Computer Engineering and Applications Vol. 15, No. 2, June 2026

Progressive Ordinates GAN with Centroid Fuzzy Ray-Tracing for Scene Image Detection from Thermal Imaging

R. Rajeswari[1*], Dr. S. Kannan[2]

[1*] *Asst. prof in CS Dept, Saiva Bhanu Kshatriya College, SBK School Rd, Vasantham Nagar, Aruppukkottai, Tamil Nadu 626101, Research Scholar in Madurai Kamaraj University, Water Tank Rd, Vettunimadam, Nagercoil, Tamil Nadu 629001*
scholar.rajeswari213@gmail.com
[2] *Professor, Dept of Computer Application, Madurai Kamaraj University, Water Tank Rd, Vettunimadam, Nagercoil, Tamil Nadu 629001*
*kannan921@yahoo.com

ABSTRACT

The security threats from terrorism and illegal migration raised security concerns, insisting the use of thermal cameras in surveillance systems due to their night vision and all-weather capability. However, thermal images suffer from variations in surface roughness, texture, and radiation, hindering accurate object detection. To address this, a Progressive Ordinates GAN with Centroid Fuzzy Ray-Tracing model is proposed for the efficient object detection using thermal images. The existing object detection algorithms struggle in detecting thermal signatures, as they depend only on temperature while neglecting emissivity differences from diverse surface characteristics of thermal images. To overcome this problem, a novel Progressive Ordinates Generative Adversarial Network (POGAN) is introduced, for more accurate detection and characterization of objects with diverse surface characteristics having varying emissivity, mitigating the challenges of inconsistent thermal signatures. Moreover, the shadow casting of objects from thermal image hinders identifying the object boundaries. The existing algorithms struggle to determine the shadow casting as they are not aware of the related parameters such as geometric properties, spatial relationships and the angle of incidence, as they operate on fixed-size image patches or regions. Hence, Centroid Ray-Tracing Fuzzy Clustering (CRFC) is introduced, which effectively acclimates to varying scene complexity by dynamically adjusting cluster centroids, thus enables a better understanding of geometric properties, spatial relationships, and angle of incidence within the scene. The analysis on this work validates that the proposed model achieves better performance with improved accuracy, sensitivity, mean Average Precision (mAP) and recall, with minimized Mean Average Error (MAE).

Keywords: Scene Image, Thermal Object Detection, Thermal Signatures, Generative Adversarial Network, Fuzzy Clustering.

1. INTRODUCTION

Thermal object detection is a pivotal technology in computer vision that enables the identification and tracking of objects within a scene using thermal imaging data. By harnessing the infrared radiation emitted by objects, thermal cameras provide a

unique perspective, especially in low-light or adverse weather conditions where traditional cameras struggle. The process begins with capturing thermal images of the scene and converting heat signatures into digital data. Subsequently, algorithms analyze these images to detect and localize objects based on temperature variations. Unlike visible light-based detection systems, thermal object detection is less affected by factors like illumination changes or camouflage, making it highly reliable for surveillance, security, and industrial applications. Common approaches include deep learning techniques such as convolutional neural networks (CNNs), which excel at extracting features from thermal images to recognize diverse objects with high accuracy. Integration with other sensor modalities, such as visible light or radar, further enhances detection capabilities, enabling comprehensive situational awareness. Moreover, detecting thermal signatures is an effective process in object detection from scene images in thermal datasets [1-4].

Thermal signature refers to the unique thermal radiation pattern emitted by an object, which is captured and analyzed to extract valuable information about its properties and behavior. Several factors influence thermal signatures, including the object's material composition, surface texture, shape, size, and temperature distribution. Additionally, environmental conditions such as ambient temperature, humidity, and atmospheric interference play significant roles in modulating thermal emissions. Deep neural network (DNN) algorithms have emerged as powerful tools for extracting meaningful insights from thermal signatures. CNNs are particularly well-suited for this task. However, implementing DNN algorithms for thermal signature analysis poses several challenges, including the scarcity of labeled thermal data for training, the need for large computational resources to train complex models, and the interpretability of the learned features. Emissivity is one such factor that affects an object’s thermal signature, and it is determined by how efficiently it emits thermal radiation [5-8]. Objects with high emissivity emit and absorb thermal radiation effectively, resulting in stronger thermal signatures, while those with low emissivity exhibit weaker thermal signatures.

To address emissivity variations in thermal signature analysis, algorithms such as emissivity correction algorithms and material classification algorithms are used. Emissivity correction algorithms aim to adjust thermal signatures based on known emissivity values of materials, while material classification algorithms leverage spectral and spatial features to identify materials and infer emissivity. Challenges include accurately estimating emissivity for different materials, handling variations in emissivity across surfaces, and developing robust algorithms capable of real-time processing in diverse environments. Moreover, the complexity of material interactions and the need for precise calibration pose additional hurdles in achieving accurate emissivity correction and material classification in thermal signature analysis. Additionally, the thermal signature detection for better classification of scenic objects is affected by the shadow casters, significantly affecting the boundaries of objects in thermal imaging [9-11].

When a shadow is cast onto an object, it alters the temperature distribution, leading to inaccuracies in object boundary detection and classification. Factors influencing shadow casting include the position, size, and shape of the object relative to the light source and the surface onto which the shadow is cast. Additionally, the spatial context of the scene, including the layout of objects and the geometry of the environment,

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

influences shadow-casting dynamics. Existing methods to mitigate shadow effects include shadow removal algorithms that use image processing techniques to identify and remove shadow regions, as well as context-aware algorithms that leverage spatial information to infer object boundaries and distinguish between shadows and actual objects. Challenges include accurately distinguishing between shadows and objects, especially in complex scenes with overlapping shadows, as well as handling variations in lighting conditions and surface properties that can affect shadow appearance. Moreover, achieving real-time performance and scalability in shadow removal algorithms for large-scale scenes poses additional computational challenges. [12-15].

The main contribution of this work is explained as follows:

- To identify the thermal signatures from the thermal object in the scene, a novel POGAN is introduced, which efficiently reduces the difficulties caused by inconsistent thermal signatures with diverse surface properties and emissivity.
- To delineate the object boundaries casted by shadow, a novel CRFC algorithm is introduced, which efficiently adjusts cluster centroids to adapt to changing scene complexity, comprehending geometrical characteristics, spatial relationships, and angle of incidence inside the scene.

This work is organized in the following manner, section 2 defines the literature review of the existing works on thermal object detection, section 3 explained the proposed concept of the work under Progressive Ordinates GAN with Centroid Fuzzy Ray-Tracing, section 4 discusses the overall result that obtained with their effectiveness and section 5 concludes the work.

2. LITERATURE SURVEY

Dai et al. [16] proposed a unique object detection method, called TIRNet, based on CNNs for accurate and efficient object detection in thermal infrared (TIR) images. Rather than relying on the slow and expensive deep CNN backbone (ResNet, ResNeXt), the lightweight feature adoptor (VGG) was used. The Residual Branch was introduced to obtain robust and discriminating features for precise box regression and classification. More remarkably, it doesn't exist outside of the training phase since every computation was contained within a single network, this TIRNet tested and optimized end-to-end. Additionally, a continuous information fusion technique was suggested to enhance detection performance. This strategy successfully addressed issues such as complicated backgrounds and occlusion and produced detection results that were more accurate and smoother. This work on multispectral object detection and enhancing TIRNet fails to merge the properties of TIR and color images.

Tu et al. [17] presented a multi-interactive dual-decoder to mine and model the multi-type interactions for accurate RGBT SOD. To be more precise, this work first encodes two modalities into feature representations that were multi-level and multi-modal. Next, to perform the interactions of multi-level features, two modalities, and global contexts, this work constructs a novel dual-decoder. Because of these interactions, this technique performed effectively even when there was an invalid modality, in a variety of difficult settings. Ultimately, this work conducts

comprehensive tests using publicly available RGBT and RGBD SOD datasets, and the findings demonstrate that the suggested approach outperforms cutting-edge methods. Integrating features from multiple modalities (RGB and Thermal) at different levels while preserving their distinct characteristics and effectively capturing their complementary information is challenging.

Miller et al. [18] examined the thermal cameras and compared its efficiency to dash cameras for increasing the accuracy of detection of objects in dangerous roads. Initially, data sets were taken and annotated with thermal and dash camera footage under several climatic conditions. A deep learning object detection model was designed with the help of YOLOv8 and Roboflow. Every camera was trained with distinct models and then fused to balance the mutual shortcomings. The dash camera was noted to be susceptible to occlusions and diverse lighting, meanwhile the thermal camera had excellent low-light settings. The thermal camera detected cyclists as well as pedestrians, thereby proving its effectiveness to advance road safety. Although the fused system showed strong potential for improving road safety, it remained limited by computational complexity and was not fully optimized for real-time deployment on edge devices.

Kernbauer et al. [19] presented PanoTherm, which was the first panoramic thermal camera with the target of continuous 360◦ surveillance of the location. The calibration and stitching procedure were explained and described the usage of camera with the existence of vehicles. A reference for detecting and tracking the objects was provided in this work. The method enabled object detection in the calamity situations as well as in monitoring the trespassers and maintaining pipe inspection, with the help of a uniform and calibrated view. The model failed to evaluate the performance of the MOT module nor did it examined Pano Therm under adversarial environmental conditions.

Kuehnel et al. [20] modeled an automatic traffic light system depending on thermal detection to vary signal duration dynamically for each person with walking disability or mobility burdens individually. The auditory signal was triggered for visually impaired individuals to enable each humane to mingle and progress their path without any barriers. Finally, thermal dataset was constructed for people having mobility restrictions which assisted capturing various scenarios for pedestrians, especially under diverse weather conditions like during lightning, crowded areas, and with climate changes. Though thermal imaging enable privacy and robustness in tough situations, it added intrinsic hurdles to detect objects due to its colorless and fine texture details, which was solved using YOLO-Thermal. It was an integration of feature extraction and attention mechanisms. Yet, despite these strengths, the reliance solely on thermal data limited semantic richness, making classification of object types less reliable.

Wang et al. [21] developed a novel Cross-Guided Fusion Network (CGFNet) for RGB-T salient object recognition, which aims to accomplish full integration of cross-modality information and deep mining of the distinctive features of single modalities. In particular, a module called Cross-Scale Alternate Guiding Fusion (CSAGF) was suggested to mine the high-level semantic data and offer global context assistance. To accomplish enough cross-modality fusion, this work then constructs a Guidance Fusion Module (GFM) that uses one modal as the primary guidance and the other

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

modal as an auxiliary. The primary decoding block, the Cross-Guided Fusion Module (CGFM), was finally introduced. Additionally, each decoding block was divided into two sections, each of which has two modalities of information that serve as the primary guidelines: global auxiliary enhancement (GAE) and cross-shared cross-level enhancement (CLE). The primary distinction between the two sections was the GFM's use of various modalities as its primary guidance. The thorough experimental findings demonstrate that this approach outperforms the most advanced salient detection techniques. However, this work lacks in mining high-level semantic information and providing global context support.

Du et al. [22] presented a novel application of substantially occluded vehicle detection in the dynamic wild background of aerial weak infrared camera images with over 50% of the vehicles obscured. Yolov4 was the detection model that was employed. Model performance was improved by using secondary transfer learning from the visible dataset to the infrared dataset (AP). First, the model was trained on the UCAS_AOD visible dataset. It was then moved to the VIVID visible dataset, and lastly, it was moved to the VIVID infrared dataset for a second training session. Concurrently, the YOLOv4 model was enhanced with a hard negative example mining block, which reduces complicated background disturbance and therefore lowers the false detection rate. This work further can be enhanced by improving the detection accuracy of the detection model in this scene by changing the structure of the model and the loss function.

