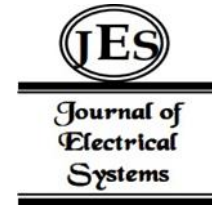


¹* YiJin Wu¹ Dan Liu

Research on the Application of Image Processing and Image Recognition Algorithms in Digital Media in Content Editing and Production



Abstract: - Image processing and image recognition algorithms are essential to content generation and editing in the digital age. They provide creative ways to improve visual quality, streamline processes, and customize user interfaces. In-depth methods for incorporating cutting-edge image processing and image recognition algorithms into digital media content production and editing workflows are covered in this study. **Methods:** To maximize visual quality, the research first focuses on pre-processing and improving digital images using the median filtering technique. The Histogram of Oriented Gradients (HOG) technique is applied to locate and identify the objects in images, that enabling customized and interactive content modification. After that, the images are segmented using the watershed approach, and the precise classification of the images is achieved by applying the Mayfly Optimized Spatial Graph Recurrent Neural Network (MOSGRNN), which improves content organization and retrieval. **Results:** The results of the experiments demonstrate that the suggested strategy performs well in all aspects of image processing in terms of accuracy (94.91%), recall (92.70%), recognition speed (44 FPS), and f1-score (93.7%). **Conclusion:** Content production, editing, and delivery across a variety of platforms might be significantly transformed by research on image processing and image recognition algorithms in digital media.

Keywords: Image Processing, Image Recognition, Histogram of Oriented Gradients (HOG), Watershed, Mayfly Optimized Spatial Graph Recurrent Neural Network (MOSGRNN)

I. INTRODUCTION (*HEADING 1*)

The enormous growth of internet-based image graphs and image-sharing internet pages has made accurate image forensics equipment increasingly important, however the speed at which these methods have developed additionally renders to verify the authenticity of images stored online.

The Development of Digital Image graph: Digital devices are ubiquitous in today's environment and generate massive volumes of digital images in several forms. This expansion necessitates a deeper examination of image graphic forensics technology, namely digital image processing [1]. Forensic experts have always concentrated on specific image reports, due to the rapid advancement of internet and digital imagegraphing; digital images are readily accessible internet [2]. This discovery highlights how important to assess, validate, and corroborate electronic visual information using excellent image forensic expertise.

Obstacles in Digital Image Verification: Digital imaging and wireless connections have enabled the sharing of images, the abundance of image processing and editing programs on the market poses an array of difficulties [3]. Rapid advancements in digital image manipulation and augmentation have brought up concerns regarding privacy of information and the improper usage of edited images in various situations. This propensity casts doubt on the reliability and validity of image graphs used in official records, legal proceedings, and studies in science, highlighting the necessity of stringent image authorization processes [4].

Implications of Techniques for Image Processing: The incorporation of computational imaging and recognition techniques has revolutionized the creation, alteration, and enhancement of images in digital formats. These techniques are necessary in automated image-editing tasks, enhancing the appearance of images, and changing complicated information [5]. However, when digital image manipulation can lead to erroneous conclusions, unethical conduct, and challenges maintaining the authenticity of images, technology creates moral and cultural concerns whenever technology grows used [6]. Important advances in processors for signals and processors during the 2000s made processing images widely used method for mapping, healthcare, and image graphic purposes.

Identity Digital Media's Role: False accounts on websites have grown prevalent across a range of industries, professions, and social groupings in the contemporary digital world. Issues alongside authentication and reliability in online contacts are raised, it is to fabricate a false identity using digital communication technology [7, 8]. Moreover, criminal groups purposefully use these advances to compromise the safety of people and national security. In order to solve these problems and prevent theft of identities along with other detrimental activities on the internet, strong cyber security rules and processes must be established [9].

¹ Changzhou College of Information Technology, Changzhou, 213000, China
Corresponding author: YiJin Wu, wuyijin@czcit.edu.cn
Copyright © JES 2024 on-line : journal.esrgroups.org

Technological Developments and Their Consequences on Society Balancing: As technologies develop, it is imperative to weigh its advantages against any possible disadvantages. While digital breakthroughs increase productivity and ease of use, but also bring up difficult questions about safety, confidentiality, and responsible application of digital assets [10]. To mitigate the adverse effects of innovation on societal progression, it is imperative to fortify regulatory structures, encourage conscientious utilization of technology, and enhance digital literacy [11].

Objective of the study: The study implements Multi-Order Spectral Graph Recurrent Neural Network (MOSGRNN) to achieve accurate image classification, thereby improving the structure and efficiency of content organization and retrieval tasks by leveraging advanced graph-based learning techniques.

Contributions of the study:

1. The research utilized an MS-COCO dataset to improve content organization and retrieval.
2. Utilizing median filtering in pre-processing, the study effectively enhances the visual quality of digital images, reducing noise and improving clarity.
3. The adoption of the HOG technique enables accurate object localization and identification within images, facilitating targeted and interactive content modifications based on detected objects.
4. The watershed approach applied to image segmentation helps in dividing images into meaningful regions.

The remaining research is divided into categories: The literature review is presented in Part II, the methodologies are classified in Part III, comparative analysis and discussions are shown in Part IV and the conclusion is presented in Part V.

II. LITERATURE REVIEW

A thorough examination of the different image processing and recognition algorithms used in digital media content generation and editing is shown in Table 1. These tactics include deep learning (DL)-based content creation and automatic color correction. They have greatly increased the quality and productivity of digital media creation, improving user experiences and workflows across sectors including entertainment, advertising, and education, despite obstacles like algorithmic accuracy and computing complexity.

Table 1: Hybrid and Deep Learning (DL) Models for Different Image Processing and Recognition

Ref	Methods	Objective	Limitations	Results
[12]	DL Convolution Neural Network (DL-CNN)	To identified and analysed different retinal layers, gather pertinent data, monitor for any new deviations, and predict eye abnormalities.	The approach might perform better in settings when memory is constrained, although at some performance cost.	The study compared to manual ophthalmological diagnosis, experimental testing performed exceptionally well, with 0.965 classification accuracy, 0.960 sensitivity, and 0.986 specifications.
[13]	Deep Convolution Neural Network (DCNN)	To discharge farming responsibilities by identifying leaf diseases that promotes the growth of healthy plants.	Rely on labelled data for training; they are limited in their ability to reliably categorize new, undiscovered diseases or alterations that can impair healthy plant growth.	Across 32 epochs and 10 plant classes, the DCNN-based approach yielded an accuracy rate of 96.02% with a logarithmic loss of 0.01.
[14]	Salp swarm algorithm and particle swarm optimization (SSPSO)	The proposal proposes an improved version of the SSO that utilizes PSO to achieve superior classification performance over blood cell image databases.	The emphasized negatives are that, although most of the hyper-parameters of any of these pre-trained CNNs are set in stone, some of them require to be modified.	The proposed method improved performance with high precision and F1-score and 99% total classification accuracy.
[15]	Deep neural network (DNN)	To accelerate image processing on Internet of Things (IoT) devices, the study examines the usage of FogFlow for coordinating DNNs.	It is inconvenient to transfer streams to a cloud image processing service due to limited connectivity.	The detection timings of 1000 successive detections of several DNN models discovered that, on average, most of them took less than 55 ms, except mobile net v1, which took 28 fps on average.
[16]	Multistage DL (MSDL)	Image recognition-based network intrusion detection.	Due to computational complexity and latency problems across several	Achieves 99.8% accuracy in detecting generic attacks and 99.7% accuracy in detecting

