

Analyzing satellite images by apply deep learning instance segmentation of agricultural fields

Atheer Joudah¹, Souheyl Mallat², Mounir Zrigui³

¹ Department of computer sciences University of Monastir, Faculty of Monastir, Tunisia

² Research Laboratory in Algebra, Numbers theory and Intelligent Systems

³ Department of computer sciences University of Monastir, Faculty of Monastir, Tunisia

ABSTRACT

This novel research focuses on multi-exposure satellite images of agricultural fields using image analysis and deep learning techniques. The development of image edge smoothening system using CNN is in hot pursuit, with special attention being given to the smoothening of all the edges of image. Given its high propensity to meta-size, going hand in hand with severe decreases in preservation rates, and the high inter-edge variability in image appearance, as well as a strong requirement on the training of the physician properly de-noising an image can be considered a daunting task. The purpose of this advance research is to use a deep learning and image analysis pipeline for multi-exposure satellite image for the segmentation of edges in an image using with hybrid techniques in deep learning and imaging. The literature review of different papers was conducted with different imaging model architectures. The CNN custom model was created for the task, and deep learning technique (CNN) was used with different levels of fine tuning of hybrid satellite image analysis techniques. Screening for high edge filter to identify edges at high accuracy has been under debate. The custom deep learning model architectures were designed to represent different depths. Additionally, deep learning CNN model was created to represent traditional automated image analysis approach. The study also attempts to find solutions to practical deep learning challenges such as low training speed and lack of transparency with an accuracy of 98.17% absolutely.

Keywords: Satellite imagery, deep learning, agriculture fields, optimization, CNN, image analysis, segmentation, detection, computer vision, population, surface.

Corresponding Author:

Atheer Joudah
Department of computer sciences University of Monastir
Faculty of Monastir, Tunisia
E-mail: atheermubs@gmail.com

1. Introduction

Image satellite image analysis and edge smoothening is a trend that has been driven by the fact that human beings are known to be susceptible to filtering mistakes enabled a 97% correct simultaneous identification of both males and females of all the 18 species in an independent test as mentioned in [1]. Research by [2] proposed a method for precise extraction of edge region which is based on deep learning methodologies. Results show that learning the optimum kernel combination of multiple features vastly improves the performance, from 55.1% for the best single feature to 72.8% for the combination of all features. To lessen noisy artifacts, the input image, after pre-processing, was applied to CNN. CNN produced a segmentation mask which detected the area of the image edge. Using some post processing operations, the quality image of the mask is being further improved. Their input images were produced by customary cameras; henceforth, those were pre-processed in order to manage noisy artifacts. They used a filter to decrease the noise of the images As shown by the experimental results, the excellent accuracy is obtained by using our method, i.e., 98.46% mean intersection-over-union and

99.15% mean pixel accuracy on the BH-rail-dataset as mentioned in [3]. They created two patches, one is local patch, and another is global patch. The local patch is illustrated as a window around each pixel in result showed accuracy 90.67% as mentioned by [4]. Whereas the global patch is used to show the global structure of the area. After that, both the local and global textures are sent into the proposed CNN architecture as the training input. The experimental outcomes showed that their suggested method can outpace other architectures to detect lesions. An image colors may consist of an image signal processor (ISP), the image analysis algorithm and the controller. A traditional image signal processor processes the RAW output of an image sensor and produces a compressed image that can be used by a vision application. Typical ISP stages include smoothening, de-noising, white balancing and color mapping, gamut mapping, tone mapping, and compression. Those stages, albeit standard, require excessive computational intensity to produce a high-quality output image, which in the case of computer vision applications is not necessarily essential. It shows the local texture around the middle pixel shows that our combined model (Xception-LSTM) has the highest classification accuracy of 90.79% for blood cell images compared to other models as given by [5] and can be approximated. Approximate computing is a concept that relies on building systems with acceptable behavior from inexact hardware or software components. It allows to trade-off application accuracy in order to achieve considerable performance and energy gains the recovery accuracies for the two training sets are 0.893 and 0.887.as given by [6] at design time. A key challenge in approximate computing is the identification of those sections either in hardware or software that can be approximated, as there is always a risk of crashing an application if a critical component is being approximated. This led to a discussion of which features were most fundamental to the process of surveillance: Reliability¹, of the system, Maintenance (frequency and cost), which directly affects the time that the system is not available and the cost of operation/maintenance for processing satellite images, So Human resources, which is a limited and expensive factor in the surveillance methodology and Versatility of missions, since it is expected that the developed system could be a powerful aid for other types of missions besides surveillance making it a more cost-effective investment, in Overview of the area, since it would permit better route identification in order to grow the crops on the fields, Precise identification of the crop area, as it minimizes resources while permitting an efficient and fast localization and intervention for fielding the land then Performance on rough terrain, as most of the forest and agricultural areas are located in rough terrain making it crucial that the system can easily overcome this issue, Adequate performance during the night, since the most dangerous fires start during that time, usually intentionally, representing around 30% of the total fires. This is since during the night they go undetected for longer, causing the fire to grow in size. This paper is organized into 3 sections:

In the first section, in order to ensure that the developed system is adequate for its surveillance role from satellite images. In the Second section, we choose open data then make training and preprocessing for it and will choose split data from our dataset after that will do extraction and selection feature and make normalization of all data by CNN algorithm then segmentation if test true will experimentally result of recognition otherwise will re-tanning and check again. In the third section We show the graphs for the CNN algorithm on satellite images from which we learned a threshold of 50 for this purpose using an 8-bit encoding that yields values ranging from 0 to 255.

¹in order to ensure that the developed system is adequate for its surveillance role from satellite images;

²Integration of an image preprocessing pipeline and approximate computing in a baseline deep learning based CNN framework.

³A CNN stands for convolutional neural networks originated from this work whose network has a similar architecture cognition through which 70% data is used for training, 20% for testing and remaining 10% for validation.

³To exploit the pre-trained CNN for image edge smoothening of the different datasets, a top model is to stacked on the deep learning with an accuracy of 98.17%.

⁴Not only are these residual blocks hypothesized to facilitate training by learning the residual mapping, they also decrease the number of parameters while allowing a multi-layered top model.

