

Cross-Domain Outdoor Scene Understanding Using PC-GAN and TL-SZKSE-DenseNet in Adverse Weather Conditions

Deepa Mane*

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, SPPU, Pune, India

Department of Information Technology, Pune Institute of Computer Technology, SPPU, Pune, India

Email: deepabmane@gmail.com

ORCID iD: <https://orcid.org/0000-0003-3026-451X>

*Corresponding author

Sandhya Arora

Department of Computer Engineering, MKSSS's Cummins college of Engineering for women, SPPU, Pune, India

Email: sandhya.arora@cumminscollege.in

ORCID iD: <https://orcid.org/0000-0002-6412-7831>

Received: November 1, 2025; Revised: January 1, 2026; Accepted: June 29, 2026; Published: October 8, 2026

Abstract: In autonomous driving, Outdoor Scene (OS) understanding in unfavorable weather conditions is pivotal for enabling safe navigation. Conventional studies failed to capture the intricate pairwise object relationship within scenes in OS classification. To mitigate these challenges, this research introduces a dual framework bringing together a Panoptic Cycle-Generative Adversarial Network (PC-GAN) with a Transfer Learning-based Squeeze-and-Zero-Knowledge Spatial Excitation DenseNet (TL-SZKSE-DenseNet). The suggested PC-GAN achieves realistic cross-weather image translation with semantically aligned representations in clear and degraded scenes. The TL-SZKSE-DenseNet improves feature learning with the addition of a zero-knowledge spatial excitation mechanism that selectively enhances weather-independent and context-dependent features for enhanced discrimination in crowded scenes. This generative-discriminative hybrid method effectively overcomes structured and unstructured driving domain gaps, as demonstrated on benchmark datasets such as BDD and IndiaScene365. Primarily, the OS images are generalized to adverse conditions using Panoptic Cycle-Generative Adversarial Network (PC-GAN). After that, using saliency mapping, the most significant regions in the generalized output are highlighted. In the meantime, the significant features are extracted utilizing the Oranguan Optimization Algorithm (OOA). In addition, from the detected objects, the Object Frequency and Relationship (OFR) is analyzed. Later, using TL-SZKSE-DenseNet, the OS are classified based on the OFR and the selected features. The performance of our proposed approach was tested on our newly developed IndiaScene365 dataset as well as various benchmark datasets such as BDD100k. Our newly developed dataset, IndiaScene365, includes 3000 images that entail three different weather conditions, which are foggy weather, rainy weather, and poor light conditions along with clear day weather. The proposed approach records 94.87% accuracy, 93.24%, and 94.18% sensitivity on our dataset.

Index Terms: Outdoor Scene, Autonomous Driving, Cross Domain, Multi-Scale Analysis, Object Detection, Transfer Learning, and Deep Learning.

1. Introduction

OS understanding has turned out to be one of the most fascinating and essential requirements in several applications like autonomous driving, intelligent traffic monitoring, disaster response, smart surveillance, and environmental analytics [1]. Detecting objects and understanding their contextual arrangement within the given environment are the fundamental requirements of such understanding [2]. Particularly, OS understanding enhances road safety and prevents accidents in the field of Autonomous Vehicles (AV). These advancements in AV are slowly replacing traditional automobiles with fully functional intelligent machines. As per the forecast for the AV market size from 2022-2032, the market is projected to increase from 23.80 million units in 2022 to 80.43 million units [3]. For

effective OS understanding, AV highly depends on the Visual Perception (ViP) system to sense the surrounding environment and OS [3,4]. Therefore, for enhancing safety, precise ViP of the OS is essential [5, 6]. In general, the ViP system's performance is degraded by the AWCs, leading to critical safety concerns [7]. Therefore, advanced technologies are developed for enhancing safety in poor weather conditions. Advanced image processing and Artificial Intelligence (AI) are deployed to optimize the ViP, particularly in the case of AWCs [8, 9]. Also, to provide a detailed real-time understanding of the OS, various sensors are deployed. A sensing technology that senses the OS through the emission of laser pulses and evaluates the return time to calculate precise distances is known as LiDAR. Similarly, RaDAR enhances object detection capability by emitting radio waves. Several conventional works focused on the OS classification under clear weather conditions [10, 11]. Additionally, to mitigate the rain effect in an OS image, the authors in [12] utilized de-raining software. Also, to improve the scene classification accuracy, traditional methods employed BRISQUE [13, 14, 15]. Nevertheless, conventional works failed to concentrate on the intricate pairwise object relationship within the scenes. Analyzing the object relationship within the scenes in AWCs and classifying the OS effectively are the motivations of the proposed system, thereby promoting safety and giving valuable context to make appropriate decisions. So, a TL-SZKSE-DenseNet-based OS classification in AWCs is proposed in this research.

1.1. Significant gaps and our contributions to these aspects

None of the prevailing works effectively captured the intricate pairwise object relationship within the scenes, thereby diminishing the reliability and accuracy of scene classification. Traditional works [16, 17] completely depended on the overall image features and ignored the presence of specific objects within the scene, thus limiting the contextual understanding and decreasing the classification accuracy. In the prevailing work [18], the scene detection accuracy was reduced due to only training on specific conditions, like clear-day images. Nevertheless, the existing work failed to generalize the adverse conditions like nighttime settings. Owing to the task-specific training data dependency, the study in [19] struggled to generalize across varying environmental conditions, leading to poor performance when deployed in real-world scenarios with diverse and previously unseen conditions. In the existing works, dependency on single-scale analysis restricted the ability to simultaneously capture the global structures and fine-grained textures, which further reduced the performance under challenging environmental situations. Traditional works often struggled under difficult weather conditions, including fog, snow, rain, or lower light, thereby leading to the loss of visual details, reduced contrast, and introduced noise. Even though several algorithms and datasets have made their mark in the research community related to atmospheric weather classification, these datasets/algorithms face some limitations: As stated in the previous studies, most research activities related to atmospheric cross weather scene classification were carried out using public/private datasets. These datasets were limited in terms of either their scale or quantity to train deep learning algorithms traditionally. Even though the Image2Weather dataset proposed by researchers [20] considers 255 837 images to be sufficient for weather classification tasks, it is not available in the public domain for research purposes. However, our Indiace365 dataset considers various atmospheric weather challenges that include weather conditions to test basic baseline performance in such complex scenarios. Compared with the existing datasets, our dataset is more representative of the important atmospheric and weather conditions that occur in outdoor environments. It is also important to mention that, although many studies have been conducted on atmospheric and weather classification using CNNs, most of these studies are based on existing CNN architectures that have been pre-trained on the ImageNet dataset [21, 22], either by fine-tuning or transfer learning for classification [23, 24, 25], or by using these CNNs as feature extractors and then combining the extracted features along with traditional hand-crafted features and classical classifiers [26, 27, 28]. Only a few studies have proposed new CNN-based frameworks specifically designed for atmospheric and weather classification [29, 30, 31]. In addition, these studies usually involve several pre-processing steps, such as sky or boundary region detection [32, 33, 34], and thus their performance is highly dependent on the accuracy of the pre-processing steps. Inspired by the above, in this paper, we introduce a novel framework for classifying atmospheric and weather-degraded image sequences. The proposed framework is built using variants of skip-connection-based CNN building blocks to enhance robustness in adverse environmental conditions.

We summarize contributions of our paper below:

- A dual framework consisting of a combination of Panoptic Cycle-GAN and the concept of Transfer Learning based Squeeze and Zero-Knowledge Spatial Excitation DenseNet was proposed to comprehend images from traffic scenes involving extreme changes in the illumination/weather conditions.
- We have collected samples for forming a challenging image dataset named IndiaScene365, which provides 3000 images classified into three atmospheric or lighting conditions. The data set is accessible at Mendeley data for utilizing it for any kind of research work related to ITS. To the best of our knowledge, this is the first attempt that created transfer Learning dataset for traffic scene images captured on Indian roads where structured/ unstructured traffic was also at the same location but captured under varied atmospheric conditions.
- We compare the performance of the best existing CNNs of various architectures, pre-trained from various datasets, whose deep features are extracted from different layers.

The paper is structured as follows: The associated research work is explained in Section 2, a concise proposed framework is displayed in Section 3, the implementation of the proposed approach's performance is analyzed in Section 5, and finally, the paper is concluded with future scope in Section 6.

2. Literature Survey

Many atmospheric and weather conditions have been discovered and classified using various computer vision techniques. This section will discuss the most relevant studies related to atmospheric and weather classification, which can be considered as a type of scene recognition. It will also summarize the publicly and privately available datasets that have been commonly used by researchers for atmospheric and weather classification. [35] employed multi-class scenes under adverse weather conditions utilizing a CNN. Here, the OS image pre-processing was carried out. Further, to evaluate the significance of kernels in CNN, the Entropy Guided Mean (EGM) was utilized. Lastly, CNN accurately categorized the OS in different weather conditions and clear day conditions. However, due to the ignorance of specific objects within the scene, the classification accuracy was degraded. [36, 37, 38] recommended a Deep Learning (DL) methodology to recognize and classify OS images from video recorders in challenging weather conditions. Primarily, the OS image was collected; after that, the image's damaged areas were excluded. After using a BPNN, the OS was accurately classified from the remaining image. However, the model failed to generalize about adverse conditions, such as nighttime settings. References [39, 40, 41, 42] deployed a DL-centric object detection and scene perception under difficult weather situations. At first, the OS scene images were pre-processed. Thereafter, the pre-processed image was converted into a grid. Then, the OS in each grid was effectively detected utilizing You Only Look Once (YOLO) for rainy and regular weather conditions. Nevertheless, owing to the dependence on task-specific training data, YOLO struggled to generalize across varying environmental conditions. [35, 43, 44] represented transformer self-supervised depth estimation for unpaired multi-domain learning. Initially, the daytime images were gathered. Later, the depth information across various domains was learned from the daytime image features by using the Residual Network (ResNet) [45–46]. Finally, based on the dynamically significant depth features, ResNet accurately classified the OS. However, the effects of dynamic elements in complex scenes weren't eliminated. [47–48] explored a multi-domain associated-guided network for image-based long-term visual localization [49–51]. Initially, different scene images were collected. After that, the discriminative dominant-invariant features in different scenes were extracted through a perceptual network [52–53]. Further, to obtain more dominant features, the associated guided model connected the extracted features. Lastly, the localization of different scenes was improved using the Multi-Domain Association-Guided Network (MDAG-Net) [54–55]. Nevertheless, the model's performance was degraded under dynamic scenes. [56] introduced a cross-modal attention-based multimodal grounding for autonomous driving. Initially, five specialized encoders, namely text, emotion, vision, context, and cross-model, extracted the linguistic nuances within the command. Later, the five encoders were fused by a multi-modal decoder, leading to multi-modal prediction. Hence, Context-Aware Visual Grounding (CAVG) [56] effectively improved the visual grounding of AV. However, CAVG was highly susceptible to AWCs. [57] investigated automatic, model-centric dataset generation for high-level vision tasks under hazy weather conditions. Here, the real-world images of different scenes in hazy weather were acquired. Next, GAN generated synthetic images. Subsequently, centered on the supervised learning strategy, the model HazeGAN ensured whether the generated image could be useful or not in the high-level vision task performance. Hence, the vision performance in haze weather was improved by HazeGAN [58]. However, due to the training instability of GAN, the computational cost was high. [59] implemented a hybrid feature-enhancement network for robust object detection under AWCs. Primarily, the environmental data was collected by incorporating a camera and RADAR. Further, using a class-balancing approach, each class in the data was balanced. Lastly, YOLO efficiently detected the objects under AWCs. However, the adopted approach faced difficulties in object detection for weather conditions different from the training data. [60] explored a self-attention-based for unsupervised image dehazing. Primarily, the OS images were collected. Then, cycleGAN [61] attained data augmentation of foggy images. Thereafter, the sense of hierarchy in the fog effect was refined and the loss was assessed. At last, ResNet accurately classified the OS from the improved images. Nevertheless, due to memory overhead and optimization difficulty, the cost was increased. [62] recommended cognitive robustness in adverse weather with DL and thermal camera fusion. Here, using LiDAR and thermal sensors, the OS scene images were collected. Afterward, from the OS images, the features were extracted. Also, the scenes were effectively detected by YOLO. In contrast, as YOLO struggled to capture very small objects, it generated inaccurate Bounding Boxes (BBs).