Ying et al. [23] built the first large-scale and high-diversity benchmark for visible-thermal small object detection referred to as RGBT-Tiny. The dataset had 115 paired sequences, 93K frames, and 1.2M manual annotations. It contained a huge categories of objects (a total of seven) and various scenarios. Many of the annotated objects were smaller than 16×16, and paired bounding box annotations with tracking IDs were provided to enable challenging and practical applications such as in RGBT image fusion, object detection, and tracking. In addition, a scale adaptive fitness (SAFit) measure was employed to demonstrate high robustness for evaluating both small and large objects. Based on RGBT-Tiny, extensive evaluations were conducted using IoU and SAFit metrics on 32 recent state-of-the-art algorithms across visible generic detection, visible SOD, thermal SOD, and RGBT object detection tasks. However, despite its richness, the dataset still suffered from issues of misaligned image pairs and a lack of specialization for real-world ITS applications, thereby limiting its direct deployment value.

Zhou et al. [24] suggested an efficient and reliable feature fusion network (ECFFNet) for RGB-T SOD by merging thermal and RGB pictures. An efficient cross-modality fusion module in ECFFNet completely fuses features of matching sizes from the thermal and RGB modalities. After that, the bilateral fusing of foreground and background data was carried out by a bilateral reverse fusion module, allowing for the complete extraction of salient object boundaries. Lastly, to acquire complementary information, a multilevel consistent fusion module merges features from many levels. Extensive tests on three RGB-T SOD datasets demonstrate that, for various evaluation markers, the suggested ECFFNet performs better than twelve cutting-edge techniques. This research needs the most recent and cutting-edge

techniques for RGB-T SOD that are more accurate and efficient to handle scenarios that are more challenging and complicated.

Shyam, P. and Yoo [25] studied the thermal camera applications in enhanced perception capabilities in extreme climatic conditions such as during snow, night-time driving. A distinct object detector was utilized for providing low latency to RGB and thermal. A learnable projection function was employed to convert the thermal images into RGB color space, which provide less alterations to the primary image detector. In addition, super-resolution was linked with object detection to make use of low-resolution and low-cost uncooled thermal imaging sensors. A resolution optimization algorithm was presented inside the network structure to optimize inference efficiency while adhering to the constraints of resource-limited devices. A bidirectional feature pyramid network was presented to enhance the feature representation. Moreover, the reliance on dataset-specific optimizations could restrict generalization to broader autonomous driving.

The above statements have discussed that [16] fails to merge the properties of TIR and color images, whereas [17] has challenges in the integration of RGB and Thermal features at different levels while maintaining their unique characteristics and effectively capturing their complementary information. In [18], limited by computational complexity and was not fully optimized for real-time deployment on edge devices. [19] lack of MOT module performance evaluation and investigation under adversarial environmental conditions, [20] dependency on thermal data limited semantic richness, making classification of object types less reliable. [21] lacks in mining high-level semantic information and providing global context support. [22] can be enhanced by improving the detection accuracy by changing the structure of the model and the loss function. [23] Lack of specialization for real-world ITS applications, limited its direct deployment value, [24] need the most recent and cutting-edge techniques for RGB-T SOD with more accuracy and efficiency to handle scenarios, and [25] reliance on dataset-specific optimizations could restrict generalization.

3. PROGRESSIVE ORDINATES GAN WITH CENTROID FUZZY RAY-TRACING

Object detection in scene analysis from thermal images is a critical task with applications spanning surveillance, search and rescue, industrial inspection, and more. Object detection algorithms analyze thermal signatures to identify and localize objects of interest within a scene. However, the thermal signatures of each object vary with their surface properties such as texture, roughness, and color, which affect how objects absorb and emit thermal radiation. In addition, the lightweight models such as MobileNet and EfficientNet, suitable for conventional RGB object detection, are limited to intensity-based features and failed to deal with the unique challenges of thermal imaging, such as emissivity variations, reflections, shadows from radiative transfer, or incidence angle geometry. To overcome these challenges, an innovative technique is introduced called “Progressive Ordinates GAN with Centroid Fuzzy Ray-Tracing for Scene Image Detection from Thermal Imaging”. Challenges arise when objects with varying emissivity values are present in the scene, which leads to inconsistencies in the detected thermal signatures. Objects with highly reflective

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

surfaces exhibit lower apparent temperatures in thermal images due to their lower emissivity, while objects with matte or absorbent surfaces appear warmer. This means that they emit less thermal radiation relative to their temperature, leading to lower apparent temperatures in thermal images. Hence, detecting and accurately characterizing objects with diverse surface characteristics is challenging, as their thermal signatures do not accurately reflect their true temperature or thermal behavior. The existing algorithms struggle to detect these thermal signatures, as they solely classify objects based on their apparent temperature in the thermal image, without considering underlying emissivity differences. Hence, to accurately detect and characterize objects using the thermal signature, a novel algorithm is introduced called Progressive Ordinates Generative Adversarial Network (POGAN), where the Discrete Ordinates Method better captures the complex interactions between thermal radiation and surfaces with varying emissivity values. Progressive Attention Generative Adversarial Networks (PAGAN) allow the network to learn hierarchical features at multiple scales to capture fine details and gradations in thermal images, and further enable the network to focus on relevant regions of the image, allowing it to prioritize areas with significant emissivity variations. Thus, POGAN effectively allows for more robust detection and characterization of objects with diverse surface characteristics having varying emissivity, mitigating the challenges associated with inconsistent thermal signatures.

Moreover, thermal objects cast shadows that affect their thermal signatures and boundary appearance. Thermal radiation undergoes reflection and refraction when it interacts with surfaces and materials in the scene. When thermal radiation reflects off surfaces, it creates patterns that closely resemble objects or shadows, leading to ambiguity in interpretation. The formation and appearance of these reflections are influenced by the scene geometry. Moreover, the angle of incidence affects the direction and intensity of reflected thermal radiation, leading to variations in shadow appearance and complexity. Hence, understanding the geometric properties and spatial relationships along with the angle of incidence within the scene is essential for accurately identifying shadow casters, and distinguishing between true objects and reflections, to accurately identify the object boundaries in the scene. The existing algorithms struggle to understand the geometric properties and spatial relationships along with the angle of incidence, as they typically operate on fixed-size image patches or regions without considering the broader spatial context that is crucial for identifying shadow casters. Hence, to identify the shadow casters and to distinguish the true object from reflections, a breath-taking algorithm named Centroid Ray-Tracing Fuzzy Clustering (CRFC), where the Ray-Tracing algorithm accurately predicts the direction and intensity of reflected thermal radiation by considering the geometry of the scene, including the arrangement of objects, surfaces, and light sources. Moreover, the Centroid-Link Fuzzy Clustering is adopted, where the centroid-link method often considers the spatial relationships between data points and inherently captures the broader spatial context, which is crucial for understanding scene geometry, and Fuzzy clustering allows pixels to have partial membership in multiple clusters, accommodating variations in thermal signatures due to changes in angles and reflections. Thus, CRFC effectively adapts to varying scene complexity by dynamically adjusting cluster centroids, thereby enabling the algorithm to better understand geometric properties, spatial relationships, and angle of incidence within

the scene. The POGAN by embedding the Discrete Ordinates Method, models emissivity effects and employs progressive attention for fine-scale thermal variations. Similarly, the CRFC module incorporates scene geometry and reflection modeling, which are not captured by MobileNet or EfficientNet. Hence, these models were selected as they are well suited in thermal scene understanding than generic lightweight backbones.

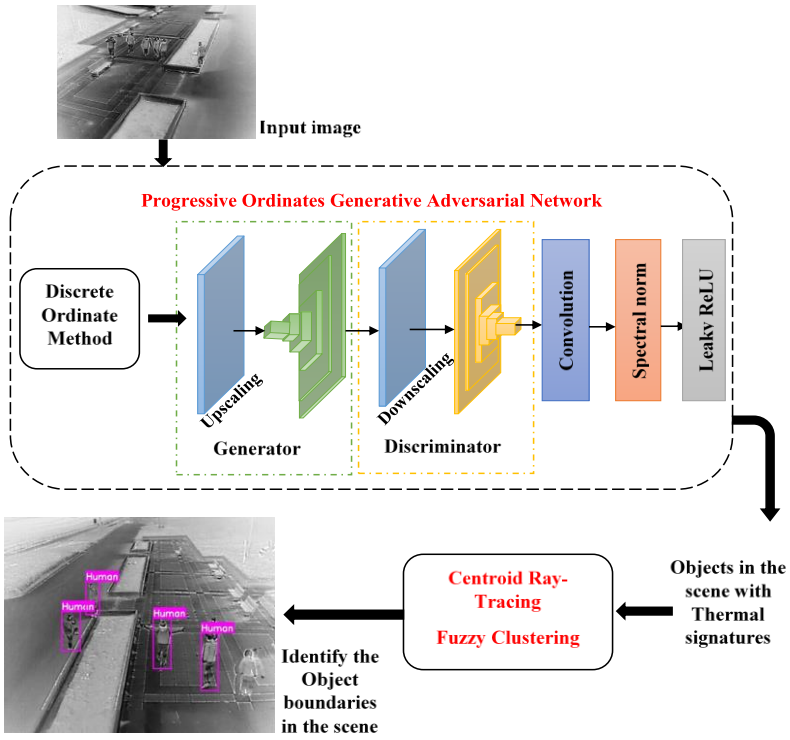


FIGURE 1. Block Diagram of The Proposed Model.

Figure 1 explains the block diagram of the proposed work, where the thermal image is given as the input for the scene object detection. Initially, the thermal signatures in the scene image which are affected by the surface properties are detected using POGAN, aiding in precise detection and characterization of objects under the objects with different surface characteristics. Then, the object boundaries impacted by shadow casters are effectively identified by CRFC empowering the algorithm to understand efficiently the spatial relationships, geometric properties, and incidence angle within the scene. By doing so, the proposed work effectively identifies the scenic objects from thermal imaging with better prediction accuracy.

3.1 PROGRESSIVE ORDINATES GENERATIVE ADVERSARIAL NETWORK

To accurately detect and characterize objects using the thermal signature, POGAN is introduced, which uses DOM and PAGAN. DOM is a numerical technique used to solve the radiative transfer equation (RTE), which describes how thermal radiation propagates through a medium. It discretizes the radiative transfer equation into a set of equations representing radiation transfer along discrete directions in space. The

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

working of DOM is explained as follows: In the context of thermal object detection, the DO method is applied to model the interaction between thermal radiation and surfaces with varying emissivity values. The radiative transfer equation (RTE) describes the transport of radiant energy through a medium and is given by Equation (1).

$$\frac{dI_\lambda}{ds} = -\chi_\lambda I_\lambda + \epsilon_\lambda N_\lambda(Tem) \tag{1}$$

Where, the spectral intensity is given as I_λ, s is the path length, the absorption coefficient is χ_λ, emissivity is ϵ_λ and the spectral radiance (Planck's function) at temperature Tem is $N_\lambda(Tem)$. The Discrete Ordinates method discretizes the angular space and solves the RTE numerically for a set of discrete directions (or ordinates). The equation for the Discrete Ordinates method is stated as per Equation (2).

$$\frac{dI_\lambda(s)}{ds} = -\chi_\lambda(s) I_\lambda(s) + \epsilon_\lambda(s) N_\lambda(Tem) \tag{2}$$

Where, the spatial coordinate is given as s and the wavelength dependency is stated as λ. By considering the directionality of radiation transfer, DOM better captures the complex interactions between thermal radiation and surfaces with varying emissivity values. It allows for more accurate modeling of how different surface properties affect the emission and absorption of thermal radiation.

