# Identification of Classroom Sheet Exam Cheating Trials using a Deep Learning Approach

Getasew Abeba
School of Computing
Woldia University, Woldia, Ethiopia

Amare Genetu
School of Computing
Woldia University, Woldia, Ethiopia

Sefinew Getenet
School of Electrical and Computer Engineering,
Woldia University, Woldia, Ethiopia

Dires Bitew
School of Computing
Woldia University, Woldia Ethiopia

Fasil Alemu
School of Computing
Woldia University, Woldia Ethiopia

## ABSTRACT
When a student tries to obtain academic credit in a dishonest, discourteous, careless, unreliable, or unjust manner, that behavior is called cheating. It has a variety of effects on the nation, education, as well as the student themselves. One is that cheating causes schooling to become less effective. There are various ways that researchers try to spot exam cheating, but they focus much of their efforts on finding instances of online exam cheating. However, little research has been done on the issue of classroom paper exams. In order to categorize a given classroom exam image as cheating or not, we model the detection of cheating trials as a classification task in this article. The model includes fundamental elements including image preprocessing, image classification, and evaluation approaches to identify cheating trial images. Following many experimental analyses, CNN exhibits the best accuracy of 92% for images with a size of 300 by 300. Finally, we advise considering this research to be a major issue that necessitates an in-depth investigation of dataset preparation. Therefore, we advise researchers to collect cheating trial datasets from various perspectives on cheating cases in order to enhance the model performance.

## Keywords
Exam cheating, sheet exam cheating trial, image classification, Convolutional Neural Networks

## 1. INTRODUCTION
Cheating is occurred when a person misleads, deceives or acts dishonestly on purpose to get unfair advantage [1]. Cheating may happen at school, at home or while playing sport. Cheating at school can happen in examination. Cheating in examination is widespread in all academic institution. Students are more likely to cheat answers during exam time. In examination, students cheat through different ways such as : copying from another student , copying from another student without her/ his knowledge , using unpermitted notes as cheat sheets , using electronic devices like phones as cheat aid ,asking for and giving answers verbally ,Morse code: using coughing or sneezing to communicate an answer ,passing notes on which answers to questions are written ,distracting invigilators to help others cheat ,writing notes on body parts ,writing notes on desks[1] .

Many scholars' researches results show that student cheating on exams requires mechanisms or tool for assuring the quality of educations through detecting cheating on exams [1]. Therefore, many researches have done on detections of cheating exam such as, a study by [2]. Proposed data mining method to identify students (persons) that commit cheat in online assessments (cyber cheats). In this study, the authors used starting times, local or remote IP addresses, finishing times and, the student's behavior: frequency of visits as attribute for determines the student were fraudulent behavior or not. But, their works difficult to identify cheating using data mining and only identified online assessment cheating.

In study by [2]. proposed text mining methodology and decision tree algorithm to detect cheating on open-ended exam. This study takes a document analysis.

However, most pervious researches were conducted mainly focus on detecting online exam cheating. Therefore, detecting online cheating exams is not enough for assuring the quality of education in related to detecting student cheating exam. They must be including offline sheet cheating exam detections which occur while in classroom test. In offline sheet exam, student used many ways of cheating techniques during exam time such as student uses sign language for sharing answers. they also use view on paper and mobiles [1].

Related works by [2],[4], and [6] identifies cheating trial for online exams. so, the dataset behavior is fully differed from offline classroom exam cheating trials. Works by [5] tried to identify cheating for offline exam, but the method used is document similarity measure. So, it lacks speed and flexibility when the number of comparisons is huge.

This article identifies the gap in offline cheating trial categories and uses deep learning to solve the problem in cheating trial detection.

## 2. STATEMENT OF PROBLEM
Student cheating in academic institutions affects the country, education, as well as the students in several ways. One is that cheating leads to a decrease in the quality of education. The main reasons for the difficulty of identifying cheating trials in offline exams are; Due to the position of invigilator, students perform cheating, Due to various types of cheating trials, the invigilator is not able to detect cheating. When different students perform cheating at the same time, the invigilator is unable to identify all the cheating trials at the same time, even when using video record analysis, it is difficult to analyze all videos because it is labor intensive and not perfect [3].

Researchable questions: To solve the above problems and to achieve the research objectives, the following research questions were proposed.