The ablation study clearly shows that the proposed framework delivers over 16% absolute accuracy improvement over a simple baseline, confirming that the added components are necessary rather than redundant.

3. Proposed PC-GAN And TL-SZKSE-DENSENET Approaches-Based Cross-Domain Outdoor Scene Classification In Adverse Weather Conditions

Conventional baselines (e.g., direct YOLO/DenseNet pipelines without domain adaptation, multi-scale analysis, or contextual modeling) were empirically insufficient due to (i) severe domain shift between clear and adverse weather

scenes, (ii) loss of fine-grained object details under noise and blur, and (iii) limited contextual reasoning from object presence alone. Each module in the proposed framework addresses a specific failure mode observed in baseline models: PC-GAN mitigates weather-induced distribution mismatch; CLVNAHEE and TDWT enhance degraded visual cues; IGD-GMM improves foreground-background separation; ASRPN-NMSYOLO enhances object localization; OFR captures contextual object relationships; OOA reduces feature redundancy; and TL-SZKSE-DenseNet ensures robust and transferable scene classification. We have added a dedicated baseline comparison and ablation study Table 1, demonstrating that removing or replacing individual components with simpler alternatives results in consistent performance degradation. These results confirm that simpler pipelines are insufficient to handle the compounded challenges of adverse weather, while the proposed modular design achieves improved robustness and generalization. Importantly, the framework is modular rather than monolithic, allowing individual components to be replaced or omitted depending on application constraints. This design prioritizes robustness and interpretability over unnecessary architectural complexity. TL-SZKSE-DenseNet extends DenseNet with a *zero-knowledge spatial excitation mechanism* that differs from conventional SE or Zoneout by learning spatial importance without relying on explicit domain or weather supervision. This acts as a domain-invariant regularizer rather than a standard dropout-style stochastic masking. Ablation results Table 1 demonstrate consistent performance degradation when SZKSE is removed, confirming that it is not a trivial renaming. Similarly, PC-GAN is not a generic GAN with segmentation guidance. The proposed panoptic cycle constraint explicitly enforces semantic structure preservation across weather domains, which is absent in standard CycleGAN or segmentation-guided GANs. This design choice is theoretically motivated by domain alignment theory and empirically validated through cross-domain performance gains. To validate the necessity and non-trivial contribution of the proposed components, we conduct systematic ablation experiments by incrementally removing or replacing each module with its closest well-known counterpart. All experiments are performed under identical training and evaluation settings on adverse-weather datasets.

Table 1. Performance comparison of different CNN architectures evaluated on the IndiaScene365 dataset for atmospheric and weather condition classification.

Mod el No	No. of Layers	No. of Convolution Layers	No. of Convolution Filters in Each Layer	Kernel Size of Convolution Layers	No. of Pooling Layers	Kernel Size of Pooling Layers	No. of Fully Connected Layers	No. of Learnable Parameters	Learning Accuracy (%)
M1	6	2	32, 64	3×3	2	2×2	1	1.25 M	82.73
M2	8	3	32, 64, 128	3×3	2	2×2	1	2.10 M	85.96
M3	10	4	32, 64, 128, 256	3×3	3	2×2	2	3.85 M	88.74
M4	12	5	64, 128, 256, 256, 512	3×3	3	2×2	2	6.92 M	91.32
M5	14	6	64, 128, 256, 256, 512, 512	3×3	4	2×2	3	9.84 M	92.87

Here, the OS are classified by generalizing the data to AWCs using TL-SZKSE-DENSENET and PC-GAN approaches. Fig 1. exemplifies the proposed approach's block diagram.

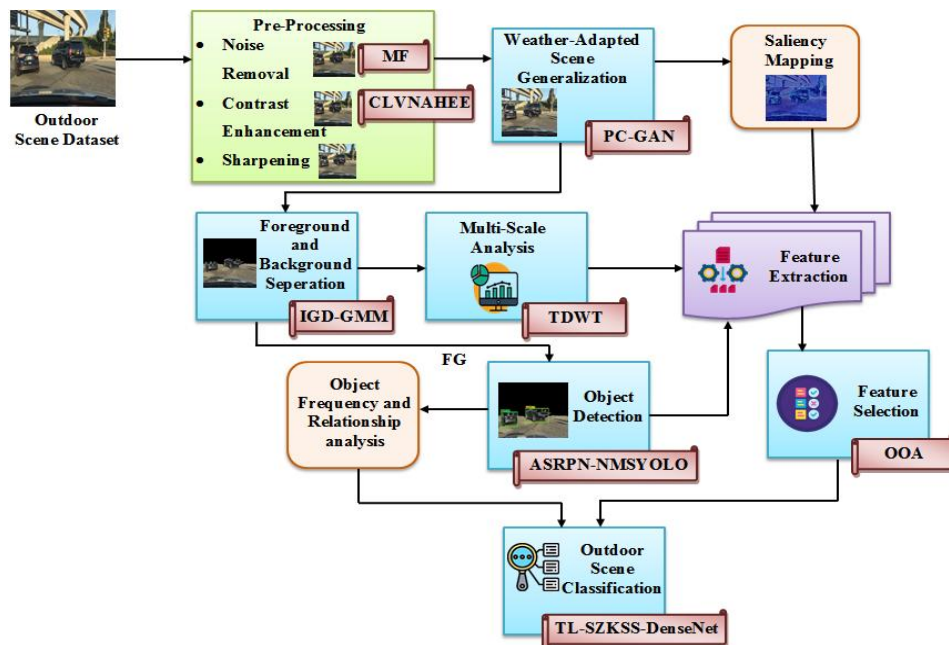


Fig. 1. Block diagram of the proposed framework.

3.1. Outdoor Scene Dataset

Initially, from the Berkeley Deep Drive 100K (BDD100K) dataset, OS images (d) are collected to classify the OS. This dataset encompasses images from diverse locations and integrates several environmental conditions.

3.2. Pre-Processing

Here, to eliminate noise and improve the image quality, the OS images are pre-processed. (d) The steps involved in pre-processing are given as follows,

❖ Noise Removal

Here, using the Median Filter (MF), the noise (j) present in (d) is removed to improve the image quality. MF replaces the pixel ($\tilde{d}$) value at the center with the median (δ) of the neighboring pixels.

$$\delta = \begin{cases} \tilde{d}_{\varphi} & \text{if } \tilde{d} \text{ is odd} \\ \frac{\tilde{d}_{\varphi 1} + \tilde{d}_{\varphi 2}}{2} & \text{if } \tilde{d} \text{ is even} \end{cases} \quad (1)$$

Where, the middle values of the sorted pixels are signified as ($\tilde{d}_{\varphi}$), ($\tilde{d}_{\varphi 1}$), and ($\tilde{d}_{\varphi 2}$). Therefore, the noise is eliminated by replacing (δ) at the center of the pixel values. Here, the noise-removed output is represented as ($\tilde{d}$).

❖ Contrast Enhancement

Further, utilizing CLVNAHEE, the contrast of ($\tilde{d}$) is enhanced for improving the visibility of objects and details in the image. Through limiting the amplification of noise, the contrast is enhanced by Contrast Limited Adaptive Histogram Equalization (CLAHE). Nevertheless, in CLAHE, the improper setting of the clip parameter makes the image over-enhanced or under-enhanced. To address this, the Von Neumann Entropy (VNE) technique sets the clip limit by measuring the complexity or information content in the tiles of the image.

Initially, ($\tilde{d}$) is divided into small tiles ($\tilde{d}_s$) of size ($a \times b$). After that, a histogram is computed for each tile with the intensity ($\mathfrak{R}$) and is defined as,

$$H[\tilde{d}_s(\mathfrak{R})] = \sum n'(\mathfrak{R}) \quad (2)$$

Where, the histogram of the tile with intensity is notated as ($\mathfrak{R}$) ($H[\tilde{d}_s(\mathfrak{R})]$), the number of tiles with intensity ($\mathfrak{R}$) is given as ($n'(\mathfrak{R})$), and the tile's height and width are represented as (a) and (b), correspondingly. Thereafter, to avoid over-amplification, VNE sets the clip limit by measuring the randomness in ($\tilde{d}_s$) using the density matrix (ρ) and trace operation (T).

$$\Xi = -T(\rho \log \rho) \times \frac{a}{b} \left[1 + \frac{w}{100} (g_{max}()) \right] \quad (3)$$

Where, the clip limit set by VNE is indicated as (Ξ), the clip parameter is illustrated as (w), and the maximum allowable slope is exhibited as ($g_{max}()$). Afterward, the ($H[\tilde{d}_s(\mathfrak{R})]$) that exceeds (Ξ) is clipped, and the clipped histogram is signified as ($\hat{H}$). Thereafter, to analyze the frequency distribution after clipping, the Cumulative Distribution Function (CDF) (μ) is computed for the clipped histogram.

$$\mu(\hat{H}) = \sum f(\hat{H}) \quad (4)$$

Where, the frequency distribution in ($\hat{H}$) is exemplified as (f). Lastly, the probability of the pixel intensity in ($\mu(\hat{H})$) is assessed to reassign ($\mathfrak{R}$) for creating a contrast-enhanced image with a more uniform intensity distribution.

$$\tilde{d}_\lambda = \mu(\hat{H}) \times \mathfrak{R} \quad (5)$$

Where, the contrast-enhanced image with intensity (λ) is illustrated as ($\tilde{d}_\lambda$). Then, ($\tilde{d}_\lambda$) is input into the succeeding operation.

