chosen so as to provide good discrimination capabilities between the predefined classes. This means that the most meaningful and relevant information must be obtained through a feature extraction technique.

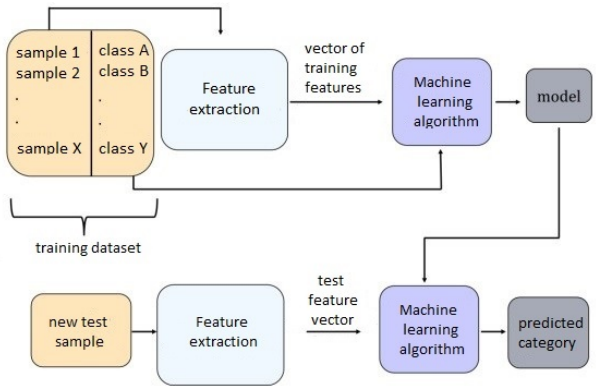

FIGURE 6. Workflow diagram.

Image classification refers to the prediction of the corresponding category for new image samples that have not been previously seen by the classification system. Figure 6 shows a block diagram that describes the components of a classic supervised machine learning approach to image classification that includes two stages, training and actual prediction.

4. Discussion

The training results can be observed by analysing the fluctuation of precision and recall rate. If the fluctuation is not very large, the training yield is better, as can be seen in Figure 7, which assumed training the model using 100 photos.

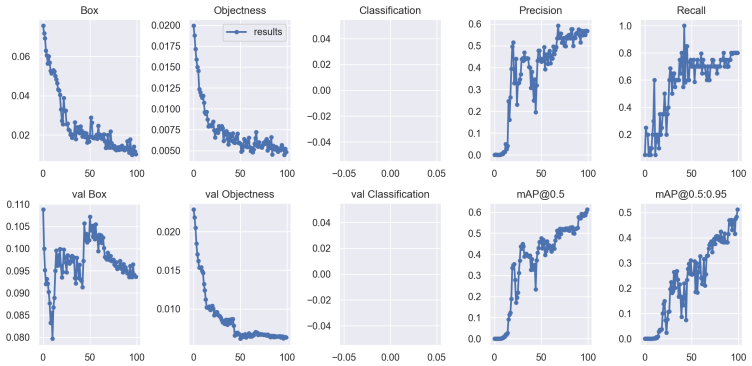

FIGURE 7. Training performance – 100 images as input.

To show the performance of the model in relation to the input data, the authors decided that it was necessary to test the neural network training with a larger data set. Thus, the next training was done on a set of 500 photos, resulting in an improved

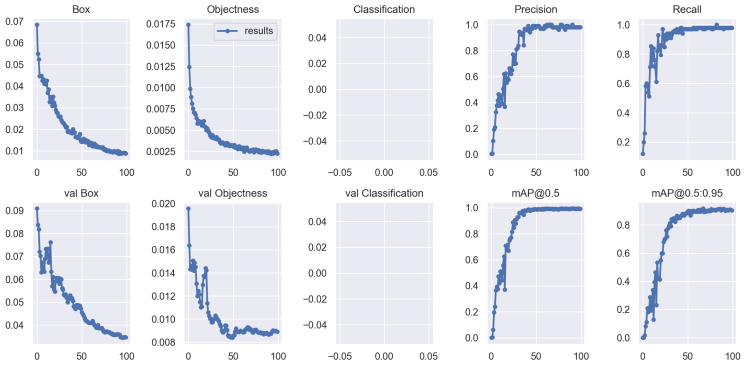

FIGURE 8. Training performance – 500 images as input.

performance of the algorithm compared to the previous experimental one, as shown in Figure 8.

To ensure the classifier's reliability and performance, a post-training evaluation phase is conducted. This involves labelling a newly curated set of images that the model has not previously encountered, thereby testing its ability to generalize beyond the training dataset. The true labels of these images are systematically compared with the predict-ed labels generated by the classifier.

Compared to other AI models (Zhang 2023), which were trained on the FaceScrub database, the model described in this paper demonstrates a notable performance with an accuracy of 90% in correctly identifying the searched sub-object. The FaceScrub database, widely recognized for its diverse and high-quality dataset of celebrity face images, has been extensively used to train and evaluate face recognition models, particularly in unconstrained environments. Models trained on this database often benefit from the variability in pose, lighting, and expression, which enhances their robustness in real-world scenarios

In contrast, the model described in this paper, trained on a significantly smaller dataset of only 500 images, achieves comparable accuracy levels despite the limited training data. This result highlights the efficiency and effectiveness of the model's architecture and training methodology. The ability to achieve 90% accuracy with such a reduced dataset suggests that the model employs advanced techniques, such as feature extraction or optimization strategies, to maximize performance even with constrained resources. This is particularly significant when considering the challenges associated with face recognition in the wild, where variations in image quality, occlusion, and environmental factors can severely impact recognition accuracy. While the models trained on the FaceScrub database may have the advantage of a larger and more diverse dataset, the model in this paper demonstrates that high accuracy can still be achieved with fewer training samples, provided the model design is optimized. This finding underscores the potential for developing efficient face recognition systems that are less reliant on extensive datasets, making them more accessible for applications where data collection is limited or resource-intensive.