Once after detecting the emissivity difference among diverse surfaces, the object detection and characterizing using thermal signature is made through PAGAN, which is a progressive growing Generative Adversarial Network (GAN). In this, the generator and discriminator are modeled as a composite function as $G = ge_{R+1}, \ldots., ge_1$ and $D = di_1, \ldots., di_{R+1}$, where every function ge_x and di_x for $x \in [1, R+1]$ resembles to a module of the generator and discriminator.

Generator: The concatenated input $[\pi, \phi(x), V_{t:t+\tau-1}]$ from a length sequence τ to sequence length 8 is stated as $\tilde{Y}_0$, employing average pooling, as a pre-processing step. The generator ge_x first layer applies the main function p as Equation (3).

$$\tilde{Y}_0 \mapsto \tilde{Y}_1 = o.f(\tilde{Y}_0) \tag{3}$$

An input sequence $\tilde{Y}_{x-1}$ is mapped as ge_x for $x \in [2, R]$ to an output sequence $\tilde{Y}_x$ using the input sequence upscaling (US) and the function $o.f$, which is represented in Equation (4).

$$\tilde{Y}_{x-1} \mapsto \tilde{Y}_x = o.f\left(US\left(\tilde{Y}_{x-1}\right)\right) \tag{4}$$

ge_x output is concatenated back to $V_{t:t+\tau-1}$, which is time features and is forwarded to the subsequent block. At last, the multivariate sequence $\tilde{Y}_R$ is reshaped

by the final layer of the generator ge_{R+1} to a univariate time series $\tilde{Y}_{x,t,\tau}$ of length $\tau = 2^{R+3}$ utilizing a spectral normalisation and one-dimensional convolution.

Discriminator: The discriminator's design is identical to the generator's architecture. The generator's output $\tilde{Y}_{x,t,\tau}$ and the time features $V_{t:t+\tau-1}$ is mapped by this to a score l. A one-dimensional convolution u_1 and a LeakyReLU activation function is used in the first module of the discriminator di_{R+1}, which is stated in Equation (5).

$$\left(\tilde{Y}_{R+1}, V_{t:t+\tau-1}\right) \mapsto \tilde{Z}_R = SN(LRL\left(u_1\left(\tilde{Y}_{R+1}, V_{t:t+\tau-1}\right)\right)) \tag{5}$$

Where, SN is the spectral normalization operator, and LRL is LeakyReLU operator. The module di_x smears a downscale operator $(DOWN)$ and the main function o, For $x \in [R+1,2]$, and is given in Equation (6).

$$Z_x \mapsto Z_{x-1} = DOWN(o(Y_x)) \tag{6}$$

The input sequence is turned into a score based on the last module h_1, and is given in Equation (7).

$$Z_1 \mapsto Z_0 = SN\left(FC\left(LRL\left(SN\left(u(o(Z_1))\right)\right)\right)\right) \tag{7}$$

Where, the fully connected layer is FC. Thus, PAGAN capture fine details and gradations present in thermal images, including variations caused by different surface properties such as texture, roughness, and color. Further, effectively capture both global and local characteristics of objects in thermal images by learning features at multiple.

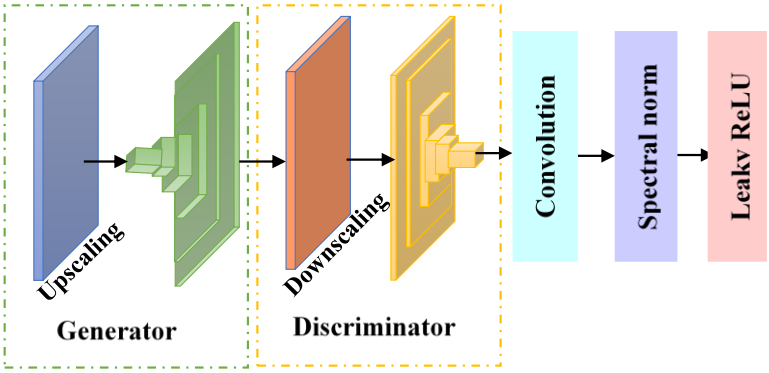


FIGURE 2. Block Diagram of PAGAN.

Figure 2 explains the architecture of PAGAN, where in the generator the upscaling and downscaling operations are done, and through fully connected layer and leaky ReLU, it captures the course details and gradations existing in thermal images, with variations triggered by diverse surface properties with varying emissivity, thereby efficiently detect and characterize the object from the scene.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Hence, the proposed POGAN accurately simulate thermal radiation by implicitly learn to distinguish between objects based on delicate differences in their thermal signatures, leading to more robust and accurate object detection results.

3.2 CENTROID RAY-TRACING FUZZY CLUSTERING

Moreover, for detecting the object boundaries in scenic image from thermal imaging, the Centroid Ray-Tracing Fuzzy Clustering is introduced. By leveraging centroid link-based fuzzy clustering, the algorithm efficiently identifies clusters of pixels with similar thermal characteristics while accounting for uncertainty and ambiguity, particularly in regions affected by shadows and reflections. Integrating ray tracing techniques allows for accurate simulation of thermal radiation interactions with scene geometry, enabling the algorithm to precisely determine the direction and intensity of reflected thermal radiation. The algorithms related to this section is explained as follows:

The foundation of ray-tracing techniques is field representations under high-frequency assumptions, meaning that the wavenumber increases to infinity. Where astigmatic wavefronts are often taken into consideration, the rays show the directions of electromagnetic wave propagation. The propagation of waves is described as the electric E and magnetic fields H, and these concepts using Equation (8) and (9).

$$\mathrm{E} = \hat{e}E_0\sqrt{\frac{\rho_1\rho_2}{(\rho_1+d)(\rho_2+d)}}e^{-jkd} \tag{8}$$

$$\mathrm{H} = \sqrt{\frac{\epsilon}{\mu}}\hat{k}\times\mathrm{E} \tag{9}$$

Where, the distance for which the wave propagates from a reference point is referred as $d \in \mathbb{R}$, the radii of curvature of the astigmatic wave at the reference point is ρ_1 and ρ_2, the wavenumber is k, the dielectric permittivity and the magnetic permeability of the propagation medium, respectively is given as ϵ and μ, and the orthonormal vectors which characterize the electric field polarization and the propagation direction, respectively is stated as $\hat{e}$ and $\hat{k} \in \mathbb{R}^3$. The output obtained from this algorithm is the clear understanding of information regarding the direction, intensity, and distribution of reflected radiation.

Then the next step is the centroid-link based fuzzy clustering, which adopts for better understanding of the geometric properties and spatial relationships within the scene, ultimately leading to more accurate object identification and boundary delineation. The centroid-link based fuzzy clustering is explained as follows: A key component of unsupervised machine learning, fuzzy membership classifiers (FCM) give a degree of membership to each class using a fuzzy membership, allowing a single piece of data to belong to two or more clusters. The partition matrix, the goal function, and the fuzzy membership function are the three fundamental operators that make up the FCM approach. With a dataset and a cluster number, the technique iterates to minimise the subsequent objective function that stated in Equation (10).

$$J(P,Q) = \sum_{i=1}^{C}\sum_{j=1}^{Nu} g_{ij}^{w} \left\|z_i - h_j\right\|^2, s.t. \sum_{j=1}^{Nu} g_{ij} = 1 \; \forall i \tag{10}$$

Where, any real number greater than 1 is w, the Euclidean norm is $\|.\|$, the membership function is $P = \{g_{ij}\}$ and is given in Equation (12), the cluster center is denoted by $Q = \{h_j\}$ and is stated in Equation (11). A local minimum for J is got by using the Lagrange multiplier approach, provided that update P and Q in accordance with procedure (1) at each step k.

$$h_j^{(k+1)} = \frac{\sum_{i=1}^{C}\left(g_{ij}^{(k)}\right)^{w} z_i}{\sum_{i=1}^{C}\left(g_{ij}^{(k)}\right)^{w}} \tag{11}$$

$$g_{ij}^{(k+1)} = \left(\sum_{l=1}^{Nu}\left(\frac{\left\|z_i - h_j^{(k+1)}\right\|}{\left\|z_i - h_l^{(k+1)}\right\|}\right)^{\frac{2}{w-1}}\right)^{-1} \tag{12}$$

Here, the centroid link is involved, which connect centroids of neighboring clusters based on spatial proximity. This can be represented by a connectivity matrix W, where $wt_{ij} = 1$ if centroids i and j are connected, and $wt_{ij} = 0$ otherwise. This work incorporates centroid linking into fuzzy c-means by modifying the objective function to include spatial information. This is achieved by introducing a penalty term that encourages neighboring centroids to have similar membership degrees. The modified objective function is given in Equation (13).

$$J_{CL}(P,Q) = J(P,Q) + \lambda \sum_{i=1}^{C}\sum_{j=1}^{C} wt_{ij} \sum_{l=1}^{Nu}\left(g_{il} - g_{jl}\right)^2 \tag{13}$$

Where, weighting parameter controlling the influence of centroid linking is λ and the connectivity between centroids i and j is wt_{ij}. This objective function encourages neighboring centroids to have similar membership degrees, promoting spatial coherence in the clustering result, thus facilitating a comprehensive understanding of scene geometry in thermal imagery analysis.

Figure 3 depicts the process flow of proposed Centroid Ray-Tracing Fuzzy Clustering, where the thermal image is initially subjected to ray-tracing that determines the valuable thermal information for subsequent processing, then the next step is centroid link-based fuzzy clustering where FCM understands the geometric properties and spatial relationships, then the centroid link connects the neighboring centroids based on spatial proximity, which provides the final output with accurate object identification and boundary delineation from thermal images.

Overall, POGAN is adopted to detect the thermal signatures from the object in the scene irrespective of their surface properties, which hence aids in the precise detection and characterization of objects. Then the CRFCis adopted to detect the object

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

boundaries that disrupted by the shadows or reflection from other objects, thereby aids in accurate classification of objects in the scene with increased performance.

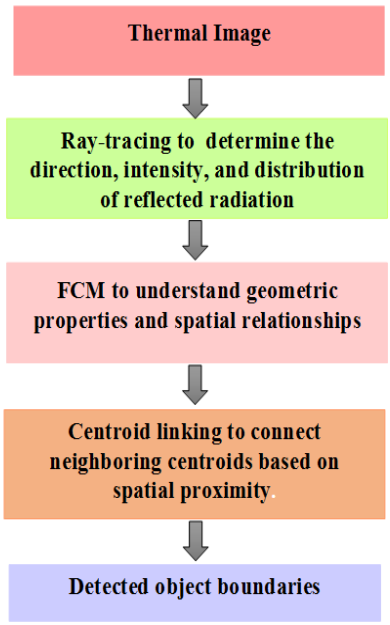


FIGURE 3. Process Flow of the Centroid Ray-Tracing Fuzzy Clustering.