1.1. Problem statement

Satellite image uses multiple cameras for additional safety during navigation. A use case of a vision-based lateral control example using a single camera is studied in this project, where the camera output is being processed by an image analysis application and the autonomous functionality is maintained by a controller that actuates based on the input from the camera. The camera produces 60 frames per second and the rest of the application needs to achieve real time performance in order to process all those frames as given by [7]. The camera is the sensor and is attached to the system. Each frame from the camera sensor goes through a series of processing stages combined with complementary advances in the field, OHEM leads to state-of-the-art results of 78.9% and 76.3% mAP on PASCAL VOC 2007 and 2012 respectively as mentioned by [8]. Initially, the frames need to be processed by an image signal processor (ISP), which converts the RAW output of the camera sensor to a format that is useful for the human and computer vision. The resulting image is, then, ready to be processed by the image analysis stage, which performs feature extraction and provides information about the image analysis environment to the image-based controller, average identification accuracy of 98.9%, and the Cifar10 model achieves an average accuracy of 98.8%. The improved methods are possibly improved the accuracy of maize leaf diseases mentioned by [9]. The controller uses the visual feedback from the camera and actuates the required steering angle in order to allow the car to drive autonomously. In principle, a fundamental parameter in the control design is the sampling period. It depends on the duration the application needs in order to finish the required computations, from sensing to actuating. Typically, shorter sampling periods are necessary to achieve real time performance and maintain high quality-of-control as mentioned by [10]. For the purpose of improving the quality-of-control of this image-based control application, seven different approximated ISP pipelines are developed by [11]. This allows to analyze the trade-off between runtime improvement and quality degradation with respect to quality-of-control for different degrees of approximation.

Table 1. Decision matrix, comparison between different agriculture field surveillance methods through satellite

PARAMETERS	IMPORTANCE (1-10)	CONTROL TOWER	SURVEILLANCE	PATROLIN G	NAN
Reliability of the system	9	4	4	5	6
Low maintenance	6	6	6	5	5
Low field resources	5	2	7	3	5
Safe for human life	10	4	7	4	7
Versatility of missions	8	4	4	7	7
Overview of the field area	9	5	4	3	7
Method of deterrence	7	4	4	7	6
Precise identification of the agricultural field	9	4	4	6	6
Performance independent of terrain	8	4	4	4	7
Detection during the night	9	4	5	5	7
Total		41	49	49	63

1.2. Related work

Approximate computing is a technique that leverages from the tolerance of applications to errors or inexact computations that reduce the quality in a controllable and acceptable manner. A large number of modern applications can tolerate those inexact computations and boost performance and energy efficiency as mentioned by [12]. Software techniques such as loop perforation, memorization, precision scaling, task dropping, and data

sampling or hardware techniques such as over scaling, clock over-gating, body-biasing and refreshing rate, may yield benefits of possibly up to 50% in execution time and analogous improvements in energy efficiency.

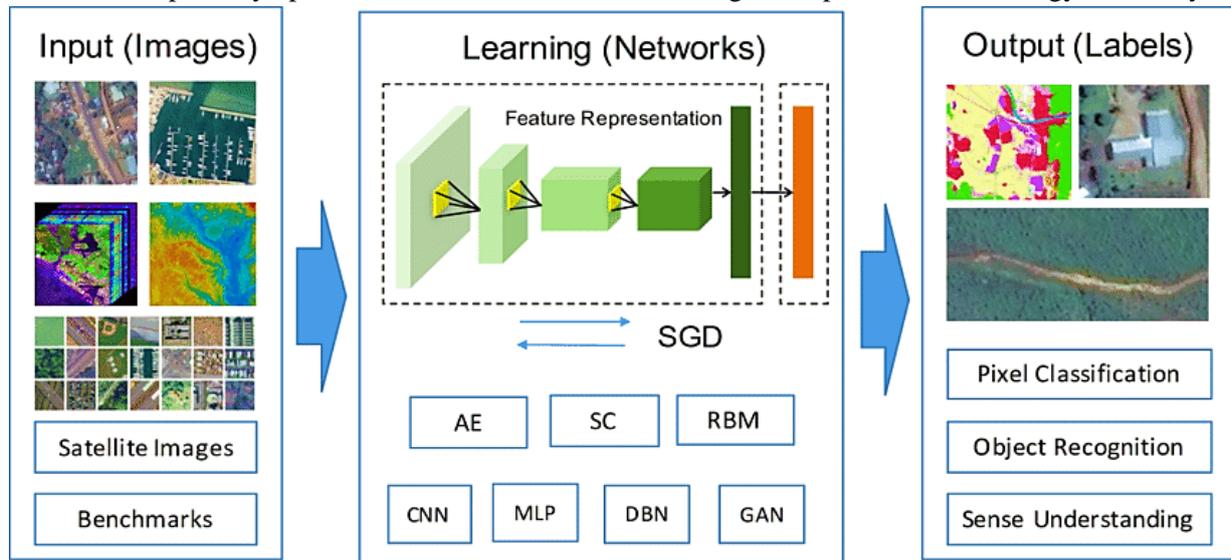


Figure 1. Different deep learning-based models and learning approaches for agricultural field processing [12]

The satellite image analysis with deep learning is used to validate the data in artificial intelligence-based systems, interpretable machine may even teach humans about how to make better decisions as mentioned by [13]. To address the transparency problem, visualization techniques can be used to identify the image regions most important for the model's prediction. It is also possible to highlight the shapes and textures the network sees by visualizing pixel-wise impact on prediction score. Another visualization technique is called guided backpropagation as mentioned by [14]. The RBN network uses multiple beacon layers as well as 3 fully connected layers. The classification of agricultural fields using RBN is not easy with sample images as they are quite different from classification in ImageNet. ImageNet requires classification of everyday objects such as animals, household items, etc., while agricultural fields sample images are medical data where the objective is detecting the agricultural fields. Therefore, it is safe to assume the fully connected AE and SE classification layers were trained from scratch and it is highly likely the deeper convolutional layers were replaced as well. This is where looking back at the training evaluation can be useful and; indeed, the observation that smaller training losses still correlated with improvement in testing performance do not permit any conclusion towards an overfitting hypothesis. The DBP is very expensive technique to follow up for segmentation. Active devices have high accuracy and low complexity, but they can be uncomfortable since it needs physical contact between the user and the device. While the MLP technique is computationally demanding, they don't require physical contact, and they are friendly compared to active devices. It can be observed that the general tendency is for imaging networks to better fit the data than feedforward ones. Secondly, the GAN also shows that networks that had two layers retrained could reach lower training losses than the model in which a single layer was retrained. The scenes in the physical world include brusque lightening circumstances leading to highlights (over-exposed areas) or shadows (under-exposed areas) in images taken digitally.