This evaluation process, which can be represented through a confusion matrix, allows the application to measure its accuracy, precision, recall, and overall effectiveness in object detection. Based on the performance metrics obtained, further refinements can be applied, including dataset augmentation, hyperparameter tuning, or additional training iterations to improve recognition accuracy and reduce false positives or negatives.

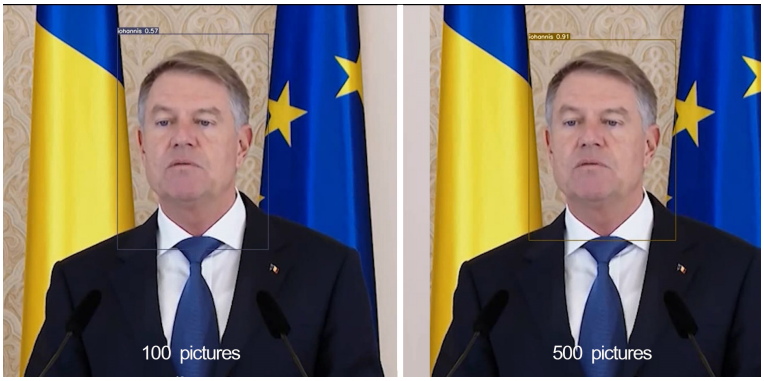

FIGURE 9. Comparison of identification results.

The identification result can be observed by the user through the real-time visualization of detected faces, each enclosed within a labelled bounding box corresponding to the class for which the model was trained. This ensures immediate feedback on the detection process and allows the user to verify the accuracy of the recognition system. From Figure 9, the performance difference between two training scenarios-one with a larger dataset and another with fewer training images-can be analysed, highlighting the impact of dataset size on the model's accuracy and robustness.

Once the neural network model was fully trained, it was integrated into the final video archive management application. This integration involved embedding the trained classifier into a comprehensive software system designed to automate and optimize the process of searching, categorizing and retrieving video content. The application includes a user-friendly interface, as illustrated in Figure 10, where users can define key parameters for the search process. These parameters include selecting the source of the video files, specifying the object or class to be identified within the video archive and choosing the destination folder where successfully identified and validated video files should be moved.

The application operates by continuously analysing video frames, extracting relevant features, and comparing them against the trained model's stored feature representations. Each frame is processed using the convolutional neural network, which applies the trained classification algorithm to detect and recognize objects of interest. The detection results are then dynamically updated within the user interface, providing the user with an intuitive and real-time overview of the classification results.

By automating the retrieval and classification of objects in video archives, this application significantly enhances the efficiency of media asset management, addressing several critical challenges faced by broadcasters and content managers. Traditional

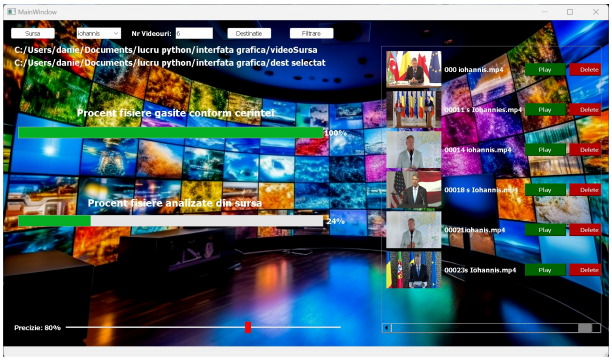

FIGURE 10. User interface.

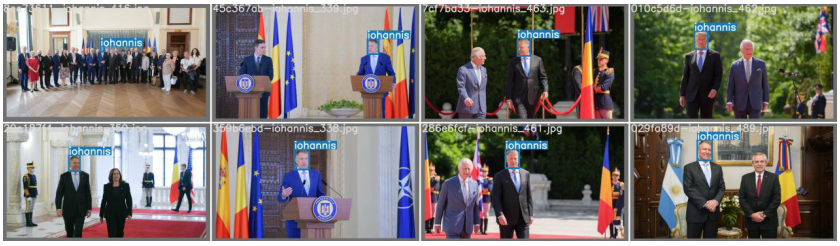

FIGURE 11. Video file list containing the identified class.

methods of manually indexing and retrieving archived video content are labour-intensive, error-prone, and time-consuming, particularly when dealing with extensive media libraries. The proposed solution leverages convolutional neural networks (CNNs) and object detection algorithms, such as YOLOv7, to streamline this process by automatically identifying and categorizing relevant objects, faces, or scenes within archived footage. One of the primary issues in TV archive management is the difficulty of accurately retrieving specific content from vast collections of unstructured video data. Conventional keyword-based search mechanisms often rely on incomplete or inconsistent metadata, limiting the precision of retrieval. By integrating AI-driven visual recognition, this application allows content to be indexed based on actual scene content rather than relying solely on manually entered descriptions. This improves search accuracy, ensuring that relevant material is retrieved quickly, regardless of variations in tagging conventions.