4. RESULTS AND DISCUSSIONS

This section comprises a systematic discussion of the implementation results, as well as the performance of the proposed system, and a comparison section to validate the performance of the proposed POGAN and CRFCmethods in thermal object detection from the scenic image.

4.1 DATASET DESCRIPTION

The data set used for testing the proposed system is the FLIR ADAS v2 3K dataset, a comprehensive collection to support research in Advanced Driver Assistance Systems (ADAS) and autonomous vehicles. It contains both thermal and visible spectrum images, making it ideal for developing object detection systems using convolutional neural networks (CNNs) and for exploring RGB-T (visible + thermal) sensor fusion techniques to enhance automotive safety and efficiency. The dataset includes 26,442 fully annotated frames with approximately 520,000 bounding boxes across 15 object categories such as person, bike, car, motorcycle, bus, train, truck, traffic light, fire hydrant, street sign, dog, skateboard, stroller, scooter, and other vehicles. The images are divided into 9,711 thermal and 9,233 RGB frames, with 7,498 frames recorded from videos at 24 Hz, providing a one-to-one match between thermal and visible images. Annotations are available in MSCOCO and Conservator JSON formats. The thermal images were captured using a Teledyne FLIR Tau 2 camera (640x512 resolution, 13mm f/1.0 lens), while the visible images were acquired using a Teledyne FLIR Blackfly S BFS-U3-51S5C camera with a 52.8° HFOV lens.

This dataset is particularly valuable for training and evaluating models in object detection, classification, sensor fusion, and low-light or nighttime driving scenarios, supporting the development of robust ADAS and autonomous vehicle systems. The data set is taken from https://www.kaggle.com/datasets/niteshc7r/datasets-for-object-detection-night-and-thermal.

4.2 EXPERIMENTAL SETUP

This section provides a detailed description of the implementation findings and the performance of the proposed system, which is simulated in Python. It also includes a comparison section to verify the outback of the proposed system.

- Software : Python
- OS : Windows 10
- Processor : Intel i5
- RAM : 16GB

4.3 EXPERIMENTAL RESULTS

The section emphasizes the outcomes of the proposed model's innovative design setup when implemented in python for proving the accuracy in thermal image detection.

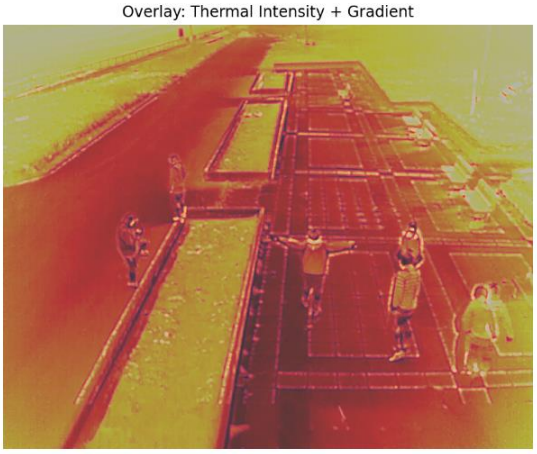


FIGURE 4. Image Representation of Thermal Intensity with Gradient.

The figure 4 depicts the image exhibiting thermal intensity and gradient in which thermal intensity of the base image depicts the raw radiative emission patterns of objects and surfaces, and highlights differences in temperature throughout the scene. Meanwhile, the gradient features capture spatial intensity transitions, depicting edges, object boundaries, and structural features that may appear less pronounced in pure thermal images. By combining intensity and gradient information, the overlay provides a richer representation where both emissivity-driven temperature variations and fine structural details are visible simultaneously. This fusion improves the visibility of object boundaries while reducing uncertainty caused by emissivity variations and diffuse radiation.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

FIGURE 5. Progressive Feature Representation in The POGAN For Thermal Scene Analysis.

Figure 5 depicts intermediate feature maps extracted at different progressive levels in the POGAN framework. Each level of features demonstrates infinite emissivity-based and structural characteristics of the thermal scene objects. POGAN Feature 1 shows global emissivity differences based on coarse-level patterns of thermal intensity that are captured. Whereas Feature 2 enables a better differentiation between the foreground and background elements in the scene by enhancing object boundaries and structural variations. At the same time, POGAN Feature 3 allows for a differentiation between objects with reflective versus matte surfaces, based on learning tiny temperature gradient, and fine-scaled emissivity variations. Finally, to support proper localization and recognition of identified objects in the scene, POGAN Feature 4 presents a high-level, refined semantic representation that encapsulates contextual and emissivity based details. Thus the POGAN efficiently models emissivity effects and differences with this progressive feature extraction.

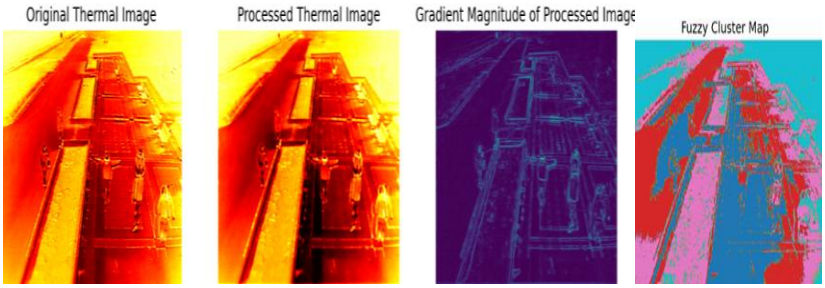


FIGURE 6. POGAN-CRFC Framework for Thermal Scene Object Detection.

The entire POGAN-CRFC framework for thermal scene object detection is illustrated in Figure 6 where the first image is input raw thermal image having objects with varying emissivity values and reflective surfaces. After applying the POGAN module, which has the Discrete Ordinates Method to model emissivity effects, while the Progressive Attention mechanism extracts hierarchical features at multiple scales, enhancing fine-grained detection of emissivity variations. Then the CRFC Module incorporates geometric and radiative transfer modeling where the ray-tracing simulates thermal radiation reflection and refraction considering incidence angle and scene geometry, while fuzzy clustering dynamically adapts cluster centroids, enabling

partial pixel memberships to better distinguish true objects from reflections and thermal shadows. Hence, an accurate object localization and boundary delineation in thermal scenes is obtained as output which is robust against emissivity inconsistencies, reflections, and shadow artifacts.

FIGURE 7. Histogram of The Original Image and The Processed Image.

The histogram of the original image and that of the processed image is illustrated in Figure 7 which shows the original raw thermal image has fairly smooth, broad intensity distribution that exhibits multiple peaks and valleys. This indicates emissivity inconsistency, diffuse reflections and environmental background interference which would reduce contrast between objects against backgrounds. Meanwhile, the histogram of processed image at the right is sharply peaked with well-defined, well separated peaks. The height of the frequencies represents contrast between peak intensities where the higher frequency demonstrates improved delineation and contrast of the objects against their background. Particularly, the processed histogram illustrates less overlap in the intensity ranges between object and background intensity, allowing for greater feature separability during the detection and clustering stage.

FIGURE 8. The Intensity Distributions of The Original and Processed Image.

The intensity distributions of the original image as well as the processed image is represented using Figure 8. Here, the original image has relatively smooth intensity distribution but sometimes it exhibits abrupt spikes which is due to reflections and emissivity distortions hindering detection of object boundaries and reduce reliability. Meanwhile, the processed image has more consistent background levels and well-defined peaks at object locations, indicating improved boundary contrast and reduced influence of emissivity-related noise. This is due to the proposed model's reduction

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

of baseline intensity fluctuations while preserving sharper transitions at object boundaries.

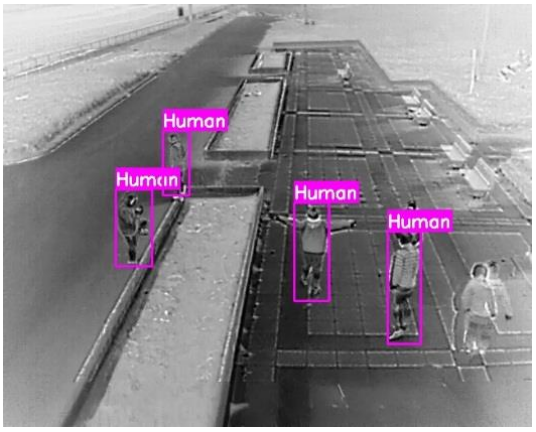


FIGURE 9. The Output Image Predicting Object Boundaries.

The final output of image detection in thermal imaging in detecting the object boundaries is depicted in Figure 9. This shows the effectiveness of the proposed framework in alleviating emissivity variations, which usually influence apparent temperatures, suppressing reflections and shadows through geometry aware ray-tracing and fuzzy clustering and enhancing the definition of boundaries through gradient base feature extraction. This image illustrates accurate image detection and is a proof of testing with real-world images ensuring the robustness from this proposed model.

4.4 PERFORMANCE ANALYSIS OF PROPOSED WORK

In this section, a detailed clarification of the efficiency of the proposed technique and the achieved outcome are enlightened. The deployment metrics used are

Accuracy: It describes the degree to which a measurement or prediction is correct or true.

$$\frac{\mathrm{TP}+\mathrm{TN}}{(TP+FN)+(FP+TN)} \tag{14}$$

F1-score is a performance metric for classification models that calculates the harmonic mean of precision and recall, providing a balanced measure of both.

$$\mathrm{F1}-\mathrm{score}=\frac{2*\mathrm{TP}}{2*TP+FN+FP} \tag{15}$$

Mean accuracy refers to the overall correctness of a classification model, calculated by dividing the total number of correct predictions (True Positives and True Negatives) by the total number of observations.

Mean Absolute Error (MAE) a statistical metric that measures the average magnitude of errors in a set of data, calculated by taking the absolute difference between predicted and actual values, summing these differences, and dividing by the total number of observations.

$$MAE = \frac{1}{n}\sum_{1=1}^{n}|x_i - x| \tag{16}$$

Mean Average Precision (mAP) is a common metric for evaluating the performance of information retrieval and object detection models, measuring the average precision across all queries or classes. AP_k is average precision of class k.

$$\text{mAP} = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{17}$$

Intersection of Union (IoU) is a performance metric that evaluate the accuracy of annotation, segmentation, and object detection algorithms. Mean IoU of the image is calculated by taking the IoU of each class and averaging them.

$$\text{IoU (Jaccard Index)} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{18}$$

Precision measures the trustworthiness of positive predictions by calculating the ratio of True Positives (TP) to the total number of positive predictions (TP + False Positives).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{19}$$

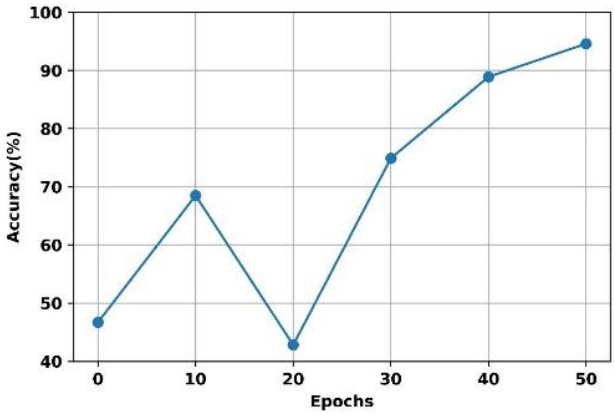
Recall measures the model's capability to capture all relevant objects in the image.

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{20}$$

Time: It measures real-time object detection, which is the ability to detect and localize objects in a continuous stream of data.

