International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 18s,2025
https://theaspd.com/index.php

A Comparative Analysis Of CNN Models For Object Detection In Thermal Imaging

[1] **Mr. Mandar Digambar Khatavkar,** [2] **Dr. B. D. Jadhav**
[1]Electronics & Telecommunication Engineering (Ph. D. Scholar), RSCOE Tathwade, Pune, Maharashtra, India. Email: mandar.khatavkar@gmail.com Orchid Id: https://orcid.org/0000-0002- 9243-8469
[2]Electronics & Telecommunication Engineering (Ph. D.), RSCOE Tathwade, Pune, Maharashtra, India. Email: bdjadhav_entc@jspmrscoe.edu.in

Abstract
Object detection in thermal imaging is a prevalent method within the fields of surveillance, medicine, and health. Thermal images represent the thermal reflections emitted by objects or bodies, captured using long- wavelength infrared (LWIR) cameras or thermal cameras. This paper reviews and compares four convolutional neural network (CNN) models—Faster R-CNN (FRCNN), Efficient Det, Single Shot Detector (SSD), and You Only Look Once (YOLO)—for object detection in thermal images. FRCNN demonstrates high accuracy, particularly for complex objects and small-scale features, but it is slower due to its two-stage process. Efficient Det is computationally efficient while maintaining high accuracy. SSD predicts object classes and bounding boxes in a single step without generating region proposals, whereas YOLO is highly efficient and suitable for real-time object detection. Precision, recall, and F1 score performance parameters are employed for comparison. It is observed that among EfficientDet, SSD, and FRCNN, the YOLOv8 method provides superior precision at 70.6%, recall at 43.6%, and F1 score at 53.9% for the given thermal images. These findings assist in selecting the appropriate model for object detection in thermal imaging with enhanced performance.
Keyword: *Deep Learning, Object Detection, Thermal Image.*

1. INTRODUCTION

Thermal cameras have become increasingly prevalent across various fields, particularly following the COVID-19 pandemic, where their ability to detect elevated body temperatures was widely leveraged for health monitoring and screening purposes [1]. These cameras detect infrared radiation emitted by objects and humans, and their classification is based on the wavelength range they operate in: Long-Wave Infrared (LWIR), Mid-Wave Infrared (MWIR), and Short-Wave Infrared (SWIR) [2]. Table 1 provides an overview of these infrared bands and their respective wavelength ranges. Thermal imaging offers several distinct advantages over conventional visible-light cameras: Camouflaged objects that are undetectable in the visible spectrum can be identified using LWIR cameras [3]. Objects in complete darkness remain invisible to standard cameras and may only be partially detected using night vision devices, which rely on ambient light and may not function effectively in bright daylight. In contrast, LWIR cameras perform reliably in both day and night conditions [4]. Thermal cameras are capable of detecting objects through light rain, fog, or smoke, scenarios where traditional cameras fail to deliver clear imagery [5]. Despite their advantages, thermal images are often degraded by environmental conditions. Image blurring may result from the movement of heat-emitting objects, which distorts the surrounding air or surfaces. Additionally, steam, ambient heat from surrounding areas, and thermal reflections from nearby walls, floors, or glass surfaces can further reduce image clarity. Image quality is also influenced by two major factors:First is the presence of hot air, particularly during midday in summer, can introduce distortions and blur in thermal images [6], and other is continuous usage can increase the internal temperature of the thermal camera, causing heat to build up within the camera body and further deteriorating image quality [7].
Moreover, the high cost of high-resolution thermal imaging equipment presents a significant challenge in broader adoption. A typical object detection pipeline using thermal imagery involves the following key stages:
Pre-processing – This includes image deblurring and super-resolution enhancement to improve the quality of thermal images.
Colorization – Applying color to grayscale thermal images to facilitate better visual analysis and model performance.
Object Detection – Utilizing computer vision or deep learning methods to identify and classify objects in the thermal images.

Table 1: thermal camera and wavelength [8]

Sr. No.	Thermal camera	Wavelength
1	SWIR	0.9 – 1.7 µm
2	MWIR	3.0 – 5.0 µm
3	LWIR	8.0 – 14.0 µm

Deep Learning: Deep learning, a part of machine learning (ML), which uses deep neural networks (DNNs) composed of many layers. Motivated by the human brain's structure and functionality, these networks process large datasets and can learn in unsupervised or semi-supervised settings. By automatically extracting meaningful features from input data, deep learning models have become helpful in fields like image and speech recognition, as well as natural language processing. Each layer in a deep learning model transforms the input data into increasingly abstract and complex representations. For instance, in image recognition, the initial input may be a matrix of pixels. The layer might identify basic features like edges; the next layer might recognize patterns by combining these edges; the following layer could detect elements such as eyes and a nose; and a higher layer might recognize the overall shape of a face. Importantly, deep learning models can automatically determine the most usable features to extract and represent at each level.
The structure of Deep learning involves layers as shown in fig. 1. The input layer receives the input which is then applied to hidden layers which performs mathematical computations and provide results to output layers, each hidden layer consist of Neurons, the neurons may have multiple inputs and multiple outputs. Generally activation functions are used at last steps of hidden layers.