- How to identify cheating trial types for off-line paper-based exams?
- What are the previous works in cheating trial detection and their methods used to solve the problem?
- How to develop a model for classifying images in exam, whether it is a cheating trial or not.
- Which deep learning classification algorithms work best in the model?

## 3. METHODOLOGY

In this research, we have used experimental methodologies, which are broadly used in computing science. This methodology uses two phases to evaluate new solutions for problems. The first is the exploratory phase, in this phase the researcher is taking measurements that help to identify what are the questions that should be asked about the system under evaluation. Then the second phase, which is an evaluation phase, answered these questions. A well-designed experiment started with a list of the questions that the experiment is expected to answer [7]. We use an experiment tool to evaluate the research.

### 3.1 3.1 Data collection Methodology

The dataset we used in this research is an image that confirms the cheating trial. In order to get the cheating trial image, we used a recorded video in the exam and trials available on the web. In order to annotate the image, whether it is a cheating trial or not, we used

literature reviews. To enhance and incorporate more cheating trial types, we did analysis on cheating trial cases in mid and final exams by Woldia University students.

### 3.2 3.2 System development methodology

Currently Machine learning and Deep learning techniques helps us for Identification problem using large dataset as input for model training. Deep Learning has emerged as a new area in machine learning and is applied to a number of signal and image applications [8]. So. In this research we transformed the cheating trial identification problem us a classification problem. In this research we evaluated the state-of-the-art techniques in image classification problem experimentally.

Proposed Model Architecture Implementation flow: In this research we proposed a model for cheating trial Identification by incorporating different components.