			processing steps, MSDL can be inefficient in real-time, high-throughput network traffic detection.	Distributed Denial of Service (DDoS) attacks and normal traffic.
[17]	Hybrid model based on CNN and Random Forest (RF)	The process involves transferring the retrieved features from a CNN into a RF for categorization.	The hybrid CNN model transfers features from CNN to RF for categorization, potentially introducing complexity and overhead.	In the study, both the classification accuracy and generalization capacity of the hybrid model have greatly outperformed those of the RF model.
[18]	Support vector machine (SVM) and CNN	The use of image recognition to interpret sign language.	The performance subtleties for multi-class classification or regression issues cannot be well captured by performance measures.	The results showed that CNN performed better in terms of precision but poorer in terms of accuracy and recall, with a score of 96% in each of the three categories.
[19]	SVM and k-nearest neighbor (KNN)	The study utilized median and mean filters to reduce bands on Indian Pines, Pavia University, and Salinas datasets.	The study's incapacity to manage intricate spectral interactions can limit the use of median and mean filters for band decreases.	Using image pre-processing and feature selection results in significantly higher rates of accuracy for SVM and KNN classification.
[20]	SVM and You Only Look Once, version 3 (YOLO _{v3})	The approach makes use of image segmentation to pinpoint certain fruits.	It could have trouble correctly recognizing fruits that are overlapping or densely grouped in a single image.	The YOLO _{v3} technique achieved a 93% accuracy rate after training and testing the dataset model on a range of fruits. This is greater than the 90% accuracy rate of previous approaches.
[21]	MRA-CNN	The integration of attention blocks into the backbone enhances the visualization of features and accurate classification by reinforcing significant qualities.	MRA-CNN can raise model complexity and add computational cost, which might result in longer training durations and more memory use.	Tests reveal that the model works better than RA-CNN and MACNN and that MRA-CNN can be used to segment images, lower computing costs, and propose learning regions in the future.
[22]	Improved Back Propagation neural network (IBPNN)	The suggested method combined innovative bill information recognition technology with image processing and recognition theory to identify tickets.	Due to their reliance on pre-programmed elements, they can be less accurate when designing complex tickets.	The experimental findings showed that, in terms of classification and recognition, the revised technique performs better than existing algorithms for data collecting and recognition.
[23]	YOLO _{v5}	To ensure appropriate mask wear in public areas, the most sophisticated objection detection algorithm is being developed for real-world implementation.	Despite its strong object identification skills, YOLO _{v5} struggles to identify mask wear in busy or dynamic public environments because of obstacles, different mask styles, and different lighting conditions.	The research's experimental findings demonstrated the algorithm's effectiveness in identifying facemasks and enhancing staff monitoring efficiency.
[24]	CNN	The study enhanced the advantages of sustainable eating habits by creating a carbon footprint tracking system.	The entire ecological cost of food consumption as well as any mistakes in carbon footprint calculations resulting from using different data sources.	The model's 94.79% accuracy in identifying food types and a 22.25% decrease in participants' carbon footprint during two weeks of testing was found.

A. Motivation of the study

The creation and editing of digital media material has advanced significantly, but there are still several challenges facing image processing and image recognition algorithms, such as those related to computational complexity, accuracy, and format flexibility. By guaranteeing seamless integration across digital media platforms, streamlining content editing processes, and improving image identification accuracy via the use of innovative techniques and state-of-the-art algorithms, this study aims to address these problems. The primary goal is to set higher standards for results, dependability, and adaptability in the manufacturing of technological material. The methodology attempts to address current problems and encourage more developments in this area by means of careful examination, evaluation, and improvement.

III. CONTENT EDITING AND PRODUCTION IN DIGITAL MEDIA

The MOSGRNN is a technical framework, which utilizes spatially graphing techniques and recurring neural networks topology to categorize images. Its capacity to recognize complicated connections among image information and to outperform on measures including recognizing items, scenario comprehension, and categorizing images might be advantageous towards sophisticated analysis of images. Figure 1 represents the proposed block diagram.

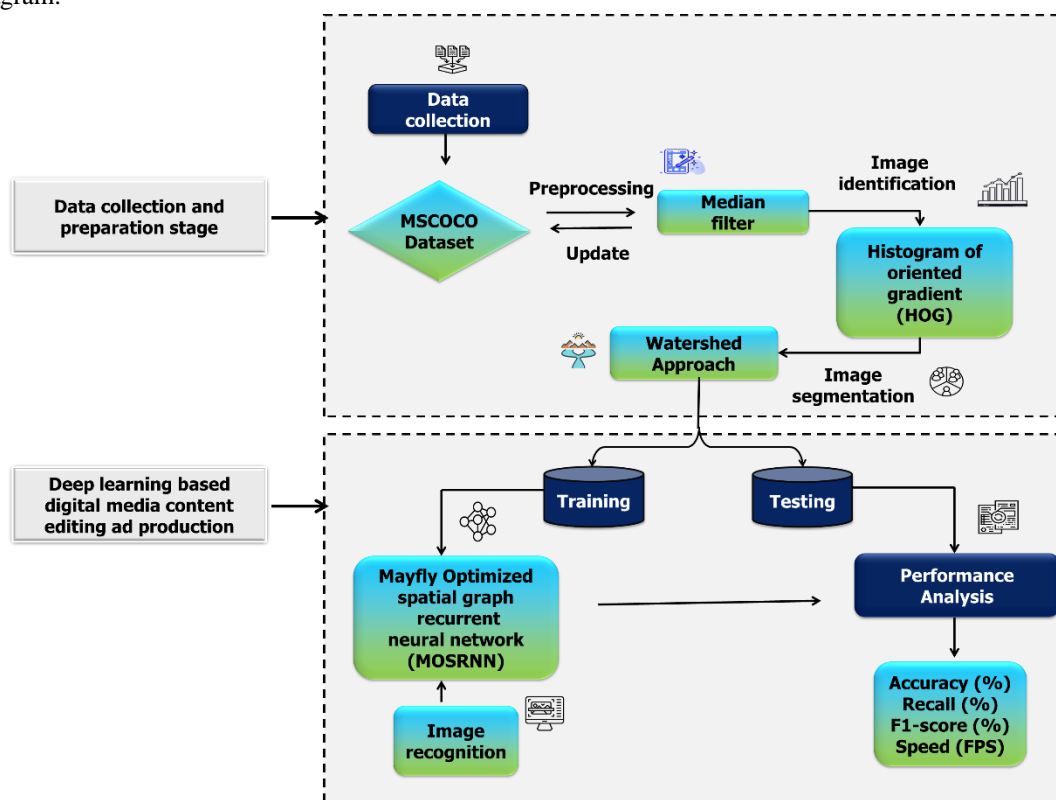


Figure 1: Block Diagram of our Proposed

A. Data collection

In the research, we collected data from kaggle (<https://www.kaggle.com/datasets/mbkinaci/fruit-images-for-object-detection>). The object detection dataset provides a substantial quantity of data for testing and training object identification models, with 240 images in the train folder and 60 images in the test folder.