Sensitivity is the ability of a model to correctly identify positive cases.

$$\text{Sensitivity } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$



ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

FIGURE 10. Performance of Accuracy for Proposed Approach

Figure 10 defines the performance regarding the accuracy metrics for the proposed system under varying epoch values. Accuracy increases with the increases in the epoch value. From the graph, it is clear that the accuracy of proposed achieves minimum value of 42.97 when the epoch is at 20, and achieves the maximum value of 94.6 when the epoch is at 50. The proposed POGAN techniques provide contextual understanding of meaningful feature representations and capture important characteristics of thermal objects and their surroundings, leading to improved accuracy.

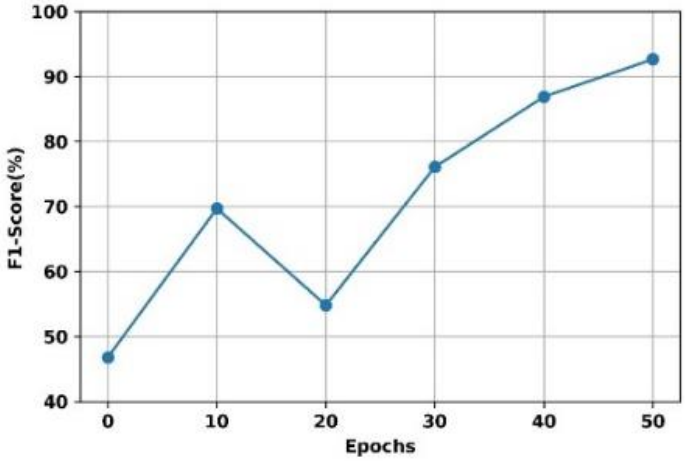


FIGURE 11. Performance of F1-Score for Proposed Approach.

Figure 11 outlines the performance regarding the F1-score metrics for the proposed system under varying epoch values. F1 score increased with the increases in the epoch value. From the graph, it is clear that the F1-score of proposed achieves minimum value of 48.57 when the epoch is at 0, and achieves the maximum value of 92.67 when the epoch is at 50. The proposed CRFCconsiders intensity values along with spatial and geometric characteristics, which helps better discriminate between true object boundaries and spurious thermal features, thus boosting the F1-score performance.

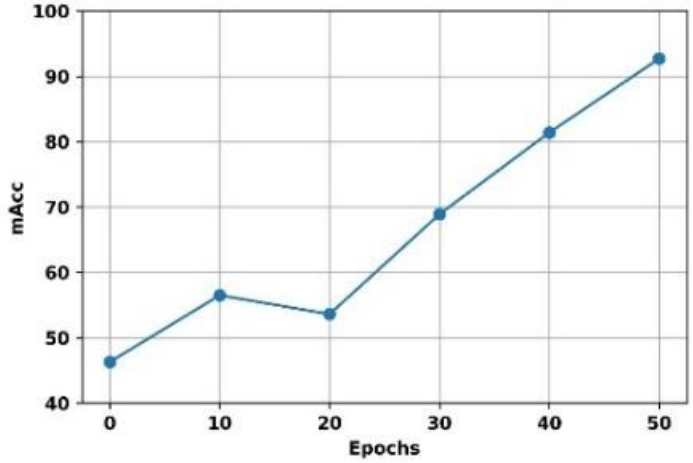


FIGURE 12. Performance of Mean Accuracy for Proposed Approach.

Figure 12 delineates the performance regarding the mean accuracy metrics for the proposed system under varying epoch values. mAcc increases with the increases in the epoch value. From the graph, it is clear that the mean accuracy of proposed achieves minimum value of 47.8 when the epoch is at 0, and achieves the maximum value of 92.7 when the epoch is at 50. The proposed CRFCtechnique trace the paths of thermal rays that interact with surfaces and objects for understanding the geometric properties and spatial relationships within the scene, thereby improves the mean accuracy performance.

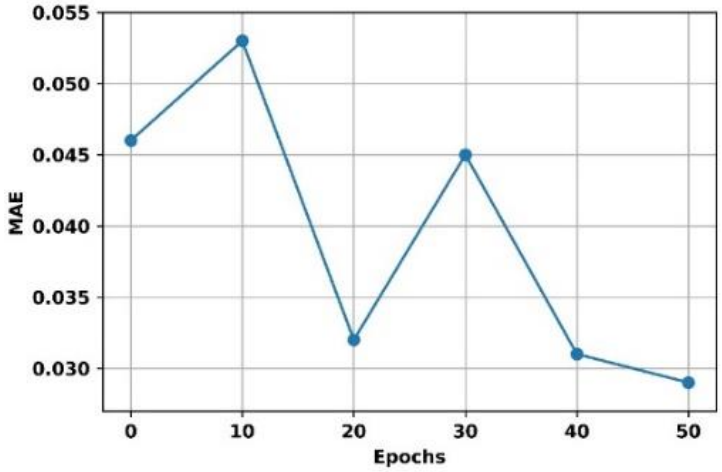


FIGURE 13. Performance of Mean Absolute Error for Proposed Approach.

Figure 13 expresses the performance regarding the MAE metrics for the proposed system under varying epoch values. The MAE decreases with the increases in the epoch value. From the graph, it is clear that the MAE of proposed achieves maximum value of 0.047 when the epoch is at 0, and achieves the minimum value of 0.029 when the epoch is at 50. The proposed POGAN technique is trained to generate realistic thermal images, which implicitly learns to represent thermal objects by capturing finer details and context-specific features relevant to object detection tasks, thereby reducing the mean average error.

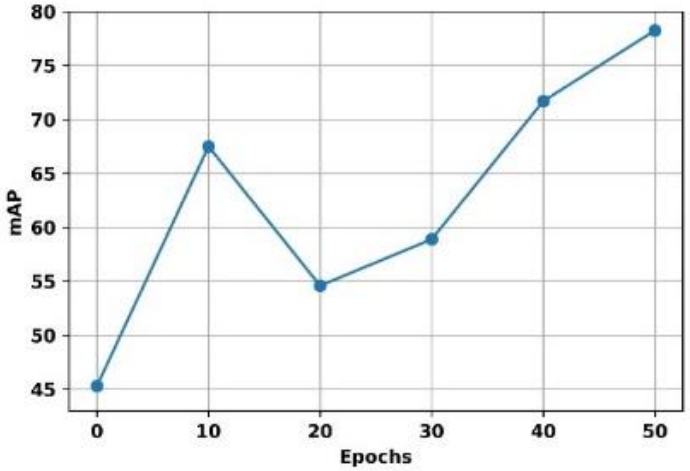


FIGURE 14. Performance of Mean Average Precision for Proposed Approach.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Figure 14 states the performance regarding the mAP metrics for the proposed system under varying epoch values. mAP it increases with the increases in the epoch value. From the graph, it is clear that the mAP of proposed achieves minimum value of 45.23 when the epoch is at 0, and achieves the maximum value of 78.97 when the epoch is at 50. The proposed CRFCallows for the flexible assignment of pixels to clusters, accounting for uncertainty in boundary delineation, enabling the algorithm to capture the gradual transition between object and background regions, thereby enhancing the mAP.

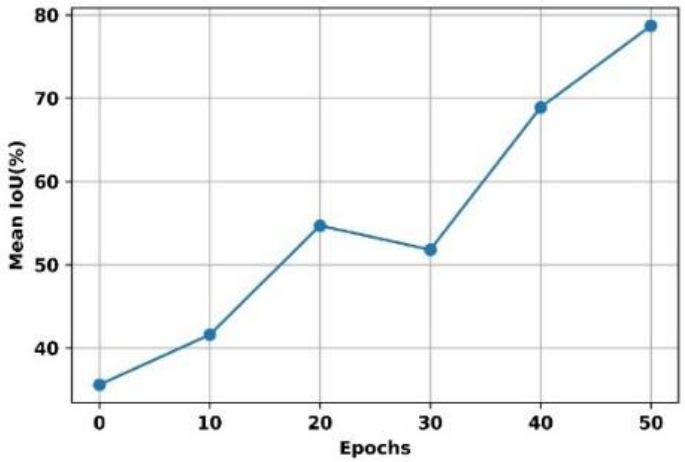


Figure 15. Performance of Mean Intersection of Union For Proposed Approach.

Figure 15 defines the performance regarding the mean IoU metrics for the proposed system under varying epoch values. The mean IoU increases with the increases in the epoch value. From the graph, it is clear that the mean IoU of proposed achieves minimum value of 36.17 when the epoch is at 0, and achieves the maximum value of 78.7 when the epoch is at 50. The proposed CRFCtechnique considers the spatial relationships between pixels by linking centroids based on their proximity accurately delineating thermal object boundaries and shadow, which improves the IoU score.

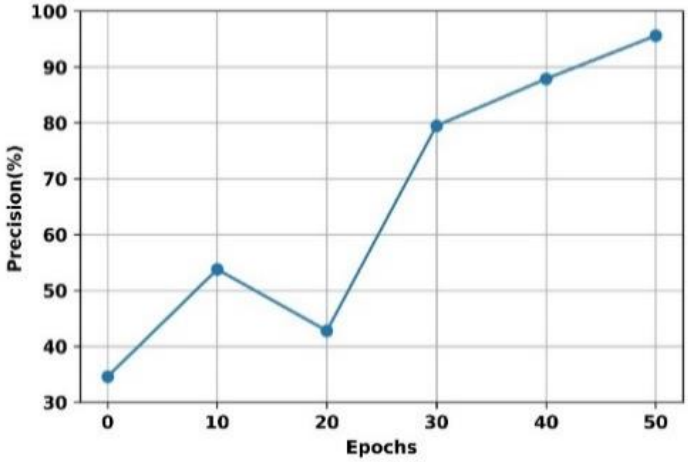


FIGURE 16. Performance of Precision for Proposed Approach.

Figure 16 defines the performance regarding the precision metrics for the proposed system under varying epoch values. The precision increases with the increases in the epoch value. From the graph, it is clear that the precision of proposed achieves minimum value of 34.25 when the epoch is at 0, and achieves the maximum value of 95.63 when the epoch is at 50. The proposed POGAN technique with progressive training allows the model to learn intricate details and features of thermal objects and scenes at different scales, leading to improved precision in object detection.

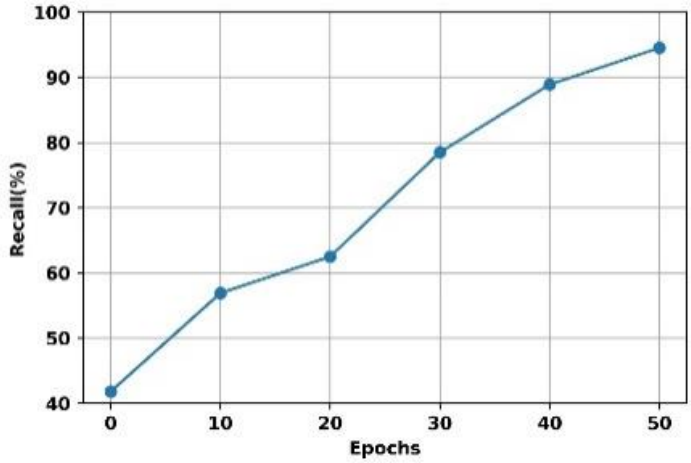


FIGURE 17. Performance of Recall for Proposed Approach.

