International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

HUMAN AND FACE DETECTION OF MULTI INTENSITY IMAGES USING DEEP LEARNING TECHNIQUES

Elukapelly Vijay Kumar[1], Dr. B. Prabhakar[2]
[1]M. Tech Student, [2]Professor, [1,2] Department of Digital Systems and Computer Electronics, Jawaharlal Nehru Technological University Hyderabad, University College of Engineering, Nachupally, Kodimial Mandal, Jagtial Dist, 505501, Telangana, India.

ABSTRACT
In standard infrared illuminators used in nighttime surveillance systems, inadequate lights can result in the misidentification of distant gadgets, at the same time as excessive illumination might cause overexposure of nearer objects. To solve these problems, we use the MI3 images data set; have created the use of multi-tee ir-iluminations (MIIR) as our scale for modern object detection techniques. First, we supply complete annotations for Mi3, because its existing floor reality is insufficient. Subsequently, we hire these more intensity of illuminated infrared movies to evaluate the diverse received items, namely SSD, Yolo, and faster R-CNN and mask R-CNN by exploring the effective range of different light intensities. The proposed approach can create a brand new trends for facial detection and items at different distances by incorporating the tracking system and providing a new fusion approach for different lighting levels to increase performance.

Key words: Face Detection, Object Recognition, Computer Vision, Image Preprocessing, Real-Time Detection

Received: 25-07-2025 Accepted: 29-08-2025 Published: 08-9-2025

1. INTRODUCTION

In the night watching of the video, there are usually challenges from the fluctuations of the surrounding lighting. The detection of intruders at a huge distance below insufficient lighting is demanding, but the recognition of objects in close range is difficult due to excessive exposure in intense light. IRLUMINATOR IR with more intensity was designed in [1] to at the same time solve the problems of underexposure and excessive exposure by providing often variable lighting intensity. Chan et al. [2] they then created a Mi3 database, which includes video sequences with variable brightness from various indoor and outdoor scenarios. Two types of ground truths are provided: people count and categorize Photolových pixels in the foreground that exclude any statistics of border boxes. Despite Mi3 showing stimulating effects, it requires strong prerequisites, such as the absence of the foreground in preliminary standing. In addition, the truths about the floor in the MI3 data set often combine numerous matters, such as a bag inseparable from an individual that holds it, while some ground truths are either incomplete or doubtful.

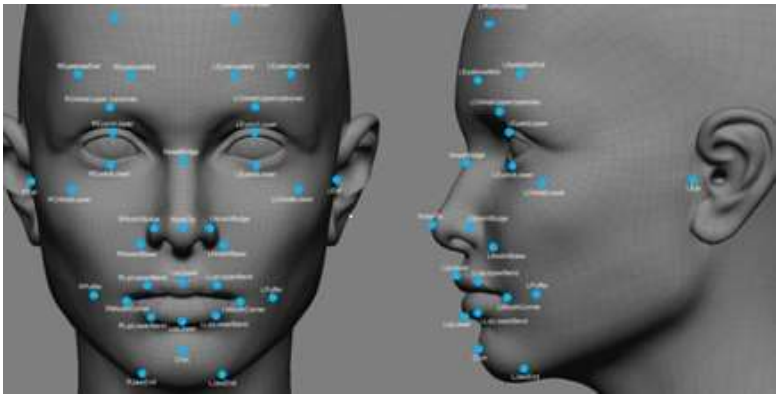

Figure 1. Face detection

In a "Gaussian mixture model (GMM)" is utilized for the popularity of foreground items in multi-intensity infrared movies. However, this strategy is generally inadequate for reliably dealing with complex foregrounds. Conversely, these prior studies simply illustrate qualitatively that superior image nice of distant (proximal) objects can be obtained with elevated (reduced) intensity tiers by multi-intensity lighting fixtures. These studies may also offer a quantitative assessment of the complimentary results throughout videos with varying illumination intensities, called channels. In mild of the increasing traits in utilizing deep learning for item identity, we are able to appoint MI3 as a benchmark dataset to evaluate item detectors for specific sceneries and lighting conditions.

