# Modified Convolutional Neural Networks (CNN) with MobileNet-v2-186 Architecture in Crescent Moon Image Detection

2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) | 979-8-3315-7930-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICIMCIS68501.2025.11326934

1st Muhammad Habib Yuhandri
*Department of Information Technology*
*Universitas Putra Indonesia YPTK*
Padang, Indonesia
mhabibyuhandri@upiyptk.ac.id

2nd Yuhandri
*Department of Information Technology*
*Universitas Putra Indonesia YPTK*
Padang, Indonesia
yuyu@upiyptk.ac.id

3rd Sumijan
*Department of Information Technology*
*Universitas Putra Indonesia YPTK*
Padang, Indonesia
sumijan@upiyptk.ac.id

***Abstract*—The use of image processing in rukyatul hilal (crescent moon observation) is highly important, since the observed crescent images often suffer from quality degradation. This indicates the presence of defects or noise, resulting in excessively high contrast and image blurring. Digital image processing operations are essential for addressing challenges encountered in the observation of crescent moon imagery. This study aims to detect the appearance of crescent moon imagery in real time by developing and modifying Convolutional Neural Networks (CNN). The development and modification of the CNN model were based on a series of experiments, including the adaptation of the architecture using MobileNet-v2-186. The research dataset consisted of 27 crescent moon observation videos, comprising a total of 5,241 training frames and 1,042 testing frames. Experimental results indicated that the MobileNet-v2-186 model achieved an accuracy of 99.6835%, a precision of 82.1152%, a recall of 82.2122%, and an F1 score of 80.5725%. This development was undertaken to enhance the detection accuracy of crescent moon images, which exhibit extremely subtle visual characteristics and are difficult to identify manually.**

***Keywords—Image Processing, Detection, Crescent, Convolutional Neural Networks (CNN), MobileNet-v2-186.***

## I. INTRODUCTION

Astronomy is a scientific discipline that studies and observes the motion of celestial bodies and natural phenomena occurring beyond Earth's atmosphere [1]. These celestial bodies include the Sun, planets, Moon, stars, meteoroids, asteroids, and others [2]. Additionally, astronomy encompasses the investigation of past events related to the origin and formation of the solar system and the universe [3]. In Indonesia, the study of astronomy fundamentally falls under the duties and responsibilities of a Non-Ministerial Government Institution (LPND) known as the Meteorology, Climatology, and Geophysics Agency (BMKG). The BMKG's scope of responsibilities encompasses the fields of meteorology, climatology, air quality, and geophysics [4]. BMKG also holds a specific responsibility in the monitoring of the crescent moon (hilal) at the beginning of each month in the Hijri calendar [5]. Additionally, BMKG is authorized to provide scientific recommendations to relevant stakeholders in determining the commencement of the Hijri month [6]. The crescent moon (hilal) monitoring conducted by BMKG has now adopted image processing as a technical innovation in the practice of rukyatul hilal [7]. The application of image processing techniques forms part of the advancement of astrophotography methods, aimed at enhancing the clarity of lunar imagery [8]. Crescent moon images that are difficult or even impossible to interpret accurately through astrophotography alone can be more effectively analyzed through the application of image processing techniques [9]. The application of image processing in rukyatul hilal (crescent moon observation) is of critical importance, as the observed hilal images often suffer from quality degradation [10]. Digital image processing is typically employed to enhance image quality, making it more interpretable to the human eye, and to extract and process data within the image for the purpose of automated object recognition thus facilitating the interpretation of compromised crescent moon imagery [11]. Image processing is a technology designed to render images clearer, brighter, and cleaner without altering their original structure [12]. It cannot generate nonexistent objects while it can refine faint or low-quality imagery, it cannot reveal objects that are not present. Therefore, image processing cannot manifest what is beyond the realm of physical visibility [11]. Previous studies have indicated that Odeh formulated the criteria for crescent moon (hilal) visibility based on two key factors: the altitude of the hilal (ARCV) and its width (W). The altitude serves as a reference for the contrast threshold between the hilal and the Sun, while the width is used to determine the percentage of hilal illumination. Classification is conducted according to observer conditions, dividing hilal visibility into several zones. Ultimately, Zone E was established, which requires the application of image processing techniques when the hilal's altitude ranges between 6 and 8 degrees [13]. Other studies have demonstrated that artificial neural networks are capable of performing pattern recognition for predicting the visibility of the new crescent moon, achieving an accuracy rate exceeding 72% [14]. Similar studies have also described the process of identifying the appearance of the crescent moon using Computer Vision (CV) algorithms [9]. The implementation of Computer Vision (CV) has played a significant role in the interpretation of identification results, achieving a validation accuracy of 97.8% [15]. Further research has also presented the performance of Artificial Neural Networks (ANN) in determining the lunar phases within the Hijri calendar, yielding an accuracy of 72%. The critical role of the detection process, particularly through the adoption of the MobileNet semantic segmentation model, has been proven to be highly effective in enhancing detection capabilities. Building upon prior research, this study conducts real-time detection of the crescent moon (hilal) through the application of image processing techniques. The hilal detection process plays a pivotal role in determining the commencement of Ramadan and Shawwal, particularly within the Indonesian context. This is grounded in the fact that the determination of the crescent