Conventional digital cameras usually fail capturing details in over as well as under exposed areas in the case when high or low exposure configurations are set in camera as mentioned by [15]. To illustrate the scene as realistically as possible, some modifications are made on the hardware side, however, to reduce the costs, some are made on software side, where most of the problems were solved from this perspective. The local contrast which is rapidly reduced as soon as the under-exposed image has been started to capture with foreground and background focused as mentioned by [16]. Secondly, slight variation and robustness with entropy measure detected. furthermore, at the over-exposed image the CNN feature start to increase and gradually decreased. the other features ranging between (no affected to light effect) with the light conditions mentioned by [17]. The

main challenge for this research is to find the agriculture field-wise solution for illumination estimation with human interference with CNN technique by separating different instances of multi-exposure satellite images. Traditional segmentation methods such as the seeded watershed algorithm, which is implemented in satellite images, are slow and sometimes make errors as mentioned by [18]. In addition, the user needs to be experienced to use these algorithms as parameter tuning is required. The work of Zhanget [19]. This is especially true for news topics chosen by the user. This work is created utilizing the rationale-augmented convolutional neural network (CNN) [20]. In this research vehicle detection system from infrared images using YOLO (You Look Only Once) computational mechanism [21]. In this research Data clustering is an important machine-learning topic. It is useful for variety of applications one of them is image segmentation. [22]. Some of the many classification models are SVM (support vector machine), KNN (K- Nearest Neighbors), Decision tree, Logistic Regression and ANN (Artificial Neural Network) back propagation. For this paper we would consider different procedure and method of early detection of the glaucoma disease using the MATLAB Deep Convolutional Neural Network (DCNN) [23].

One of the topics that occupies many articles is the improvement of medical diagnostic processes, where many articles have begun to elicit algorithms to increase the efficiency of disease diagnosis [24]. In this research for 5-class grouping assignment, we report 88.4% exactness. For 4-class grouping undertaking to recognize carcinomas we report 92.3% exactness, 96.2%, and affectability 94.5 by 87.2% at the high-affectability working point. As far as anyone is concerned [25].

1.3. Aim of contribution

The aim of this paper is contributed to the following key aspects of the use of deep learning and image analysis for smoothening of edges in an image. The main contributions of this research include the following:

- Integration of an image preprocessing pipeline and approximate computing in a baseline deep learning based CNN framework.
- Characterization of accurate and approximate algorithms to identify the regions-of-interest for approximation.
- Study and characterization of approximation choices of the satellite imagery & deep learning algorithm with respect to agriculture fields.
- Approximation of satellite image without taking into account the timing impact.
- Approximation of satellite image taking into account the timing impact on sampling period.
- Application profiling to obtain execution times and to compute optimized sampling periods for segmentation of agricultural fields.
- Trade-off analysis between approximation and satellite image for optimized sampling period designs for image analysis with CNN.
- Development of a toolchain written in a high performance language that allows the exploration of different approximations and their impact on quality-of-control.
- Identify the highly performance quality measure approach with CNN.
- Identify the optimal combination between the quality measure approaches that provide the best outcomes.

2. Methodology

In Figure 2 we choose open data then make training and preprocessing for it and will choose split data from our dataset after that will do extraction and selection feature and make normalization of all data by CNN algorithm then segmentation if test true will experimentally result of recognition otherwise will re-tanning and check again.

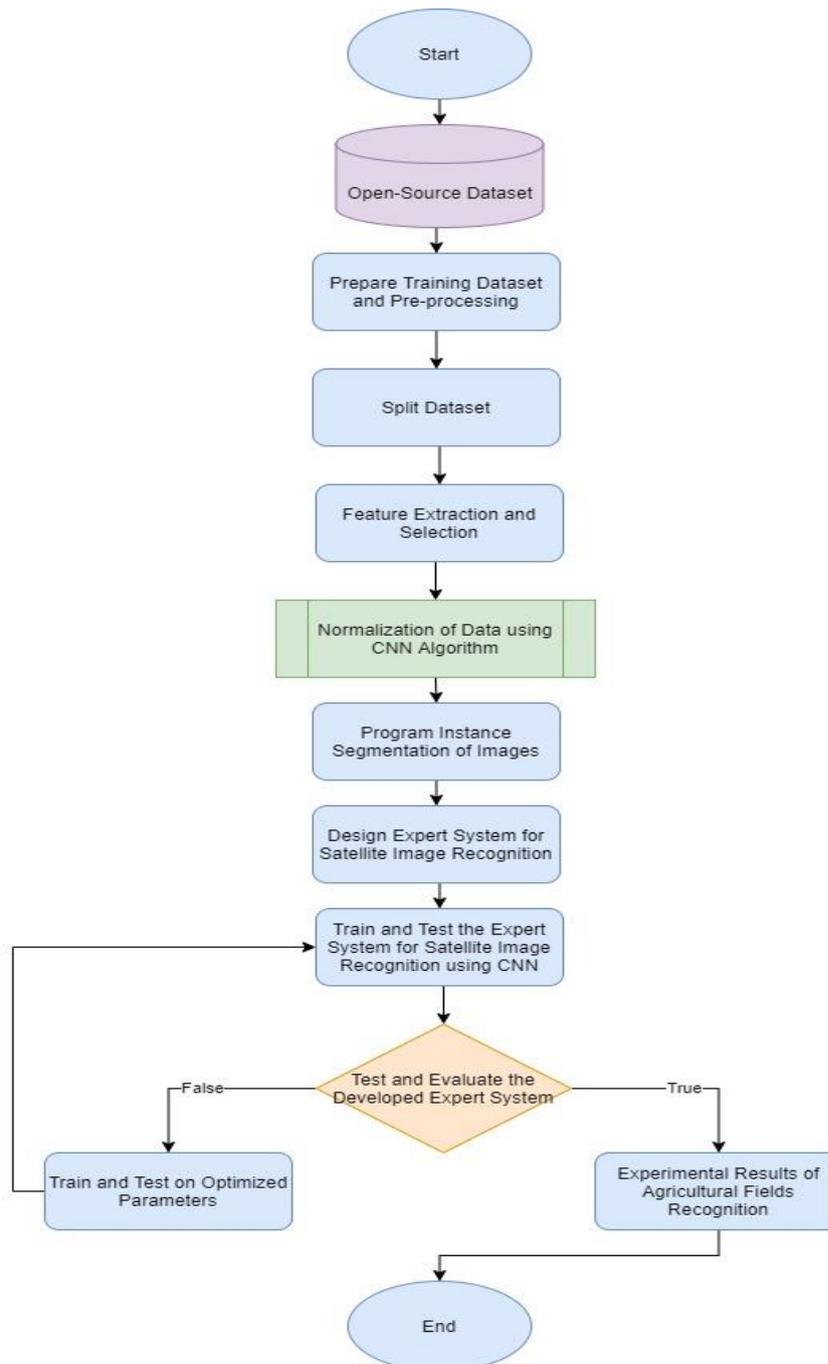


Figure 2. Flowchart of proposed approach being followed in this research work

The examination will make commitments to the agriculture field-wise estimation remuneration and multi-exposure imaging, which comprises of testing images of light in different stances from arrangement. We will refine and finish the recursive filtering to incorporate multi-exposure satellite image analysis, all things considered, and impediments for order utilizing Convolutional Neural Network (CNN), In this architecture, the CNN are responsible for feature extraction, where the features to which they will respond are determined through a learning process of the variable input connections. Simple features, such as specific orientations of lines and edges, will be extracted by lower-level CNN models, while higher-level CNN models extract more global features (e.g. parts of learned patterns). The CNN model receives their inputs through fixed connections from CNN models in the previous layers and are responsible for a robust pattern recognition (i.e. decreasing