❖ Sharpening

After that, for improving the overall visual quality of the image, the edges of $(\tilde{d}_\lambda)$ are sharpened. Sharpening defines the boundaries by boosting the contrast between the adjacent pixels.

$$d'' = \tilde{d}_\lambda + \varpi[J \times \tilde{d}_\lambda] \quad (6)$$

Where, the sharpened output is denoted as (d'') , the sharpness gain is notated as (ϖ) , and the high-pass filter is represented as (J) .

3.3. Weather-Adapted Scene Generalization

Here, to avoid the over-fitting issue in OS classification, (d'') is generalized to various weather conditions using PC-GAN. Cycle-Generative Adversarial Network (C-GAN) refers to a technique that encompasses two generators and two discriminators to effectively perform image-image translations. However, even though the image looks realistic, the image generated lacks semantic similarity. To eliminate this, a Panoptic Segmentation (PS) map is deployed. PS map optimizes the semantic similarity by assigning a class label and an instance Identification (ID) for distinct object instances.

At first, to generate a fake image, the first generator (G_1) maps (d'') from the source domain (X_m) to the target domain (Y_m) .

$$G'_1 = G_1 \times d''(X_m, Y_m) + u_1 + v_1 \quad (7)$$

Here, the fake images generated using (G_1) are notated as (G'_1) , and clear, rainy, overcast, snowy, partly cloudy, and foggy weather conditions are encompassed in (X_m) . Similarly, rainy, clear, snowy, overcast, foggy, and partly cloudy weather conditions are included in (Y_m) . The weight and bias of (G_1) are indicated as (u_1) and (v_1) , respectively, and $(d''(X_m, Y_m))$ specifies that (d'') are mapped from domain (X_m) to domain (Y_m) . Next, the PS map assigns a class label (h) and instance ID (i) for each object in (G'_1) .

$$G''_1 = O(G'_1)[h, i] \quad (8)$$

Here, the PS mapped output is signified as (G''_1) , and indicates the objects in , where and are assigned $(O(G'_1))(G'_1)(h)(i)$. Then, the discriminator (D_1) distinguishes the real image and (G''_1) .

$$D'_1 = \begin{cases} \text{If } (G''_1 \cup d'') \text{ then } 1 \\ \text{else } 0 \end{cases} \quad (9)$$

Where, the output of the discriminator is denoted as (D'_1) , and the higher likelihood of (G''_1) and (d'') is notated as $(G''_1 \cup d'')$. Subsequently, the adversarial loss functions for (G_1) and (D_1) are assessed as,

$$\xi_1 = \left(-\frac{1}{p}\right) \times \sum \log(D'_1) \quad (10)$$

$$\xi_1 = \left(-\frac{1}{p}\right) \circ \left[\sum \log(D'_1[1]) + \left(-\frac{1}{p}\right) + \sum \log(1 - D'_1[0]) \right] \quad (11)$$

Where the loss functions of (G_1) and (D_1) are represented as (ξ_1) and (ξ_1) , respectively, the total number of samples generated is indicated as (p) , and the higher and lower likelihoods of (D'_1) are denoted as $(D'_1[1])$, and $(D'_1[0])$, correspondingly. In the meantime, the generator (G_2) and discriminator (D_2) undergo the same process and generate a fake image (G'_2) by translating from (Y_m) to (X_m) . Thereafter, for both translations, cycle consistency loss is evaluated to ensure that the translated images retain the original structure of (d'') .

$$\mathring{\mathbf{x}} = d'' \left[\left\| G'_2 \left(G'_1(X_m) \right) \right\| - 1 \right] \quad (12)$$

$$\mathring{\mathbf{x}} = d'' \left[\left\| G'_1 \left(G'_2(Y_m) \right) \right\| - 1 \right] \quad (13)$$

Here, the cycle consistency loss while translating from (X_m) to (Y_m) and vice-versa is denoted as $(\mathring{\mathbf{x}})$ and $(\mathring{\mathbf{x}})$. After completely fooling the discriminator, similar images like (d'') are generated by the generators.

$$\dot{G} = (G_1 + G_2) \times d''(\mathring{\mathbf{x}} \times \mathring{\mathbf{x}})[(\xi_1 + \xi_2) + (\xi_1 + \xi_2)] \quad (14)$$

Where, the generated realistic image is signified as $(\hat{G})$, and the adversarial loss functions of (G_2) are notated as (ξ_2) and $(\check{\xi}_2)$.

3.4. Saliency Mapping

Here, utilizing Saliency Mapping (SM), the most prominent regions in $(\hat{G})$ are highlighted. The most significant parts of the image are visualized by SM to assist in understanding the decision-making process of the complex models. Primarily, to generate a saliency map, the probability score (P) and loss function (L) for $(\hat{G})$ are calculated.

$$P = \sum \sigma(\hat{G} \times k + c) \quad (15)$$

$$L = -(x \cdot \log P + (1 - \hat{G}) \times \log(1 - P)) \quad (16)$$

Where the true label of $(\hat{G})$ is indicated as (x) , the weight and bias of $(\hat{G})$ are specified as (k) and (c) , respectively, and the sigmoid activation function is elucidated as (σ) . Further, to identify the features that strongly influence the OS classification, a gradient is computed. The gradient (∇) is computed as,

$$\nabla \hat{G} = \frac{\partial L}{\partial k} = \frac{\partial L}{\partial P} \times \frac{\partial P}{\partial k} \quad (17)$$

Here, the partial differentiation of (L) regarding (k) and (P) is shown as $(\partial L / \partial k)$ and $(\partial L / \partial P)$, respectively, and the partial differentiation of the (P) concerning (k) is illustrated as $(\partial P / \partial k)$. Lastly, the saliency map is generated as,

$$\lambda = \nabla \hat{G} \times P \quad (18)$$

Where, the generated saliency map is specified as (λ) . Later, (λ) is subjected to a feature extraction process.

3.5. Foreground and Background Separation

Here, using IGD-GMM, the FG and background pixel regions within $(\hat{G})$ are separated from the surrounding to improve the understanding of the scene's content. The Gaussian Mixture Model (GMM) is highly adaptive to background changes. Also, it effectively assigns input to the Gaussian Distribution (GD) (E) . However, the fixed number of GD makes the Gaussian components underfit or overfit the scene, which may not be ideal for complex scenes with multiple background layers. To solve this, Infinite Gaussian Dirichlet (IGD) permits an infinite number of GDs and determines the optimal number of GDs adaptively from the data, thereby eliminating the need for predefining the number of distributions. Primarily, $(\hat{G})$ is converted into pixels $(\hat{G})$. Thereafter, using the IGD function, a base GD (E°) is initialized adaptively from $(\hat{G})$, which is mentioned in equation (19).

$$E^\circ [\tilde{\mu}, \sum V, \alpha] = \frac{1}{(2\pi)^{\frac{z}{2}} \cdot \sqrt{|\Sigma V|}} \exp\left(-\frac{1}{2}(z - \tilde{\mu}) \sum -1(z - \tilde{\mu})\right) + \alpha \quad (19)$$

Where, the mean, concentration parameter, and covariance matrix of (E°) are depicted as $(\tilde{\mu})$, (α) , and (V) , respectively, the total number of objects in $(\hat{G})$ is indicated as (z) , and (π) specifies a numerical value (i.e., 3.14). Further, the responsibility is assessed to determine whether $(\hat{G})$ can be clustered to (E°) or not.

$$P' = \frac{\eta I(\hat{G} | \tilde{\mu}, \Sigma V)}{\sum \eta_\infty I(\hat{G} | \tilde{\mu}_\infty, \Sigma V_\infty)} \quad (20)$$

Where, the responsibility of $(\hat{G})$ is represented as (P') , the mixing coefficients of (E°) and an infinite number of GDs are given as (η) and η_∞ , respectively, the probability density function is signified as (I) , and the mean and the covariance matrix of an infinite number of GDs are indicated as $(\tilde{\mu}_\infty)$ and (V_∞) , respectively. Next, $(\hat{G})$ is assigned to (E°) based on the responsibility; likewise, all the pixels are clustered to various GDs.

$$E = P' \times \sum E^\circ [\tilde{\mu}, \sum V, \alpha] + \hat{G} \quad (21)$$

Here, the clustered output is indicated as (E) . Lastly, after convergence, each GD signifies the distinct clusters of the pixel features within the scene. Thereafter, based on the average intensity $(\bar{I})$ of $(\hat{G})$, a threshold value (Ψ) is determined to separate FG and background pixels from all the clustered outputs of $(\hat{G})$.

$$\Psi = \bar{I}(\hat{G}) \times \sigma \quad (22)$$

Where, the standard deviation of (E) is notated as (σ) . Further, the FG and background pixels are separated as,

$$\phi = \begin{cases} \text{If } (E(\mathfrak{H}) \geq \Psi) \phi' \\ \text{else } \phi'' \end{cases} \quad (23)$$

Where, the weight of (E) is represented as $(\mathfrak{H})$, the FG and background separated output is exhibited as (ϕ) , the FG pixel regions are given as (ϕ') , and the background pixel regions are denoted as (ϕ'') . The FG pixels are the pixels with weight values above the threshold, while the background pixels are pixels with weight values lower than the threshold. Hence, the FG and background pixel regions are effectively separated by the IGD-GMM.

3.6. Object Detection

After that, using ASRPN-NMSYOLO, the objects in (ϕ') are detected to understand the semantic information of the scene. YOLO examines the object just once to identify and locate it. Nevertheless, it struggles to identify the precise location of small objects. Therefore, the Region Proposal Network (RPN) is employed. RPN effectively generates predefined anchor BBs of different scales and aspect ratios. However, the BBs do not always match the scale and aspect ratio, resulting in missed detections. To eliminate this, Adaptive Spatial (AS) Feature Fusion is deployed that fuses multi-scale features at each spatial location and handles objects of different sizes, thereby generating precise BB. In addition, each grid in YOLO is limited to a fixed number of BBs. Also, as there are more objects in the grid, some objects in the grid are missing. Thus, to remove the overlapping BB, the Non-maximum Suppression (NMS) technique is utilized, thereby capturing all the objects in the grid without missing. Primarily, (ϕ') is divided into grids $(\tilde{\phi})$. Each cell in the grid is accountable for detecting the objects. Thereafter, $(\tilde{\phi})$ is fed into the convolutional layer to extract the significant features.

$$x' = \sigma(k' \times \tilde{\phi}) + c' \quad (24)$$

Here, the feature map from $(\tilde{\phi})$ is indicated as (x') , and the weight and bias of the network are signified as (k') and (c') , respectively. Next, the coordinates of the BB that enclose the objects within $(\tilde{\phi})$ are predicted based on (x') to guess the initial localization of the object.

$$\dot{y} = \dot{x} \cdot (\ddot{a} \times \ddot{b}) \quad (25)$$

Where, the BB predicted output is exemplified as $(\dot{y})$, and the BB's height and width are depicted as $(\ddot{a})$ and $(\ddot{b})$, respectively. Then, RPN generates BB with a confident score $(\tilde{\mathfrak{N}})$ and probability scores $(\dot{f}(\tilde{r}))$ for each object to precisely locate the small objects. In RPN, the AS technique extracts the necessary features $(\dot{x})$ from $(\dot{y})$ to improve the BB generation accuracy. The BB generation is defined as,

$$\ddot{a} = \dot{x} \cdot \sum \tilde{\mathfrak{N}} + \dot{f}(\tilde{r}) \quad (26)$$

$$\dot{x} = \sum \tilde{\lambda} \times \dot{y} \quad (27)$$

Where, the generated BB is indicated as $(\ddot{a})$, and the learnable parameter of the features is denoted as $(\tilde{\lambda})$. Then, the NS method chooses the BB with the highest $(\tilde{\mathfrak{N}})$ and suppresses other overlapping BBs to capture all objects and avoid repeated prediction of the same object.

$$\tilde{R} = \sum \hat{y} \times \frac{A'''(\hat{y}_1 \cap \hat{y}_2)}{A'''(\hat{y}_1 \cup \hat{y}_2)} \quad (28)$$

Where, the area of $(\hat{y})$ is given as (A''') , the generated BB, which is the required objects detected in (ϕ') , is notated as $(\tilde{R})$, and the area of intersection and union between two BB is offered as $(\hat{y}_1 \cap \hat{y}_2)$ and $(\hat{y}_1 \cup \hat{y}_2)$, respectively. The pseudo-code for ASRPN-NMSYOLO is given further.