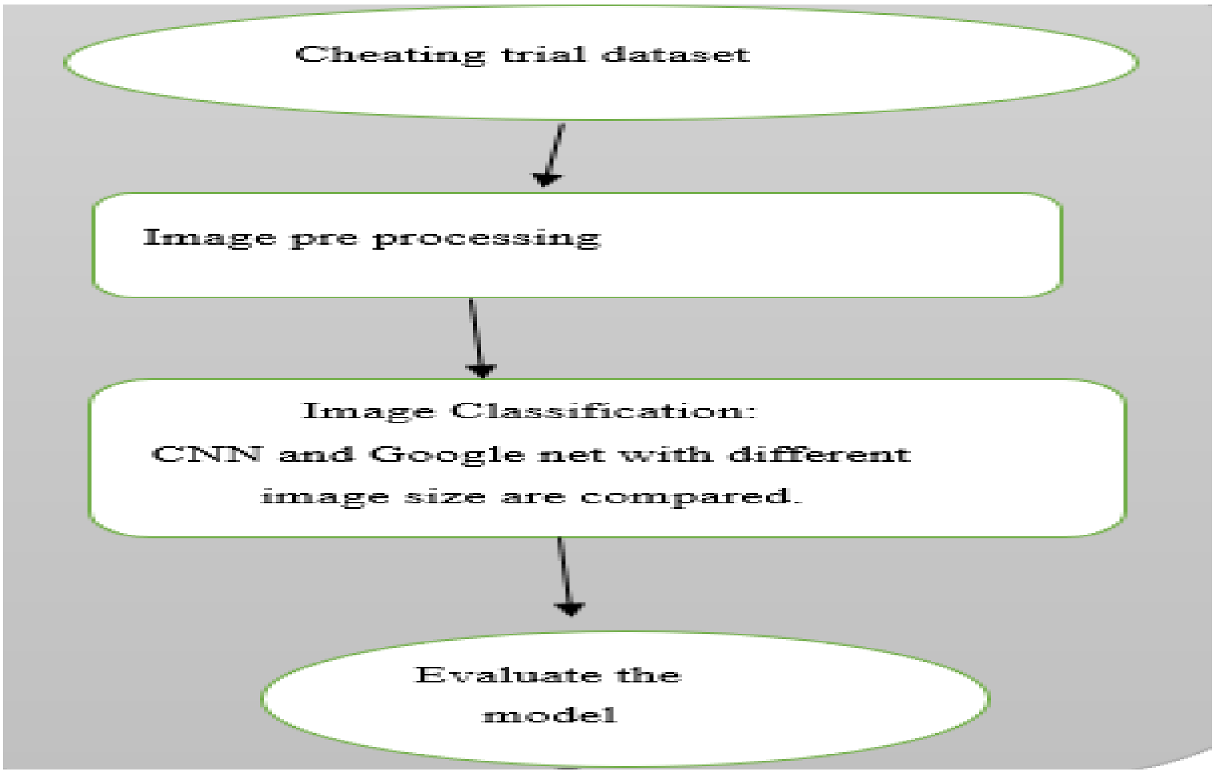


**Fig 1 Proposed Model Flow**

**Image acquisition**: The dataset used in this research is a crucial aspect of the exam cheating detection model. It consists of images that confirm exam cheating actions as well as non-cheating actions. We manually collected the image dataset from Woldia university students using Samsung A30 mobile phone cameras with 16 megapixels and Sony lens cameras with 16.1 megapixels. We also gathered images from the internet that depict exam cheating actions to expand the dataset further. In total, we collected approximately 1236 images for the cheating class and 1565 for the non-cheating class. However, since it was challenging to capture real images during an exam, we utilized augmentation techniques to increase the size of the dataset.

The augmented dataset includes 2663 images for the cheating class and 2623 images for the non-cheating class, totaling 5286 images used for building the model. Collecting and preparing the dataset was a crucial step in ensuring the accuracy and reliability of the model. By utilizing a diverse range of images and techniques, we were able to create a comprehensive dataset that accurately represents real-world scenarios of exam cheating. Overall, the dataset played a vital role in building the deep learning model used for detecting exam cheating.

Figure 2 displays a selection of sample images used in the dataset for exam cheating detection. These sample images provide a clear understanding of the types of cheating actions included in the dataset. By using diverse examples, we were able to create a comprehensive dataset that accurately represents real-world scenarios of exam cheating. This helped to ensure that the deep learning model developed for detecting exam cheating is reliable and effective in identifying various forms of cheating.

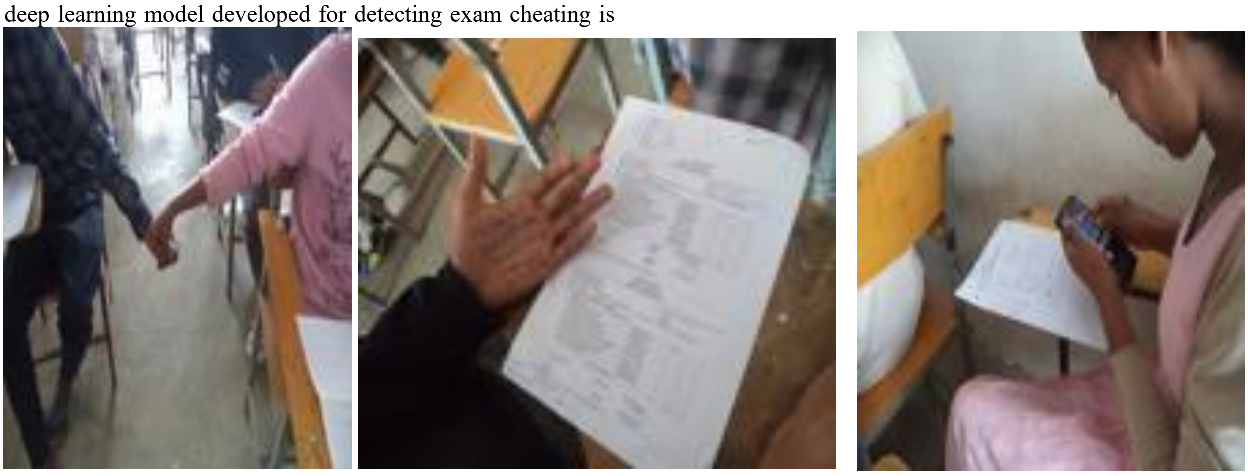

**Fig 2 Cheating Images**

**Image preprocessing**: is the steps taken to format images before they are used by model training and inference. In this research we did Convert recorded video into image frame, resizing image, remove noise and segmentation. In order to effectively detect and analyze the images, several image preprocessing techniques were employed in this research. The first step was to remove any artifacts and enhance the contrast of the acquired image to facilitate the detection power. This is crucial as it improves the quality of the original image and enables better detection of the image. Bag was identified as a noise to be removed in the dataset, and various techniques were used to address this issue [9].