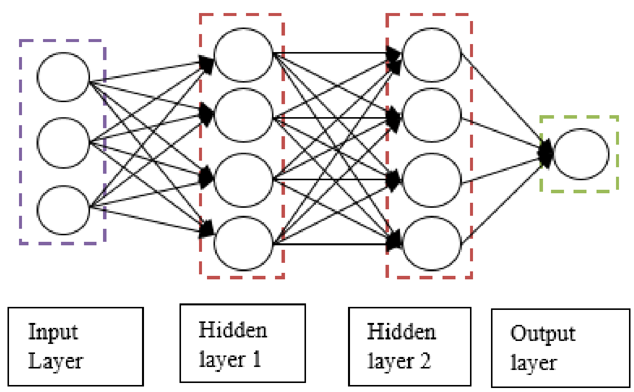

Fig. 1: *Structure of Deep Learning*

Section 1 represents introduction, followed by Section 2 as related works. In Section 3 covers the standard CNN models and evaluation steps for object detection. Section 4 presents comparative results. Section 5 provides conclusion, followed by references.

2. RELATED WORK

2.1 Object Detection

Before the object detection phase, an essential step is background subtraction. James W. Davis and Vinay Sharma hosted a technique that utilizes a contour-based fusion approach for combining thermal and visible imagery [9]. This method is usually to both thermal and visible domains, leveraging region-based and gradient based processing to emphasize significant contours across the two modalities. The process begins with low-level operations, including stream registration, identifying regions of interest (ROIs), generating a contour saliency map, and applying thinning and thickening techniques. The network proposed by the author [10] incorporates both a fusion module and an MRCNN module. In this approach, the fusion module includes an encoder and decoder structure. The resulting fused image is then used to

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 18s,2025
https://theaspd.com/index.php

detect objects more effectively. Specifically, the MRCNN algorithm is applied for object detection in nighttime images using thermal imagery. Given that MRCNN alone struggles with detecting features in such conditions, an image enhancement module is employed as a pre-processing step to emphasize object features in nighttime images. Within the fusion module, the encoder is responsible for extracting relevant features, working on pairs of source images to capture salient details, although depthwise convolution is not used in the initial layer due to the grayscale nature of the images. Instead, a standard CNN layer generates multiple channels before introducing depthwise convolution into the network. There are various methods that can be used for object detection in thermal images, some methods are based on neural network and some are based on conventional methods. In this methods we are using some neural based methods for object detection. The conventional methods further categorized as Background Subtraction (Detecting objects based on the difference between the frames and models of current and background resp.), HOG based method [11], thresholding[12] (Segmenting objects by applying a threshold to the thermal intensity values), Contour Detection (Identifying object boundaries using contours). The Machine Learning Methods are categorized as Support Vector Machines (SVM)[13]. Random Forests
[22] (Employing ensemble learning techniques for detection and classification), Deep Learning Methods [18], Convolutional Neural Networks (CNNs) (Training CNNs specifically on thermal images for object detection). The popular architectures include YOLO 19, SSD, and Faster R- CNN, etc. Transfer Learning: uses pre-trained models on visible spectrum images and fine-tuning them on thermal images [20-25]. It can applied to infrared imagery [26][29], it has various applications [27-28]

3. CNN BASED METHODS

3.1 Faster R-CNN (FRCNN) is a widely used and highly influential object detection framework that builds upon earlier versions like R-CNN and Fast R-CNN. It introduces a Region Proposal Network (RPN) that significantly improves the speed and efficiency of generating region proposals, making it suitable for real-time applications. The Faster R-CNN architecture can be summarized in the following components:

- Backbone Network: A convolutional neural network (e.g., VGG16, Res Net) is used as the backbone to extract feature maps from the input image.
- Region Proposal Network (RPN): The RPN takes the feature maps from the backbone network and slides a small network over them. At each sliding window location, the RPN predicts
  - The coordinates of bounding boxes (relative to anchor boxes).
  - The objectness score indicating whether the box contains an object or background.
- RoIPooling: Region of Interest (RoI) pooling is applied to the proposed regions to extract fixed-size feature maps. These feature maps are used for the final classification and bounding box regression.
- Classification and Bounding Box Regression: The RoI- pooled feature maps are fed into fully connected layers for classification and bounding box regression. The network predicts the class of the object and refines the bounding box coordinates. Fig. 2 shows block diagram of FRCNN architecture.