Pseudo-code for ASRPN-NMSYOLO

Input: Input image I ; YOLO backbone parameters θ ; Confidence threshold τ_c ; IoU threshold τ_{nms} ; ASRPN scaling parameters α, β

Output: Final detections D

-
1. Resize input image I to YOLO input resolution
 2. Extract multi-scale feature maps $F \leftarrow \text{Backbone}_{\text{YOLO}}(I; \theta)$
 3. Initialize empty proposal set $R \leftarrow \emptyset$
 - for each** feature map $F_k \in F$ **do**
 - Generate anchor boxes A_k
 - Compute spatial weights $W_k \leftarrow \sigma(\alpha \cdot \text{GAP}(F_k) + \beta)$
 - Refine anchors $A'_k \leftarrow A_k \odot W_k$
 - Update proposal set $R \leftarrow R \cup A'_k$
 - end for**
 4. Apply YOLO detection head on proposals R
 5. Predict bounding boxes, object nesses scores, and class probabilities
 - Compute confidence score $s \leftarrow s_o \times \max(p_c)$
 - Remove predictions where $s < \tau_c$
 6. Sort remaining detections in descending order of confidence
 7. Initialize final detection set $D \leftarrow \emptyset$
 - while** detections are not empty **do**
 - Select highest-confidence box b_i
 - Add b_i to D
 - for each** remaining box b_j **do**
 - Compute $\text{IoU}(b_i, b_j)$
 - if** $\text{IoU}(b_i, b_j) > \tau_{nms}$ **and** $W(b_j) < W(b_i)$ **then**
 - Suppress b_j
 - end if**
 - end for**
 - end while**
 8. Return final detection set D
-

3.7. Object Frequency and Relationship Analysis

Here, to determine the intricate pairwise object relationship within the scenes, the OFR is analyzed for $(\tilde{R})$. OFR analysis renders a spatial distribution of the detected objects and finds the distance between them. Primarily, the BB's centroid is calculated via the centroid formula.

$$\xi(X', Y') = \tilde{R} \times \left(\frac{\tilde{a}_1 + \tilde{a}_2}{2}, \frac{\tilde{b}_1 + \tilde{b}_2}{2} \right) \quad (29)$$

Where, the coordinates of the centroid(ξ) are indicated as (X', Y') , and the coordinates of the top-left corner and bottom-right corner of $(\tilde{R})$ are specified as $(\tilde{a}_1, \tilde{a}_2)$ and $(\tilde{b}_1, \tilde{b}_2)$, respectively. Further, using the Euclidean distance formula, the distance between the centroids of the pair of $(\tilde{R})$ is assessed.

$$\hat{S} = \sum \sqrt{(Y' - X'')^2 - (Y'' - X'')^2} \quad (30)$$

Where, (ξ) of the neighboring BB coordinates are denoted as (X'', Y'') , and the distance between the pair of BB is given as $(\hat{S})$. After that, the OFR histogram $(H'[\tilde{R}(\hat{S})])$ is computed for BB.

$$H'[\tilde{R}(\hat{S})] = \sum \hat{S} \quad (31)$$

Lastly, the pairwise object relationship within the scenes is analyzed based on $(H'[\tilde{R}(\hat{S})])$, and the OFR analyzed output is indicated as (B) .

3.8. Multi-Scale Analysis

Afterward, TDWT captures the global structures and fine-grained textures in $(\hat{\phi})$ to enhance TL-SZKSE-DenseNet performance under challenging environmental conditions. A technique that provides multi-resolution decomposition of an image for enabling analysis at different scales is known as the Discrete Wavelet Transform (DWT). Nevertheless, the grid is fixed at each level of decomposition and doesn't align with the natural image structures. The Tetrahedral Interpolation (TI) technique is deployed to eliminate that issue. TI is a digital processing method that converts a 3D grid into a tetrahedral structure, aligning with natural images.

Primarily, the decomposition level is set to $(\hat{q}^{th})$ level to decompose $(\hat{\phi})$ into different time and frequency components. Also, to obtain the approximation coefficients $(\hat{e})$, a low-pass filter $(\hat{w})$ is applied to $(\hat{\phi})$.

$$\hat{e} = \hat{\phi} \times \hat{w} \quad (32)$$

Likewise, the detailed coefficients $(\hat{e})$ are acquired for extracting the high frequency components of $(\hat{\phi})$ by applying a high-pass filter to $(\hat{\phi})$.

$$\hat{e} = \hat{\phi} \cdot \hat{\lambda} \quad (33)$$

Thereafter, to discard the unwanted samples in $(\hat{\phi})$, both $(\hat{e})$ and $(\hat{e})$ are decomposed until the $(\hat{q}^{th})$ level, thus mitigating the data size while preserving the crucial information at a coarser scale.

$$Q' = \hat{n} \cdot (\hat{e} \cdot \hat{e}) \quad (34)$$

Here, the decomposed output is exhibited as (Q') . Eventually, TI is employed for converting (Q') into a tetrahedral grid $(\hat{Q})$ and then aligning $(\hat{Q})$ with the original image structure, thereby capturing all fine details from the input.

$$\bar{w} = \hat{Q} \times \sum \hat{\phi} \cdot (\hat{\mathfrak{J}} * j') \quad (35)$$

Where, the interpolated output that captures the fine details of $(\hat{\phi})$ is depicted as $(\bar{w})$, and the barycentric weights and vertices of $(\hat{\phi})$ are represented as $(\hat{\mathfrak{J}})$ and (j') , respectively.

3.9. Feature Extraction

Here, from (λ) , the significant features like saliency area ratio, saliency hist features, saliency entropy, aspect ratio, eccentricity, numblobs, and avgblobsize are extracted. Later, from $(\bar{w})$, features like mean, standard deviation, entropy, and energy are extracted. In the meantime, from $(\bar{R})$, features like average object area, object density, mean aspect ratio, and total coverage are extracted. These extracted features are exemplified as $(\bar{q})$.

3.10. Feature Selection

In this section, using OOA, the significant features required for OS classification are selected from $(\bar{q})$ to mitigate the training time of OS classification. OOA refers to a nature-inspired meta-heuristic algorithm inspired by the orangutan's foraging as well as nesting behaviors. As OOA effectively balances exploration and exploitation, it is highly suitable for real-world scenarios. First, the population members in the search space are initialized. Here, the orangutan (Ψ) population is signified as $(\bar{q})$. The initial position of the (t^{th}) orangutan in the (o^{th}) dimensions is expressed as,

$$\psi_o^t = \bar{\ell} \times (u'' - \bar{l}) + \bar{l} \quad (36)$$

Where, the random number between (0 and 1) is given as $(\bar{\ell})$, and the upper and lower bound variables are specified as (u'') and $(\bar{l})$, respectively. Next, the fitness is evaluated by considering the maximum accuracy (ϕ) .

$$\mathfrak{R} = \max(\phi) \quad (37)$$

Here, the fitness of the population is indicated as $(\mathfrak{R})$. In the succeeding phases, the position of the population is updated based on $(\mathfrak{R})$.

➤ Foraging phase

Here, the members' positions are updated through the foraging behavior, which results in substantial exploration in the environment, thus permitting the population to discover several areas in the search zone.

$$\bar{\psi}_o^t = \psi_o^t + \bar{\ell} \cdot (\check{\kappa} - \psi_o^t) \quad (38)$$

Where, the selected food source of (ψ) is expressed as ($\check{\kappa}$), and the new position of (ψ) is illustrated as ($\bar{\psi}_o^t$). Further, the position is updated centered on the following condition,

$$\bar{\psi}_o^t = \begin{cases} \bar{\psi}_o^t & \text{If } \Re(\bar{\psi}_o^t) \geq \Re \\ \psi_o^t & \text{otherwise} \end{cases} \quad (39)$$

Where, the fitness function of ($\bar{\psi}_o^t$) is indicated as ($\Re(\bar{\psi}_o^t)$). Here, the new position is updated only if the objective functions of ($\bar{\psi}_o^t$) are improved.

➤ Nesting Phase

Further, the member's position is updated by the nesting behavior of the orangutan. OOA improves the algorithm's exploration capability and also enhances the fine-tuning of solutions by simulating the nesting skills of an orangutan.

$$\hat{\psi}_o^t = \psi_o^t + (1 - 2\bar{\ell}) \cdot \frac{(u'' - \bar{\ell})}{\bar{n}} \quad (40)$$

Where, the number of iterations is depicted as ($\bar{n}$), and the exploitation phase's new position is represented as ($\hat{\psi}_o^t$). Next, the position is updated centered on the condition given below,

$$\hat{\psi}_o^t = \begin{cases} \hat{\psi}_o^t & \text{If } \Re(\hat{\psi}_o^t) \geq \Re \\ \psi_o^t & \text{otherwise} \end{cases} \quad (41)$$

Lastly, after convergence, the best features ($\hat{A}$) are selected and fed to the OS classification.

3.11. OS classification

Finally, the TL-SZKSE-DenseNet classifies OS based on ($\hat{A}$) and (B) to improve the understanding of the complex environment in autonomous driving. ($\hat{A}$) and (B) are signified as (M). To improve performance, DenseNet utilizes dense connections between the layers. It also facilitates the network to reuse features, making the network effective for examining complex scenes with varying textures and objects. In contrast, during extremely deep implementations, gradient explosions can still occur. To solve this, the Zoneout regularization technique, which applies a drop-out mechanism to maintain a consistent learning process across time steps, is utilized. Furthermore, the dense mechanism might lead to redundant feature learning, particularly in OS having repeated patterns. To overcome this, Selective Kernel Squeeze-and-Excitation (SK-SE) blocks are introduced in DenseNet. The Kernel Squeeze-and-Excitation (SK-SE) blocks are introduced into the DenseNet framework to enable the adaptive selection of convolution kernels of varying sizes, which helps the network learn local details and contextual information effectively. The TL-SZKSE-DenseNet's step-by-step process is given below. Fig 2. displays the architecture of the proposed TL-SZKSE-DenseNet classifier.