Another major challenge is the scalability of archive management systems. As broadcasters continuously produce and store large amounts of video content, traditional methods struggle to keep up with the growing volume of data. The AI-based solution proposed in this application enables automated content labelling and categorization, al-lowing for real-time indexing without requiring extensive human intervention. This capability ensures that newly acquired footage is immediately searchable and easily accessible for reuse, improving overall workflow efficiency. In addition to improving efficiency, the application enhances security and compliance in video archives. Automated facial recognition and object detection can be leveraged

to identify restricted content, copyrighted materials, or sensitive imagery, ensuring compliance with broadcasting regulations. Furthermore, in security applications, this system can facilitate rapid forensic analysis by identifying individuals or objects across multiple video sources, aiding investigative processes.

Beyond retrieval and security, the proposed solution enhances content monetization opportunities. By offering broadcasters and media companies an efficient way to categorize and repurpose archived footage, this system enables the seamless integration of historical content into modern productions, documentaries, or digital platforms. AI-driven tagging also supports personalized content recommendations, optimizing audience engagement across streaming services.

Overall, this AI-powered video archive management application not only reduces manual effort and search time but also ensures scalability, accuracy, and compliance, making it an invaluable tool for broadcasters, security agencies, and digital content man-agers. Through advanced machine learning techniques, it transforms the way media assets are stored, retrieved, and utilized, ultimately revolutionizing the efficiency of video archive management in the modern broadcasting and content production landscape.

5. Conclusions

The implementation of artificial intelligence in the management of television archives represents a revolutionary step in optimizing the processes of searching, organizing and accessing audio-visual materials. Advanced technologies, such as convolutional neural networks (CNN) and object detection algorithms, such as YOLOv7, allow the automation of the processes of identifying and indexing video content, eliminating the dependence on traditional manual, time-consuming and resource-consuming methods.

One of the major advantages of this application is the increased efficiency in searching and retrieving content. By using a pre-trained model and a transferable learning strategy, the system can automatically recognize objects, people, places and texts from video materials, thus allowing the rapid identification of relevant fragments. This reduces the time required for cataloguing and archiving content, providing immediate access to historical or recent materials. In addition, by integrating an advanced speech and text recognition function, the application can also index audio content, facilitating searches based on dialogues or keywords.

In addition to being efficient, the solution also ensures high accuracy in content classification and filtering. Since the system can continuously learn, adapting to indexing requirements, archive management becomes dynamic and scalable.

One of the future development directions is related to offering a solution to automate the verification and compliance processes essential to media institutions, especially for identifying copyrighted content, inappropriate language or sensitive materials. In this regard, media regulations will be respected, reducing human risks and ensuring compliance with legal and ethical norms.

Last but not least, the adoption of this solution based on artificial intelligence opens up new opportunities for analysing and monetizing media content. TV archives can be capitalized on more effectively through personalized recommendations or identifying trends.

In conclusion, the AI-powered TV archive management application offers a significant set of benefits, including increased operational efficiency, cataloguing accuracy, automated regulatory compliance, optimized performance, and new opportunities for monetizing media content. By integrating advanced technologies such as AI-specific convolutional neural networks, this solution redefines the way audio-visual materials are archived, searched, and reused, representing a key innovation for the modern media industry.

References

[1] M.A.H. Akhand, S. Roy, N. Siddique, M.A.S. Kamal, T. Shimamura, Facial emotion recognition using transfer learning in the deep CNN, *Electronics* **10** (2021), no. 9, 1036.

[2] S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, *International Conference on Engineering and Technology (ICET)*, Antalya, Turkey, 2017, 1–6. DOI: 10.1109/ICEngTechnol.2017.8308186

[3] H. Du, H. Shi, D. Zeng, X.P. Zhang, T. Mei, The elements of end-to-end deep face recognition: A survey of recent advances, *ACM Computing Surveys (CSUR)* **54** (2022), no. 10s, 1–42.

[4] H. Ge, Z. Zhu, Y. Dai, B. Wang, X. Wu, Facial expression recognition based on deep learning, *Computer Methods and Programs in Biomedicine* **215** (2022), 106621.

[5] D. Gurin, V. Yevsieiev, S. Maksymova, A. Alkhalaileh, Using Convolutional Neural Networks to Analyze and Detect Key Points of Objects in Image, *Multidisciplinary Journal of Science and Technology* **4** (2024), no. 9, 5–15.