sensitivity to a deformation or location shift of a pattern). Each CNN model receives inputs from a number of CNN models ('the input window') in one cell plane, i.e. from CNN models that extract the same feature, but at different spatial locations. By normalizing and thus reducing the range of values of the input, the covariant shift is reduced. As batch normalization can be utilized between hidden layers, it also reduces the need for the current layer to adjust to the previous one. Each layer is made slightly more independent, since a hidden layer no longer needs to handle varying ranges coming from the previous layer. It can be viewed as an input layer, where the input features are the output of the previous layer. It would gain the same advantages of normalization as the input layer. This has shown to speed up learning even further.

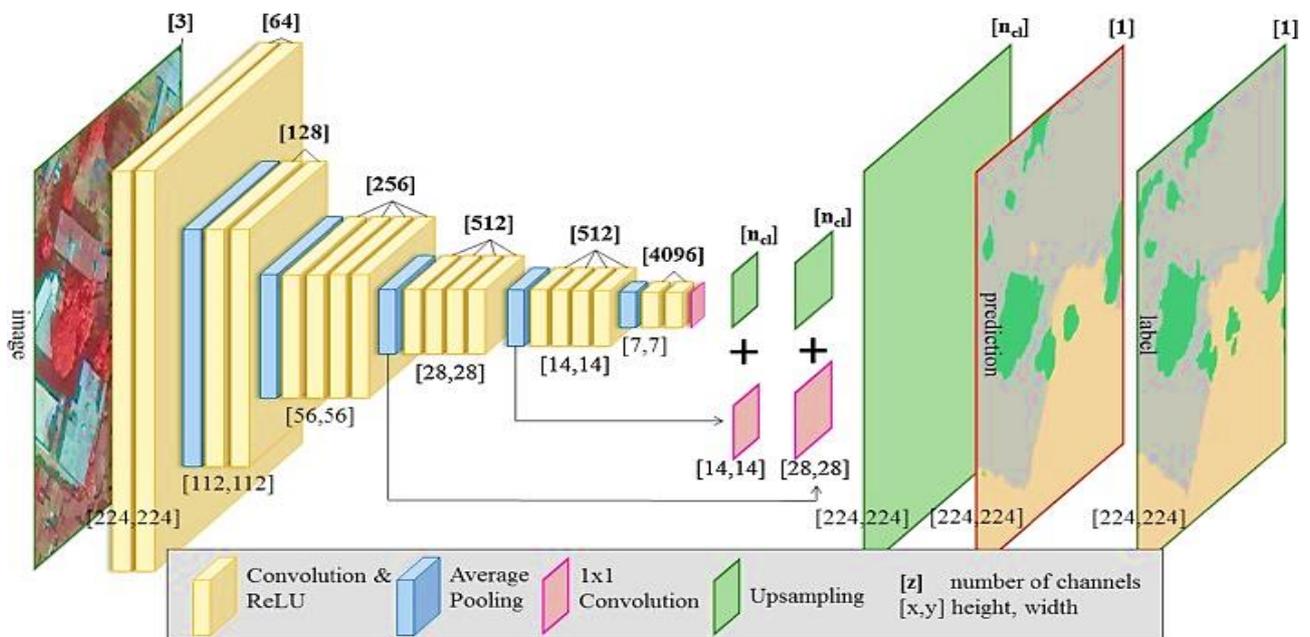


Figure 3. A CNN architecture for processing the satellite images of fields

The models will fire, if it receives an input from at least one of the models. Consequently, the feature will be detected even under a shift in the location of the input features, rendering the system less sensitive to the exact feature locations. The behavior of the models can however also be interpreted from an alternative point of view. The input windows for the different models strongly overlap thus models can be considered as performing a spatial blur on the excitatory signals they receive from the models. This spatial pooling is obtained by averaging these signals from the input window, which are models that perceive the same feature at slightly different locations. Furthermore, the excitatory cell input window is often framed by a small inhibitory region. Furthermore, the training of the aforementioned deep CNNs generally required multiple GPU's and was a time-consuming process (up to 3 weeks). Thus, it is clear that training the same, high-performance CNN architectures on real-life data is in most cases unfeasible. The CNNs trained to recognize the ImageNet dataset have nevertheless a strong capability to extract very distinctive features from natural images, rendering them useful for datasets of other images.

2.1. Convolutional neural network (CNN)

A convolutional neural network does not necessarily only consist of convolutional layers. Even though using kernels has reduced the number of parameters substantially, it is often useful to reduce the parameter count even further. Thus downsampling is used. Downsampling can be done in many different ways. We can reduce the output image simply by making the stride bigger than 1. A natural downsample happens if padding is not utilized and also techniques such as dilation can reduce the output image.

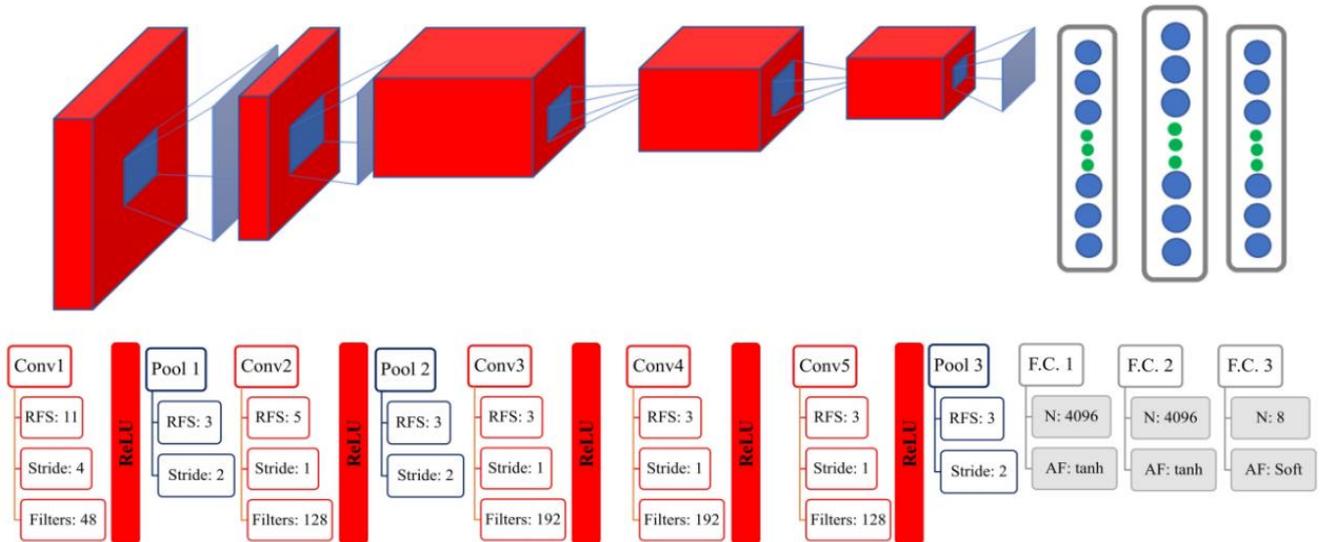


Figure 4. Representation of basic building blocks of convolutional neural network