2. LITERATURE REVIEW

Vision -based visual systems have a qualified increase in popularity. However, its functionality is significantly limited during midnight due to insufficient visibility due to insufficient lighting. General infrared (IR) Illuminators set up when used with night vision cameras, often lead to imaging problems with excessive exposure or the camera cannot capture proper exposure of objects positioned too near or distant from it. We utilize progressive infrared infrared illuminator technology to expand camera observation spectrum range and this document describes the development of MI3 databases (multi -phase -fed infrared lighting). A diverse set of video sequences with different intensity levels shows all interior and outdoor scenes in the database. The majority of research packages offer floor journals which include people counting along with foreground labeling. The respective algorithm obtains a specified effectiveness for performance assessment purposes. In the night video supervision, details of remote objects are often difficult to find out because of insufficient lighting, while the areas of nearby items may also seem washed out due to excessive exposure. In order to solve 2 problems in the middle of the night video supervision, we enforce a new multi -year infrared illuminator as an auxiliary light source that regularly provides different levels of lighting. The illuminator focused on the diverse intensity of lighting allows clean and nearby objects to be clear. This work proposes two methods for the automatic recognition of foreground objects at varying distances in image sequences recorded with a multi-depth infrared illuminator. Experimental outcomes indicate that each techniques attain an average accuracy of 90% in foreground object popularity, albeit with various computing difficulties.

The effectiveness of video monitoring is significantly reduced in low light conditions and insufficient visibility during the night. However, the requirement for night supervision is equivalent to the requirement of daily supervision due to the increased frequency of accidents that occur at night. Fixed intensity light source (IR) "works efficiently to a certain extent, which is the main to underexposure or excessive exposure problems when the item is placed too much or too close to power supply. Based on the License ID and finding, the design of a high dynamic range is

merged, which increases the visualization of the sightseeing marks even in the background.

In nocturnal video surveillance, enough illumination is crucial for picture exceptional. Conventional IR illuminators with set intensity frequently render distant objects hard to see due to insufficient illumination, at the same time as proximal items may also revel in overexposure, leading to subpar quality inside the image's foreground and historical past. This painting affords a revolutionary video summarizing technique that employs an modern multi-depth infrared illuminator to provide photos of human activities under varying illumination levels. Utilizing a GMM-based foreground extraction method for photos captured at each light level lets in for the choice and integration of foreground gadgets of the highest quality with a predetermined depiction of a static background. The outcome produces a coherent video precis for dynamic foregrounds, which is typically unattainable in nighttime surveillance footage. The object detection networks predict object positions through regional-based concept application. These detection networks operated at lower timescales when researchers unveiled the boundary limitations through innovation with SPPNET along with Rapid R-CNN. A "RPN) neural network merges convolutionary features with entire picture inputs while using detecting algorithm which reaches near-zero computational costs. RPN functions as a conventional network because it performs simultaneous boundary detection tasks along with item scoring activities in all target areas. RPN learns to identify goals using the end-to-end presentation strategy adopted by fast R-CNN for detecting targets. The RPN factor determines emphasis rules for United Network by using "attention" operations from modern neural network language while sharing convolutional functions to build a single combined network with RPN and fast R-CNN. Our detection method built with deep VGG-16 model processes three hundred image designs to achieve 5fps frame speed on GPU at the same time maintaining 100% accuracy against Pascal VOC 2007-2012 and MS Coco data sets. Multiple categories in both Ilsvrc and Coco 2015 competitions received their main winning awards through R-CNN and RPN. Introducing Yolo, innovative methodology to detect objects. Previous research of detection of objects adapts classifiers for detection tasks. We concept to detect items as a problem with regression with spatially different border boxes and the appropriate probability of the class. The unique nerve community predicts the boundary boxes and the probability of the class at the same time from the whole picture in one rating. The entire pipe detection capacity as a unique network that allows you to direct the end-to-end-based detection-based performance. Our included architecture is extraordinarily rapid. The foundational YOLO model analyzes snap shots in real-time at a rate of 45 frames in line with second. The compact version of the community, rapid YOLO, methods an impressive one hundred fifty five frames consistent with 2d even as achieving double the "mean average Precision (mAP)" of alternative real-time detectors. In comparison to detection systems, YOLO has, a higher frequency of localization mistakes however demonstrates a discounted propensity for generating fake positives in background contexts. Ultimately, YOLO acquires relatively generalized representations of gadgets. It surpasses alternative detection algorithms, such as DPM and R-CNN, in generalizing from natural images to other domain names, including artwork.