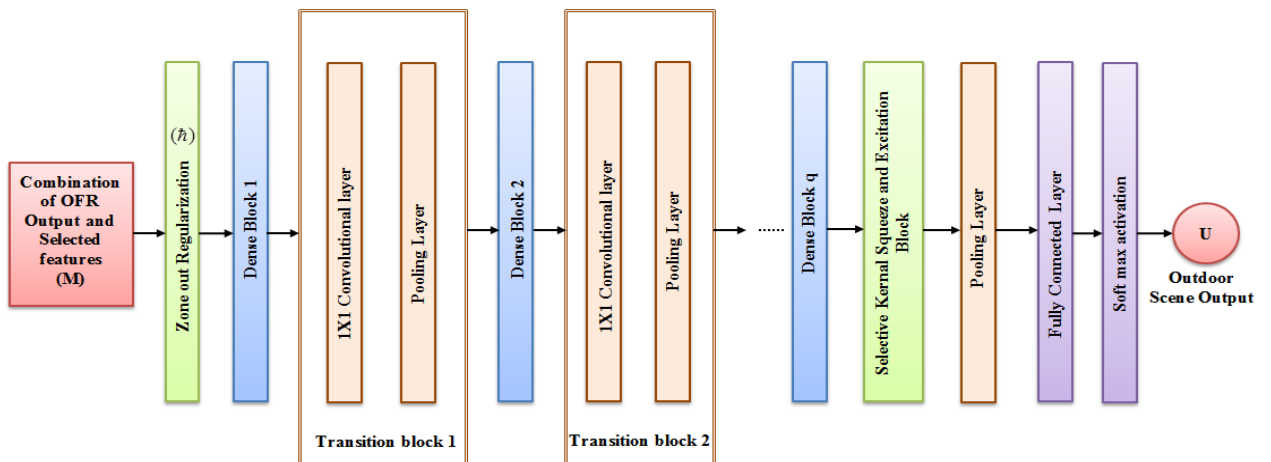


Fig. 2. Architecture of proposed TL-SZKSE-DenseNet classifier.

Primarily, using the Zoneout regularization technique, (M) is regularized to resolve the gradient issue.

$$\hat{h} = M \times \hat{g} \otimes (1 - \hat{g}) \quad (42)$$

Where, the Zoneout-regularized output is illustrated as $(\hat{h})$, and the vector of random Bernoulli variables is represented as $(\hat{g})$. Thereafter, regularized output $(\hat{h})$ is fed into dense blocks, which encompass multiple convolution layers; each of which receives feature maps from all preceding layers. First, from $(\hat{h})$, the hierarchical features are extracted in convolution layers, which employ filters to scan the input. To enable direct feature reuse, each convolution layer's output is concatenated to the subsequent layers.

$$\ddot{z} = \sum \sigma \cdot (\hat{k} \times \hat{h}) + \hat{c} \quad (43)$$

Here, the feature maps are denoted as $(\ddot{z})$, the sigmoid activation function is notated as (σ) , and the weights as well as biases of the convolution layers are represented as $(\hat{k})$ and $(\hat{c})$, respectively. Further, $(\ddot{z})$ is subjected to the transition block that encompasses a 1*1 convolution layer and a pooling layer. 1*1 convolution layer acts like a bottleneck to control the model complexity and computation cost by reducing the feature map.

$$\dot{z} = \sigma \times (k'' \times \ddot{z}) + c'' \quad (44)$$

Where, the reduced feature map output is denoted as $(\dot{z})$, and the weight and bias of a 1*1 convolution layer are represented as (k'') and (c'') , respectively. After that, $(\dot{z})$ is fed into a pooling layer to decrease the high dimensions of the features, and the pooled output $(\tilde{\beta})$ is articulated as,

$$\tilde{\beta} = \max(\dot{z}) \quad (45)$$

Further, $(\tilde{\beta})$ is fed to the SK-SE blocks, where the most relevant features are selected to avoid redundant feature learning. These blocks allow the network to dynamically adjust the Receptive Field (RF) $(\dot{S})$ of convolutional kernels based on the input features of $(\tilde{\beta})$.

$$s''' = \dot{S} \times \left[\hat{h} \left(\sum \tilde{\beta} \right) \right] \quad (46)$$

Here, the SKSE output is specified as (s''') , and the composite function performed at $(\tilde{\beta})$ is given as $(\hat{h})$. Subsequently, (s''') is subjected to the pooling layer for spatial downsampling.

$$\beta'' = \max(s''') \quad (47)$$

Where, the SKSE pooled output is exhibited as (β'') , and it is converted into 1D information for obtaining a flattened input $(\tilde{e})$. Lastly, each neuron in the fully connected layer is connected to $(\tilde{e})$ values by sharing the weight and bias.

$$U = \text{softmax}(\tilde{k} \times \tilde{e}) + c''' \quad (48)$$

Here, the OS classified output is depicted as (U) , and the weight and bias values are indicated as $(\tilde{k})$ and $(\tilde{c})$, respectively.

$$U = \{U_1, U_2, U_3, U_4, U_5, U_6\} \quad (49)$$

Where, the city street, highway, residential, parking lot, tunnel, and gas stations are represented as (U_1) , (U_2) (U_6) , respectively. The algorithm for TL-SZKSE-DenseNet is offered as follows,

Pseudo-code for TL-SZKSE-DenseNet

Input: Input image I ; pretrained DenseNet weights θ_{TL} ; Number of classes C ; learning rate η Batch size B ; number of epochs E

Output: Predicted class label $\hat{y}$

1. Initialize Network

Load pretrained DenseNet backbone with parameters θ_{TL}

Freeze initial convolution and early dense blocks

for each dense block DB_k in DenseNet do

\quad Insert SZKSE module after DB_k

end for

Replace final classification layer with a fully connected layer of size C

2. Forward Pass

Resize and normalize input image I

Extract feature maps

$$F \leftarrow DenseNet_{TL}(I; \theta_{TL})$$

3. SZKSE Feature Recalibration**for each feature map F_k do**

\quad Compute channel squeeze

$$z_k \leftarrow GAP(F_k)$$

\quad Compute zero-knowledge spatial excitation weights

$$w_k \leftarrow \sigma(W_2 \cdot \delta(W_1 \cdot z_k))$$

\quad Recalibrate features

$$\tilde{F}_k \leftarrow F_k \odot w_k$$

end for

Concatenate recalibrated feature maps

Apply global average pooling

Compute class probabilities using Softmax

4. Training Phase

Compute categorical cross-entropy loss

Update trainable parameters using Adam optimizer with learning rate η

5. Inference Phase

Predict class label

$$\hat{y} \leftarrow \arg \max(\text{Softmax}(\cdot))$$

Return predicted label $\hat{y}$

Therefore, the OS in AWCs is effectively classified by the TL-SZKSE-DenseNet. The proposed work accurately categorizes OS and captures the intricate pairwise object relationship within scenes in AWCs. The proposed work's performance is discussed in section 5.1.

4. Implementation**4.1. Dataset Description**

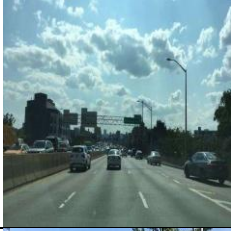
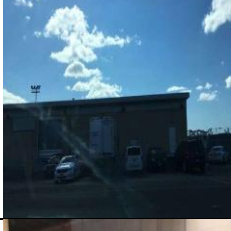
The IndiaScene365 dataset is a pioneering effort at creating an annotated dataset for object detection in challenging atmospheric and weather conditions such as fog, haze, rain, and low-light conditions. The details of the data acquisition process have been documented in our previous publication.

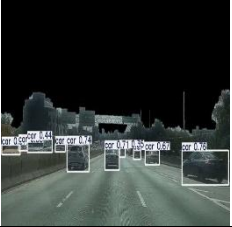
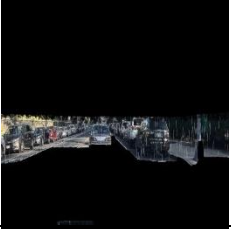
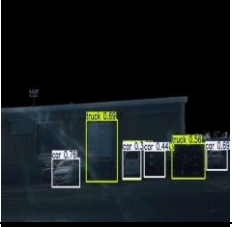
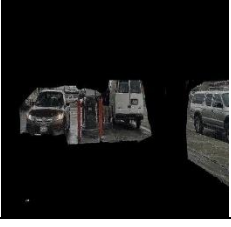
Every frame of the IndiaScene365 contains objects, and most of the scenes considered in the database are shot outdoors, specifically in the city environment, which has high surface variety because of the presence of buildings, roads, plants, and pedestrians. Such properties of the database make it highly useful for object detection and high-level outdoor vision.

In this paper, the authors have used images from IndiaScene365 for atmospheric and weather conditions. 3,000 images were randomly picked with an equal number of samples for each category, which included fog, haze, rain, poor lighting conditions, and clear day images. The IndiaScene365 dataset was divided into three sets based on the conditions that were taken into consideration. From the randomly selected images, about 1,000 low-light and daylight images, 1,000 rainy images, and 1,000 fog images were picked for a detailed annotation task, which finally resulted in a set of 3,000 images. To reduce temporal redundancy, frames have been sampled at controlled intervals from video

sequences of every weather condition for the training, validation, and test sets. By using OS images from the IndiaScene365 dataset, the performance evaluation is done. A dataset constitutes 3000 OS images of different classes. A link to data source is provided in the references. The proposed work utilizes 80% (2400) and 20% (800) of the data for training and testing, respectively. Table 2 displays the proposed work’s image results.

Table 2. Image results.

OS classes	Input	Noise Removal	Contrast Enhancement	Sharpening
City street				
Highway				
Residential				
Parking Lot				
Tunnel				
Gas Stations				

OS classes	Weather-Adapted Scene Generalization	Saliency mapping	Foreground and Background Separation	Object Detection
City street				
Highway				
Residential				
Parking Lot				
Tunnel				
Gas Stations				

The image results of the proposed framework in classifying the OS in AWCs regarding noise removal, contrast enhancement, sharpening, weather-adapted scene generalization, saliency mapping, FG and background separation, and object detection are indicated in Table 2.

4.2. Experiment

All experiments were implemented in PyTorch and conducted on a workstation equipped with an NVIDIA RTX 30900 GPU (24 GB VRAM), Intel Xeon CPU, and 64 GB RAM. Training and evaluation were performed under identical settings across all ablation and comparison experiments to ensure fairness.