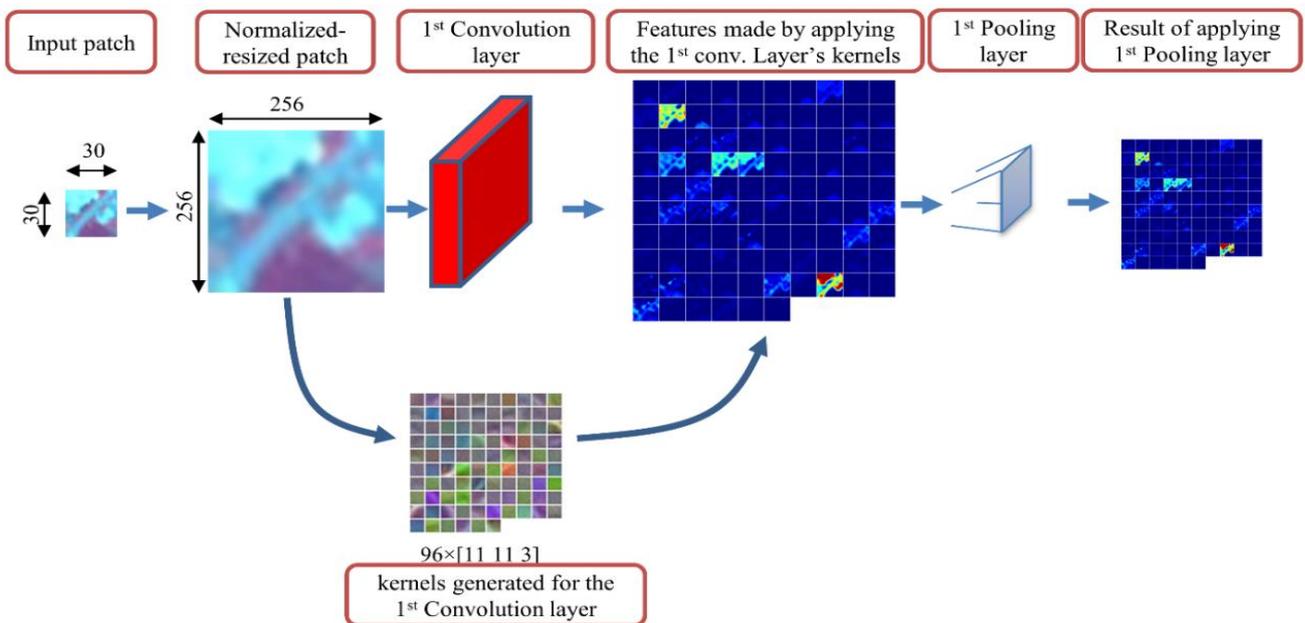


Figure 5. Load over deflection, where the blue points represent elastic mode, and after 100 mm of deflection, the beam entered in the plastic mode (orange points)

The transferability of different levels of features was touched upon in this work with results confirming that the deepest layers are most task-specific and better results can be obtained by using features from levels upwards in the model. This was confirmed by the work, in which furthermore was suggested that pooling of features as well as combinations of features may improve performance.

2.2. Mathematical modeling and feature extraction using CNN

With the extracted convolutional neural network (CNN) feature map for satellite images, as we computed the local visibility and maps of consistency for the determination of weight map for multi-exposure satellite image analysis where the image segmentation mathematical modeling is being acquired from [Chen et al., 2018]. The CNN cannot generate the output of the MEF; rather than that, it exploits characteristics of pre-trained CNNs for the calculation of weight maps for simple and sufficient MEFs.

Let $L_i, i = 1, 2, 3, \dots, K$ be a set of K images of multi-exposure.

The feature map of every one of the source images is obtained with the use of the following equation:

$$F_i(x, y) = CNN(I_i)(x, y), \tag{1}$$

CNN(.) represents a pre-trained deep NN at particular layer. For a pixel (x, y) , it is possible obtaining a feature vector, and the dimension of that vector is the number of filters that have been utilized in that convolutional layer.

Measuring pixel $l_i(x, y)$ visibility as the L_1 normalization of $V_i(x, y)$:

$$V_i(x, y) = \|F_i(x, y)\| \tag{2}$$

The feature vector strength $V_i(x, y)$ may be utilized for the determination of pixel $l_i(x, y)$ visibility.

Motion detection may be implemented easily through the calculation of Euclid distance of 2 feature vectors that have been normalized. Denote by $\bar{F} = \frac{F}{\|F\|}$ the L_2 normalized $F(x, y)$ feature for removing the effect of the difference of exposure.

The distance of 2 feature vectors at a pixel of 2 can be calculated with the use of the following formula:

$$s_{ij}(x, y)^2 = \|\bar{F}_i(x, y) - \bar{F}_j(x, y)\|_2^2 \tag{3}$$

$s(x, y)^2$ measures similarity of two vectors. A smaller $s(x, y)^2$ represents stronger temporal consistency.