Multitude state features purportedly enhance the accuracy state "Convolutional Neural Networks (CNNs)". Empirical evaluation state feature combinations on extensive datasets, along with theoretical validation cutting-edge the consequences, is necessary. Certain features are limited to specific fashions and issues or are limited to small-scale datasets; conversely, features like batch normalization and residual connections are extensively applicable across most models, tasks, and datasets. We posit that these commonplace traits encompass "Dear residual connection (WRC), partial connection to movement (CSP), moving mini-dude normalization (CMBN), self-versat training (SAT)" and MISH activation. We use new features: “WRC, CSP, CMBN, SAT, MISH Activation, Mosaic Records Augmentation, Dropblock Regulation And Ciou Loss and integrate

many of them to achieve the results: 43.5% average accuracy (65.7% common accuracy in 50)" for MS Coco data file in second. We suggest a way of detecting objects in pictures using a unique deep neural network. Our Methodology, called SSD, discretizes the output space of a modern day limiting containers to the set modern starting fields in numerous elements and scale for each position of the function map. During the prediction, the community evaluation is issued indicating the probability of a top object in each default field and making a field adjustment to accurately agree to the item form. The community integrates predictions from the characteristic maps of North Ultra-modern resolution for efficient control of objects indicates a variety of size. The SSD model offers a simplified approach due to its network-based implementation, which omits all requirements for design generation and pixel determination as well as feature selection. The convenience of training SSD becomes better when addressing detection problems within specific structures. The dataset shows that SSD reaches higher accuracy than object detection techniques utilizing subsequent steps of the concept while demonstrating faster performance during training and evaluation processes. SSD achieves superior accuracy compared to one-level alternative methods both with short photographic duration. SDS reaches 58 frames during 300 × 300 entry on its way to obtain 72.1% of the average accuracy (MAP) at VOC2007 testing. In current years, deep networks have confirmed exquisite overall performance in laptop vision, in particular in face recognition. The performance state a deep community-based face popularity version is in the main inspired with the aid of interrelated factors: 1) The architecture state the deep neural network, and a pair of) the amount and quality modern day schooling information. In practical applications, versions in mild are a crucial thing that prtoday'soundly affects the efficacy latest face popularity systems. Deep network models can best gain the desired performance if there is a enough amount state training data consisting of various illumination intensities. Nonetheless, this latest training facts is hard to gain in real-world scenarios. This study addresses the difficulty modern illumination variation by proposing a deep network model that incorporates each visible light and near-infrared pics for face identification. Near-infrared imagery has extensively decreased sensitivity to illumination. Seen mild facial photos provide huge textural records that is surprisingly useful for facial identity.

Therefore, we develop an adaptive score fusion method that minimizes information loss, alongside the nearest neighbor set of rules for final type. The experimental results indicate that the model is highly effective in real-global circumstances and outperforms other models regarding illumination variation. Traditional night research of facial detection mainly uses mild cameras or thermal cameras near the infrared (NIR), which show resistance to fluctuations in environmental lighting and low lighting conditions. With the NIR digital camera, however, it becomes a modification of the depth and attitude of the additional Ilir in relation to its distance from the item. The thermal camera is very price used as a tracking device. As a result, we propose a night technique of facial reputation using deep learning using a singular sensor of visible light. In the night image for a lengthy distance, the facial detection is difficult to blur noise and image. Introducing a faster "Convolutional neural network based on the region (R-CNN) using images preliminary images through a histogram alignment (HE)". Our system identifies the body and facial areas in two stages, first then the locating face in a confined body area. While preserving accuracy in facial detection, our technique reduces the processing time of faster R -CNN. We have demonstrated that the suggested approach overcomes further existing facial detection methods using an independently assembled database named "DNFD-DB1": DNFD-DB1" along the available database from Fudan College. Using outstanding accuracy compared to those who lacked our pre-workout in the detection of the night face, the proposed step faster R-CNN gave the single faster R-CNN and our solutions.

3. PROPOSED SYSTEM

This article will explore one-stage detectors, including SSD and Yolov4, as well as phase detectors such as R-CNN and R-CNN mask. Due to various infra -red images, universal IR data file cannot be determined; As a result, the MI3 data file is tested in predetermined pre -pre -mentioned models of the above detectors to create a default value for a quantitative assessment of the impact of illumination of more intensity. For example, analysis of the reliability values of deep -based objects based on learning may indicate the desired wide range of lighting intensities essential for detection of objects in an extended distance range. In addition, we will define the optimal range in which large detection results can be achieved using one or several light intensities of more deep IR -IRIMINATOR lighting. A tracking methodology is introduced to enhance face detection outcomes, hence improving the F-degree of face detection.

A fusion method is presented to efficiently integrate data from several channels to enhance accuracy in object and face detection.