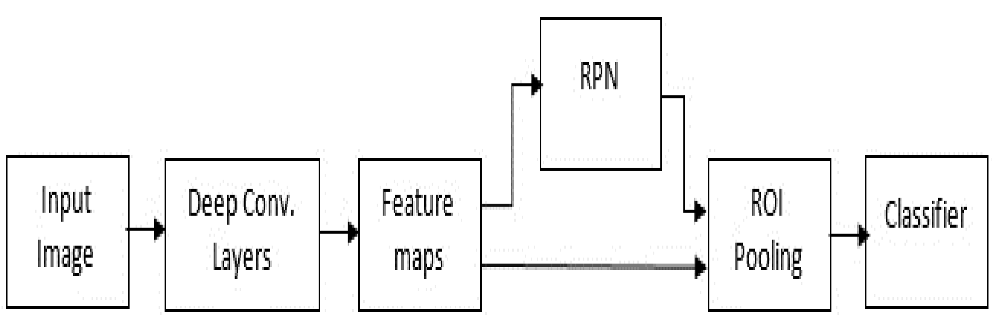

Fig. 2: *Simplified architecture of Faster R-CNN object detection method.*

3.2 The EfficientDet is a series of object detection models created by Google Research, aimed at balancing accuracy with computational efficiency. It is based on the EfficientNet architecture, which applies a compound scaling approach to consistently adjust depth, width, and resolution. Below is a detailed explanation of the EfficientDet approach: Efficient Net Backbone: Efficient-Det uses the Efficient-Net architecture as its backbone network. Efficient-Net is a family of convolutional neural networks that use a compound scaling method to achieve high accuracy with fewer parameters. This makes Efficient-Net a strong and efficient feature extractor for object detection.

- BiFPN (Bidirectional Feature Pyramid Network): One of the key innovations in Efficient-Det is the Bidirectional Feature Pyramid Network (BiFPN). Traditional feature pyramid networks (FPN) merge features at different resolutions to capture both high-level semantic information and low-level detail.BiFPN improves upon this by allowing feature fusion. BiFPN is also bidirectional, meaning it merges features in both top-down and bottom-up pathways.
- Compound Scaling: To scale the model Efficient- Det uses a compound scaling method. This method scales the backbone network, BiFPN, as well as box/class prediction networks uniformly based on a single compound coefficient. This scaling ensures that the model balances accuracy and efficiency across different model sizes (from EfficientDet-D0 to EfficientDet-D7)
- Weighting Mechanism: Efficient-Det introduces a weighted sum approach to combine features, allowing the network to learn the optimal way to merge different scales of features. This mechanism helps in improving the efficiency and performance of the feature fusion process. Fig.3 shows block diagram of Efficient Det architecture.

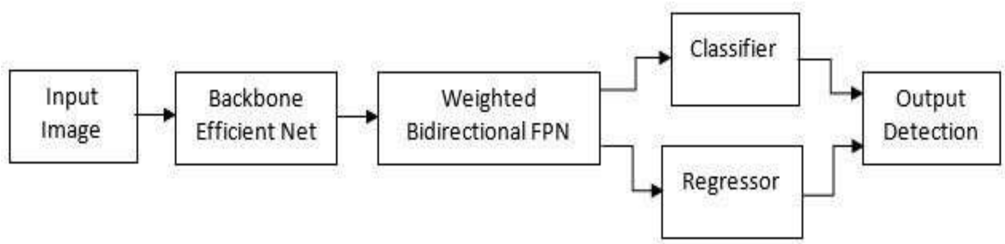

Fig. 3: *Simplified architecture of Efficient-Det Object detection method*

3.3 SSD is a popular algorithm known for its efficiency and high performance. Unlike some other object detection methods that require multiple stages (like Faster R-CNN). The SSD architecture can be summarized in the following components:

- **Base Network:** SSD employs a base network (such as VGG16, ResNet, or MobileNet) from the input image. This base network is usually a convolutional neural network that has been pre- trained.
- **Additional Feature Layers:** Extra convolutional layers are added to the base network to generate feature maps at various scales. These additional layers progressively reduce in size, capturing broader receptive fields.
- **Multi-Scale Feature Maps**: SSD extracts feature maps from both the base network and the additional layers, which are used to predict bounding boxes and class scores across different scales.
  - o **Default Boxes and Predictions**: At each point in the feature maps, SSD applies a set of default boxes with varying aspect ratios and scales. The network then predicts: The offsets for each default box to refine the predicted bounding box
  - o The class scores for each default box.