4.2.1. Datasets and Preprocessing

Experiments were conducted on IndiaScene365 [58] datasets. Images were resized to 640 × 640 pixels. Standard normalization was applied using ImageNet mean and standard deviation. Data augmentation included random

horizontal flipping ($p = 0.5$), random scaling ($\pm 10\%$), and color jittering to improve robustness under adverse weather conditions.

4.2.2. PC-GAN Training

PC-GAN was trained using an Adam optimizer with a learning rate of 2×10^{-4} and $\beta_1 = 0.5$, $\beta_2 = 0.999$. The generator and discriminator were trained for 200 epochs with a batch size of 4. Cycle consistency, adversarial, and panoptic semantic losses were weighted empirically to preserve structural semantics across weather domains.

4.2.3. TDWT Configuration

TDWT employed 3 decomposition levels with adaptive thresholding. Tunable parameters were optimized on the validation set to suppress weather-induced noise while retaining structural information.

4.2.4. ASRPN-NMSYOLO

YOLO was used as the base detector with anchor refinement enabled via adaptive region proposal scoring. Adaptive NMS replaced standard NMS to reduce false suppression under heavy occlusion. Detection confidence threshold was set to 0.25, and NMS IoU threshold to 0.6.

4.2.5. TL-SZKSE-DenseNet Training

The classifier backbone was initialized with ImageNet-pretrained DenseNet weights. Fine-tuning was performed using AdamW with a learning rate of 1×10^{-4} for the backbone and 1×10^{-3} for newly added SZKSE layers. Training was conducted for 80 epochs with early stopping based on validation accuracy. Cross-entropy loss was used for scene classification. Summary of Hyperparameters is as shown in Table 3.

Table 3. Hyperparameter Summary Table.

Component	Parameter	Value
Input Resolution	Image size	640 × 640
Optimizer (GAN)	Adam	$\beta_1 = 0.5$, $\beta_2 = 0.999$
GAN Learning Rate	Generator / Discriminator	2×10^{-4}
GAN Batch Size	—	4
GAN Epochs	—	200
TDWT Levels	Decomposition levels	3
YOLO Confidence Threshold	—	0.25
YOLO NMS IoU Threshold	Adaptive NMS	0.6
Backbone	DenseNet	ImageNet pretrained
Optimizer (Classifier)	AdamW	—
Backbone LR	—	1×10^{-4}
SZKSE LR	—	1×10^{-3}
Batch Size (Classifier)	—	16
Epochs (Classifier)	—	80
Loss Function	Classification	Cross-entropy
Hardware	GPU	NVIDIA RTX 3090

5. Results and Discussions

Here, the performance of the proposed framework is shown. The conventional statistical measures, mean, and standard deviation of each evaluated metric are used to evaluate the predictive power of the TL-SZKSE-DenseNet model. For the performance analysis of this model, the test set of the IndiaScene365 image sequences, including a total of 3,000 images as shown in Table 3, which were not used for training, was used. As shown in Table 9, after removing 5% of the convolutional kernels using the proposed kernel removal scheme presented in Section 4.4, the proposed network still maintains its high performance. The proposed model classifies the respective images with the average accuracy, specificity, as well as the sensitivity of the proposed model of 94.87%, 93.24%, & 93.18%, respectively.

5.1. Performance Evaluation

Here, the performance of the proposed TL-SZKSE-DenseNet, PC-GAN, ASRPN-NMSYOLO, TDWT, and CLVNAHEE are analyzed.

In Fig 3, the performance of the proposed TL-SZKSE-DenseNet classifier concerning accuracy, precision, recall, F-Measure, sensitivity, and specificity is displayed. By employing zoneout regularization and SK-SE blocks, the proposed TL-SZKSE-DenseNet improved accuracy, thus attaining 94.87% accuracy, 93.18% precision, 94.24% recall,

94.11% F-Measure, 94.18% sensitivity, and 94.92% specificity. The existing DenseNet, SqueezeNet, Deep CNN (DCNN), and CNN had average accuracy, precision, recall, F-Measure, sensitivity and specificity of 90.06%, 92.59%, 92.59%, 92.09%, 92.59%, and 93.52% respectively. As per these outcomes, the proposed TL-SZKSE-DenseNet classified the OS effectively, outperforming other prevailing classifiers.

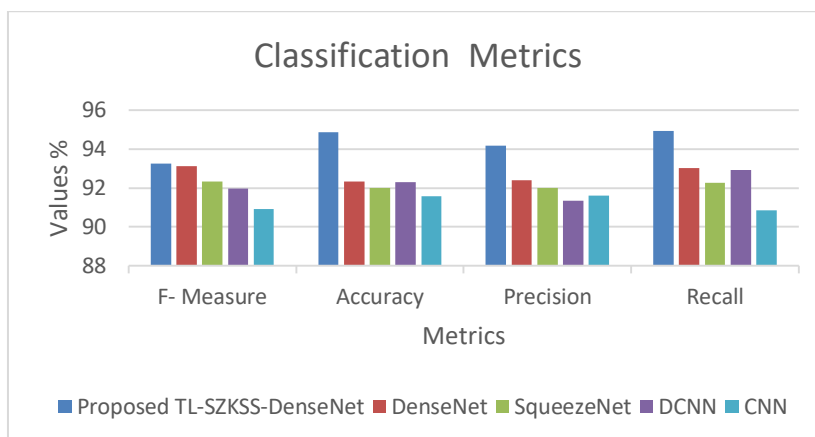


Fig. 3. Graphical plot for OS classification assessment.

Table 4. Numerical metrics-based performance assessment of TL-SZKSE-DenseNet.

Models	FNR	FPR	PPV	NPV	TPR	TNR
Proposed TL-SZKSE-DenseNet	1.071	0.754	93.24	94.87	94.18	94.92
DenseNet	2.964	1.598	93.14	92.33	92.40	93.03
SqueezeNet	4.731	2.997	92.35	92.00	92.00	92.26
DCNN	5.069	4.658	91.96	92.31	91.34	92.93
CNN	9.128	8.370	90.91	91.58	91.62	90.87

TL-SZKSE-DenseNet achieves the lowest false rates, with FNR = 1.071% and FPR = 0.754%, indicating superior robustness in minimizing both missed detections and false alarms. In contrast, conventional CNN exhibits substantially higher error rates (FNR = 9.128%, FPR = 8.370%), highlighting its limited discriminative capability. In terms of reliability, the proposed model records the highest PPV (93.24%) and NPV (94.87%), confirming that its positive and negative predictions are highly trustworthy. Sensitivity and specificity are also maximized, with TPR = 94.18% and TNR = 94.92%, outperforming DenseNet by approximately 0.8–1.9% and significantly exceeding SqueezeNet, DCNN, and CNN. Overall, the results demonstrate that integrating transfer learning with SZKSS attention mechanisms substantially enhances feature discrimination and generalization, leading to lower misclassification rates and more stable predictions than both lightweight and conventional deep CNN architectures.

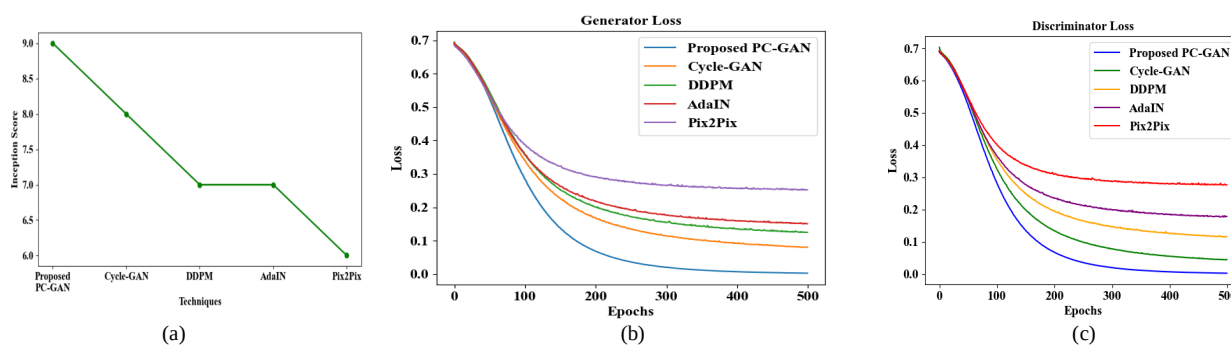


Fig. 4. Graphical representation of the proposed PC-GAN's performance based on (a) Inception score, (b) Generator loss, and (c) Discriminator loss

Fig 4. shows the performance of the proposed PC-GAN based on inception score, generator loss, and distribution loss. As the semantic similarity was improved by the PS map, PC-GAN attained an inception score of 9.258. Similarly, an average inception score of 7.197 was attained by traditional methods like C-GAN, Denoising Diffusion Probabilistic Model (DDPM), Adaptive Instance Normalization (AdaIN), and Pixel-to-Pixel (Pix2Pix). Additionally, the generator and discriminator losses over 500 epochs were also displayed. By attaining a lower generator loss (0.03) and discriminator loss (0.025), the proposed PC-GAN outperformed other existing approaches, thus indicating a stronger

convergence. Hence, the proposed PC-GAN's effectiveness in accurately generalizing the OS images to AWCs was demonstrated.

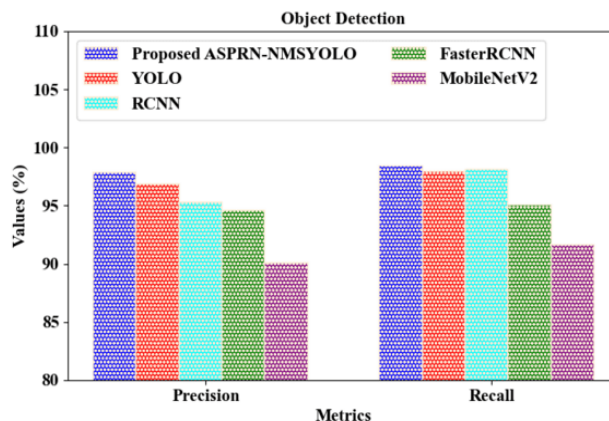


Fig. 5. Object Detection Analysis.

The proposed ASPRN-NMSYOLO's performance assessment regarding precision and recall is depicted in Fig 5. The proposed ASPRN-NMSYOLO demonstrated considerably higher precision (97.90%) and recall (98.50%). In the proposed approach, the performance improvement was due to the usage of RPN, AS feature fusion, and the NS technique in YOLO. Nevertheless, the existing approaches like YOLO, Faster Region-based CNN (Faster RCNN), MobileNet Version 2 (MobileNet V2), and RCNN obtained an average precision of 94.25% and a recall of 95.73%. Therefore, the proposed approach detected the object in autonomous driving effectively.