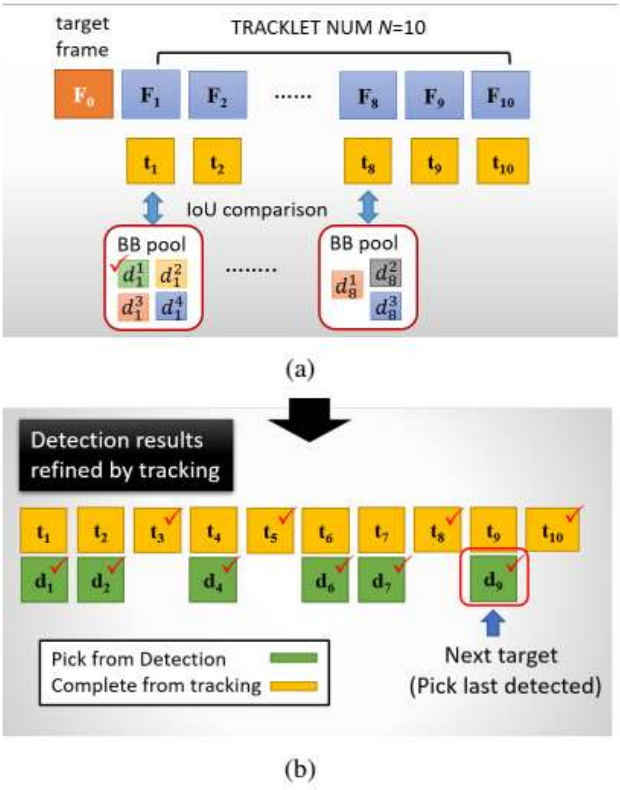

Figure 2. Proposed System architecture

YOLOV5 – YOLO This abbreviation operates only during the first usage period. Ultralitics released version 5 in June 2020 thus becoming the highest-evolved algorithm for object identification. A new fixed nervous network called CNN operates in real time to discover objects with superior accuracy levels. The algorithm distributes the whole image to a single nerve network that segments the image into smaller parts while simultaneously making predictions regarding the boundary boxes and component probabilities. The border boxes receive a weight based on their expected opportunity. The image "only once" is processed by neuron network predictions, which advance along a forward direction. The implementation of NE-MAX oppression following processing allows the detection algorithm to identify objects only once in the frame. Yolov8 - Ultralitics represents the most recent development of the Yolov 8 Yolo object detection model and image segmentation technology. The condition -off -Airt Sota model provides Yolov8 with all benefits from the previous versions while adding new features to

enhance both performance abilities and operational flexibility and efficiency. Yolov8 presents itself as a high-performance option for different AI functions because it focuses on speed, size and accuracy. The design includes Network and a new split-free head together with new damage capabilities.

Jolov3-Jolov3 A real-time object detection algorithm inside version 3 enables it to identify specific items in video, live channels or any images. The Yolo Machine Learning algorithm detects objects through a deep fixed nerve tight feature it retrieves. Yolo Joseph Redmon and Ali Farhadi developed the version 1-3 of Yolo Machine Learning, which the third version provides superior accuracy beyond the original ML algorithm.

MaskRNN - R-CNN mask stands today as one of the leading modern models, which built upon R-CNN. The firm neural network architecture of Rapid R-CNN uses this region as its base [2] to provide boundary pixels with each object and class label detection area. R-CNN works through an architecture that functions in a two-step process before discussing R-CNN mask and its modern model status.

During the first phase the system operates two networks known as spine (resane, VGG, incoming, etc.) and field design networks. The networks operate just once to generate field design for pictures. The dissolution of the area corresponds to the mapped area where an object exists.

The second phase determines boundaries for each potential area, which was proposed during the network process in Phase 1. The proposed areas can hold different dimensions from one another but fully connected layers need specific vector sizes. The proposed areas undergo size-fixing by means of ROI processing which resembles the MaxPooling approach or Royling mechanisms..

FasterRCNN – Faster R-CNN is a deep convolution network used to detect objects that seem to be the only end-to-end unified network. The net can accurately and quickly predict the location of different objects.

SSD – The SSD training system compares anchor boxes with their related dimensions to boundary boxes that represent each object present in the visual field during the matching phase. The prediction task belongs to the anchor box showing the most overlap with the analyzed object.

4. RESULTS

The figure 3 appears to be the homepage/dashboard of a machine learning or deep learning project interface. It could be a prototype for a web-based tool allowing users to access ear detection features, possibly for research or biometric purposes.

Figure 3. Welcome to Dashboard Interface

The figure 4 is a sign up form for new users to create an account on the system. Once filled and submitted, it would likely store user details in a database for authentication and later use. From the sequence of your previous image and this one, this form seems to be linked from the "Sign up"

link on the earlier dashboard homepage. This is a typical front-end UI for user registration in a locally hosted web application.