B. Data Pre-processing using a median filter

Median filtering is a pre-processing technique that improves the visual quality of digital images by reducing noise and clarity. The median filter is an inexpensive way of using a non-linear filter to reduce noise. This substitutes the nearby median value for the targeted noisy pixels. It replaces individual pixel values with the median value of neighboring pixels within a specific window or kernel, resulting in more visually appealing, detailed, and smooth images. The size of the filtering window determines how many neighbors there are. In a sorted series, the mid value is the median value.

$$\begin{aligned}
 Median(O) &= Med\{O_j\} \\
 Median(O) &= Med\{O_j\} \\
 &= \frac{O_j(l+1)}{2}, l \text{ is odd} \\
 &= \frac{1}{2} \left[O_j\left(\frac{l}{2}\right) + O_j\left(\frac{l}{2} + 1\right) \right], l \text{ is odd}
 \end{aligned} \tag{1}$$

Even $O_1, O_2, O_3 \dots \dots O_l$. The neighboring Pixel Sequence is denoted by O_{jl} . Before applying a filter, every pixel in the image must be organized in an ascending or descending order. The image pixel sequence that results from sorting is $O_{j1} \leq O_{j2} \leq O_{j3} \leq \dots \dots O_{jl}$, where l is often odd. Figure 2 depicts the outcomes of (a) before and (b) after using median filtering.

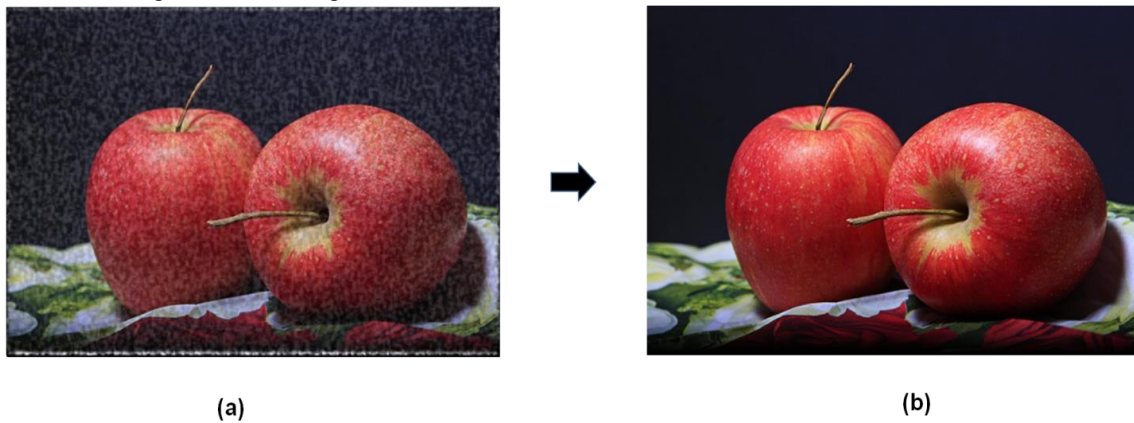


Figure 2: (a) Before and (b) After Filtering

C. Histogram of Oriented Gradients (HOG) for identifying and locating objects in the image

HOG is a feature classifier that is useful in computer vision systems for object recognition, particularly in the detection of pedestrians. In addition to providing gradient information and well-designed histograms, it extracts target edge distribution information. Because HOG can effectively identify the geometry and texture of altered images, it is very helpful in identifying copy-move fakes. Figure 3 depicts the (a) after filtering to (b) object identified and localized the image.

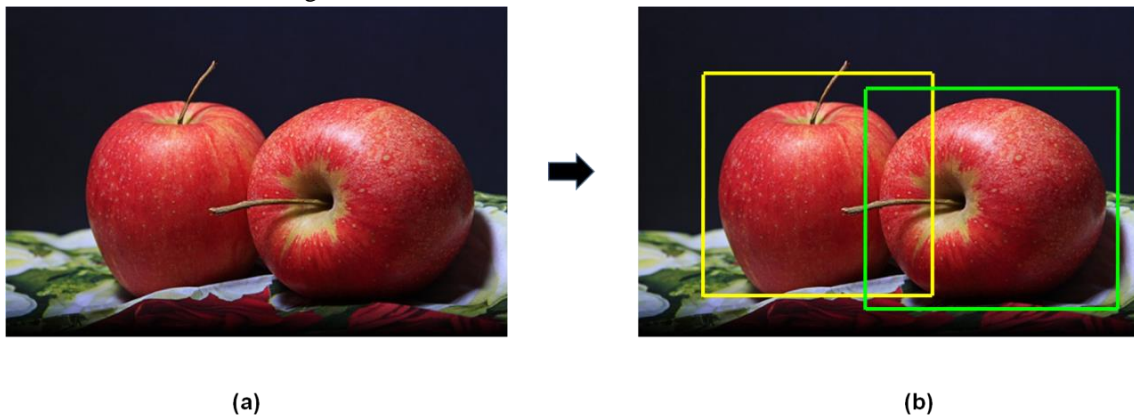


Figure 3: Object Identify and Localization

HOG was utilized in this piece to symbolize the characteristics of fake images. Slider windows are positioned over the regions where evidence of fabrication was found after the image has been divided into overlapping sub-blocks. Every window represents a subset of the fake image, allowing corresponding HOG feature vectors to be computed to represent the gradient and edge data inside that area.

A conventional to determine each point's image gradients value in both directions, use the gradient operator $[-1,0,1]$. By organizing the gradient operators with the image, one can extract the variations at a given slope (w, z) in image J :

$$H_w = [-1,0,1] * J(w, z) \text{ and } H_z = [-1,0,1]^S * J(w, z) \tag{2}$$

The image detection window, referred to as "cells," is divided into tiny geographical regions. The change in gradient or edge orientations is collected to create a local 1-D histogram for each pixel in each cell. By casting votes using the gradient magnitudes of each pixel inside the cell, the orientation of the histogram is decided. An

arrangement of cells is called a "block" To find the gradient's magnitude at a given place (w, z) , use the following formula:

$$H(w, z) = \sqrt{H_w(w, z)^2 + H_z(w, z)^2} \tag{3}$$

The following formula is used to determine the edge's orientation at a given position (w, z) :

$$\theta(w, z) = \frac{\tan^{-1}H_z(w, z)}{H_w(w, z)} \tag{4}$$

Through applying normalization computations, local responses can be contrast-normalized to give invariance to light and shadowing. Building up a local histogram energy measure throughout a block is the method. For a cell to occur multiple times with different normalizations in the final output vector; the result is used for normalizing each cell within the block. The HOG defines the homogenized block descriptor.

1. The last step is gathering HOG descriptors from every block inside the detection window and assembling them into a feature vector that can be used for identification.