Gauss kernel has been utilized for mapping similarity between $\bar{F}_i(x, y)$ and $\bar{F}_j(x, y)$ to the range $[0, 1]$ using the following equation:

$$S_i(x, y) = \sum_{j=1}^K \exp \frac{-s_{ij}(x, y)^2}{2\sigma^2} \tag{4}$$

Where, σ represents a constant number and has been assigned the value of 0.05. A bigger weight has to be assigned to pixels that have temporal consistency. With similarity and visibility weight maps S and V , it is possible getting the ultimate weight map $W_i(x, y)$ as follows:

$$W_i(x, y) = \frac{V_i(x, y) \times S_i(x, y) \times M_i(x, y)}{\sum_{j=1}^K V_i(x, y) \times S_i(x, y) \times M_i(x, y) + \alpha} \tag{5}$$

where $M_i(x, y)$ represents the mask of exposure that is calculated on the intensity of the pixel. For the sake of avoiding the division by 0, a small coefficient α is added to the value of 10^{-10} . The commonly utilized mask is the hat function which can be represented with the use of the following equation:

$$W_i(x, y) = \begin{cases} 1, & \beta < I_i(x, y) < 1 - \beta, \\ 0, & \text{else,} \end{cases} \tag{6}$$

where $\beta \in [0, 1]$ represents the parameter that controls the quality of exposure when normalizing input images. Where $\beta = 0.2$ in the implementation. Finally, the images are fused, based on the following equation for the production of MEF output I_f :

$$I_f(x, y) = \sum_{i=1}^K I_i(x, y) \times W_i(x, y) \tag{7}$$

2.3. Experimentation and evaluation

A CNN stands for convolutional neural networks originated from this work whose network has a similar architecture cognition through which 70% data is used for training, 20% for testing and remaining 10% for validation. It translates the behavior of the models in mathematical formulations, respectively as convolutions and subsampling. The network primarily distinguishes itself however from the cognition in its training process. Whereas the neo-cognition relies on a layer-wise, unsupervised training process of the cell layers and supervised training of the output layer, CNN is trained to find a global minimum over all the parameters through back-propagation. An implemented method that is very similar but uses a more involved filtering strategy. Among all connected components in the pixel's not classified as boundary, those whose mean image intensity is above a threshold are defined as nucleus. The threshold can be learned using the training data. We show the graphs for

the CNN algorithm on satellite images from which we learned a threshold of 50 for this purpose using an 8-bit encoding that yields values ranging from 0 to 255.

This makes it difficult to create robust algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite images can show even stronger intra-class variations. More details concerning every one of the sequences, like name, number of source images, and spatial resolution. Amongst the 8 sequences of test, two of them have been considered in the moving object scenes and the others have been considered in the static scenes. In addition to that, the first 5 sequences. The levels of exposure are set manually according to the camera tools. For uncontrolled outdoors environment, moving fields such as, trees moving due to wind, make acquiring well-aligned sequences quite a challenge. However, all the images in each dataset were registered in order to align the frames to be in the same angle and direction, so the content of each image will not vary according to the used measure. Therefore, this will be not variable in our experiments. The dataset can be downloaded² from the link

6.1. Dataset description

The dataset has been acquired from an open-source repository known as the airbus satellite image dataset. The physical appearance (shape, size, texture, color, etc.) of fields in images can vary depending on many physical and human geographical factors, both from the ground perspective as well as in satellite imagery. Possible sources of variation include the cultivated species, plant status, topography, soil properties, weather conditions, cultivation methods and other human influences. Due to such local variations, the task of accurately differentiating between several field fields can become ambiguous. This makes it difficult to create robust algorithms that recognize all kinds of field instances without overfitting to the specific scene. In addition to the variation within a single scene, satellite images can show even stronger intra-class variations. More details concerning every one of the sequences, like name, number of source images, and spatial resolution. Amongst the 8 sequences of test, two of them have been considered in the moving object scenes and the others have been considered in the static scenes. In addition to that, the first 5 sequences. The levels of exposure are set manually according to the camera tools. For uncontrolled outdoors environment, moving fields such as, trees moving due to wind, make acquiring well-aligned sequences quite a challenge. However, all the images in each dataset were registered in order to align the frames to be in the same angle and direction, so the content of each image will not vary according to the used measure. Therefore, this will be not variable in our experiments. The dataset can be downloaded from the link⁵.

Table 2. Details of dataset images that were used for processing

Number of Images	Standard Image Size	Image Format
700	1921 x 1080 px	JPEG (.jpg)

3. Results and discussion

The goal of each satellite image agriculture field-wise illumination estimation experiment is to obtain a set of features for each image that describes its sequence. This set of features is called the image profile. Image profiles can be analyzed and compared to each other. In an agriculture field-wise illumination estimation experiment for example, illumination estimation of images that were treated can be compared against profiles of images in the control set to quantify important matrix changes using CNN. For example, image patches can reveal an image sequence state or can be used for classification in image states such as phases of the image cycle or hematopoietic differentiation. Agriculture field-wise illumination estimation using CNN can be of very different kinds. Examples include expression profiles that quantify the transcription of genes and morphological profiles that quantify the shape of the image and its compartments. Agriculture field-wise illumination estimation is an

⁵: <https://www.intelligence-airbusds.com/satellite-image-gallery/>

important tool in morphological profiling and captures the images used to obtain morphological image profiles. In this experiment, the original datasets were used, without adding any further effects on them. Moreover, in order to produce the proposed image, a combination of all of the used quality measure were implemented together.