- **Loss Function**: The SSD loss function is a combination of
  - o Localization loss
  - o Confidence loss
  - o The loss function ensures that the network learns to predict accurate bounding boxes and class scores.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 18s,2025
https://theaspd.com/index.php

Fig. 4 shows block diagram of SSD architecture.

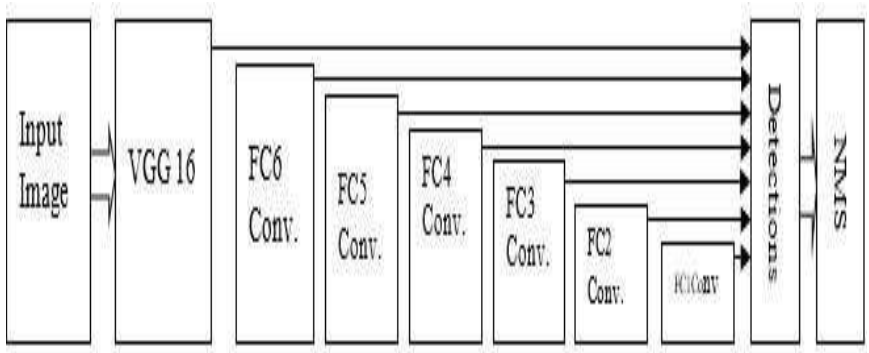

Fig. 4: *Simplified architecture of Single Shot detection method.*

3.4 **YOLO8:** YOLOv8 (You Only Look Once, version 8) is an advanced real-time object detection model in the YOLO family, which aims to improve speed and accuracy over previous versions. YOLO models are known for their capability to perform object detection in a single pass through the neural network, making them very fast and suitable for real-time applications. Here's an overview of the YOLOv8 method:

- **Single-Stage Detection**: YOLOv8, similar to earlier versions. This contrasts with two-stage detectors (like Faster R-CNN) that first generate region proposals before classify them.
- **Architecture Improvements:** YOLOv8 introduces architectural improvements over previous versions, such as more efficient backbones and improved head designs for better feature extraction and bounding box regression. These architectural enhancements help to achieve better accuracy while maintaining high speed.
- **Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN):** YOLOv8 employs feature pyramid networks and path aggregation networks to enhance multi-scale feature representation.
- **Anchor-Free Detection:** Moving towards an anchor-free detection paradigm, YOLOv8 reduces the complexity associated with predefined anchor boxes. Instead, it uses point-based predictions where the model predicts key points or center points of objects directly, simplifying the detection process and improving efficiency
- **Loss Functions**: YOLOv8 uses advanced loss functions to optimize the bounding box regression, classification, and objectness scores. These loss functions ensure that the model learns to predict accurate and well-aligned bounding boxes.
- **Augmentation Techniques:** Advanced data augmentation techniques, such as mosaic augmentation and mix up, are used during training to improve the model's robustness and generalization ability. **Post-Processing Enhancements:** YOLOv8 employs improved non- maximum suppression (NMS) techniques, like Soft-NMS and DIoU-NMS, to refine the final set of detected bounding boxes, ensuring that overlapping boxes are handled more effectively.

Fig. 5 shows block diagram of YOLO 8 architecture

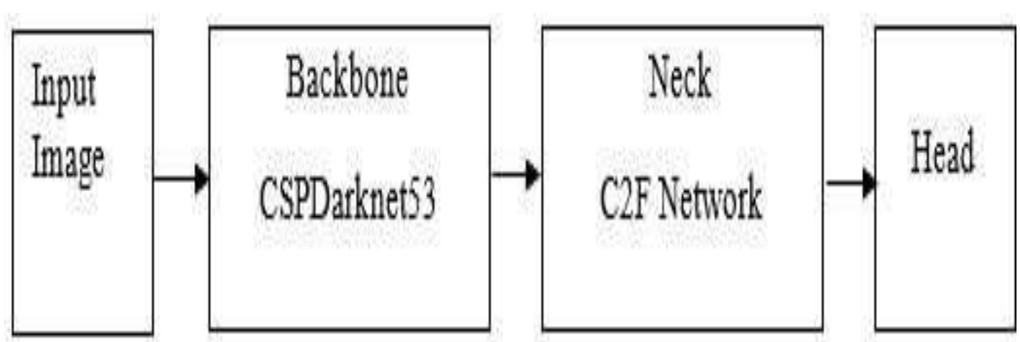

Fig. 5: *Simplified architecture of YOLO 8 object detection method.*

4. PERFORMANCE MEASUREMENT:

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 18s,2025
https://theaspd.com/index.php