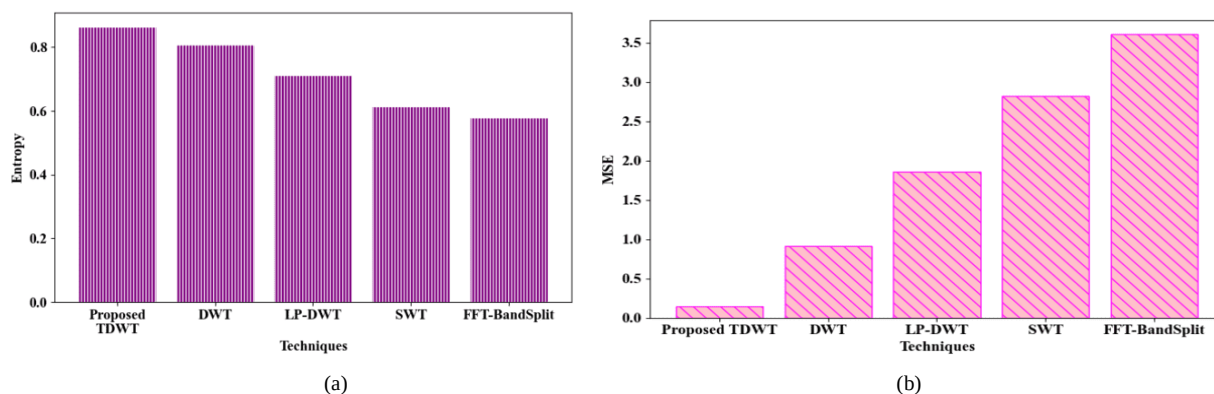


Fig. 6. Graphical representation of proposed TDWT in multi-scale analysis (a) Entropy and (b) MSE.

Fig 6. shows that a lower Mean Squared Error (MSE) (0.1475) and higher entropy (0.865) were obtained by the proposed TDWT. In contrast, the existing DWT, Low Pass DWT (LW-DWT), Stationary Wavelet Transform (SWT), and Fast Fourier Transform (FFT)-BandSplit achieved higher MSE and lower entropy values. The existing FFT-BandSplit achieved a higher entropy (3.6051) and a lower MSE (0.578). Hence, when contrasted to the conventional techniques, the proposed approach performed better by employing the TI technique to improve the image alignment.

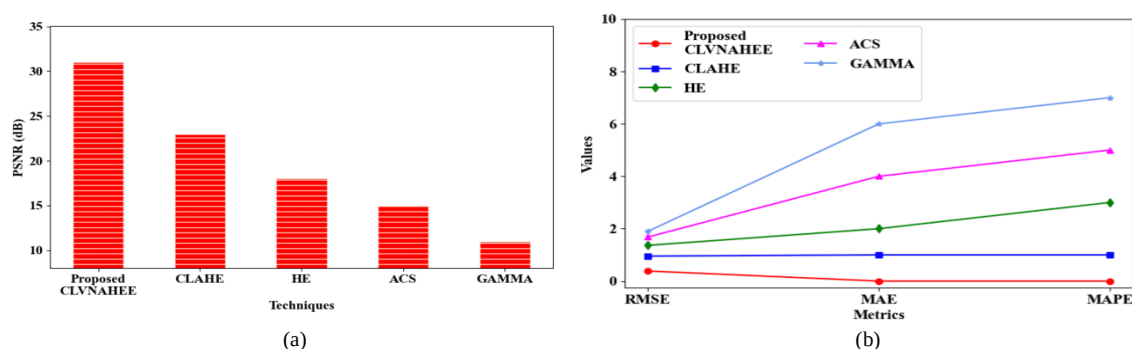


Fig. 7. Graphical plot of proposed CLVNAHEE concerning (a) PSNR and (b) RMSE, MAE, and MAPE.

Table 5. Numerical assessment-based performance evaluation of CLVNAHEE.

Techniques	PSNR	RMSE	MAE	MAPE
Proposed CLVNAHEE	31.154	0.3841	0.476	0.435
CLAHE	23.216	0.9539	1.214	1.026
HE	18.498	1.3631	2.145	3.21
ACS	15.187	1.6802	4.032	5.421
GAMMA	11.295	1.8987	6.145	7.256

The performance of the proposed CLVNAHEE based on the metrics like Peak Signal to Noise Ratio (PSNR), Root Mean Square Error (RMSE), Mean Average Error (MAE), and Mean Average Precision Error (MAPE) is demonstrated in Table 5 and Fig 7(a) and 7(b). The proposed CLVNAHEE obtained PSNR of 31.154, RMSE of 0.3841, MAE of 0.476, and MAPE of 0.435. In the meantime, the existing approaches like CLAHE, Histogram Equalization (HE), Automated Control System (ACS), and GAMMA (factor) attained an average PSNR of 17.049, RMSE of 1.473, MAE of 13.536, and MAPE of 4.228. Hence, the proposed CLVNAHEE deployed the VNE technique for setting the clip parameter. Also, it showed superior performance in improving the OS images' visual details.

Table 6. Comparison with the existing works.

Study	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Work	TL-SZKSE-DenseNet	94.87	94.18	93.24	94.21
[13]	Attention-based Cross Fusion Module (ACFM)	-	84.9	76.8	-
[17]	YOLOv5	84.9	70	71	71.4
[20]	Adverse WeatherNet (AWNNet)	-	92.31	97.58	-
[13]	Faster R-CNN	-	73.7	-	-
[22]	Faster-RCNN	-	73.2	-	-

In Table 6, the comprehensive analysis of the proposed work and several conventional methodologies is shown. The OS was effectively classified by the proposed work with accuracy, precision, recall, and F-Measure of 94.87%, 95.18%, 93.24%, and 94.21%, respectively. Next, the traditional ACFM attained a precision of 84.9% and a recall of 76.8%. Nevertheless, as the adverse weather images were prone to artifacts, the conventional ACFM's performance was degraded. Then, the existing YOLOv5 obtained 84.9% accuracy, 70% precision, 71% recall, and 71.4% F-Measure. However, the speed of YOLOv5 was minimized due to the increased dimensions of the input. Moreover, the prevailing AWNNet achieved a precision of 92.31% and a recall of 97.58%. Nevertheless, as LiDAR was sensitive to AWCs, the existing AWNNet's performance was mitigated. Furthermore, the existing [13] and [22] utilized Faster R-CNN, obtaining precision values of 73.7% and 73.2%, respectively. However, it displayed a poor performance in detecting the small objects. By employing modern methodologies and accurately classifying the OS, the proposed model vanquished those problems.

5.2. Limitations

There are certain limitations related to the proposed framework despite its strong performance on the BDD100K and IndiaScene365 datasets. First, the model exhibits reduced accuracy in very low-visibility scenes such as dense fog, heavy rain at night, and severe motion blur due to seriously degraded semantic cues. Second, this work comes at the price of higher computational overhead brought by increased network depth and additional modules, which might also limit real-time deployment on resource-constrained embedded platforms. Third, although transfer learning facilitates generalization, domain shift across unseen geographic regions and sensor configurations can still degrade performance. Error analysis demonstrates that most of the misclassifications are between those scene categories which are visually similar, for instance, highway versus urban roads in poor lighting conditions. This typically evidences difficulties with fine-grained contextual discrimination. Qualitative inspection of failure cases shows that wrong predictions often result from occlusions, extreme illumination variation, and typical cluttered traffic scenes in adverse weather. Sometimes, panoptic segmentation also fails to capture small or far objects, and this propagates errors to the classification stage. These findings suggest that although the framework is robust under moderate conditions, it remains sensitive to severe environmental distortions.

6. Conclusion and Future Scope

Atmospheric and weather degradations in outdoor environments cause significant deterioration in computer vision performance. Hence, precise classification of adverse weather scenes is critical for effective scene understanding. In this paper, we presented a CNN based transfer learning model, named TL-SZKSE-DenseNet, for the classification of cross-weather outdoor scenes and weather conditions. Experiments performed on the IndiaScene365 dataset and other benchmark datasets show that the proposed TL-SZKSE-DenseNet performed better than conventional pretrained CNNs

and the latest state-of-the-art approaches in terms of robust predictive performance in adverse weather conditions. In addition, TL-SZKSE-DenseNet has great potential for practical applications in real-time systems, such as outdoor video surveillance and intelligent transportation systems. In this research, the network concentrates on scenes degraded by single weather factors, such as fog, haze, dust, rain, poor lighting, and clear scenes; future work should extend the framework to address compound and blended weather conditions toward improvement of robustness in a more complex real-world environment. Future work will focus on reducing computational complexity through model pruning and lightweight attention mechanisms to enable real-time deployment. Incorporating temporal information from video sequences and multi-sensor fusion (e.g., LiDAR or radar) is expected to improve robustness under extreme weather. Additionally, self-supervised and continual learning strategies will be explored to better handle domain shifts and unseen conditions.

All the Declarations and Statements

Author Contributions Statement

Deepa Mane – Conceptualization, Methodology, Data preprocessing and Data analysis and Writing original draft.
Sandhya Arora – Methodology, supervised project execution.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The datasets generated and analyzed during this study are publicly available in the Mendeley Data repository at <https://data.mendeley.com/datasets/pwffhg6nhz/2>

Ethical Declarations

The work adheres to ethical publishing standards.

Acknowledgments

NA

Declaration of Generative AI in Scholarly Writing

During the preparation of this work, the authors used ChatGPT to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Abbreviations

NA

References

- [1] Aloufi, N., Alnori, A., & Basuhail, A. (2024). Enhancing Autonomous Vehicle Perception in Adverse Weather: A Multi Objectives Model for Integrated Weather Classification and Object Detection. *Electronics (Switzerland)*, 13, 1–20. <https://doi.org/10.3390/electronics13153063>
- [2] Chow, T. Y., Lee, K. H., & Chan, K. L. (2023). Detection of Targets in Road Scene Images Enhanced Using Conditional GAN-Based Dehazing Model. *Applied Sciences (Switzerland)*, 13, 1–26. <https://doi.org/10.3390/app13095326>
- [3] Dreissig, M., Scheuble, D., Piewak, F., & Boedecker, J. (2023). Survey on LiDAR Perception in Adverse Weather Conditions. *Arxiv*, 1–8. <https://doi.org/10.1109/IV55152.2023.10186539>
- [4] Guan, D., Huang, J., Xiao, A., Lu, S., & Cao, Y. (2022). Uncertainty-Aware Unsupervised Domain Adaptation in Object Detection. *IEEE Transactions on Multimedia*, 24, 2022. <https://doi.org/10.1109/TMM.2021.3082687>