Image resizes: One of the key techniques used was image resizing, which involved standardizing the total pixel count of the images. This was done to reduce processing time and computational costs, as the collected images had different sizes, ranging from 100 to 2000 pixels in width and 100 to 3000 pixels in length. Resizing the images to a similar size was essential for the proposed method, as large pixel sizes consumed too much computational cost and time. To determine the optimal image size, we compared the effects of resizing the dataset into 128x128, 224x224, and 300x300 pixels.

For exam cheating image preprocessing, both cheating and non-cheating images were resized. and converted from RGB to Grayscale. We also balanced the image intensity using histogram equalization and performed noise minimization using a median filter. These techniques helped to enhance the quality of the images and improve the accuracy of the proposed method for detecting cheating actions.

*Converting RGB into Grayscale*: Converting RGB images to grayscale is an important step in many image processing tasks. Since RGB images are three-channel color format, they contain a lot of redundant information that may not be useful for some applications. Therefore, to reduce the complexity and improve the processing speed, converting RGB images into grayscale format is often preferred. The following figure shows the original RGB image and grayscale images respectively.

***Histogram equalizer*:** Histogram equalization is a widely used technique to balance the intensity of an image. This technique is useful for enhancing the appearance of images by adjusting the intensity of the pixels. When an image is predominantly dark, its histogram is skewed towards the lower end of the grey scale, which compresses all the image details into the dark end of the histogram. This makes the image less clear and hard to interpret. However, by stretching out the grey levels at the dark end of the histogram to produce a more uniformly distributed histogram, the image can become much clearer and easier to analyze.

*Median filter*: In digital image processing, the use of a median filter has been adopted to minimize the effect of noise. The median filter is a type of smoothing filter used to reduce noise in an image, similar to a mean filter. However, it outperforms the mean filter in preserving important details in the image. It is highly effective in eliminating impulse noise, which appears as random high-frequency features that are bright and/or dark across the image. Impulse noise usually lies outside the peak of the distribution of any given pixel neighborhood, making the median filter well-suited to learn where impulse noise is absent and to remove it through exclusion. The median filter is demonstrated to be superior to other algorithms in noise removal, as it preserves edges for a fixed window size. Thus, median filtering is widely used in digital image processing [10].

**Image classification**: involves the extraction of features from the image to observe some patterns in the dataset [8]. We compared deep learning image classification algorithm CNN with different parameter and predefined model like Google Net.

**Model evaluation**: We evaluated the model using the confusion matrix and accuracy. The relationship between the predicted value and the actual value is called accuracy. Accuracy measures how close the predicted value is to the actual value.

# 4. RESULT AND DISCUSION

## 4.1 Dataset preparation

To train and evaluate the proposed system, a custom dataset is prepared as there were no available datasets. The dataset was prepared by collecting images of both exam cheating actions and non-cheating situations from Woldia University, as well as from the internet. The dataset was divided into training and testing sets, with 80% of the data used for training and 20% used for testing. Our system's performance was evaluated using accuracy, confusion matrix, and time metrics. Additionally, we compared the accuracy of our system during training with other experimental scenarios.

### *4.1.1 Experimental Tools*

We determined the values for hyper parameters based on our dataset, taking into account training time and performance. Assigned hyper parameter value are; epoch = 50, batch size = 32, iteration = 30, kernel = 3 * 3, learning rate= 0.001, dropout = 0.5.

### *4.1.2 Evaluation Result*

The proposed was trained multiple times using different activation functions (ReLU), datasets (original and preprocessed data), and image sizes (128 x 128, and 300x300). Finally, we compared our model with other architectures such as Google Net. The reason behind selects this range of dataset size is because of the actual image size.

The proposed model results with image size 128 x 128: The model was trained using the Adam optimizer, the ReLU activation function, and a learning rate of 0.001. In the first experiment, the original data was used without any preprocessing except for resizing the images. The results obtained were a training accuracy of 99% and a test accuracy of 91%, as depicted in the accompanying figure.

The proposed model results with image size 300x300: The experimental results show that the model achieved a training accuracy of 99% and a validation accuracy of 92%. This performance was attained using the Adam optimizer, ReLU activation function, and a learning rate of 0.001, as depicted in the accompanying figure.

The figures below provide a comprehensive overview of the experiment. They contain all the details and results obtained during the study.