moon (hilal) is traditionally conducted manually, leading to differing interpretations of religious texts (nash) concerning rukyatul hilal. In response to this issue, the present study aims to detect the appearance of crescent moon imagery in real time by developing and modifying Convolutional Neural Networks (CNN). The development and modification of the CNN are expected to enhance the accuracy of detection outcomes while introducing novel approaches to the detection process. This innovation is reflected in the architectural models produced through a series of experiments and in-depth analyses. The resulting detection outputs are anticipated to contribute to the determination of the beginning of Ramadan and Shawwal, thereby serving as a scientific reference for decision-making by the government, particularly the Ministry of Religious Affairs of the Republic of Indonesia.

## II. METHOD

The detection process of crescent moon (hilal) imagery in this study is conducted in real time through the development and modification of Convolutional Neural Networks (CNN). The detection workflow comprises several structured stages, beginning with the collection of hilal observation videos, followed by pre-processing, and culminating in the detection phase through the development of the MobileNet V2-186 architectural model. The preprocessing stage is designed to adjust video duration, extract frames from the video into image data, define the Region of Interest (ROI) for Ground Truth imagery, and resize the extracted frames. The detection process incorporates the development of the MobileNet V2-186 architectural model, representing a novel approach aimed at producing precise and accurate detection outcomes. The complete sequence of the detection workflow is illustrated in the research framework presented in Fig. 1.