Mean Average Precision (mAP): The Mean Average Precision is a widely used metric for assessing the performance of object detection models. It measures the balance between precision and recall across various confidence thresholds. For object detection, precision and recall are computed separately for each class, and the Average Precision (AP) is determined for each class individually. The mAP is the mean of these AP values across all classes.

Precision: Precision assesses the accuracy of the model's positive predictions, indicating how many of them are actually correct. Precision is calculated using the formula:

Precision = TP / (TP + FP) (1)

A high precision value suggests that the model's positive predictions are generally accurate, though it does not account for any false negatives.

Recall (Sensitivity): Recall, also referred to as sensitivity or the true positive rate (TPR), This metric reflects the model's capacity to capture all positive examples within the dataset.:

Recall = TP / (TP + FN) (2)

F1-score: The F1-score is balanced metric that reflects model performance by accounting for both false positives and false negatives. It ranges from 0 to 1, with a score of 1 representing an ideal balance between precision and recall, indicating optimal performance.

F1-score = 2 * (Precision * Recall) / (Precision + Recall) (3)

5. RESULT AND DISCUSSION

FRCNN method: The faster R CNN is state-of-art method implemented in google colab platform. As stated earlier this method comprises of region proposal and classifier stages. In the development, the resnet-50 as backbone network is used. The total network consist of backbone network, Layer 1 to 4 each one having 3 convolution layer and 3 batch normalization layer ReLu function and down sampling, layer 1 consist of 64-256 input channels and 64-256 output channel, layer 2 consist of 128-512 input channels and 128-512 output channel, layer 3 consist of 256-1024 input channels and 256-1024 output channel, layer 4 consist of 512-2048 input channels and 512-2048 output channel. The feature pyramid network consist of different convolution layers with 256-2048 input channels and 256 output channels. The region proposal network consist of convolution layer and ReLu function, RoI pooling and box predictor layers. The result is shown figure 6.

Fig. 6: *Result of object detection using FRCNN method.*

Efficient-Det: The architecture consist of backbone network for feature extraction for better balance and efficiency BiFPN is used, separable convolutions are used to reduce computation cost, activation function ReLu used, batch normalization layer and optimization layers are used for bounding box refinement NMS layers are used. The obtained result shown in Fig. 7

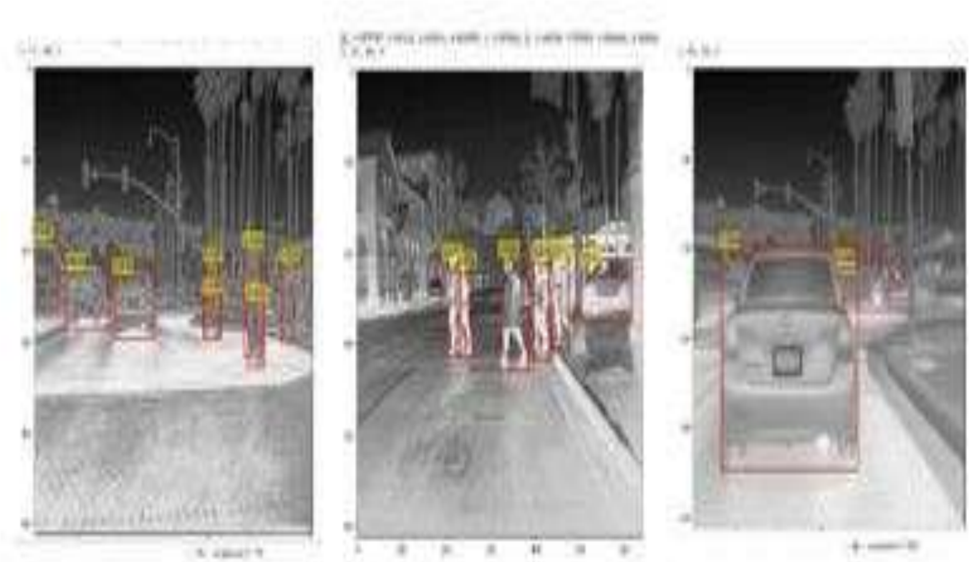

Fig. 7: *Result of object detection using Efficient-Det*

SSD: The SSD Architecture starts with input layer of 5 different blocks, block 1 and 2 having two layers of convolution and one layer of max pooling, whereas remaining blocks consist of three layers of convolution and one layer of max pooling. At the final stage dense layer used for predictions. The result is as shown in Fig. 8