Figure 17 defines the performance regarding the recall metrics for the proposed system under varying epoch values. The recall increases with the increases in the epoch value. From the graph, it is clear that the recall of proposed achieves minimum value of 41.38 when the epoch is at 0, and achieves the maximum value of 94.53 when the epoch is at 50. The proposed CRFC technique effectively ability to accurately capture thermal object boundaries, even in challenging scenarios with shadows and reflection, thereby increasing recall rates in thermal imagery analysis.

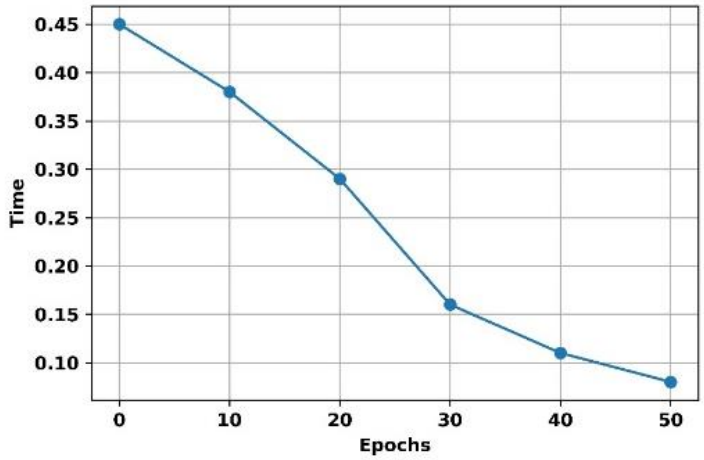


FIGURE 18. Performance of Time for Proposed Approach.

Figure 18 expresses the performance regarding the time metrics for the proposed system under varying epoch values. The time decreases with the increases in the epoch value. From the graph, it is clear that the time of proposed achieves maximum value of 0.45 when the epoch is at 0, and achieves the minimum value of 0.08 when the epoch is at 50. The proposed POGAN techniques with ordinates-based representation

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

educes the computational complexity of the network by simplifying the input data while preserving essential thermal characteristics, thereby achieving faster execution times without sacrificing detection accuracy.

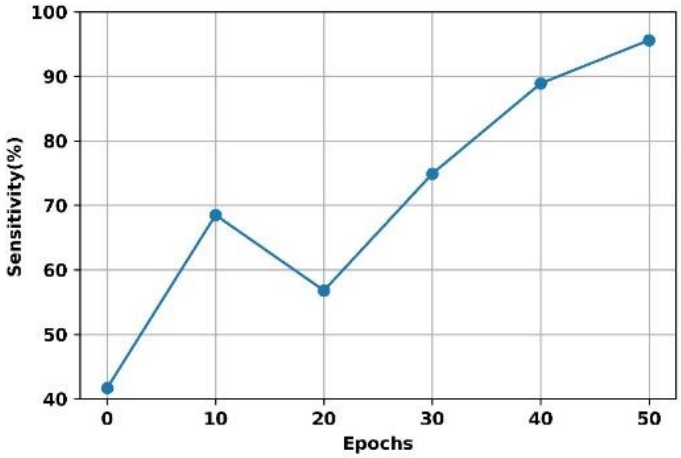


FIGURE 19. Performance of Sensitivity for Proposed Approach.

Figure 19 defines the performance regarding the sensitivity metrics for the proposed system under varying epoch values. The sensitivity increases with the increases in the epoch value. From the graph, it is clear that the sensitivity of proposed achieves minimum value of 42.97 when the epoch is at 20, and achieves the maximum value of 96.81 when the epoch is at 50. The proposed POGAN technique with progressive training allows the network to learn increasingly detailed features and spatial relationships in the data, thereby increasing the sensitivity performance in object detection task.

4.5 COMPARATIVE ANALYSIS OF PROPOSED WORK

This section emphasizes the effectiveness of the proposed model by comparing it with the outcomes of existing methodologies and illustrating their outcomes based on several metrics. The comparisons are made from the previous techniques with various metrics such as accuracy, precision, recall, sensitivity, mAP, MAE, mean Accuracy, time, mIoU and F1-score. Comparisons are made with the existing techniques such as Global Convolutional Network (GCN), Dense Upsampling Convolution (DUC), Discriminative Feature Network (DFN), Edge-Conditioned Convolutional Neural Network (EC-CNN) [19], Full-Resolution Residual Networks (FRRN), Bilateral Segmentation Network (BiSeNet), Segmentation High-Resolution Network (SegHRNet), RGB-thermal fusion network (RTFNet) [23], Progressively Complementarity Aware Fusion Network (PCF), Adaptive Fusion Network (AFNet), Depth-induced Multi-scale Recurrent Attention Network (DMRA), cross modality Saliency Generative Adversarial Network (cmSalGAN) [17], Single Shot multibox Detector (SSD), Region-based Fully Convolutional Networks (R-FCN), Path Aggregation Network (PANet) and Thermal Infrared Network (TIRNet) [16].

Figure 20 defines the comparative analysis on accuracy of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves an accuracy of 92.8%, 93.1%, 93.5% and 93.8%, respectively,

the proposed model achieves higher accuracy value of 94.6%. The proposed method achieves maximum accuracy while the GCN achieves minimum accuracy.

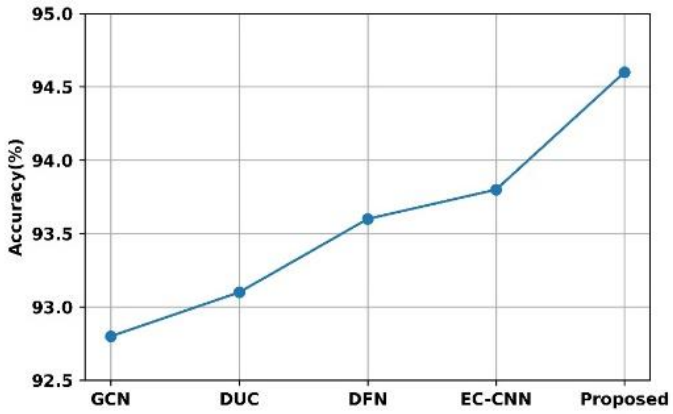


FIGURE 20. Comparison on Accuracy of Proposed Approach.

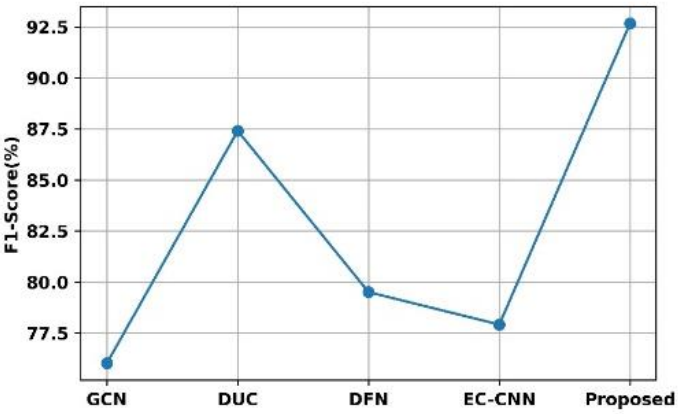


FIGURE 21. Comparison on F1-Score of Proposed Approach.

Figure 21 describes the comparative analysis on F1-score of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a F1-score of 75.9%, 87.43%, 79.26% and 78.6%, respectively, the proposed model achieves higher F1-score value of 92.67%. The proposed method achieves maximum F1-score while the GCN achieves minimum F1-score.

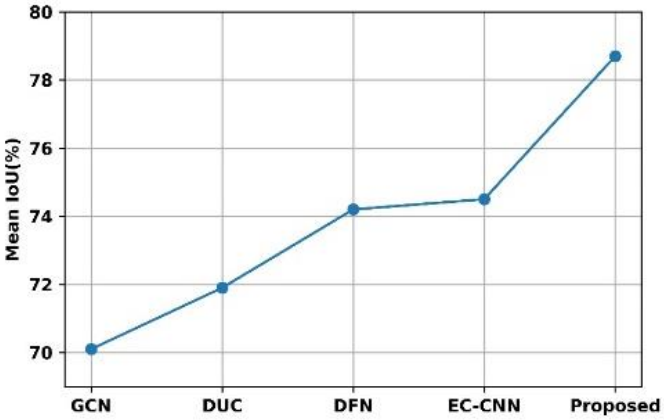


FIGURE 22. Comparison on Mean Intersection of Union of Proposed Approach.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Figure 22 describes the comparative analysis on mean IoU of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a mean IoU of 70.18%, 71.87%, 74.13% and 74.28%, respectively, the proposed model achieves higher mean IoU value of 78.7%. The proposed method achieves maximum mean IoU while the GCN achieves minimum mean IoU.

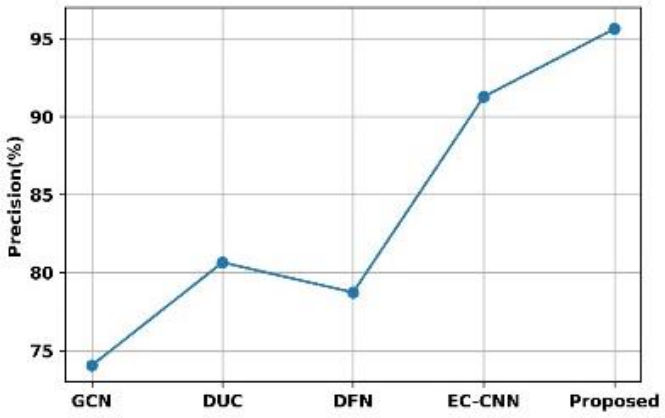


FIGURE 23. Comparison on Precision of Proposed Approach.

Figure 23 describes the comparative analysis on precision of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a precision of 75.77%, 82.13%, 78.96% and 92.46%, respectively, the proposed model achieves higher precision value of 95.63%. The proposed method achieves maximum precision while the GCN achieves minimum precision.

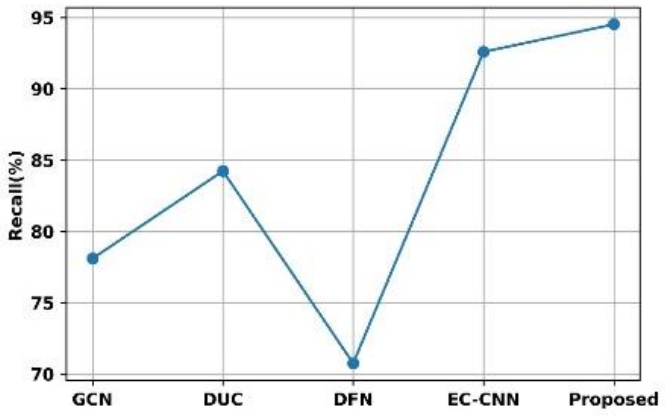


FIGURE 24. Comparison on Recall of Proposed Approach.

Figure 24 describes the comparative analysis on Recall of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a Recall of 77.98%, 84.93%, 71.31% and 93.7%, respectively, the proposed model achieves higher Recall value of 94.53%. The proposed method achieves maximum Recall while the DFN achieves minimum Recall.