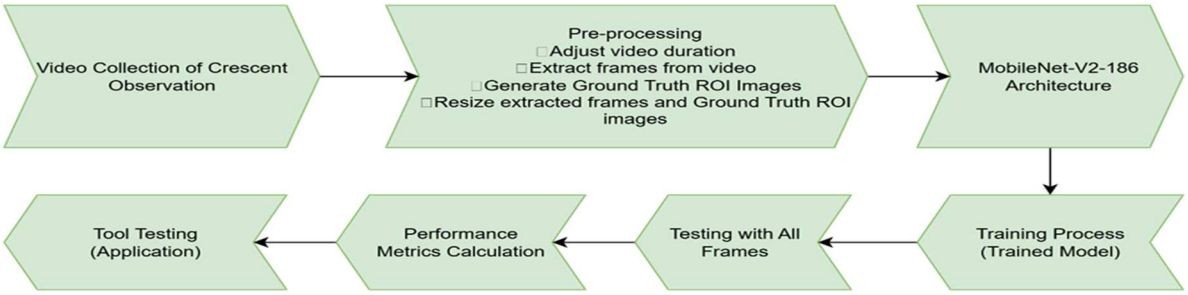


Fig. 1. Research Framework

Figure 1 illustrates that this study utilizes crescent moon (hilal) observation data provided by BMKG, accessible through 29 observation points each lunar month via the official website https://hilal.bmkg.go.id. A total of 27 hilal observation videos from the years 1443 AH to 1446 AH were collected, comprising 20 videos for training and 7 videos for testing. To strengthen the evaluation process, the test dataset was augmented with 25% of the training data, resulting in a total of 12 videos. The selected videos were then extracted into image frames, manually annotated with Regions of Interest (ROI) to mark the hilal, and resized to match the input specifications of the MobileNet-v2-186 architecture. The MobileNet-v2 architecture was modified into MobileNet-v2-186 to enable its use for semantic segmentation rather than mere classification. Adjustments were made to the number of layers, hyperparameters, and activation functions to enhance the model's sensitivity to faint and slender crescent moon (hilal) features. The training process utilized 5,241 resized frames along with corresponding ground truth annotations, employing techniques such as depthwise separable convolution, inverted residuals, and fine-tuning of pre-trained weights. Model performance was comprehensively evaluated using indicators including Accuracy, Precision, Recall, and F1-Score. The trained model was subsequently integrated into a crescent moon (hilal) detection application. This application enables users to upload observational videos and obtain automated real-time detection results. Testing outcomes indicate that the application operates optimally, efficiently, and is user-friendly, making it well-suited for practical implementation in hilal observation activities.

### A. Region of Interest Ground Truth (ROI GT)

Region of Interest Ground Truth (ROI GT) refers to manually annotated focal areas that are considered accurate representations, serving as a reference for validation purposes [16]. Accuracy evaluation in object detection or segmentation is commonly conducted at two levels: the Region of Interest (ROI) level and the pixel level [17]. ROI-level refers to the evaluation process in which each detected object such as bounding boxes or polygons is compared against the manually annotated Region of Interest (ROI) ground truth. This approach assesses the correspondence between predicted and reference regions to determine detection accuracy [18]. The ROI-level evaluation process typically employs techniques such as centroid matching, Intersection over Union (IoU), and confusion matrix pseudoregions in relation to the ground truth for pixel-level image analysis. Accuracy metrics such as precision and recall at the ROI level are more sensitive to structural errors in object detection compared to pixel-level evaluation [19].

### B. MobileNet-v2-186

MobileNet-v2 is one of the early projects that introduced a Convolutional Neural Network (CNN) architecture designed for ease of implementation and deployment across various platforms [20]. Depthwise separable convolution, as illustrated here, represents one of the key advancements in convolutional neural network architecture. This technique decomposes a single convolutional kernel into two distinct operations, enhancing computational efficiency while preserving representational capacity [21]. For instance, instead of using a 3×3 kernel, the convolution is decomposed into a 3×1 and a 1×3 kernel. This approach enables the scaling of input image dimensions from 224 to 128 pixels. Consequently, MobileNet-v2 can be trained on images sized 224×224 and later applied to images sized 128×128, owing to its use of global average pooling. Fundamentally,

MobileNet-v2 is a Convolutional Neural Network (CNN) architecture designed for image classification, in which the entire image is categorized into one of several predefined classes [22]. In this study, the MobileNet-v2 architecture was further developed into MobileNet-v2-186 to enable its application in image segmentation tasks. The modified MobileNet-v2-186 architectural model is presented in Fig 2.