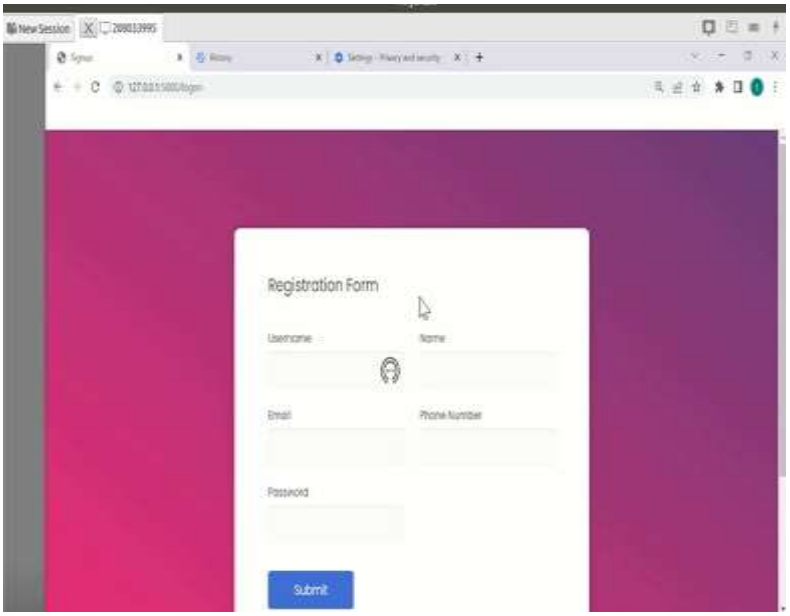

Figure 4: Registration Form Interface

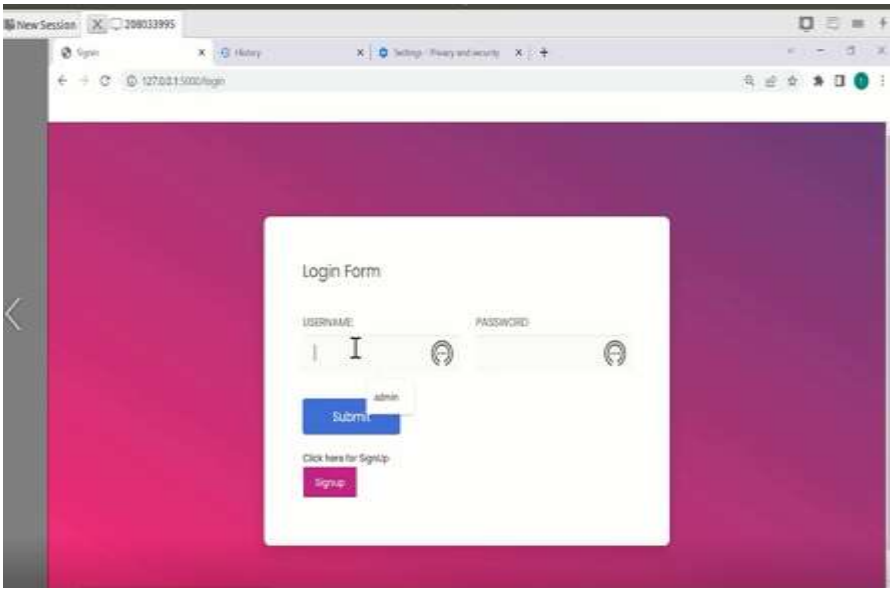

Figure 5. Login Form Interface

The figure 5 shows a Login Form web page hosted locally at http link. The background features a purple-pink gradient, and the form is centered in a white box. It has two input fields: USERNAME and PASSWORD, each with icons. Below, there's a blue Submit button to log in. A link saying *"Click here for Sign up"* with a pink Sign up button allows new users to register. The username field is shown with a suggestion drop down displaying "admin." This interface is likely part of a locally developed Flask or Django web application for user authentication.