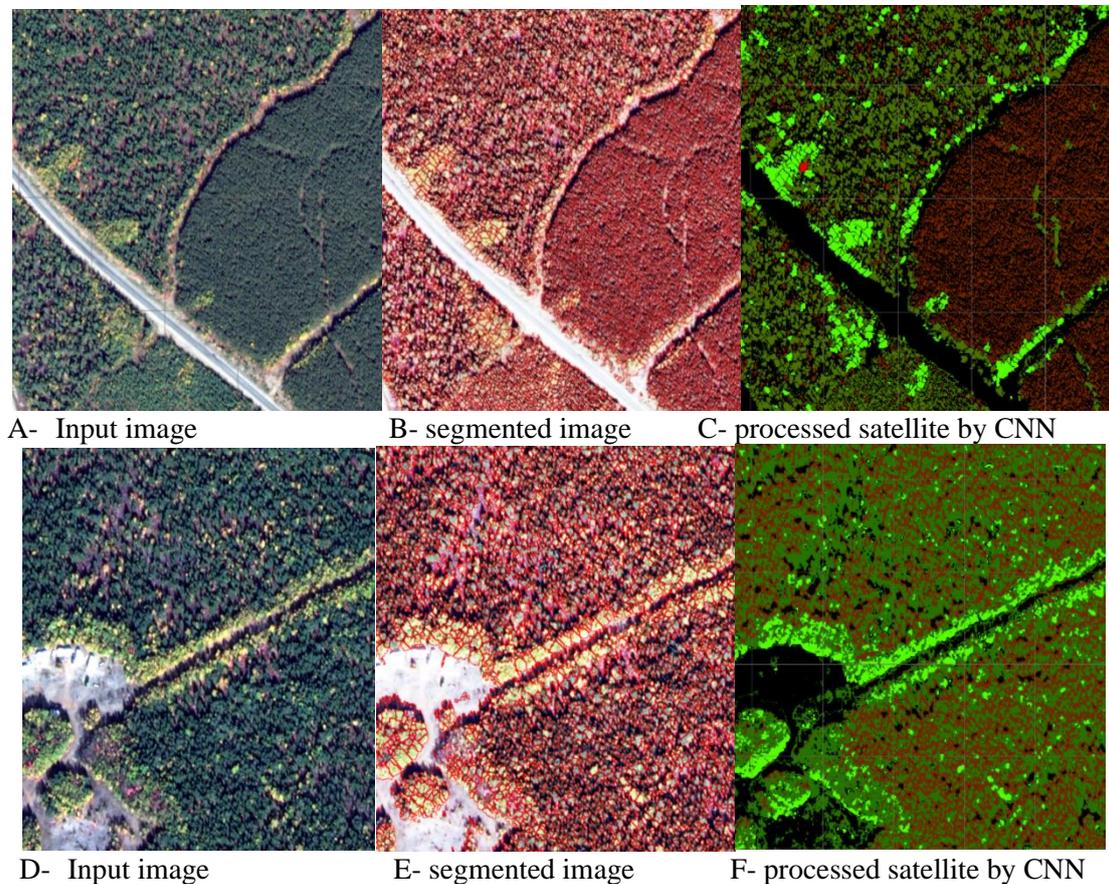


Figure 6. The input image, segmented image and processed satellite-based imagery for fields using convolutional neural network

Table 3. Summary of training and testing with respect to number of epochs

Number of Epochs	Loss	Accuracy	Validation Loss	Validation Accuracy
25	0.3301	0.8492	0.6911	0.5360
50	0.2418	0.9047	0.6942	0.5360
75	0.1891	0.9350	0.7985	0.5360
100	0.1702	0.9426	0.9698	0.5379
125	0.1384	0.9606	0.4709	0.7803
150	0.1246	0.9625	0.3569	0.8220
175	0.1020	0.9768	0.3543	0.8295
200	0.0809	0.9829	0.3598	0.8371
250	0.0747	0.9853	0.3662	0.8295
300	0.0678	0.9867	0.3709	0.8314

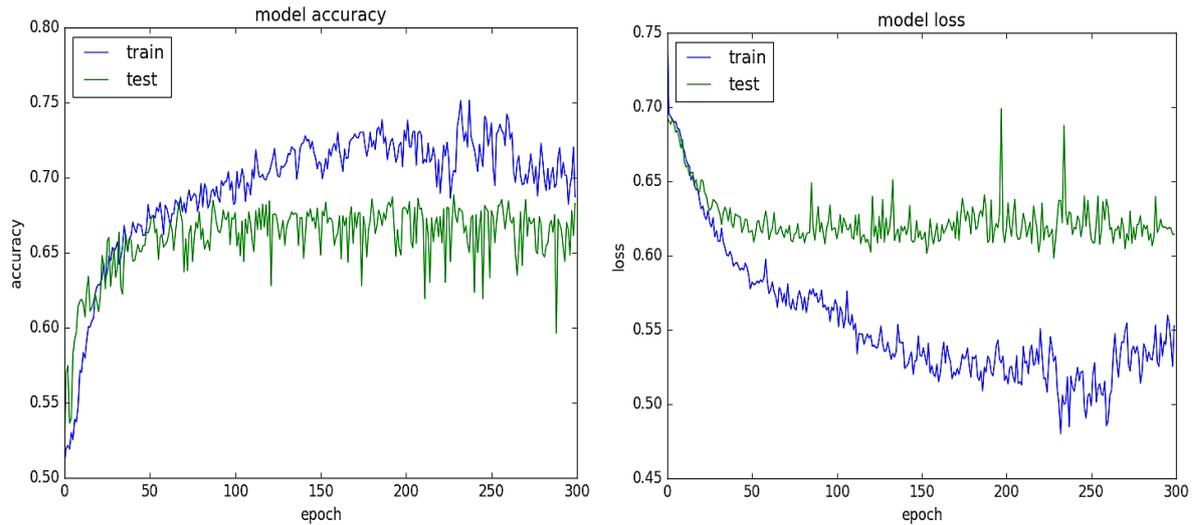


Figure 7. The CNN model accuracy (accuracy vs epochs) model loss (loss vs epochs)

This work has shown the impact of satellite image algorithmic approximation to the quality-of-control for image-based control systems for smoothed agriculture field detection. For the analysis, we used the use-case of a lateral controller that performs lane keeping for an autonomous field. The application that performs the image signal processing, lane detection and control computations, with the programmed in the domain specific language. The simulations were run on the CNN simulator, using a field and two separate tracks, namely a straight and a curved track. The application was made error-resilient by adapting the lane detection algorithm to be able to process the degree of approximation that the different approximate ISP versions required. We ran this application on an Intel i7 processor. We, initially, benchmarked the application by conducting careful and detailed profiling on a total of 8 different pipeline versions, each of which on a dataset of 200 images that were obtained in CNN. Then, we evaluated the degradation in quality-of-control that is caused by approximation, without considering the impact of approximation on runtime performance. Finally, the runtime performance gains due to approximation was considered and the improvement in quality-of-control was quantified. We used two separate metrics, namely the settling time and sum of squared errors, to evaluate the performance of the controller. Additionally, we used the energy, memory footprint and sensor-to-actuator delay to measure the improvement in the application performance. A possible explanation for the observations on the maximum length of the synthetic sequences could come from the class sequence length distribution with SVM algorithm for segmentation achieving the 96.10% of all original sequences are shorter than 200 frames, there are outlier sequences that are much longer and skew the arithmetic mean class sequence length for field segmentation. As the length of the synthetic samples is based on an RNN algorithm segmentation that makes use of these statistic, the generator might tend to produce longer sequences that help the classifier to better generalize to these rarer inputs from RNN network. When clipped, this effect cannot be fully exercised by the generated samples. The smaller the number of training samples, the smaller is the number of original outliers, which increases the positive impact. Interestingly, the maximum validation accuracy is better for the multi-class SVM when the synthetic samples are clipped to 100 instead of 200 frames for field segmentation. In this case, the data emphasizes the inputs that tend to be shorter than the average. Nevertheless, these effects might be rather random statistic correlations and it is unlikely that the positive impact comes from the varying length of the samples alone.

Table 4. The comparison of proposed work with literature that is already available

Article	Technique	Accuracy/Precision
[26]	SVM	96.10%
[27-30]	RNN	89.76%
Proposed	CNN	98.17%

Using this approach, we shed light on how real-time CNN algorithm can benefit from approximate computing. The approximated CNN pipelines achieved a maximum speedup of factor 3.5 for the sensor-to-actuator delay

and a 40% improvement in settling time. The overall performance that was achieved by the multi-dimensional application was 50% better than the baseline.