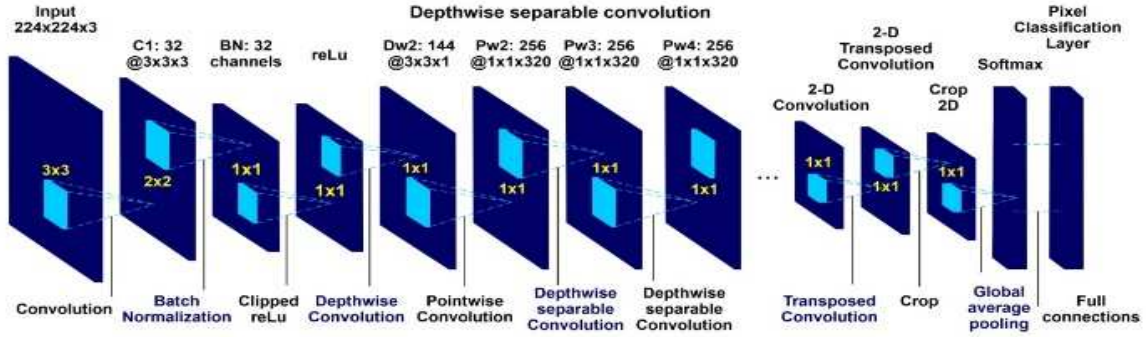


Fig. 2. MobileNet-v2-186 Architectural Model

Figure 2 illustrates that the original MobileNet-v2 model has been modified into the MobileNet-v2-186 architecture. This model presents the neural network workflow for performing semantic segmentation, which focuses on classifying each individual pixel within an image. The process begins with a 2D convolutional layer, responsible for extracting spatial features from the input image [23]. Subsequently, a 2D Transposed Convolution is applied to perform upsampling, thereby restoring the feature map to the original image resolution [24]. The 2D cropping process ensures that the output dimensions align precisely with the original input or reference data, thereby allowing direct comparison between the predicted segmentation and the pixel-level ground truth labels [25]. Subsequently, the upsampled features are processed through Global Average Pooling to condense the information into a simplified representation while preserving broader contextual relevance. This representation is then passed through a Fully Connected layer, which interfaces with the Pixel Classification Layer. Finally, a Softmax function is applied to assign class probabilities at the pixel level, enabling the model to generate a segmentation map that distinguishes regions according to the desired categories. A more detailed description of the MobileNet-v2 186 architecture is provided in Table 1 and a detailed comparison between the developed MobileNet-v2 186 architecture model and the original MobileNet-v2 is presented in Table 2.

TABLE I. DESCRIPTION OF THE MOBILENET-V2 186

| No | Layer | Operation Type | Filter | Output |
|---|---|---|---|---|
| 1 | Input Layer | Input Image | 224 × 224 × 3 | RGB (3 channels) |
| 2 | Conv1 | Convolution | 3 × 3 | 32 |
| 3 | BN1 | Batch Normalization | - | 32 |
| 4 | ReLU | Non-Linear Activation | - | 32 |
| 5 | Depthwise Conv 1 (Dw2) | Depthwise Convolution | 3 × 3 | 144 |
| 6 | Pointwise Conv 1 (Pw2) | 1 × 1 Convolution | 1 × 1 | 256 |
| 7 | Depthwise Conv 2 (Dw3) | Depthwise Convolution | 3 × 3 | 256 |
| 8 | Pointwise Conv 2 (Pw3) | 1 × 1 Convolution | 1 × 1 | 320 |
| 9 | Depthwise Conv 3 (Dw4) | Depthwise Convolution | 3 × 3 | 256 |
| 10 | Pointwise Conv 3 (Pw4) | 1 × 1 Convolution | 1 × 1 | 320 |
| 11 | Transposed Conv | 2D Transposed Convolution | 1 × 1 | - |
| 12 | Crop 2D | Cropping Layer | - | - |
| 13 | Global Average Pooling | Pooling | Global | - |
| 14 | Fully Connected Layer | Dense / FC | - | Mumber of classes |
| 15 | Softmax Output | Pixel Classification | - | Each pixel class |

TABLE II. COMPARISON ARCHITECTURE MODEL MOBILENET-V2 186 AND MOBILENET-V2

| MobileNet-v2 | | MobileNet-v2-186 | |
|---|---|---|---|
| Convolution | Input 224x224x3 | Convolution | Input 224x224x3 |
| Batch Normalization | C1:32 @3X3X3 | Batch Normalization | C1:32 @3X3X3 |
| Clipped reLu | BN:32 channels | Clipped reLu | BN:32 channels |
| Deptwise Convolution | reLu | Deptwise Convolution | reLu |
| Pointwise Convolution | Dw2:144 @3x3x3 Pw2: 256 @1x1x320 | Pointwise Convolution | Dw2:144 @3x3x3 Pw2: 256 @1x1x320 |
| Deptwise separable Convolution | Pw3: 256 @1x1x320 | Deptwise separable Convolution | Pw3: 256 @1x1x320 |
| Deptwise separable Convolution | Pw4: 256 @1x1x320 | Deptwise separable Convolution | Pw3: 256 @1x1x320 |
| | | Transposed Convolution | 2-D Convolution |
| | | Crop | 2-D Transposed Convolution Crop 2D |
| | | Global average pooling | Softmax |
| | | Full connections | Pixel Classification Layer |