2. HOG descriptions are capable of capturing local contour data, edge and gradient frameworks, and local shape characteristics. Therefore, HOG descriptions are ideal for detecting fraudulent copy-move images.

D. Image Segmentation using Watershed Approach

That segmentation would be ignored if the original segmentation map was used. We incorporate a post-diversification fusion procedure into the modified watershed methodology. This algorithm is not the same as the traditional one. The goal of this diversification fusion phase is to reduce the number of divisions without compromising the segmentation map's accuracy. It functions based on spatial criteria. One way to summarize this post-segmentation merging procedure is as follows:

$J(y, x)$ should correspond to the original image. Assume that the first divisions from the splitting of the watershed are represented by the set $Q = \{Q_1, Q_2, Q_3 \dots Q_M\}$ where N is the total number of partitions and Q_j is the j th partition.

Let N_j represent the dimension of each division Q_j .

Determine every partition Q_j 's mean intensity, and indicate this by:

$$N_j = \frac{1}{N_j} \sum_{(w,z) \in Q_j} J(w, z) \tag{5}$$

Establish a pair of measurements separating any two adjacent partitions, j and i .

The initial variable is the variation in mean intensities between partitions j and i . This can be described as:

$$N_{ji} = |N_j - N_i| \tag{6}$$

The disparity in intensity between partitions J and I constitutes the second measure.

$$A_{ji} = \frac{1}{M_{ji}} \sum_{(w_j,z_j),(w_i,z_i)} |J(w_j, z_j) - J(w_i, z_i)| \tag{7}$$

Where $(w_j, z_j) \in Q_j$ and $(w_i, z_i) \in Q_i$ in the summation are all the 8-connected pixels that are located on the boundary between partitions Q_j and Q_i . N_{ji} is the number of border pixels between partitions i and j .

Decide on a set of standards, the degree of comparability between the intensity values of two partitions, j and i , is known as D_{ju} .

$$D_{ju} = 1/2 (N_{ji} + A_{ji}) \tag{8}$$

After computing D_{ju} for each partition j and i , we ascertain the threshold S_d that D_{ju} must satisfy before partitions j and i can be combined. If D_{ju} is less than S_d , it indicates that partitions J and I are similar by the specified spatial criterion and should be merged. To find S_d , we apply the automatic thresholding method that was previously discussed in this section. Figure 4 depicts the (a) localized image to (b) segmented image.

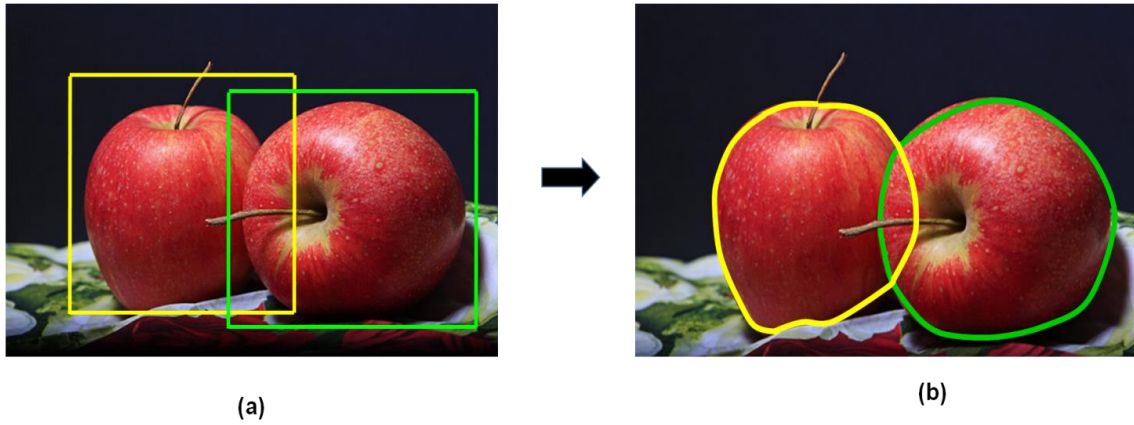


Figure 4: Image segmentation

E. *Mayfly Optimized Spatial Graph Recurrent Neural Network (MOSGRNN) for image recognition*

Recurrent neural networks (RNNs) and spatial graph structures are combined in the novel MOSGRNN image recognition system. This approach leverages both the spatial correlations between picture parts and the temporal relationships seen in image sequences. MOSGRNN is a promising development in computer vision research because it improves performance in tasks like object identification, tracking, and scene interpretation by utilizing ideas from Mayfly behavior to refine the network design.

1) *Mayfly Optimization (MO)*

Mayfly Optimization (MO) is an algorithm that imitates the fast decisions of mayflies, improving processes like image processing, video editing, and content development in digital media. Its ability to identify and prioritize important components enhances productivity, ensuring high-quality outputs in constrained timeframes, making an invaluable tool for content creators and editors. Male and female mayflies in clusters would be identified using the MO algorithm. Moreover, the optimizing performance of mayflies would be enhanced as the males would be stronger. Similar to the PSO algorithm's swarms of individuals, the MO algorithm's people would update their placements depending on each iteration's current locations $O_j(s)$ and velocity $v_j(s)$:

$$o_j(s + 1) = O_j(s) + v_j(s + 1) \tag{9}$$

All male and female mayflies to update their locations would use equation (9). Their velocities would be updated in various ways, though.

a) *Male mayfly motions*

During variations, Hordes of male the mayfly would be out on their investigation or predation. The updated speed would take into consideration both their maximum fitness values from the past $e(wg_j)$ and resent fitness values (e). Based on their present velocity, their distance to the global optimal site, and the previous optimal paths, the male mayflies would modify their speed $e(w_j) > e(wg_j)$:

$$v_j(s + 1) = h \cdot v_j(s) + b_1 f^{-\beta a_0^2} [w_{g_j} - w_j(s)] + b_2 f^{-\beta h^2} [w_h - w_j(s)] \tag{10}$$

From its maximum value to a lower value, the variable g exhibits a linear decline. The numbers need to be balanced, and the two constants are b_1 , b_2 , and β_h^2 . Two variables, s_o , and s_h , define the global optimal location in swarms, or the Cartesian separation between a person and its historical optimum placement. The second proximity array norm would correspond to the Cartesian proximity:

$$||w_j - w_i|| = \sqrt{\sum_{l=1}^m (w_{jl} - w_{il})^2} \tag{11}$$

In contrast, the male may fly were going to employ a random dance coefficient to update their current speeds if $e(z_j) < e(w_j)$:

$$v_j(s + 1) = h \cdot v_j(s) + c \cdot q_1 \tag{12}$$

b) *Female mayfly motions*

The female mayflies would alter the way they updated their velocities. The female flies would be in a hurry to locate the male mayfly so they could engage and procreate since, according to biology, they can only live for one to seven days at most. As a result, they would adjust their speed following the male mayfly they choose to mate.