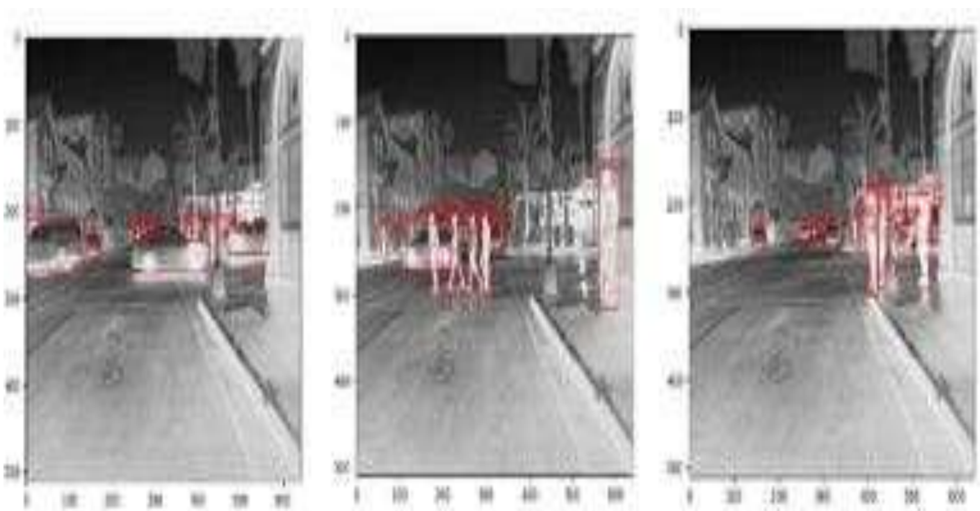

Fig. 8: *Result of object detection using SSD method*

YOLO8 Methods : YOLO8 detection model consist of 225 layers, it starts with convolution layer consist of 3 layers(convolution2d, batch normalization 2d, activation function SiLU), after this layer four convolution and C2f [CV1 (con2d,bn,act); CV2 (con2d,bn,act); Bottelneck CV1,CV2] alternatively used, after this SPPF layer is used which consist of CV1,CV2 and maxpool layer then upsample, concat layers are used, which is then connected to C2F upsample concat layer this further connected to two (C2F, convolution and Concat) layer cascaded, which the connected to C2F layer. In the final, detect stage is used which consist of CV2, CV3 layer each of these having three sequential layers [two conv (con2d,bn,act) and one conv2d] and one DFL (conv2d) layer. Result of this method is shown if Fig. 9.

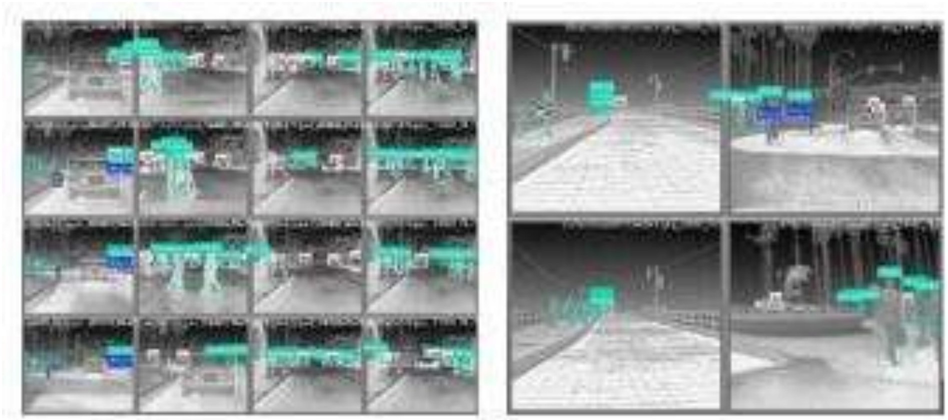

Fig. 9: *Result of object detection using YOLO 8 method.*

Comparison between different methods

Based on the performance parameter like Precision, Recall and F1 score the methods FRCNN, Efficient-Det, SSD and YOLO8 method for object detection in the thermal image is compared. The same thermal image database includes four classes (Bus, Car, Cycle, and Person) for training purpose is used, the Google colab platform is used and T4 GPU and 200 epoch configured for training. In FRCNN 55% precision 46% recall and 50% F1 score is obtained, efficient-det 67% precision 41% recall and 51% F1 score, SSD method 42% precision 43% recall and 42% F1 score and in the YOLO8 method 71% precision 44% recall and 54% F1 score is obtained. The summary of scores mentioned in Table 2 and graphically represented in Fig. 10.