Table 1 explains that the MobileNet-v2 186 architecture presented in the table describes a lightweight convolutional neural network (CNN) designed to perform efficient image classification and segmentation tasks, especially on mobile and embedded devices. The model structure emphasizes computational efficiency through the use of Depthwise Separable Convolutions, which significantly reduce the number of parameters and operations compared to standard convolutional networks. The network starts with an Input Layer that receives a 224 × 224 × 3 pixel image, representing RGB color channels. This is followed by a 3 × 3 Convolution layer (Conv1) that extracts basic features such as edges and textures. A Batch Normalization (BN) layer normalizes the output to stabilize and speed up the training process, while ReLU activation introduces non-linearity, allowing the network to learn more complex patterns. Overall, the update is a significant improvement over the previous MobileNet v2 architecture.

Table 2 presents a comparison of the MobileNet-v2 186 and MobileNet-v2 architecture models in the semantic segmentation process. The modification of MobileNet-v2 was carried out to optimize segmentation performance by manipulating the original MobileNet-v2 architecture into MobileNet-v2 186 based on the parameters presented in Table 2. The architectural modifications previously presented will be discussed further in this study regarding the detection of the crescent moon image.

## III. Result and Discussion

The object detection process for crescent moon (hilal) imagery was conducted by developing and modifying Convolutional Neural Networks (CNNs), utilizing a dataset derived from BMKG sourced observational videos. The dataset comprises 27 videos, which were extracted into a total of 6,283 image frames. These images were subsequently divided into 5,241 training images and 1,042 testing images. The representative forms of the extracted frames used in the detection process are presented in Figure 3.


Fig. 3. Extracted Image Frames

Figure 3 presents the image samples utilized in the crescent moon (hilal) object detection process, which involves modifications to the CNN architecture. These images are subsequently subjected to preprocessing, a necessary step to enhance image quality prior to the detection phase. The preprocessing procedure includes Region of Interest (ROI) extraction and image resizing, ensuring compatibility with CNN-based processing. The results of this preprocessing stage are shown in Fig. 4.


(a) Extracted Image Frame (b) ROI Output (c) Resized Output

Fig. 4. Preprocessing Results

Figure 4 presents the outcomes of the preprocessing stage employed in the crescent moon (hilal) image object detection process. Image (a) depicts a frame extracted from hilal observation videos, while image (b) shows the result of manual annotation performed by human experts to delineate the object boundaries or target areas, commonly referred to as ROI Ground Truth. Image (c) illustrates the resized output, conducted to ensure compatibility with the model's input requirements. In this study, standardization of image dimensions was necessary, and all images were resized to 240 × 424 pixels. Following the preprocessing stage, object detection was performed using a Convolutional Neural Network (CNN). The CNN architecture was modified based on MobileNet-v2 to enhance its performance, resulting in the MobileNet-v2-186 model. Initially, MobileNet-v2 was designed for global image classification; however, in this study, it was adapted for semantic segmentation to detect the presence of the crescent moon (hilal) within an image. This

required architectural modifications to address the specific challenges of the research. To support the segmentation process, MobileNet-v2 was not employed as a final classifier but rather as an encoder responsible for extracting features from the input image. MATLAB provides functions such as `deeplabv3plusLayers`, which facilitate the construction of segmentation systems using MobileNet-v2 as a backbone. This function automatically integrates additional network components, including a decoder, upsampling layers, and a pixel classification layer. These added layers are essential for reconstructing pixel label maps from the extracted features, thereby enabling precise final segmentation. The inclusion of pixel labeling and classification processes expanded the original MobileNet-v2 architecture from 153 to 186 layers. This added complexity is necessary for the network to perform semantic segmentation rather than simple classification. In this study, MobileNet-v2 serves as a feature extractor, enabling pixel-level identification of the hilal, which allows for more accurate localization of the object. The results of the CNN modification using MobileNet-v2-186, including the learning outcomes, are presented in Fig. 5.