The first mate in the MO algorithm would be the best female and male mayfly, and the second mate would be the best female and male mayfly, and so on. Thus, for the mayfly with $j = 0$, if $e(w_j) < e(w_{g_j})$:

$$v_j(s + 1) = h.v_j(s) + b_3 f^{-\beta_h^2} [w_h - w_j(s)] \tag{13}$$

In this case, v is additionally balanced by another constant b_3 . The Cartesian distance between them is represented by q_1 . Conversely, if $e(z_j) < e(w_j)$, the female mayflies would do another haphazard dance fl to update their velocities.

$$v_j(s) = h.v_j(s) + ek.q_2 \tag{14}$$

In the interval $[-1, 1]$, where q_2 is a uniformly distributed random number as well.

c) *Mating of mayflies*

Every upper-half pair of male and female mayflies would mate and produce two offspring for each parent. Their progeny would diverge at random from their parents:

$$offspring1 = K * male + (1 - K) * feamle \tag{15}$$

$$offspring2 = K * male + (1 - K) * feamle \tag{16}$$

Hence, in a Gauss distribution, L are random integers.

F. *Spatial Graph Recurrent Neural Network (SGRNN)*

DL systems use a series of L hidden layers, consisting of linear conversion, a dimensionality-reducing manager, and an integer quadratic conversion. These layers gradually extract meaningful information by combining nonlinearly local characteristics. GNNs customize these operations to the graph supporting the data. Our unique multi-relational GRN use known attributes as inputs at layer one and produces predicted labels at layer three. The output layers GRNN and GNN effort together before training the NN.

1) *First layer network*

Examining a transitional layer in our architecture, let's call it the j^{th} one. The $M \times J \times O^{(k)}$ is the layer's output. (k) Tensor $\underline{Y}^{(k-1)}$, where $O^{(k)}$ is the number of output features at J , $y_{mj}^{(k)}, \forall m, j$ with $O^{(k)}$, j are the feature vectors $O^{(k)} \times 1$. In a similar vein, the input to the layer is represented by the $M \times J \times O^{(k-1)}$ tensor $\underline{Y}^{(k-1)}$. We do not take into account a reduction in dimensionality operation in the intermediary levels as label prediction across all nodes is our objective. As a result, there are two stages to the mapping from $\underline{Y}_{jmo}^{(k)} := \sigma(\underline{Y}_{jmo}^{(k)})$. We initially develop a linear transformation to convert the $M \times J \times O^{(k)}$ tensor $\underline{Y}^{(k-1)}$ into the $M \times J \times O^{(k)}$ tensor $\underline{Y}^{(k)}$. Subsequently, a scalar linear modification $\sigma(\cdot)$ is applied elementwise on the intermediary features output $\underline{Y}^{(k)}$.

$$\underline{Y}_{jmo}^{(k)} := \sigma(\underline{Y}_{jmo}^{(k)}) \tag{17}$$

After gathering every element, we are left with the $\underline{Y}^{(k)}$ output of the j^{th} layer. Reversed linear units, or ReLUs, are used to calculate $\sigma(\cdot)$, meaning that $\sigma(d) = \max(0, d)$. Therefore, the primary objective the creation of a sequential change that corresponds to our issue scenario and translates $\underline{Y}^{(l-1)}$ to $\underline{Y}^{(k)}$. Conventional CNNs generally take into consideration a limited set of weights that can be trained and thereafter produce the convolution of the input data using this weighting yielding a linear result. By combining the values of nearby inputs, such as neighboring pixels or successive time instants, the convolution generates local neighborhood information.

By swapping out the convolution for a graph filter, the settings of which are also learned, CNNs can now process graph data because of GNNs. This minimizes the transformation's degrees of freedom while preserving locality and utilizing the graph's graphical structure. That corresponds to first explore a step that does a linear integration of the data inside a network neighborhood to identify how to accomplish this. Since the neighborhood is determined by the specific relationship that exists for the j^{th} relation.

$$g_{mj}^{(k)} := \sum_{m' \in \mathcal{N}_m^{(j)}} T_{mm'} \check{y}_{m'j}^{(k-1)} \tag{18}$$

Even if the entries of $g_{mj}^{(k)}$ rely on n 's one-hop neighbors, the operation's expanding reach when applied successively across layers would cause the information to spread across the network. Future work will address the idea of progressively applying equation (18) inside of one layer to account for multiple hops. Next, we combine the elements in $g_{mj}^{(k)}$ using (trainable) parameters to extract features from the graph data.

$$\underline{Y}_{inp}^{(k)} := \sum_{j'=1}^j Q_{jj'}^{(k)} g_{mj}^{(k)T} x_{mj'o}^{(k)} \tag{19}$$

When the outputs are mixed at separate graphs by $Q_{jj'}^{(k)}$ and the features are mixed by the $O^{(k-1)} \times 1$ vector $x_{mj'io}^{(k)}$. $O^{(k-1)} \times M \times J \times O^{(k)}$ tensor $x^{(k)}$ gathers the feature mixing weights $\{x_{mj'io}^{(k)}\}$, whereas the graph mixing weights $Q_{jj'}^{(k)}$ are collected by the $J \times O^{(k)}$ tensor $Q^{(k)}$. Its incorporation of Q as a training parameter, which gives the GRNN the capacity to learn and mix (combine) the relations stored in the multi-relational graph, is another important addition. The structure R can be constrained if previous knowledge about the dependencies among relations is available. We summarize the equations (18), and (19) as follows after gathering all the scalars $\underline{Y}_{inp}^{(k)}$ in the $M \times J \times O^{(k)}$ tensor $\underline{Y}^{(k)}$.

$$\underline{y}^{(k)} := e(\underline{Y}^{(k-1)}; \theta_Y^{(k)}) \tag{20}$$

$$\theta_y^{(k)} := [\text{vec}(\underline{X}^{(k)}); \text{vec}(\underline{Q}^{(k)})]^T \tag{21}$$

2) *Recurrent GNN layer*

The input W is gradually dispersed over the L -hop graph neighborhood by applying K GNN layers, but the pertinent neighborhood population is not necessarily obvious upfront. To give our design additional adaptability, fortunately suggest a GRNN structure that is continuous and receives W inputs at each k , therefore capturing different types of dissemination. Thus, the traditional linear process is substituted by the nonlinear matrix mapping that occurs repeatedly.

$$\underline{y}^{(k)} := e(\underline{y}^{(k-1)}; \theta_y^{(k)}) + e(\underline{W}; \theta_w^{(k)}) \tag{22}$$

When (k) encodes trainable parameters. An expanded class of network dissemination is implemented by the operator y as a change from y to $\underline{y}^{(k)}$, compared to the previous example, for instance, if $k = 3$, then the second summand creates a single-hop diffusion of W , whereas the first summand corresponds to a 1-hop diffusion of a signal that is a 2-hop (nonlinear) diffused version of W . It makes more sense to have the subsequent summary as its assurances that W 's impression on the outcome will not diminish the quantity of sheets rises. The approach uses regression-based localization to better express time-dependent labels and components that drives our upcoming research on multi-relational graph prediction of processes that change.