[6] M. Hassaballah, S. Aly, Face Recognition: Challenges, Achievements, and Future Directions, *IET Computer Vision* **9** (2015), 614–626. DOI: 10.1049/iet-cvi.2014.0084.

[7] O. Hmidani, E.I. Alaoui, A comprehensive survey of the R-CNN family for object detection, In: *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*, Marrakech, Morocco, 2022, 1-6. DOI: 10.1109/CommNet56067.2022.9993862.

[8] O.E. Olorunshola, M.E. Irhebhude, A.E. Evwiekpaefe, A comparative study of YOLOv5 and YOLOv7 object detection algorithms, *Journal of Computing and Social Informatics* **2** (2023), no. 1, 1–12.

[9] V. Pham, L.D.T. Ngoc, D.L. Bui, Optimizing YOLO Architectures for Optimal Road Damage Detection and Classification: A Comparative Study from YOLOv7 to YOLOv10, *IEEE International Conference on Big Data (BigData)*, 2024, December, 8460–8468.

[10] H. Qin, J. Wang, X. Mao, Z.A. Zhao, X. Gao, W. Lu, An improved faster R-CNN method for landslide detection in remote sensing images, *Journal of Geovisualization and Spatial Analysis* **8** (2024), no. 1, 2.

[11] R. Ranjan, S. Sankaranarayanan, C.D. Castillo, R. Chellappa, An all-in-one convolutional neural network for face analysis, *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, USA, 2017, 17-24. DOI: 10.1109/FG.2017.137.

[12] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *Proc. Adv. Neural Inf. Process. Syst.* **28**, 2015.

[13] P. Santemiz, L.J. Spreeuwers, R.N.J. Veldhuis, Automatic face recognition for home safety using video-based side-view face images, *IET Biom.* **7** (2018), 606–614. https://doi.org/10.1049/iet-bmt.2017.0203

[14] D.J. Santry, Convolutional Neural Networks. Demystifying Deep Learning: An Introduction to the Mathematics of Neural Networks, *IEEE* 2024, 111–131, DOI: 10.1002/9781394205639.ch6 https://docs.ultralytics.com/models/

[15] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, M. Karkee, Comprehensive performance evaluation of yolo11, yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments, 2024, arXiv:2407.12040.

[16] V. Strehle, N. Bendiksen, A. O'Toole, Deep convolutional neural networks are sensitive to configural properties of faces, *Journal of Vision* **23** (2023), no. 9, 5560–5560.

[17] M.M. Taye, Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions, *Computation* **11** (2023), no. 3, 52.

[18] C.-Y. Wang, A. Bochkovskiy, H.Y,M. Liao, YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, 7464–7475. DOI: 10.1109/CVPR52729.2023.00721.

[19] Q. Wang, J. Zeng, P. Qin, P. Zhao, R. Chai, Z. Yang, J. Zhang, Semi-White-Box Strategy: Enhancing Data Efficiency and Interpretability of Convolutional Neural Networks in Image Processing, *International Journal of Intelligent Systems* **1** (2023), 9227348.

[20] J. Yang et al., Neural aggregation network for video face recognition, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, Nicole.

[21] A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique, J. Henkel, A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends, *IEEE Access* **12** (2024), 41180–41218. DOI: 10.1109/ACCESS.2024.3376441.

[22] N. Zhang, A Study on the Impact of Face Image Quality on Face Recognition in the Wild, 2023. Retrieved from https://arxiv.org/abs/2307.02679.

[23] X. Zhang, T. Zhang, G. Wang, P. Zhu, X. Tang, X. Jia, L. Jiao, Remote Sensing Object Detection Meets Deep Learning: A Metareview of Challenges and Advances, *IEEE Geosci. Remote Sens. Mag.* **11** (2023), 8–44.

[24] D. Zhao, F. Shao, Q. Liu, H. Zhang, Z. Zhang, L. Yang, Improved Architecture and Training Strategies of YOLOv7 for Remote Sensing Image Object Detection, *Remote Sensing* **16** (2024), no. 17, 3321.

[25] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, M. Parmar, A review of convolutional neural networks in computer vision, *Artificial Intelligence Review* **57** (2024), no. 4, 99.

(Daniel-Gheorghe Gagiu, Dorin-Gheorghe Sendrescu, Florina-Luminita Petcu, Stefan-Irinel Cismaru, Razvan-George Dumitrascu) University of Craiova, 13 A.I. Cuza Street, Craiova, 200585, Romania

E-mail address: danielgagiu84@gmail.com, dorin.sendrescu@edu.ucv.ro, florina.petcu@edu.ucv.ro, stefan.cismaru@edu.ucv.ro, dumitrascurazvangeorge@gmail.com