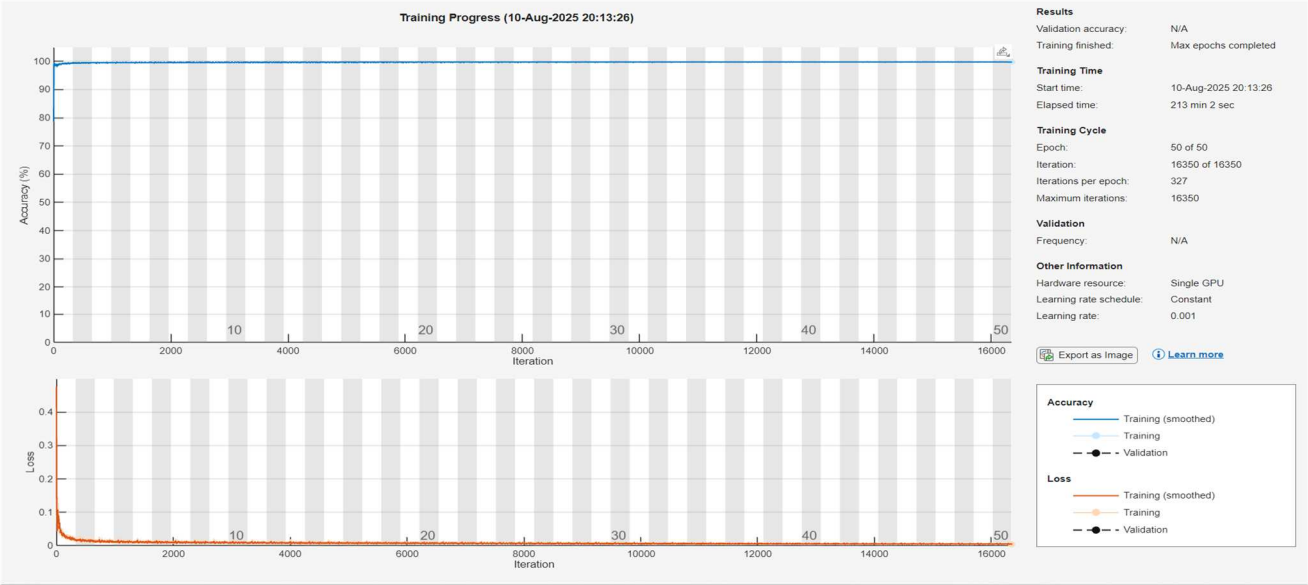


Fig. 5. Training Performance Graph of MobileNet-v2-186

Figure 5 illustrates the training outcomes of the previously modified MobileNet-v2-186 architecture. The training process was configured with 50 epochs, totaling 16,350 iterations (327 iterations per epoch), and employed a constant learning rate schedule of 0.001. This is evidenced by the validation metrics, including Accuracy, Precision, Recall, and F1 Score, evaluated on both the training and testing datasets. The detailed validation outcomes of the MobileNet-v2-186 training process are shown in Fig. 7.

| === Training === | | ===Testing === | |
|---|---|---|---|
| Accuracy | 99.6738% | Accuracy | 99.6835% |
| Precision | 84.5445% | Precision | 82.1152% |
| Recall | 85.4989% | Recall | 82.2122% |
| F1-Score | 84.0126% | F1-Score | 80.5725% |

(a) Validation of Training Results (b) Validation of Testing Results

Fig. 6. Validation Results of CNN-Based Detection Using MobileNet-v2-186

Figure 6 confirms that the performance of the CNN model modified with MobileNet-V2 186 achieved an accuracy of 99.68%, precision of 82.12%, recall of 82.21%, and an F1-score of 80.57%. These results indicate that the model demonstrates a fairly good capability in detecting the hilal during the detection process. The relatively high accuracy rate of 99.68% was obtained from the CNN learning outcomes in the semantic segmentation process. Furthermore, the evaluation results also show precision, recall, and F1-score values of 82.12%, 82.21%, and 80.57%, respectively. These precision, recall, and F1-score values are considerably lower compared to the accuracy value. This discrepancy is caused by the preprocessing results, which produce varying pixel values for each input image used in the CNN detection process. The performance results of the CNN modified with MobileNet-V2 186 in the Crescent Moon detection process can be compared with several previous studies to determine the effectiveness of the developments made. Previous studies have shown varying levels of success in the same or similar detection problems. The performance comparison is presented in Table 3.