3) *First and last layers*

The initial and final layers function in a pretty straightforward way. For layer $k = 1$, X is used to define the input $\underline{Y}^{(0)}$.

$$\check{Y}_{mj}^{(0)} = W_m(m, j) \tag{23}$$

However, the application changes to the layer yields the outcome of our network layout $K = k$'s output.

$$\hat{z} := h(\check{Y}^{(K)}; \theta_h) \tag{24}$$

Where $\hat{z}, M \times L, \hat{z}_{m,l}$ matrix, $h(\cdot)$ is a nonlinear function, θ_h is a trainable parameter, and $\check{Y}^{(K)}$ is the probability that in $= L$. The function $h(\cdot)$ is determined by the particular application; for classification issues, the normalized exponential function (softmax) is a common option. The global mapping from \check{Y} to Z , which is determined by our GRNN design and is accomplished by applying (22)-(24) sequentially, is indicated as follows for notational convenience:

$$\hat{z} := \mathcal{F}(X; \{\theta_y^{(k)}\}, \{\theta_w^{(k)}\}, \theta_h) \tag{25}$$

4) *Standardizes for developing and graphing perfection*

Thus, the weights in determined the architecture that is recommended. By reducing the difference between the supplied and projected designations, we calculate those weights. Consequently, we reach the next reduction objective.

$$\begin{aligned} \min_{\{\theta_y^{(k)}\}, \{\theta_w^{(k)}\}, \theta_h} \mathcal{L}_{tr}(\hat{Z}, Z) + \mu_1 \sum_{j=1}^J \text{Tr}(\hat{Z}^T T_j \hat{Z}) + \mu_2 \rho(\{\theta_y^{(k)}\}, \{\theta_w^{(k)}\}) + \lambda \sum_{k=1}^K \|Q^{(k)}\|_1 \\ \text{s.t. } \hat{z} := \mathcal{F}(X; \{\theta_y^{(k)}\}, \{\theta_w^{(k)}\}, \theta_h) \end{aligned} \tag{26}$$

Using $\mathcal{L}_{tr}(\hat{Z}, Z) := -\sum_{m \in L} \sum_{l=1}^L Z_{ml} \ln \hat{Z}_{ml}$ is a critical decision for the fitting cost in our classification setup, among the recognized data, the loss functional with a cross. Take note that three regularizes were additionally considered. Smooth label estimates across the graphs are encouraged by the first regularize, which is graph-base, and over fitting is avoided by using the L2 norm $\rho(\cdot)$ over the GRNN parameters. Finally, the standard deviation of $L1$ with the help of third regularize weak blending parameters is learned and a portion of the relationships at

each l are triggered. The computing difficulty of evaluating depends on the number of nonzero entries. To summarize, our suggested multi-relational GRNN design: uses R to adjust each graph; uses a straightforward but flexible recurrent tensor mapping; and incorporates multiple kinds of regularizes based on graphs. By contrast, just a single network with a particular spread subtype is used in most of the publications in the GNN research. For image identification tasks, the MOSGRNN has a number of benefits. First, it quickly updates network parameters by utilizing the Mayfly optimization technique, which imitates the brief lifespan of mayflies. This results in faster convergence and shorter training times. In addition, MOSGRNN integrates spatial graph structures, which helps to recognize spatial linkages and dependencies in pictures than typical neural networks, leading to increased accuracy and resilience in image recognition tasks.

IV. PERFORMANCE ANALYSIS

In this paper, the experimental setup is provided with an Intel core i5-8250U CPU with 34GB RAM. The comparative analysis uses existing methods such as faster region-convolution neural network (FR-CNN) [25], feature pyramid network (FPN) [25], You Only Look Once (YOLOv3) [25], and Improved You Only Look Once (IYOLOv3) [25] with the metrics of accuracy, recall, f1 score, and speed. These metrics help editors improve the quality, impact, and relevance of their content. The F1 score evaluates the editing process, recall collects crucial information, and accuracy ensures the final product aligns with the intended message. Accuracy evaluates the overall correctness of the edited text. Table 8 depicts the outcomes of existing and proposed methodologies.

Table 8: Performance Matrices

Methods	Accuracy (%)	Recall (%)	F1-score (%)	Recognition Speed (FPS)
FR-CNN	86.2	80.5	83.3	13
FPN	89.8	82.7	86.1	7
YOLOv3	87.4	81.6	84.4	39
IYOLOv3	90.3	87.6	88.9	36
MOSGRNN [Proposed]	94.91	92.70	93.7	44

Accuracy: Accuracy in digital media content creation and editing involves precise measurements, statistics, and analytics to assess the impact, efficacy, and quality of created material. It involves monitoring key performance indicators (KPIs) like as conversion, audience reach, viewer retention, and engagement rates, crucial for informed decisions, content strategy improvement, and desired results. Figure 5 depicts the graphical outcomes of accuracy. Our proposed method, MOSGRNN, yielded superior results compared to the existing methods. For instance, while FR-CNN achieved an accuracy of 86.2%, FPN reached 89.8%, YOLOV3 achieved 87.4% and IYOLOV3 achieved 90.3%, TS-WRF outperformed them all with a superior accuracy rate of 94.91%.

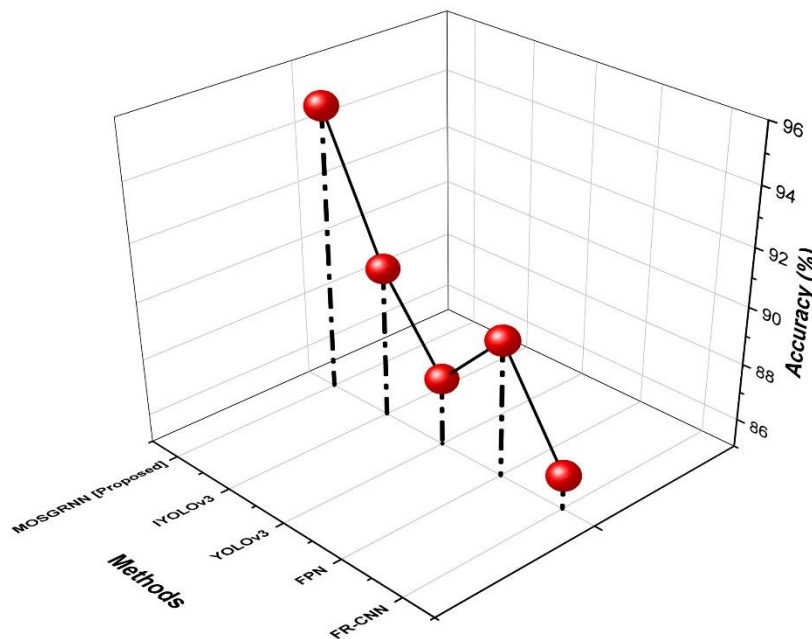


Figure 5: Graphical Outcomes of Accuracy

Recall: In the editing and production of digital media material, metric recall refers to the system's capacity to recover pertinent data from a dataset or database. It assesses the system presents the requested material from the pool of accessible resources, such as text, images, and videos. Workflows for creating and modifying information that is precise and efficient benefit from high metric recall. The recall's numerical results are shown in Figure 6. Comparing the Recall values of our suggested approach, MOSGRNN (92.70%), with those of other existing methods, including FR-CNN (80.5%), FPN (82.7%), YOLOV3 (81.6%), IYOLOV3 (87.6%), revealed higher results.