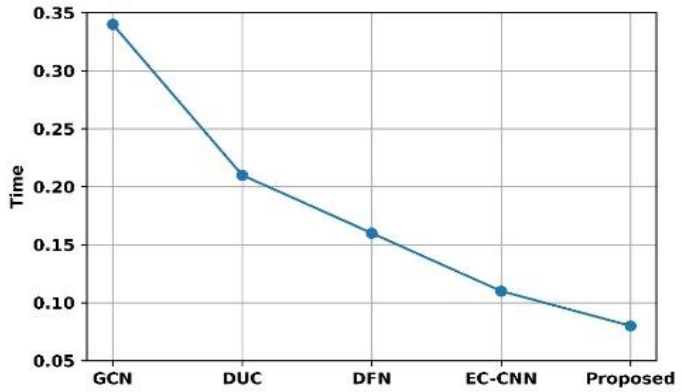


FIGURE 25. Comparison on Time of Proposed Approach.

Figure 25 describes the comparative analysis on time of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a time of 0.34s, 0.21s, 0.16s, and 0.11s respectively, the proposed model has less time value of 0.08s. The proposed method achieves minimum execution time while the GCN achieves maximum time for execution.

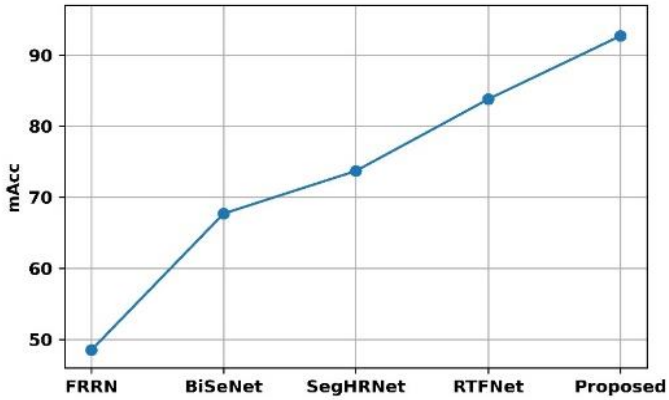


FIGURE 26. Comparison on Mean Accuracy of Proposed Approach.

Figure 26 describes the comparative analysis on Mean Accuracy of proposed model over traditional models. On comparing the traditional models such as FRRN, BiSeNet, SegHRNet and RTFNet achieves a Mean Accuracy of 49.5%, 68.9%, 74.2% and 84.7%, respectively, the proposed model achieves higher Mean Accuracy value of 92.7%. The proposed method achieves maximum Mean Accuracy while the GCN achieves minimum Mean Accuracy.

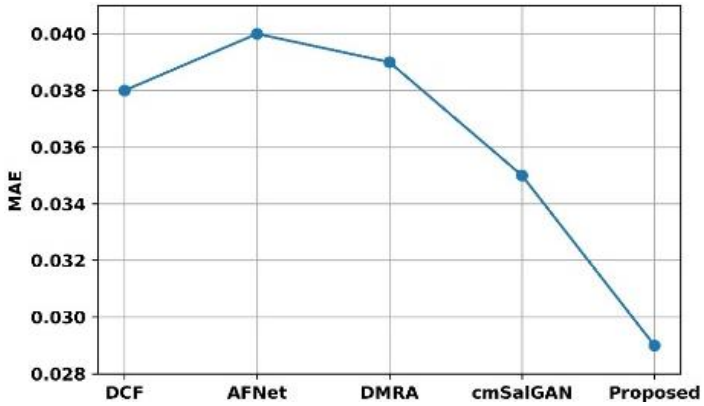


FIGURE 27. Comparison on Mean Average Error of Proposed Approach.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Figure 27 describes the comparative analysis on MAE of proposed model over traditional models. On comparing the traditional models such as DCF, AFNet, DMRA and cmSalGAN achieves a MAE of 0.038, 0.039, 0.0387, and 0.035, respectively, the proposed model achieves higher MAE value of 0.029. The proposed method achieves minimum MAE while the AFNet achieves maximum MAE.

Figure 28 describes the comparative analysis on mAP of proposed model over traditional models. On comparing the traditional models such as SSD, R-FCN, PANet, and TIRNet achieves a mAP of 57.58%, 69.53%, 73.86% and 74.7%, respectively, the proposed model achieves higher mAP value of 78.24%. The proposed method achieves maximum mAP while the GCN achieves minimum mAP.

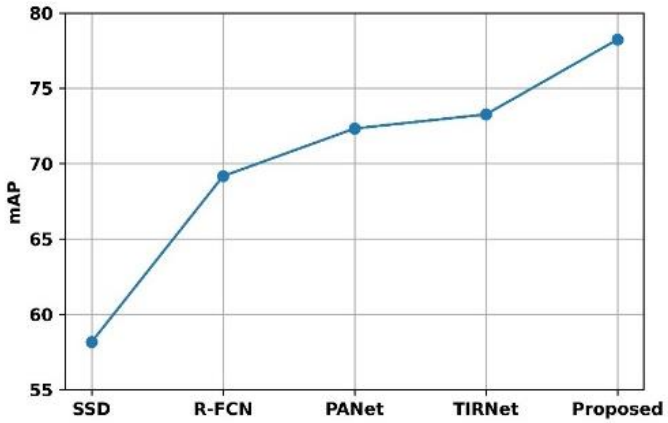


FIGURE 28. Comparison on mean Average Precision of Proposed Approach.

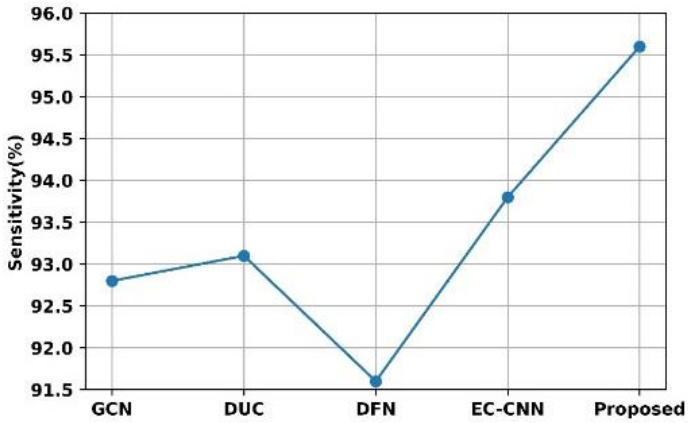


FIGURE 29. Comparison on Sensitivity of Proposed Approach.

Figure 29 describes the comparative analysis on sensitivity of proposed model over traditional models. On comparing the traditional models such as GCN, DUC, DFN and EC-CNN achieves a sensitivity of 92.78%, 93.12%, 91.58% and 93.78%, respectively, the proposed model achieves higher sensitivity value of 96.81%. The

proposed method achieves maximum sensitivity while the DFN achieves minimum sensitivity.

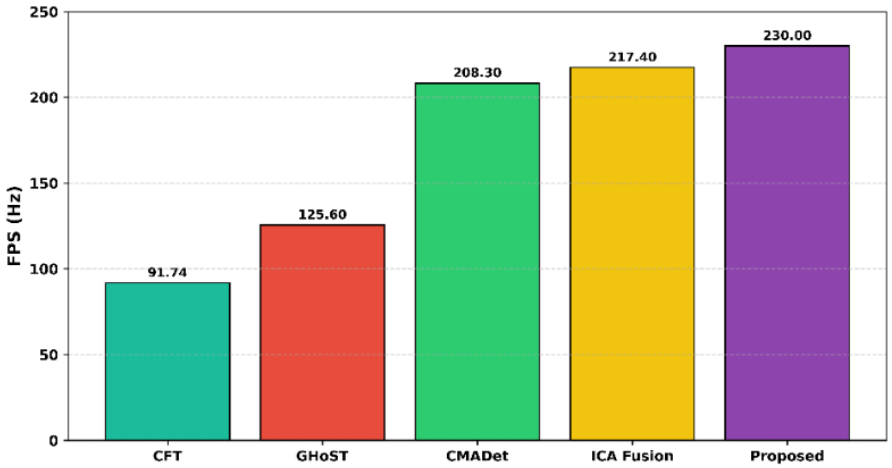


FIGURE 30. Comparison Analysis of FPS with Existing Methods.

Figure 30 examines the frame per second (FPS) in Hz of the proposed model and compared with existing techniques CFT, GHOST, CMADet, and ICA Fusion [26]. CFT has a lower FPS of 91.74Hz, GHOST and CMADet come next in having lower FPS values of 125.60Hz and 208.30Hz whereas ICA Fusion has a high FPS of 217.40Hz. Meanwhile, the proposed model has a highest FPS of 230Hz indicating its superior computational efficiency and faster processing capability. This improvement is attributed to the CRFCmodule that dynamically adapts to environmental challenges by modeling geometric properties, spatial relationships, and incidence angles, thereby improving boundary identification and the feature grouping process. This reduction in redundant computations enhances computational efficiency, resulting in the highest FPS of 230 Hz, outperforming all existing approaches.

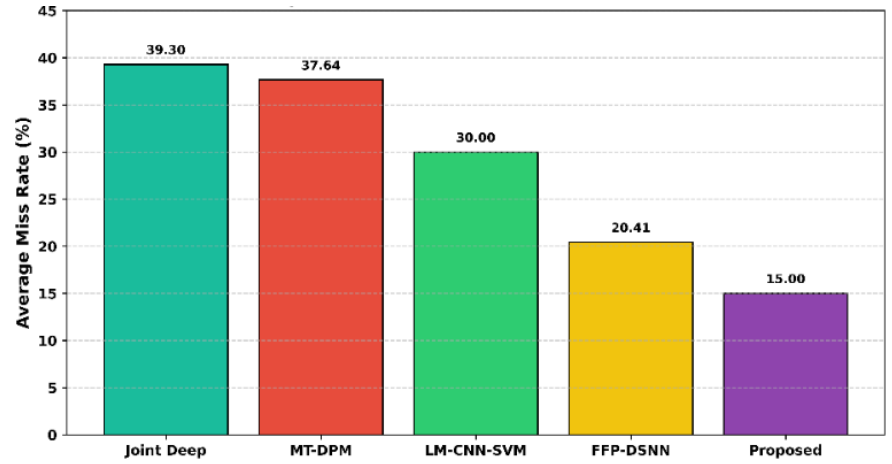


FIGURE 31. Comparison analysis of AMR with Existing Methods.

The comparative analysis of the Average Miss Rate (AMR) of the proposed method with the existing methods is represented in Figure 31 [27]. The existing method Joint Deep has AMR of 39.30%, while MT-DPM has a slightly lower AMR of 37.64%. Whereas the AMRs of LM-CNN-SVM and FFP-DSNN are even lower of

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

30% and 20.41%. On the other hand, the proposed model exhibits the least AMR of 15% which is due to POGAN that generate consistent thermal signatures for image in various scenic characteristics, to address the emissivity variation problem. This enhancement enables more accurate recognition and reduces false negatives, thereby significantly lowering the AMR to 15%, the best among all compared methods.

4.6 DISCUSSION

The overall analysis explained that the proposed method achieves better performance in thermal object detection from the scene in terms of accuracy with 94.6%, F1-score of 92.67%, mean IoU of 78.7%, Precision of 95.63%, Recall value of 94.53%, Mean Accuracy value of 92.7%, mAP value of 78.24%, sensitivity value of 96.81%, with minimized time and MAE of 0.08s and 0.029, respectively. This analysis thus validates the betterment of the proposed work over the traditional models in all aspects thus improves the object prediction from thermal images.

TABLE 1.
Comparative Analysis of Accuracy, F1-score, Mean IoU, Precision, Recall, and Time.