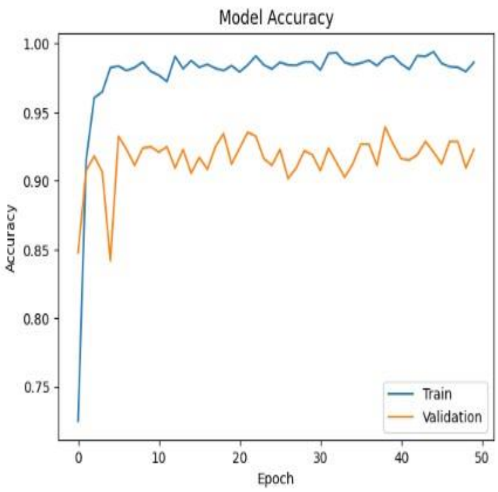


**Figure 3: train and validation results of the model with 300 X 300 image size**

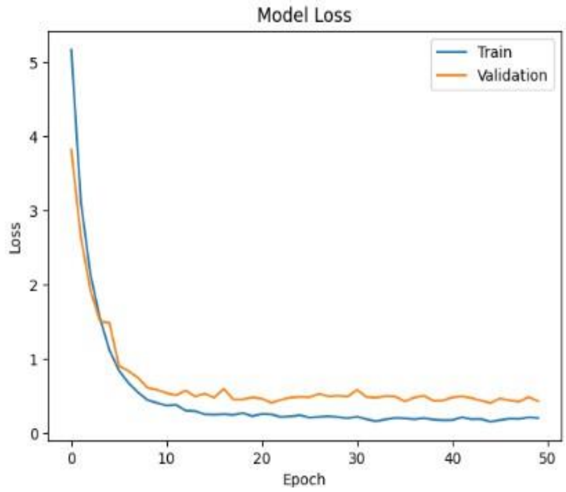


**Figure 4: train and validation loss of the model using 300 X 300 image**

Comparison of Our Model with State-of-the-Art CNN Models: In this study, we conduct a comprehensive comparison of deep neural network models using the same dataset and parameter settings while considering architectural differences. By comparing the performance of our model with GoogleNet, we aim to gain insights into the strengths and weaknesses of our approach in relation to well-established deep neural network architectures. This comparison allows us to assess the effectiveness and efficiency of our model, on architectural differences among the models.

*The google net model* results: In our evaluation, we compared the performance of the GoogleNet model using parameters similar to those used for evaluating our model, including a training duration of 50 epochs. We assessed the model's performance using evaluation metrics such as Recall, F1 Score, Accuracy, and Precision, derived from the confusion matrix. Figures 26 presents the confusion matrix, which indicates that the accuracy achieved by the GoogleNet model is 99 % for training accuracy and 89% for testing accuracy which is lower compared to our proposed model. Notably, the accuracy metric in the confusion matrix demonstrates that our proposed model outperforms the GoogleNet model. This suggests that our model exhibits higher accuracy in predicting the target classes compared to the GoogleNet architecture.

## 4.2 Summarized evaluation result

In table 4, the evaluation of cheating detection in paper-based exam classification is presented. The results are compared across GoogLeNet, considering various factors such as different image sizes (128x128 and 300x300), the ReLU activation function, a learning rate of 0.001, and training for 50 epochs. The dataset used for evaluation was preprocessed in the same manner.

**Table 1: training and testing performance comparison**

| model | Image size | Training accuracy | Testing accuracy |
|---|---|---|---|
| GoogleNet | 224 X 224 | 99% | 89% |
| CNN(proposed model) | 128 X 128 | 99% | 91% |
| CNN(proposed model) | 300 X 300 | 99% | 92% |

# 5. DISCUSSION

We conducted multiple training iterations of the model with different parameters. The evaluation results are summarized in section 4.6.5. Firstly, using a preprocessed image size of 128x128 and the ReLU activation function, we achieved 99% training accuracy and 91% testing accuracy. Secondly, with the ReLU activation function and an image size of 224x224, we obtained 95% training accuracy and 86% testing accuracy. The third experiment utilized a preprocessed image size of 300x300, resulting in 99% training accuracy and 92% testing accuracy, which outperformed the other experiments with the same dataset.

Next, we trained the dataset using the GoogLeNet model with the ReLU activation function and preprocessed image size of 224x224. This yielded 99% training accuracy and 99% validation accuracy. Based on the above results, the proposed CNN model with an image size of 300x300 and the ReLU activation function exhibited better accuracy compared to the state-of-the-art models. However, it's important to note that as the image size increased, the training time also increased due to the higher number of pixels and parameters to be trained.