TABLE III. TH COMPARISON OF MOBILENET V2 186 PERFORMANCE DEVELOPMENT RESULTS WITH PREVIOUS RESEARCH

| No | Previous Research | MobileNet-V2 186 Performance Results |
|---|---|---|
| 1 | The results of this study achieved a validation accuracy value of 97.8% by testing MobileNet-v2 on all models in image detection. | Performance of the CNN model modified with MobileNet-V2 186 achieved an accuracy of 99.68%, precision of 82.12%, recall of 82.21%, and an F1-score of 80.57% |
| 2 | The results show that GA-UNet achieves excellent performance with lung image segmentation: GA-UNet achieves an accuracy of 98.78%. | |
| 3 | Facial expression recognition accuracy is 68.62% and 95.96%, improving on the MobileNetV2 model by 0.72% and 6.14% respectively. | |

Based on the detection output of the modified CNN, it can be concluded that the development and modification of CNN with MobileNet-V2 186 have successfully provided optimal detection results. The CNN model with MobileNet-V2 186

also represents a novel approach in hilal object image detection by improving the accuracy of object identification. The findings of this study are expected to contribute to determining the beginning of Ramadhan and Syawal, particularly in Indonesia

## IV. Conclusion

The crescent moon (hilal) image object detection process, developed through the enhancement of CNN using the MobileNet-v2-186 architecture, has effectively addressed prior uncertainties by achieving an accuracy of 99.68%, precision of 82.12%, recall of 82.21%, and an F1-score of 80.57%. These results represent a novel advancement in CNN development, particularly in improving semantic segmentation performance for hilal detection. The integration of MobileNet-v2-186 into the CNN framework demonstrates its practical effectiveness and contributes to the reliability of information used in decision-making processes.

## Acknowledgment

This research forms part of the author's doctoral dissertation. The author wishes to express profound gratitude to Universitas Samudra and Universitas Putra Indonesia YPTK Padang for their invaluable support and contributions throughout the research process.

## References

[1] T. KHUZEY, "Pengaruh Aspek Astronomi Terhadap Terjadinya Perubahan Cuaca Dan Iklim Berdasarkan Pengamatan Badan Meteorologi, Klimatologi Dan Geofisika Kelas 1 Bandung," *Karya Tulis*, 2019.

[2] D. A. Prameswari, U. Karimah, E. P. Novitasari, and W. Kurniawati, "Understanding Astronomy: Characteristics Of Celestial Bodies As Knowledge In Science Learning," *Educ. J. Educ. Multimed.*, vol. 2, no. 01, pp. 98–104, 2024.

[3] V. V Goncharuk and D. K. Goncharuk, "A Novel Explanation for the Arrangement of the Universe–Solar System–Planets–Earth. Part 3," *J. Water Chem. Technol.*, vol. 43, no. 6, pp. 433–447, 2021.

[4] K. Ain, S. Soelistiono, and R. Bugatti, "Design and build of IoT based weather observation system support for BMKG (Meteorology, Climatology, and Geophysics Agency) in new normal era," in *AIP Conference Proceedings*, AIP Publishing LLC, 2023, p. 40004.

[5] B. Taman, "Refined Guidelines for Selecting Hilal Observation Points in Tropical Regions: Insights from Bengkulu City," *Al-Marshad J. Astron. Islam dan Ilmu-Ilmu Berkaitan*, vol. 11, no. 1, pp. 35–62, 2025.