Methods	Accuracy (%)	F1-score (%)	Mean IoU (%)	Precision (%)	Recall (%)	Time (s)
GCN	92.8	75.9	70.18	75.77	77.98	0.34
DUC	93.1	87.43	71.87	82.13	84.93	0.21
DFN	93.5	79.26	74.13	78.96	71.31	0.16
EC-CNN	93.8	78.6	74.28	92.46	93.7	0.11
proposed	94.6	92.67	78.7	95.63	94.53	0.08

Table 1 compares various performance metrics of the conventional CNN backbones such as GCN, DUC, DFN, EC-CNN, MobileNet, EfficientNet with the proposed model. Comparatively, the proposed method achieved higher Accuracy, F1-score, Mean IoU, Precision, Recall, and reduced Time. The conventional models learn only intensity features, omitting the intensity variations, whereas the proposed POGAN embeds the Discrete Ordinates Method to explicitly model emissivity variations, enabling consistent detection across objects with reflective, rough, or matte surfaces. In addition, the existing approaches fail to distinguish between true objects and reflected thermal artifacts. The CRFC module in the proposed framework integrates ray-tracing for accurate reflection modeling with fuzzy centroid-link clustering to incorporate spatial membership, enabling robust interpretation of shadows and reflections under varying scene geometries. The highest F1-score and Mean IoU demonstrates its superior robustness in distinguishing true object boundaries from shadows and reflections. In addition to accuracy gains, the framework is computationally efficient with shortest time of 0.08s, outperforming EC-CNN and GCN, making it suitable for real-time thermal surveillance applications.

TABLE 2
Comparative Analysis of Prior Works with The Proposed Work.

Ref No	Contribution	Limitation	Proposed Research work
[16]	CNN-based lightweight TIR detection, residual branch for robust features, continuous information fusion for background/occlusion handling.	Fails to merge complementary TIR and RGB properties, limiting multimodal effectiveness.	Proposed work integrates RGB and thermal images via MLLMs without requiring localization, enabling cross-modality reasoning and scalability
[17]	Multi-interactive dual-decoder for RGBT SOD	Integration challenges, modality imbalance	Flexible multimodal fusion achieved by in-context learning, avoiding handcrafted feature interaction pipelines.
[18]	Compared dash and thermal cameras for road safety; YOLOv8-based fusion for object detection; showed thermal excels in low-light	High computational complexity; not optimized for edge devices in real time.	Real-time inference enabled using simpler multimodal MLLM-based detection and classification pipeline
[19]	First panoramic thermal camera for 360° surveillance; useful for calamity monitoring, trespassing detection.	Lack of MOT module evaluation; not tested under adversarial weather conditions.	Robustness achieved across varied weather and ITS deployment with both IR and RGB synergy
[20]	Thermal-based dynamic traffic light system with attention + feature extraction (YOLO-Thermal); addressed mobility-impaired pedestrian scenarios	Sole reliance on thermal limited semantic richness and object classification reliability.	Multimodal image reasoning enable improving semantic interpretation across classes
[21]	Cross-guided fusion for RGB-T with CSAGF & GFM modules for RGB-T salient object detection;	Lacks effective mining of high-level semantics and global context support.	Zero-shot reasoning assists high-level scene understanding, bypassing fixed fusion blocks
[22]	Vehicle detection in weak IR aerial images (occluded); transfer learning from visible to IR datasets; YOLOv4 + hard negative mining.	Limited by model structure and loss function, and low accuracy.	Instead of YOLO, MLLMs are used to handle scene discernment and multimodal object detection.
[23]	First large-scale benchmark with SAFit	Misaligned image pairs; not specialized for ITS; limited direct deployment.	ADS/ITS context with fusion-based detection and classification, improves applicability
[24]	ECFFNet with multi-level fusion	Lacks recent advanced RGB-T SOD methods	Proposed framework introduces scalable zero-shot reasoning

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

			without reliance on handcrafted fusion architectures.
[25]	Thermal perception in extreme weather; learnable projection function, super-resolution, bidirectional feature pyramid network for lightweight optimization.	Extreme weather; learnable projection function, super-resolution, bidirectional feature pyramid network for lightweight optimization. Strong reliance on dataset-specific optimizations; limited generalization across tasks.	Proposed method uses generalizable MLLMs for multimodal detection + scene discernment, enhancing scalability and robustness.

Table 2 analyses the existing works which explored thermal-only detection ([16], [19], [20]) or RGB-T fusion with CNNs ([17], [21], [24]), they depend on handcrafted fusion strategies, dataset-specific training, or heavy architectures. Others addressed domain-specific problems such as traffic light control, panoramic monitoring, and occluded detection but did not generalize well to ITS and ADS applications. Meanwhile, the proposed work is distinct because it leverages multimodal large language models (MLLMs) with zero-shot in-context learning, enabling task adaptability without architectural redesign. The proposed integrates IR and RGB images without localization data, compensating modality weaknesses by using a reasoning-based fusion. The novel method supports scene discernment and object detection simultaneously, which is beyond most prior studies that focused on one aspect. On addition, it also provides scalability to diverse ITS or ADS scenarios while prior methods remain domain- or dataset-specific.

5. CONCLUSION

This paper proposed a Progressive Ordinates GAN with Centroid Fuzzy Ray-Tracing model for effective object detection from scene using thermal pictures. A unique POGAN was proposed to detect thermal signatures, which effectively mitigates the problems of inconsistent thermal signatures by enabling more accurate detection and characterization of objects with various surface features with varying emissivity. Furthermore, CRFCwas presented to ascertain the shadow casting from the object that influences the object boundaries, which dynamically adjusts cluster centroids to adapt to changing scene complexity, improving comprehension of geometric features, spatial relationships, and angle of incidence. According to the overall analysis, the proposed method performs better in terms of accuracy with 94.6%, F1-score with 92.67%, mean IoU with 78.7%, precision with 95.63%, recall with 94.53%, mean accuracy with 92.7%, mAP with 78.24%, sensitivity with 96.11%, along with minimized time and MAE as 0.029 and 0.08s, respectively, for thermal object detection from the scene. This analysis therefore confirms that the suggested

approach was superior to the existing models in every way, with enhanced ability to predict objects from thermal images.

REFERENCES

[1] Z. Tu, Y. Ma, Z. Li, C. Li, J. Xu and Y. Liu, "RGBT salient object detection: A large-scale dataset and benchmark," *IEEE Transactions on Multimedia*, 2022.

[2] W. Zhou, Y. Zhu, J. Lei, J. Wan and L. Yu, "APNet: Adversarial learning assistance and perceived importance fusion network for all-day RGB-T salient object detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 4, pp. 957-968, 2021.

[3] E. Collini, L. A. I. Palesi, P. Nesi, G. Pantaleo and W. Zhao, "Flexible thermal camera solution for Smart city people detection and counting," *Multimedia Tools and Applications*, pp. 1-29, 2023.

[4] F. Huo, X. Zhu, L. Zhang, Q. Liu and Y. Shu, "Efficient context-guided stacked refinement network for RGB-T salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 3111-3124, 2021.

[5] Q. Zhang, T. Xiao, N. Huang, D. Zhang and J. Han, "Revisiting feature fusion for RGB-T salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1804-1818, 2020.

[6] Z. Liu, Y. Tan, Q. He and Y. Xiao, "SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4486-4497, 2021.

[7] M. Ye, J. Shen and L. Shao, "Visible-infrared person re-identification via homogeneous augmented tri-modal learning," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 728-739, 2020.

[8] G. Li, Z. Liu, M. Chen, Z. Bai, W. Lin and H. Ling, "Hierarchical alternate interaction network for RGB-D salient object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 3528-3542, 2021.

[9] K. Fu, D.P. Fan, G.P. Ji, Q. Zhao, J. Shen and C. Zhu, "Siamese network for RGB-D salient object detection and beyond," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 9, pp. 5541-5559, 2021.

[10] A. Glowacz, "Fault diagnosis of electric impact drills using thermal imaging," *Measurement*, vol. 171, pp. 108815 2021.

[11] Z. Ma, X. Xiang, L. Shao, Y. Zhang and J. Gu, "Multifunctional wearable silver nanowire decorated leather nanocomposites for joule heating, electromagnetic interference shielding and piezoresistive sensing," *Angewandte Chemie International Edition*, vol. 61, no. 5, pp. e202200705, 2022.

[12] Z. Ma, X. Xiang, L. Shao, Y. Zhang and J. Gu, "Multifunctional wearable silver nanowire decorated leather nanocomposites for joule heating, electromagnetic interference shielding and piezoresistive sensing," *Angewandte Chemie International Edition*, vol. 61, no. 15, pp. e202200705, 2022.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[13] H. Li, X.J. Wu and J. Kittler, "RFN-Nest: An end-to-end residual fusion network for infrared and visible images," *Information Fusion*, vol. 73, pp. 72-86, 2021.

[14] Y. Dai, Y. Wu, F. Zhou and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9813-9824, 2021.

[15] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341-1360, 2020.

[16] X. Dai, X. Yuan and X. Wei, "TIRNet: Object detection in thermal infrared images for autonomous driving," *Applied Intelligence*, vol. 51, pp. 1244-1261, 2021.

[17] Z. Tu, Z. Li, C. Li, Y. Lang and J. Tang, "Multi-interactive dual-decoder for RGB-thermal salient object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 5678-5691, 2021.

[18] A. Miller, Y. Marikar, A.Yousif, H. Sadreazami and M.Amini, "Pedestrian and Cyclist Object Detection Using Thermal and Dash Cameras in Different Weather Conditions," *In 2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS),* pp. 1340-1343, 2024, August. IEEE.

[19] T. Kernbauer, P. Fleck and C. Arth, "PanoTherm: panoramic thermal imaging for object detection and tracking," *Proc. Copyright*, vol. 98, pp.109, 2024.

[20] X. Ni, C. Kuehnel, and X. Jiang, "Thermal Detection of People with Mobility Restrictions for Barrier Reduction at Traffic Lights Controlled Intersections," *arXiv preprint arXiv:2505.08568*, 2025.

[21] J. Wang, K. Song, Y. Bao, L. Huang and Y. Yan, "CGFNet: Cross-guided fusion network for RGB-T salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2949-2961, 2021.

[22] S. Du, P. Zhang, B. Zhang and H. Xu, "Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4," *IEEE Access*, vol. 9, pp. 25671-25680, 2021.

[23] X.Ying, C. Xiao, W. An, R. Li, X. He, B. Li, X. Cao, Z. Li, Y. Wang, M. Hu, and Q. Xu, "Visible-thermal tiny object detection: A benchmark dataset and baselines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[24] W. Zhou, Q. Guo, J. Lei, L. Yu and J.N. Hwang, "ECFFNet: Effective and consistent feature fusion network for RGB-T salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1224-1235, 2021.

[25] P. Shyam, and H. Yoo, "Lightweight thermal super-resolution and object detection for robust perception in adverse weather conditions," *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 7471-7482, 2024.

[26] C. Liu, X. Ma, X. Yang, Y. Zhang and Y. Dong, "COMO: Cross-mamba interaction and offset-guided fusion for multimodal object detection," *Information Fusion*, vol. 125, pp.103414, 2026.

[27] F. Nezhadalinaei, L. Zhang, M. Mahdizadeh and F. Jamshidi, "Motion object detection and tracking optimization in autonomous vehicles in specific range with optimized deep neural network," *In 2021 7th international conference on web research (ICWR),* pp. 53-63, 2021, May. IEEE.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)