Table 2: *Result of comparison between different Object detection methods.*

Methods	Precision	Recall	F1 Score
FRCNN	0.55	0.46	0.50
Efficient-Det	0.67	0.41	0.51
SSD	0.42	0.43	0.42
YOLO-8	0.71	0.44	0.54

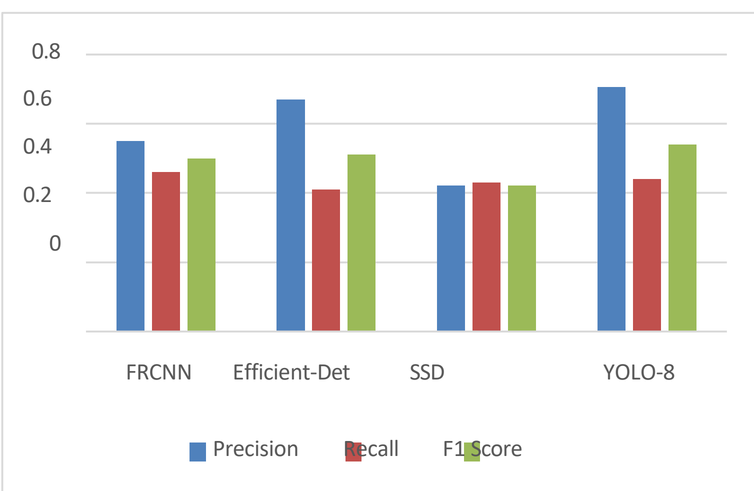

Fig. 10: *Result of comparison between different Object detection methods.*

The experiment performed for different epoch values from 50-200 as shown in Table the precision value increased from 24.6% to 70.6%. the recall value increased from 17.2% to 43.6%, F1 score value increased from 20.2 % to 53.9%, mAP@50 value increased from 14.5% to 54% and mAP@50-95 value increased from 0.6% to 23.8% which shows better results compared to state-of-the-art methods.

Comparison of parameters for object detection using YOLO8 method for different epoch

Table 3: *Result of comparison of parameters for object detection using YOLO8 for different epoch*

Epoch	Precision	Recall	F1 Score	mAP50	mAP50-95	Time in hours
50	0.246	0.172	0.202	0.145	0.060	0.036
0.716	0.305	0.428	0.347	0.144	0.074	
0.78	0.409	0.537	0.508	0.226	0.113	
0.706	0.436	0.539	0.54	0.238	0.144	

6. CONCLUSION

The result comparison is done between some CNN based method for object detection. It is found that out of Efficient Det, SSD, FRCNN the YOLO8 method provides better precision value of 70.6%, recall value 43.6% and F1 scores value of 53.9% for given thermal images. FRCNN method gives high accuracy, especially for complex objects and small-scale features but it is slower due to its two-stage process. Efficient Det provides result closer to that of YOLOv8 architecture, It also efficient in terms of computation speed, SSD directly predicts object classes and bounding boxes in a single step without generating region proposals.

REFERENCES:

[1] Togneri, R., Pokrajac, D., & Jampani, V. (2021). Thermal imaging for fever detection during COVID-19 pandemic: A review of the effectiveness and challenges. Biomedical Signal Processing and Control, 68, 102759. https://doi.org/10.1016/j.bspc.2021.102759

[2] Sadeghi, S., Yoo, Y., & Park, C. (2020). Classification of infrared sensors: SWIR, MWIR, and LWIR. Optics and Lasers in Engineering, 130, 106081. https://doi.org/10.1016/j.optlaseng.2020.106081

[3] Zhang, Y., Wang, L., & He, L. (2019). Enhanced thermal image object detection based on deep learning. Infrared Physics & Technology, 96, 230–236. https://doi.org/10.1016/j.infrared.2019.01.009

[4] Chen, Y., Jiang, J., & Gao, Z. (2018). Improved object detection in thermal infrared images using spatial information fusion. Infrared Physics & Technology, 90, 14–22. https://doi.org/10.1016/j.infrared.2018.01.006

[5] Han, J., Morimitsu, T., & Tsuji, K. (2021). Thermal image processing in adverse environments. Sensors, 21(4), 1278. https://doi.org/10.3390/s21041278