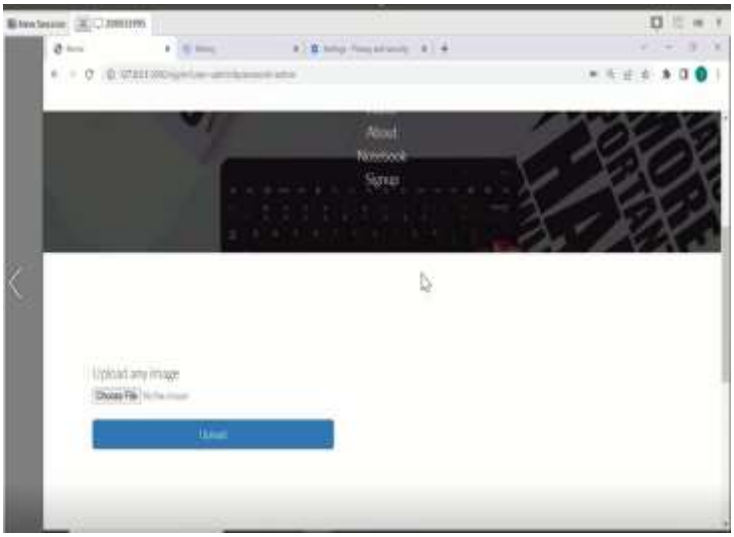

Figure 6. Upload Image File Interface

The figure 6 shows a webpage hosted locally at http. After a login attempt with credentials user=admin and password=admin. The top section retains part of the dashboard background with navigation links: Home, About, Notebook, and Signup. Below, there's an image upload interface with the label *"Upload any image"*. A Choose File button allows file selection, and a large blue Upload button submits it. No file is currently chosen. This page likely serves as an upload portal for processing images, possibly for the deep learning-based ear detection project mentioned earlier.

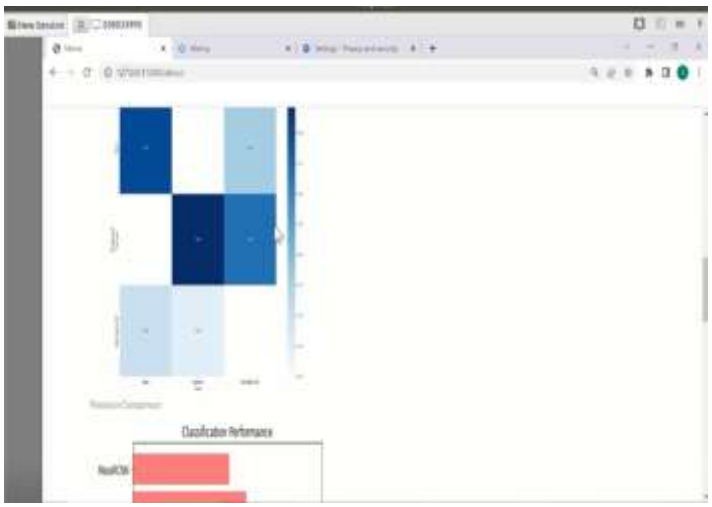

Figure 7. Classification of the Performance

The figure 7 shows the About page of a locally hosted web application (http link). At the top is a confusion matrix visualization, displaying classification results with different shades of blue to represent the number of correct and incorrect predictions for various classes. Below, there's a section labeled *"Precision Comparison"* and a Classification Performance bar chart, where "NasNetCNN" is visible as one of the models. The page likely presents performance metrics for a deep learning-based classification system, possibly related to the earlier mentioned ear detection project, showing model accuracy, precision, and classification quality.

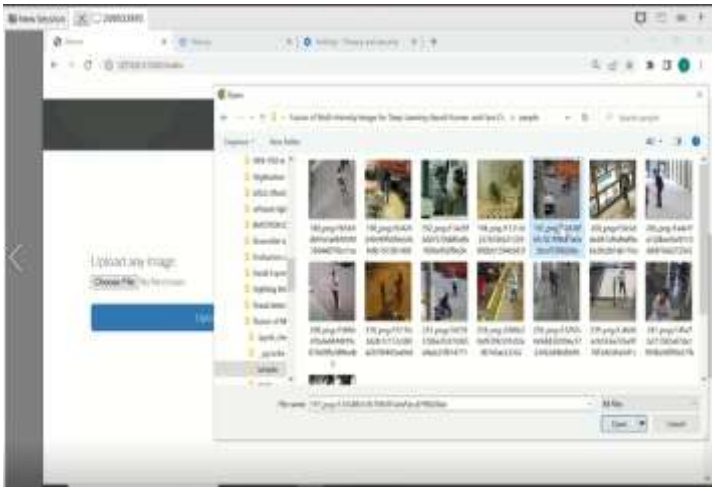

Figure 8. Choosing an Image file Interface

The figure 8 shows a file upload process on a locally hosted webpage. The page has an “Upload any image” label, a Choose File button, and a blue Upload button. A file explorer window is open, displaying a folder containing multiple surveillance-style images of people in various environments. The folder path indicates it’s part of a project titled *"Fusion of Visible-Infrared Imagery for Deep Learning-based Human and ..."* with a subfolder named sample. One image (178_jpg.rf...jpg) is selected for upload. This likely supports an AI-based detection or classification task using visible and infrared camera data.