# 6. CONCLUSION

This research concludes by addressing the critical topic of exam cheating in the educational field, especially Ethiopia. We used Convolutional Neural Networks (CNN) for feature extraction and classification using a customized dataset obtained from Woldia University and the internet, and after experiment we got the best results. The studies were out with image sizes of 128x128 and 300x300 showed growing accuracy, with the best validation accuracy of 92% being produced by the 300x300 size. Larger image sizes increased performance even though they needed more training time. Comparing our proposed model with GoogleNet our method better in accuracy, achieving a remarkable 92% accuracy with the ReLU activation function and a sigmoid classifier.

In summary, our approach using a custom dataset, CNN, and the proposed model with the ReLU activation function and sigmoid classifier showcased significant advancements in the field of exam cheating classification. By outperforming other experiments and achieving higher accuracy rates, our method presents a promising solution to combat exam cheating effectively.

# 7. CONTRIBUTION

The main contribution of this research is; the collection and annotation of our own dataset from a unique perspective, and contributes the cheating detection model for the community. Since there is no publicly available datasets specifically designed for paper-based exam cheating classification.

# 8. RECOMMENDATIONS

In future, research could be conducted in the following directions to enhance and expand the existing work: To further enhance and expand upon this research work, future researcher could explore the following directions; Alternative Algorithms, Enhanced Feature Extraction Techniques. Real-Time Detection Systems, Dataset Expansion and Diversity, Deployment and Evaluation in Real-World Settings.

# 9. REFERENCE

[1] M. A. Med, "Written Exam Cheating and Prevention and Detection Strategies: The Case Of DMU," *J. Educ. Pract.*, vol. 10, no. 7, pp. 60–70, 2019, doi: 10.7176/jep/10-7-07.

[2] J.-A. Hernándeza, A. Ochoab, J. Muñozd, and G. Burlaka, "Detecting cheats in online student assessments using Data Mining," *Conf. Data Mining| DMIN*, vol. 6, no. January, p. 205, 2006, [Online]. Available: https://www.researchgate.net/publication/220705005

[3] F. K. Chiang, D. Zhu, and W. Yu, "A systematic review of academic dishonesty in online learning environments," *J. Comput. Assist. Learn.*, vol. 38, no. 4, pp. 907–928, 2022, doi: 10.1111/jcal.12656.

[4] F. Kamalov, H. Sulieman, and D. S. Calonge, "Machine learning based approach to exam cheating detection," *PLoS One*, vol. 16, no. 8 August, pp. 1–15, 2021, doi: 10.1371/journal.pone.0254340.

[5] E. R. Cavalcanti, C. E. Pires, E. P. Cavalcanti, and V. F. Pires, "Detection and evaluation of cheating on college exams using supervised classification," *Informatics Educ.*, vol. 11, no. 2, pp. 169–190, 2012, doi: 10.15388/infedu.2012.09.

[6] P. Sharma, U. Tripathi, and U. Kumar, "AUTOMATING ONLINE PROCTORING THROUGH ARTIFICIAL INTELLIGENCE," pp. 1894–1896, 2021.

[7] N. Williams, "Non-Representational Theory Non-Representational Theo- ry / Non-Representational Geographies," 2020.

[8] D. Jaswal, S. V, and K. P. Soman, "Image Classification Using Convolutional Neural Networks," *Int. J. Sci. Eng. Res.*, vol. 5, no. 6, pp. 1661–1668, 2014, doi: 10.14299/ijser.2014.06.002.

[9] G. H. Yun, S. J. Oh, and S. C. Shin, "Image preprocessing method in radiographic inspection for automatic detection of ship welding defects," *Appl. Sci.*, vol. 12, no. 1, 2022, doi: 10.3390/app12010123.

[10] A. C. Frery, "Image filtering," *Digit. Doc. Anal. Process.*, pp. 55–70, 2013, doi: 10.1201/b10797-8.

[11] E. Bochinski, T. Senst, and T. Sikora, "Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms," *Proc. - Int. Conf. Image Process.*

[12] *ICIP*, vol. 2017-September, no. August 2018, pp. 3924–3928, 2018, doi: 10.1109/ICIP.2017.8297018.