4. Conclusion

This advance research focuses on multi-exposure satellite image using image analysis and deep learning techniques for agricultural fields. The deep learning challenge originates from the accuracy of the ground truth data sets that are used as validation and/or training data for automated image analysis algorithms for agricultural fields. These datasets usually stem from manual tracing of field boundaries. However, manual tracing of satellite image is heavily dependent on the used imagery and supplementary data. It also is a highly subjective task, inevitably leading to inaccuracies and ambiguities depending on the priorities of the operator. The ground truth dataset can be complemented by existing property parcel information that could be indistinguishable based on the imagery alone. To exploit the pre-trained CNN for image edge smoothening of the different datasets, a top model is to stack on the deep learning with an accuracy of 98.17%. We explore three different top models in this thesis to maximally make use of the labeled, supervised data. A first and obvious choice is a mere fully connected layer, applied on the flattened CNN-output with added dropout. Considering the high parametric demand this top-model places on the learning process, we opted to also investigate performance with residual blocks as top model. Not only are these residual blocks hypothesized to facilitate training by learning the residual mapping, but they also decrease the number of parameters while allowing a multi-layered top model.

References

- [1] P. Fedor, I. Malenovský, J. VaOhara, W. Sierka and J. Havel, "Thrips (Thysanoptera) identification using artificial neural networks," *Bulletin of entomological research*, vol. 98, no. 5, pp.437-447, 2018.
- [2] M. Nilsback and A. Zisserman, "Automated Flower Classification over a Large Number of Classes," 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing, 2008, pp. 722-729, doi: 10.1109/ICVGIP.2008.47.
- [3] Z. Wang et al., "Efficient rail area detection using convolutional neural network," *IEEE Access.*, to be published. DOI: 10.1109/ACCESS.2018.2883704.
- [4] C. Wu et al., "A Greedy Deep Learning Method for Medical Disease Analysis," *IEEE Access.*, vol. 6, pp. 20021-20030, 2018.
- [5] G. Liang et al., "Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell Image Classification," *IEEE Access.*, vol. 6, pp. 36188-36197, 2018.
- [6] D. Hu et al., "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access.*, vol. 6, pp. 38303-38314, 2018.
- [7] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards realtime object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
- [8] A. Shrivastava, A. Gupta and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 761-769.
- [9] X. Zhang et al., "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks," *IEEE Access.*, vol. 6, pp. 30370-30377, 2018.
- [10] C. Xie, R. Wang, J. Zhang, P. Chen, W. Dong, R. Li, T. Chen and H. Chen, "Multi-level learning features for automatic classification of field crop pests," *Computers and Electronics in Agriculture.*, vol. 152, pp. 233-241, 2018.
- [11] I. Y. Zayas and P. W. Flinn, "Detection of Insects in Bulkwheat Samples with Machine Vision," *Transactions of the ASAE*, vol. 41, no. 3, pp. 883, 2018.
- [12] P. J. D. Weeks, M. A. O'Neill, K. J. Gaston and I. D. Gauld, "Species-identification of wasps using

- principal component associative memories,” *Image and Vision Computing*, vol. 17, no. 12, pp. 861-866, 2019.
- [13] J. Cho, J. Choi, M. Qiao, C. W. Ji, H. Y. Kim, K. B. Uhm and T. S. Chon, “Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis,” *Red*, vol. 346, no. 246, pp. 244, 2017.
- [14] C. Wen, D. E. Guyer and W. Li, “Local feature-based identification and classification for orchard,” *insects*. *Biosystems engineering*, vol. 104, no. 3, pp. 299-307, 2019.
- [15] X. Li, S. Huang, M. Zhou and G. Geng, “KNN-spectral regression LDA for insect recognition,” in *Information Science and Engineering (ICISE), 2019 1st International Conference on*, 2019, pp. 1315-1318.
- [16] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [17] R. Girshick, J. Donahue, T. Darrell and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 580-587.
- [18] Y. Shen, H. Zhou, J. Li, F. Jian and D. S. Jayas, (2018). “Detection of stored-grain insects using deep learning,” *Computers and Electronics in Agriculture*, vol. 145, pp. 319-325, 2018.
- [19] J. Chen, H. Shao and C. Hu, "Image segmentation based on mathematical morphological operator", *Colorimetry and Image Processing*, 2017.
- [20] S. K. Abdulateef, S. R. A. AHMED, & M. D. Salman, "ANovel Food Image Segmentation Based on Homogeneity Test of K-Means Clustering," In *IOP Conference Series: Materials Science and Engineering*, vol.928, no.3, p.032059, 2020.
- [21] M. Ahmed, et al. An Expert System to Predict Eye Disorder Using Deep Convolutional Neural Network. *Academic Platform Journal of Engineering and Science*, 9.1: 47-52.
- [22] S. Ahmed, et al. Breast cancer detection and image evaluation using augmented deep convolutional neural networks. *Aurum journal of engineering systems and architecture*, 2.2: 121-129, 2018.
- [23] S. Ahmed; SONUÇ, Emrullah. Deepfake detection using rationale-augmented convolutional neural network. *Applied Nanoscience*, 1-9, 2021.
- [24] M. MAHMOOD, S. AHMED, M. Ahmed. Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism. In: *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. p. 022027.
- [25] M. WALEED, A. ABDULLAH, S. Ahmed. Classification of Vegetative pests for cucumber plants using artificial neural networks. In: *2020 3rd International Conference on Engineering Technology and its Applications (IICETA)*. IEEE, p. 47-51, 2020.
- [26] L. Shi, Q. Duan, X. Ma and M. Weng, "The Research of Support Vector Machine in Agricultural Data Classification", *IFIP Advances in Information and Communication Technology*, no. 1, pp. 265-269, 2011.
- [27] E. Ndikumana, D. H. T. Minh, N. Baghdadi, D. Courault and L. Hossard, "Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue France", *Remote Sens.*, vol. 10, no. 8, pp. 1217, 2018.
- [28] S. B. Tambe and S. S. Gajre, “Cluster-based real-time analysis of mobile healthcare application for prediction of physiological data,” *J. Ambient Intell. Humaniz. Comput.*, vol. 9, no. 2, pp. 429–445, 2018, doi: 10.1007/s12652-017-0562-9.
- [29] S. A. Shawkat, K. S. L. Al-Badri, and I. Al Barazanchi, “Three band absorber design and optimization by neural network algorithm,” *J. Phys. Conf. Ser.*, vol. 1530, no. 1, 2020, doi: 10.1088/1742-6596/1530/1/012129.
- [30] W. A. Oktaviani, T. Barlian, A. Muttaqin, I. Al, and I. Ahmad, “Particle Swarm Optimization Algorithm Based Optimization of Steam Turbine Generator and Gas Turbine Generator at Pt . Sriwidjaja Fertilizer,” *J. Telecommun. Control Intell. Syst.*, vol. 1, no. 1, pp. 29–35, 2021, doi: 11.11111/jtcis.xxxx.