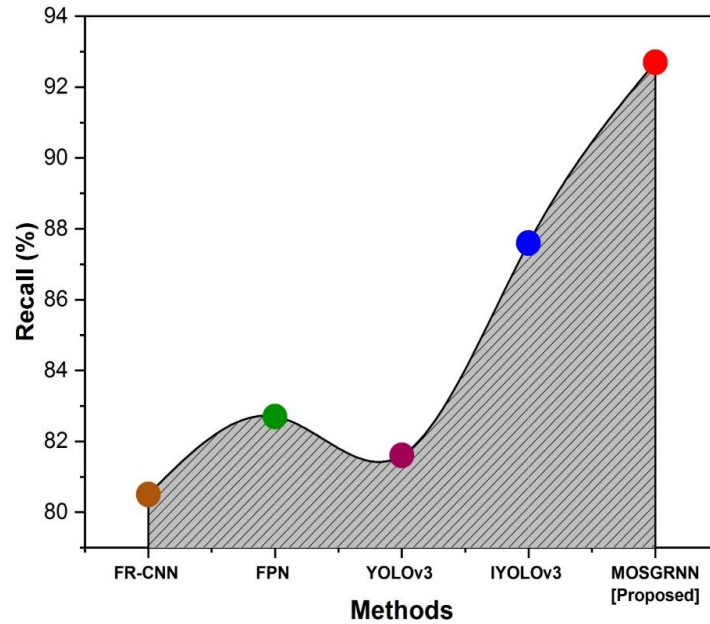


Figure 6: Graphical Outcomes of Recall

F1-value: It evaluates the capacity to produce high-quality material while avoiding mistakes and guaranteeing relevance to the target audience, highlighting effective strategies, in the space between precision and recall. Figure 7 shows that the F1-score values of our proposed technique, MOSGRNN (93.07%), were greater than those of other current approaches, including FC-RNN (83.3%), FPN (86.1%), YOLOV3 (84.4%), IYOLOV3 (88.9%).

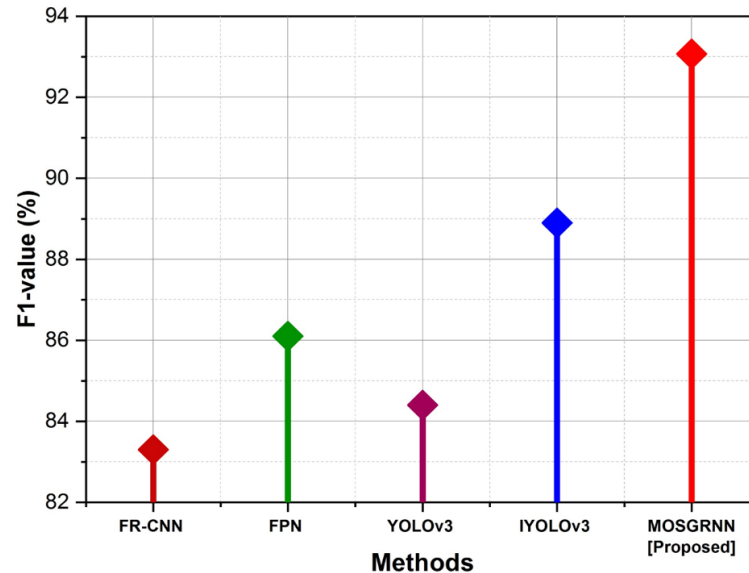


Figure 7: Graphical Outcomes of F1-score

Recognition Speed: The fast rate at which jobs like video editing, image processing, and content creation are finished in a certain amount of time, impacted by variables like editing time, software tool efficiency, and workflow optimization, is referred as metric speed in digital media content editing and production. Figure 8 shows that the speed rates of our proposed technique, MOSGRNN (44 FPS), were greater than the speed values of other current approaches, including FR-CNN (13 FPS), FPN (7 FPS), YOLOV3 (39 FPS), IYOLOV3 (36 FPS).

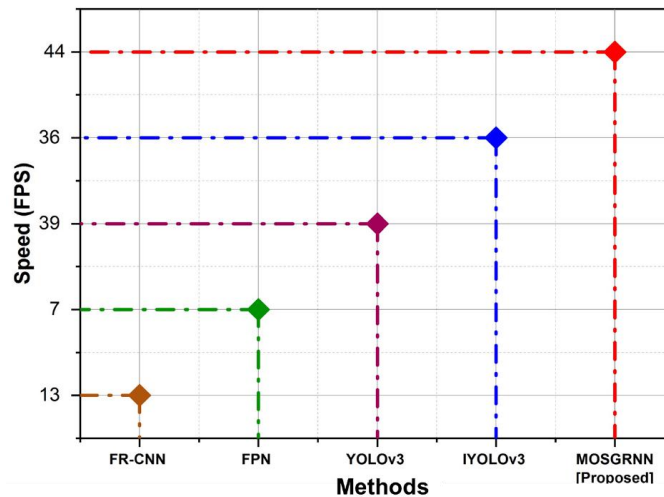


Figure 8: Graphical Outcomes of recognition speed

A. Discussion

In digital video editing, where context over time is critical, FR-CNN [25] is less successful at capturing complicated temporal links because it struggles with long-range dependencies. FPN [25] could have trouble managing visually rich and varied material, which would affect its capacity to divide and handle complex parts in manufacturing processes. YOLOv3 [25] is lacking in two areas, which can be critical for content editing tasks requiring fine-grained adjustments or object manipulations: managing small objects and precise localization. IYOLOv3 [25] despite the fact of YOLOv3's issues were recently fixed, it could continue to have trouble reliably capturing tiny optical messages, that could reduce its efficacy in tasks like semantically segmenting and tracking objects in digital media producing contexts. A useful tool for producing and editing multimedia content can be found in MOSGRNN. It is particularly good at recognizing trends in media material, keeping contextual throughout the process, and upholding intricate spatial connections. Smooth transitions and polished outcomes are produced by its versatility in handling multiple media types and its optimized operation for time and space information.

V. CONCLUSION

The enhanced efficiency, creativity, and standard of workmanship that result from using image processing and recognition techniques with the development and production of internet-based material have revolutionized the business. Algorithms can be used by expertise to eliminate time-consuming tasks, detect and resolve issues rapidly, and provide new opportunities for multimedia storytelling. Our proposed method outperforms existing methods in terms of F1-score (93.07%), recall (92.70%), accuracy (94.91%), and recognition speed (44FPS). Real-time object tracking and improved colour management represent two instances where creativity and technology combine to re-imagine the potential of digital media and open the door to more enticing and absorbing encounters in the years to come.

A. limitations and future scope

The use of sophisticated scene analysis techniques may lead to errors, handling a broad variety of image content changes can be challenging, and processing of images and recognition algorithms for images in digital formats for producing material and modifying are limited. Study on actual time based on artificial intelligence tools for smooth cooperation, better teamwork, and increased participation from viewers can be the primary objective of subsequent studies on the development and modification of content for digital media.