The figure 9 shows a file upload page from a locally hosted web application (http index). The interface displays the label *"Upload any image"* with a Choose File button, where a file named image.jpg has been selected. Below it is a large blue Upload button, and the mouse cursor is hovering over it, indicating readiness to submit. The top background hints at the dashboard’s design seen earlier. This step is part of the application’s process to allow users to upload an image, likely for AI-based processing, such as human or ear detection, as described in earlier project details.

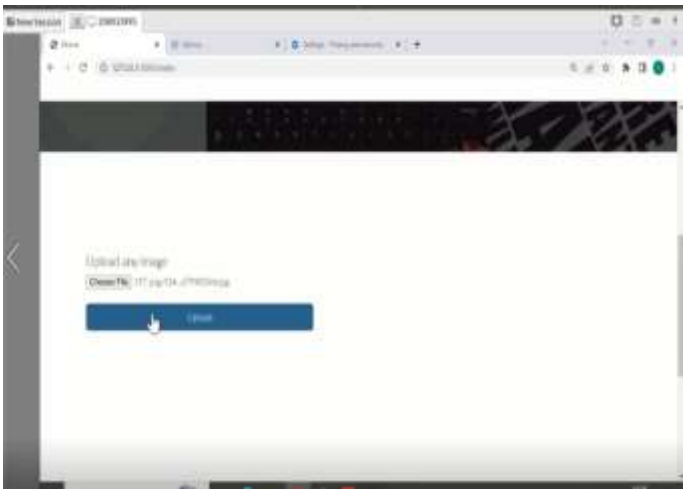

Figure 9. Upload Page File

The figure 10 shows the result of an AI-based object detection system running locally at http. The displayed surveillance photo captures a street scene where a man is walking, pushing a cart. The AI model has detected the person, marking them with a red bounding box labeled "human 0.83", indicating 83% confidence in the classification. The rest of the image remains unmarked, suggesting no other objects met the detection threshold. This output is likely part of the deep learning project mentioned earlier, demonstrating its ability to process uploaded images and

identify humans within real-world environments.

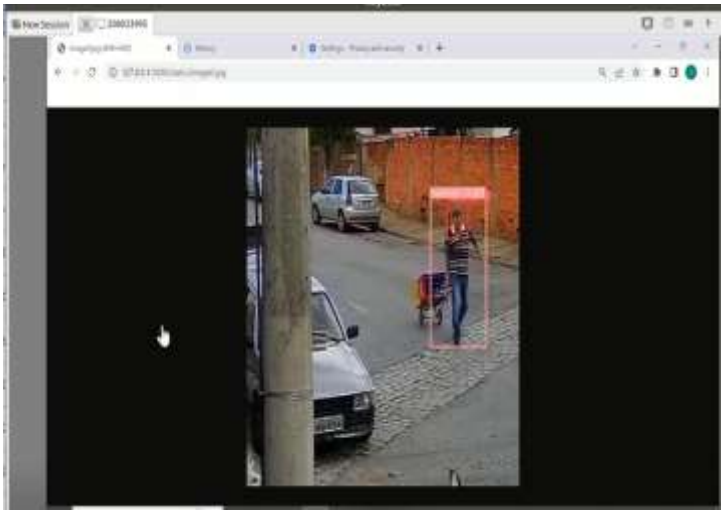

Figure 10. Human Detection Output from Uploaded Image

The figure 11 is an output from an AI-based object detection system hosted locally at http://127.0.0.1:5000. It shows a man standing on a brick-paved area, holding a phone. The detection model has identified the person, enclosing them in a red bounding box labeled "human 0.96", which indicates a 96% confidence level. The surroundings, including part of a parking area and pavement, remain unmarked. This result demonstrates the system's high-accuracy capability to detect humans in real-world scenarios from surveillance or street-level images, forming part of the deep learning project for human detection mentioned earlier.