[6] M. S. A. M. Nawawi, M. S. Faid, M. H. M. Saadon, R. A. Wahab, and N. Ahmad, "Hijri month determination in southeast asia: An illustration between religion, science, and cultural background," *Heliyon*, vol. 10, no. 20, 2024.

[7] M. M. Iqbal, M. A. Habib, and F. Z. Ramdhani, "Evaluating the Feasibility of Yanbu'ul Qur'an Observatory in Menawan Kudus for Lunar Crescent Observation," *Al-Hilal J. Islam. Astron.*, vol. 7, no. 1, pp. 117–132, 2025.

[8] A. L. A. M. Nasir *et al.*, "Comparative Analysis of Image Processing Technique in Determining the New Crescent Moon Visibility," in *Journal of Physics: Conference Series*, IOP Publishing, 2024, p. 12004.

[9] J. A. Utama, A. R. Zuhudi, Y. Prasetyo, A. Rachman, A. R. S. Riadi, and L. S. Riza, "Young lunar crescent detection based on video data with computer vision techniques," *Astron. Comput.*, vol. 44, p. 100731, 2023.

[10] N. Anggraini, Z. Zulkifli, and N. Hakiem, "Modeling Ramadan Hilal Classification with Image Processing Technology Using YOLO Algorithm," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1462–1471, 2024.

[11] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artif. Intell. Rev.*, vol. 57, no. 1, p. 11, 2024.

[12] G. K. Ijemaru *et al.*, "Image processing system using MATLAB-based analytics," *Bull. Electr. Eng. Informatics*, vol. 10, no. 5, pp. 2566–2577, 2021.

[13] M. Y. Taher and F. M. Abdulla, "Evaluating the Development of the Crescent Visibility Criteria," *Iraqi J. Sci.*, pp. 555–566, 2024.

[14] Z. T. Allawi, "Crescent Moon Visibility: A New Criterion using Deep learned Artificial Neural-Network," *Iraqi J. Sci.*, pp. 2332–2343, 2024.

[15] M. Shatnawi, N. Almenhali, M. Alhammadi, and K. Alhanaee, "Deep learning approach for masked face identification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 19, no. 6, pp. 296–305, 2022.

[16] C. Karabağ, M. A. Ortega-Ruíz, and C. C. Reyes-Aldasoro, "Impact of training data, ground truth and shape variability in the deep learning-based semantic segmentation of HeLa cells observed with electron microscopy," *J. Imaging*, vol. 9, no. 3, p. 59, 2023.

[17] F. Wahid, Y. Ma, D. Khan, M. Aamir, and S. U. K. Bukhari, "Biomedical image segmentation: a systematic literature review of deep learning based object detection methods," *arXiv Prepr. arXiv2408.03393*, 2024.

[18] C. J. Belasso *et al.*, "Bayesian workflow for the investigation of hierarchical classification models from tau-PET and structural MRI data across the Alzheimer's disease spectrum," *Front. Aging Neurosci.*, vol. 15, p. 1225816, 2023.

[19] A. Kos, K. Majek, and D. Belter, "Where to look for tiny objects? ROI prediction for tiny object detection in high resolution images," in *2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, IEEE, 2022, pp. 721–726.

[20] A. Souid, N. Sakli, and H. Sakli, "Classification and predictions of lung diseases from chest x-rays using mobilenet v2," *Appl. Sci.*, vol. 11, no. 6, p. 2751, 2021.

[21] J. Zhong, J. Chen, and A. Mian, "DualConv: Dual convolutional kernels for lightweight deep neural networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 11, pp. 9528–9535, 2022.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[23] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 7570–7588, 2021.

[24] F. Sahito, P. Zhiwen, F. Sahito, and J. Ahmed, "Transpose convolution based model for super-resolution image reconstruction," *Appl. Intell.*, vol. 53, no. 9, pp. 10574–10584, 2023.

[25] J. Anandakrishnan, A. K. Sangaiah, H. Darmawan, N. K. Son, Y.-B. Lin, and M. J. F. Alenazi, "Precise spatial prediction of rice seedlings from large scale airborne remote sensing data using optimized li-yolov9," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2024