ACKNOWLEDGMENT

This research was funded by Network Ecological Chain of Digital Creative Industry, grant number CCIT2021STIT020102.

REFERENCES

- [1] Ullah, I., Adhikari, D., Su, X., Palmieri, F., Wu, C. and Choi, C., 2024. Integration of data science with the intelligent IoT (IIoT): current challenges and future perspectives. *Digital Communications and Networks*. Elsevier <https://doi.org/10.4018/979-8-3693-0807-3.ch003>

- [2] Khan, A.A., Shaikh, A.A., Laghari, A.A., Dootio, M.A., Rind, M.M. and Awan, S.A., 2022. Digital forensics and cyber forensics investigation: security challenges, limitations, open issues, and future direction. *International Journal of Electronic Security and Digital Forensics*, 14(2), pp.124-150. <https://doi.org/10.1504/IJESDF.2022.121174>
- [3] Neale, C., Kennedy, I., Price, B., Yu, Y. and Nuseibeh, B., 2022. The case for zero trust digital forensics. *Forensic Science International: Digital Investigation*, 40, p.301352. Elsevier <https://doi.org/10.1016/j.fsidi.2022.301352>
- [4] Okolie, C., 2023. Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies*, 25(2), p.11.
- [5] Goswami, A., Sharma, D., Mathuku, H., Gangadharan, S.M.P., Yadav, C.S., Sahu, S.K., Pradhan, M.K., Singh, J. and Imran, H., 2022. Change detection in remote sensing image data comparing algebraic and machine learning methods. *Electronics*, 11(3), p.431.mdp <https://doi.org/10.3390/electronics11030431>
- [6] Bendel, O., 2023. Image synthesis from an ethical perspective. *AI & SOCIETY*, pp.1-10. Springer <https://doi.org/10.1007/s00146-023-01780-4>
- [7] Arafah, B. and Hasyim, M., 2022. Social Media as a Gateway to Information: Digital Literacy on Current Issues in Social Media. *Webology*, 19(1), pp.2491-2503. <https://doi.org/10.14704/WEB/V19I1/WEB19167>
- [8] Kim, D. and Wang, Z., 2023. The ethics of virtuality: navigating the complexities of human-like virtual influencers in the social media marketing realm. *Frontiers in Communication*, 8, p.1205610. <https://doi.org/10.3389/fcomm.2023.1205610>
- [9] Gioti, A., 2024. Advancements in Open Source Intelligence (OSINT) Techniques and the role of artificial intelligence in Cyber Threat Intelligence (CTI) (Master's thesis, Πανεπιστήμιο Πειραιώς). [dione.unipi.http://dx.doi.org/10.26267/unipi_dione/3728](http://dx.doi.org/10.26267/unipi_dione/3728)
- [10] Adiandari, A.M., 2022. Financial performance innovation since digital technology entered Indonesian MSMEs. *International Journal for Applied Information Management*, 2(1), pp.50-58.
- [11] Mohamed Abdel Razek Youssef, A., 2022. The Role of the Digital Economy in Sustainable Development. *International Journal of Humanities and Language Research*, 5(2), pp.13-25. <https://dx.doi.org/10.21608/ijhrl.2023.215933.1013>
- [12] Tayal, A., Gupta, J., Solanki, A., Bisht, K., Nayyar, A. and Masud, M., 2022. DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases. *Multimedia systems*, 28(4), pp.1417-1438. Springer <https://doi.org/10.1007/s00530-021-00769-7>
- [13] Negi, A., Kumar, K. and Chauhan, P., 2021. Deep neural network-based multi-class image classification for plant diseases. *Agricultural informatics: automation using the IoT and machine learning*, pp.117-129. Wiley Online Library <https://doi.org/10.1002/9781119769231.ch6>
- [14] Kumar, R., Joshi, S. and Dwivedi, A., 2021. CNN-SSPSO: a hybrid and optimized CNN approach for peripheral blood cell image recognition and classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 35(05), p.2157004. <https://doi.org/10.1142/S0218001421570044>
- [15] Gonzalez-Gil, P., Robles-Enciso, A., Martínez, J.A. and Skarmeta, A.F., 2021. Architecture for orchestrating dynamic and powered image processing tasks in edge and cloud devices. *IEEE Access*, 9, pp.107137-107148. <https://doi.org/10.1109/ACCESS.2021.3101306>
- [16] Toldinas, J., Venčkauskas, A., Damaševičius, R., Grigaliūnas, Š., Morkevičius, N. and Baranauskas, E., 2021. A novel approach for network intrusion detection using multistage deep learning image recognition. *Electronics*, 10(15), p.1854. mdp <https://doi.org/10.3390/electronics10151854>
- [17] Xi, E., 2022. Image classification and recognition based on deep learning and random forest algorithm. *Wireless Communications and Mobile Computing*, 2022. Hindawi <https://doi.org/10.1155/2022/2013181>
- [18] Myagila, K. and Kilavo, H., 2022. A comparative study on the performance of SVM and CNN in Tanzania sign language translation using image recognition. *Applied Artificial Intelligence*, 36(1), p.2005297. Taylor & Francis <https://doi.org/10.1080/08839514.2021.2005297>
- [19] Yaman, O., Yetiş, H. and Karaköse, M., 2021. Image processing and machine learning-based classification method for hyperspectral images. *The Journal of Engineering*, 2021(2), pp.85-96. <https://doi.org/10.1049/tje2.12012>
- [20] Shinde, P., Patil, S., Ghanwat, O., Palkar, R., Mhaske, K. and Dhawase, D., 2022. Study on Fruit Recognition Using Image Processing. *Journal of Pharmaceutical Negative Results*, pp.7595-7604. <https://doi.org/10.47750/pnr.2022.13.S07.916>
- [21] Fayou, S., Ngo, H. and Sek, Y., 2022. Combining multi-feature regions for fine-grained image recognition. *Int. J. Image Graph. Signal Process*, 14(1), pp.15-25. mecs-press <https://doi.org/10.5815/ijigsp.2022.01.02>
- [22] Hu, S., 2020. Research on data acquisition algorithms based on image processing and artificial intelligence. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(06), p.2054016. <https://doi.org/10.1142/S0218001420540166>
- [23] Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G. and Wang, W., 2020, December. Face mask recognition system with YOLOV5 based on image recognition. In 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (pp. 1398-1404). IEEE. <https://doi.org/10.1109/ICCC51575.2020.9345042>
- [24] Chiu, M.C., Tu, Y.L. and Kao, M.C., 2022. Applying deep learning image recognition technology to promote environmentally sustainable behavior. *Sustainable Production and Consumption*, 31, pp.736-749. Elsevier <https://doi.org/10.1016/j.spc.2022.03.031>
- [25] Horak, K. and Sablatnig, R., 2019, August. Deep learning concepts and datasets for image recognition: overview 2019. In Eleventh international conference on digital image processing (ICDIP 2019) (Vol. 11179, pp. 484-491). SPIE. [spiedigitallibrary](https://doi.org/10.1117/12.2019.11179)