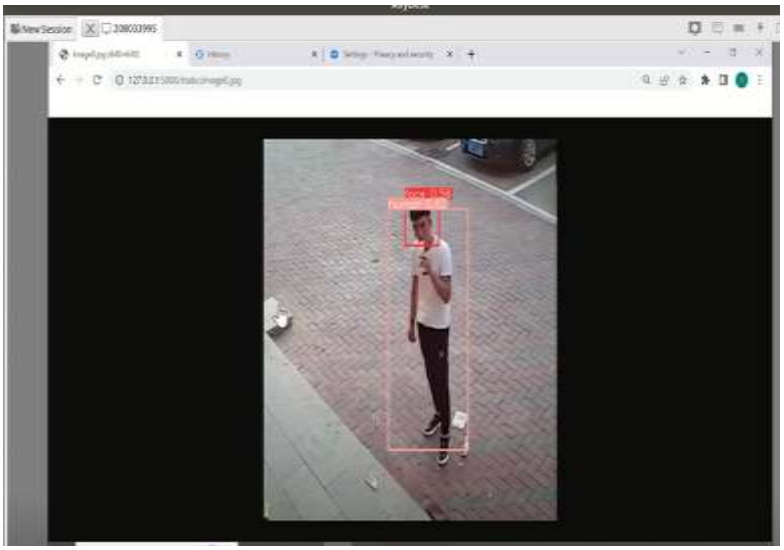

Figure 11. Human Detection with Confidence Score

5. CONCLUSION

This view of the assessment of algorithms for facial detection and facial detection and provides their overall performance metrics using a preliminary multi -red lighting data set that has been very well annotated. A basic technique is designed to use CNN pre -school detectors, a recently developed monitor and a true fusion scheme that uses additional effects of different light intensities. This article provides excellent detection and monitoring results for easy scenes; However, under investigation there are improved more complex data sets, improved fusion strategies and a scientific approach to determining the relevant parameters, consisting of a batch and the price of learning for education of a particular CNN model.

REFERENCES

[1]. W. Teng, ‘‘A new design of ir illuminator for nighttime surveillance,’’ M.S. thesis, Dept. Comput. Sci., Nat. Chiao Tung Univ., Hsinchu, Taiwan, 2010.

[2]. C.-H. Chan, H.-T. Chen, W.-C. Teng, C.-W. Liu, and J.-H. Chuang, ‘‘MI3: Multi-intensity infrared illumination video database,’’ in Proc. Vis. Commun. Image Process. (VCIP), Dec. 2015, pp. 1–4.

[3]. P. J. Lu, J.-H. Chuang, and H.-H. Lin, ‘‘Intelligent nighttime video surveillance using multi-intensity infrared illuminator,’’ in Proc. World Congr. Eng. Comput. Sci., vol. 1, 2011, pp. 19–21.

[4]. Y.-T. Chen, J.-H. Chuang, W.-C. Teng, H.-H. Lin, and H.-T. Chen, ‘‘Robust license plate detection in nighttime scenes using multiple intensity IR-illuminator,’’ in Proc. IEEE Int. Symp. Ind. Electron., May 2012, pp. 893–898.

[5]. J.-H. Chuang, W.-J. Tsai, C.-H. Chan, W.-C. Teng, and I.-C. Lu, ‘‘A novel video summarization method for multi-intensity illuminated infrared videos,’’ in Proc. IEEE Int. Conf. Multimedia Expo. (ICME), Jul. 2013, pp. 1–6.

[6]. S. Ren, K. He, R. Girshick, and J. Sun, ‘‘Faster R-CNN: Towards realtime object detection with region proposal networks,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[7]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, ‘‘Mask R-CNN,’’ in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2961–2969.

[8]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ‘‘You only look once: Unified, real-time object detection,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[9]. J. Redmon and A. Farhadi, ‘‘YOLOv3: An incremental improvement,’’ 2018, arXiv:1804.02767.

[10]. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, ‘‘YOLOv4: Optimal speed and accuracy of object detection,’’ 2020, arXiv:2004.10934.

[11]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ‘‘SSD: Single shot multibox detector,’’ in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 21–37.

[12]. K. Guo, S. Wu, and Y. Xu, ‘‘Face recognition using both visible light image and near-infrared image and a deep network,’’ CAAI Trans. Intell. Technol., vol. 2, no. 1, pp. 39–47, 2017.

[13]. S. Cho, N. Baek, M. Kim, J. Koo, J. Kim, and K. Park, ‘‘Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network,’’ Sensors, vol. 18, no. 9, p. 2995, Sep. 2018.

[14]. H. Jiang and E. Learned-Miller, ‘‘Face detection with the faster R-CNN,’’ in Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2017, pp. 650–657.

[15]. H. Nam and B. Han, ‘‘Learning multi-domain convolutional neural networks for visual tracking,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 4293–4302.

[16]. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, ‘‘The PASCAL visual object classes (VOC) challenge,’’ Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, Jun. 2010.

[17]. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, ‘‘Microsoft COCO: Common objects in context,’’ in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2014, pp. 740–755.

[18]. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[19]. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ‘‘MobileNetV2: Inverted residuals and linear bottlenecks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4510–4520.

[20]. S. Yang, P. Luo, C. C. Loy, and X. Tang, ‘‘WIDER FACE: A face detection benchmark,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 5525–5533.