[6] Kuenzer, C., & Dech, S. (2013). Thermal infrared remote sensing: Sensors, methods, applications. Springer. https://doi.org/10.1007/978-94-007-6639-6

[7] Barmpoutis, P., Papaioannou, G., & Theodoridis, T. (2021). Thermal image analysis and applications: A review. IEEE Access, 9, 51216-51241. https://doi.org/10.1109/ACCESS.2021.3069690

[8] Grujić, K. A Review of Thermal Spectral Imaging Methods for Monitoring High-Temperature Molten Material Streams. Sensors 2023, 23, 1130. https://doi.org/10.3390/s23031130

[9]J.W. Davis and V. Sharma, ``Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understand.*,vol. 106, nos. 2_3, pp. 162_182, May 2007.

[10] Heena Patel and Kishor P. Upla, "Night Vision Surveillance: Object Detection using Thermal and Visible Images", 2020 International

Conference for Emerging Technology (INCET)Belgaum, India. Jun 5-7, 2020
[11] Yan Zhang,TingZhao,JianGu, "Accurate moving object detection in thermal imagery",
[12] D. Gangodkar, P. Kumar, and A. Mittal, ``Segmentation of moving objects in visible and thermal videos," in *Proc. Int. Conf. Comput. Commun. Informat.*, Coimbatore, India, Jan. 2012, pp. 1_5.
[13] P. Kumar, A. Mittal, and P. Kumar, ``Fusion of thermal infrared and visible spectrum video for robust surveillance," in *Proc. Int. Conf. Comput. Vis.,Graph. Image Process.*, Madurai, India, Dec. 2006, pp. 528_539.
[14] W. K. Wong, H. L. Lim, C. K. Loo, and W. S. Lim, ``Home alone faint detection surveillance system using thermal camera," in *Proc. 2nd Int. Conf. Comput. Res. Develop.*, Kuala Lumpur, Malaysia, May 2010, pp. 747_751.
[15] W. Li, D. Zheng, T. Zhao and M. Yang, "An Effective Approach to Pedestrian Detection in Thermal Imagery," *8th International Conference on Natural Computation (ICNC 2012),* 2012.
[16] M. Narendran, M. Lakshmanan, S. Mohan, L. Sharma and R. P.Sharma, "A review on object detection in thermal imaging and analyzing object and target parameters," *International Journal Of Electrical, Electronics And Data Communication, ISSN: 2320-2084,*vol. 5, no. 11, 2017.
[17] M. San-Biagio, M. Crocco and M. Cristani, "Recursive segmentation based on higher order statistics in thermal imaging pedestrian detection," in *Proceedings of the 5th International Symposium on Communications, Control and Signal Processing, ISCCSP 2012*, Rome, Italy, 2012.
[18] X. Zhou, W. Gong, W. Fu and F. Du, "Application of Deep Learning in Object Detection," in *International Conference on Information Systems (ICIS 2017)*, 2017.
[19] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE, 2016
[20] Wang X , Yu K , Wu S ,et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks[J]. 2018.
[21] Tao X , Gao H , Wang Y , et al. Scale-recurrent Network for Deep Image Deblurring [J]. 2018.
[22] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in International conference on curves and surfaces. Springer, 2010, pp.711– 730
[23] J. Davis and M. Keck, ``A two-stage approach to person detection in thermal imagery," in *Proc. Workshop Appl. Comput. Vis.*, Breckenridge,CO, USA, Jan. 2005, pp. 6_111.
[24] R. Miezianko, ``Terravic research infrared database- terravic motion infrared database," in *Proc. IEEE OTCBVS WS Ser. Bench*, to be published.
[25] R. Miezianko, ``Terravic research infrared database- terravic weapon infrared database," in *Proc. IEEE OTCBVS WS Ser. Bench*, to be published.
[26] R. Bergevin, P.-L. St-Charles, and G.-A. Bilodeau, ``Mutual foreground segmentation with multispectral stereo pairs," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 375_384.
[27] K. Van Beeck, K. Van Engeland, J. Vennekens, and T. Goedeme, ``Abnormal behavior detection in LWIR surveillance of railway platforms," in*Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS),* Lecce,Italy, Aug./Sep. 2017, pp. 1_6.
[28] D. Tan, K. Huang, S.Yu, and T. Tan, ``Efficient night gait recognition based on template matching," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR),*Hong Kong, Aug. 2006, pp. 1000_1003.
[29] Z. Wu, N. Fuller, D. Theriault, and M. Betke, ``A thermal infrared video benchmark for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Columbus, OH, USA, Jun. 2014, pp. 201_208.