- [5] Humayun, M., Ashfaq, F., Jhanjhi, N. Z., & Alsadun, M. K. (2022). Traffic Management: Multi-Scale Vehicle Detection in Varying Weather Conditions Using YOLOv4 and Spatial Pyramid Pooling Network. *Electronics (Switzerland)*, 11, 1–29. <https://doi.org/10.3390/electronics11172748>
- [6] Iqbal, J., Hafiz, R., & Ali, M. (2022). FogAdapt: Self-supervised domain adaptation for semantic segmentation of foggy images. *Arxiv*, 1–13. <https://doi.org/10.1016/j.neucom.2022.05.086>
- [7] Jhung, J., & Kim, S. (2021). Behind-the-scenes (Bts): Wiper-occlusion canceling for advanced driver assistance systems in adverse rain environments. *Sensors*, 21, 1–21. <https://doi.org/10.3390/s21238081>
- [8] Kumar, D., & Muhammad, N. (2023). Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8. *Sensors*, 23, 1–29. <https://www.mdpi.com/1424-8220/23/20/8471>
- [9] Li, J., Xu, R., Liu, X., Ma, J., Li, B., Zou, Q., Ma, J., & Yu, H. (2025). Domain Adaptation Based Object Detection for Autonomous Driving in Foggy and Rainy Weather. *IEEE Transactions on Intelligent Vehicles*, 10(2), 2025. <https://doi.org/10.1109/TIV.2024.3419689>
- [10] Liao, H., Shen, H., Li, Z., Wang, C., Li, G., Bie, Y., & Xu, C. (2024). GPT-4 enhanced multimodal grounding for autonomous driving: Leveraging cross-modal attention with large language models. *Communications in Transportation Research*, 4, 1–19. <https://doi.org/10.1016/j.commtr.2023.100116>
- [11] Lin, H., Parsi, A., Mullins, D., Horgan, J., Ward, E., Eising, C., Denny, P., Deegan, B., Glavin, M., & Jones, E. (2024). A Study on Data Selection for Object Detection in Various Lighting Conditions for Autonomous Vehicles. *Journal of Imaging*, 10(7), 1–18. <https://doi.org/10.3390/jimaging10070153>
- [12] Lin, H. Y., Huang, Y. C., Lai, J. X., & You, T. T. (2025). Domain Adaptation for Vehicle Detection Under Adverse Weather. *IEEE Open Journal of Intelligent Transportation Systems*, 6, 1–11. <https://doi.org/10.1109/OJITS.2025.3563373>
- [13] Liu, B., Jin, J., Zhang, Y., & Sun, C. (2025). WRRT-DETR: Weather-Robust RT-DETR for Drone-View Object Detection in Adverse Weather. *Drones*, 9, 1–21. <https://www.mdpi.com/2504-446X/9/5/369>
- [14] Mo, T., Zheng, S., Chan, W. Y., & Yang, R. (2025). Review of AI Image Enhancement Techniques for In-Vehicle Vision Systems Under Adverse Weather Conditions. *World Electric Vehicle Journal*, 16(2), 1–20. <https://doi.org/10.3390/wevj16020072>
- [15] Nawaz, M., Khan, S., Daud, M., Asim, M., Anwar, G. A., Shahid, A. R., Ho, H. P. A., Chan, T., Pak Kong, D., & Yuan, W. (2025). Improving Autonomous Vehicle Cognitive Robustness in Extreme Weather With Deep Learning and Thermal Camera Fusion. *IEEE Open Journal of Vehicular Technology*, 6, 1–16. <https://doi.org/10.1109/OJVT.2025.3529495>
- [16] Ni, H., & Su, W. (2025). A Self-Attention CycleGAN for Unsupervised Image Hazing. *Big Data and Cognitive Computing*, 9(4), 1–16. <https://doi.org/10.3390/bdcc9040096>
- [17] Ogunrinde, I., & Bernadin, S. (2023). Deep Camera–Radar Fusion with an Attention Framework for Autonomous Vehicle Vision in Foggy Weather Conditions. *Sensors*, 23(14), 1–27. <https://doi.org/10.3390/s23146255>
- [18] Osipov, A., Pleshakova, E., Gataullin, S., Finogeev, A., & Yadav, V. (2022). Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions. *Sustainability (Switzerland)*, 14(4), 1–16. <https://doi.org/10.3390/su14042420>
- [19] Patil, S. M., Pawar, S. D., Mhatre, S. N., & Kharade, P. A. (2024). Yolov4-based hybrid feature enhancement network with robust object detection under adverse weather conditions. *Signal, Image and Video Processing*, 18, 1–31. <https://doi.org/10.1007/s11760-024-03068-6>
- [20] Piroli, A., Dallabetta, V., Kopp, J., Walessa, M., Meissner, D., & Dietmayer, K. (2023). Energy-Based Detection of Adverse Weather Effects in LiDAR Data. *Arxiv*, 1–8. <https://doi.org/10.1109/LRA.2023.3282382>
- [21] Rao, Z., Guo, J., Tang, L., Huang, Y., Ding, X., & Guo, S. (2025). SRCD: Semantic Reasoning With Compound Domains for Single-Domain Generalized Object Detection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7), 2025. <https://doi.org/10.1109/TNNLS.2024.3480120>
- [22] Roy, S. D., & Bhowmik, M. K. (2022). AWDMC-Net: Classification of Adversarial Weather Degraded Multiclass scenes using a Convolution Neural Network. *Computer Vision and Image Understanding*, 222, 1–19. <https://doi.org/10.1016/j.cviu.2022.103498>
- [23] Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B., & Fortier, P. (2022). Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions. *Electronics (Switzerland)*, 11, 1–11. <https://doi.org/10.3390/electronics11040563>
- [24] Siddiqua, M., Brahim Belhaouari, S., Akhter, N., Zameer, A., & Khurshid, J. (2023). MACGAN: An All-in-One Image Restoration under Adverse Conditions Using Multidomain Attention-Based Conditional GAN. *IEEE Access*, 11, 70482–70502. <https://doi.org/10.1109/ACCESS.2023.3289591>
- [25] Tahir, N. U. A., Zhang, Z., Asim, M., Chen, J., & ELAffendi, M. (2024). Object Detection in Autonomous Vehicles under Adverse Weather: A Review of Traditional and Deep Learning Approaches. *Algorithms*, 17, 1–36. <https://doi.org/10.3390/a17030103>
- [26] Thanh, P. T. H., Tran, T. T., Nguyen, T. H., Nguyen, M. H. V., Pham, T. V., Duy, T. V. T., & Nguyen, D. D. (2025). ULDepth: Transform Self-supervised Depth Estimation to Unpaired Multi-domain Learning. *IEEE Open Journal of Signal Processing*, 1–13. <https://doi.org/10.1109/OJSP.2025.3597873>
- [27] Wang, M., Su, T., Chen, S., Yang, W., Liu, J., & Wang, Z. (2023). Automatic Model-Based Dataset Generation for High-Level Vision Tasks of Autonomous Driving in Haze Weather. *IEEE Transactions on Industrial Informatics*, 19(8), 9071–9081. <https://doi.org/10.1109/TII.2022.3224958>
- [28] Yang, H., Carballo, A., Zhang, Y., & Takeda, K. (2023). Framework for Generation and Removal of Multiple Types of Adverse Weather from Driving Scene Images. *Sensors*, 23(3), 1–16. <https://doi.org/10.3390/s23031548>
- [29] Yuan, Y., Wei, Y., Guo, Y., Chen, J., & Jiang, T. (2025). CME-YOLO: A Cross-Modal Enhanced YOLO Algorithm for Adverse Weather Object Detection in Autonomous Driving. *Big Data and Cognitive Computing*, 9, 1–32. <https://doi.org/10.3390/bdcc9040092>

- [30] H. C. Shin et al., 'Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning', *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
- [31] S. Di, H. Zhang, C. G. Li, X. Mei, D. Prokhorov, and H. Ling, 'Cross-Domain Traffic Scene Understanding: A Dense Correspondence-Based Transfer Learning Approach', *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 745–757, Mar. 2018, doi: 10.1109/TITS.2017.2702012.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet Classification with Deep Convolutional Neural Networks', *Commun ACM*, vol. 60, no. 6, 2017, doi: 10.1145/3065386.
- [33] D. Dai, C. Sakaridis, S. Hecker, and L. van Gool, 'Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding', *Int J Comput Vis*, vol. 128, no. 5, pp. 1182–1204, May 2020, doi: 10.1007/s11263-019-01182-4.
- [34] J. Lu, M. Xu, R. Yang, and Z. Wang, 'Understanding and Predicting the Memorability of Outdoor Natural Scenes', *IEEE Transactions on Image Processing*, vol. 29, pp. 4927–4941, 2020, doi: 10.1109/TIP.2020.2975957.
- [35] H. Zhou, J. Ma, C. C. Tan, Y. Zhang, and H. Ling, 'Cross-Weather Image Alignment via Latent Generative Model with Intensity Consistency', *IEEE Transactions on Image Processing*, vol. 29, pp. 5216–5228, 2020, doi: 10.1109/TIP.2020.2980210.
- [36] M. Milford et al., 'Condition-invariant, top-down visual place recognition', in 2014 IEEE International Conference on Robotics and Automation (ICRA), May 2014, pp. 5571–5577. doi: 10.1109/ICRA.2014.6907678.
- [37] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, 'Curriculum self-paced learning for cross-domain object detection', *Computer Vision and Image Understanding*, vol. 204, Mar. 2021, doi: 10.1016/j.cviu.2021.103166.
- [38] Y. Yang, H. Dong, G. Liu, L. Zhang, and L. Li, 'Cross-Domain Traffic Scene Understanding by Integrating Deep Learning and Topic Model', *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/8884669.
- [39] C. Song, J. Wu, L. Zhu, M. Zhang, and H. Ling, 'Nighttime Road Scene Parsing by Unsupervised Domain Adaptation', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3244–3255, Apr. 2022, doi: 10.1109/TITS.2020.3033569.
- [40] C. Zhang, W. Ding, G. Peng, F. Fu, and W. Wang, 'Street View Text Recognition with Deep Learning for Urban Scene Understanding in Intelligent Transportation Systems', *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4727–4743, Jul. 2021, doi: 10.1109/TITS.2020.3017632.
- [41] M. Naseer, S. H. Khan, and F. Porikli, 'Indoor Scene Understanding in 2.5/3D for Autonomous Agents: A Survey', vol. 7, 2019, doi: 10.1109/ACCESS.2018.2886133.
- [42] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu, and C. Guan, 'Federated Transfer Learning for EEG Signal Classification', in *Proc. IEEE EMBC*, 2020, pp. 3040–3045. doi: 10.1109/EMBC44109.2020.9175344.
- [43] U. Nadeem, S. A. A. Shah, F. Sohel, R. Togneri, and M. Bennamoun, 'Deep learning for scene understanding', in *Smart Innovation, Systems and Technologies*, vol. 136, Springer Science and Business Media Deutschland GmbH, 2019, pp. 21–51. doi: 10.1007/978-3-030-11479-4_2.
- [44] X. Li et al., 'Yolov3-Pruning(transfer): real-time object detection algorithm based on transfer learning', *J Real Time Image Process*, Aug. 2022, doi: 10.1007/s11554-022-01227-x.
- [45] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, 'A survey of transfer learning', *Journal of Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.
- [46] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, 'DeCAF: a deep convolutional activation feature for generic visual recognition', in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32 (ICML'14)* 2014.
- [47] Torrey, L. and Shavlik, J. 'Transfer Learning' *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global, Hershey, 242–264 (2010) <https://doi.org/10.4018/978-1-60566-766-9.ch011>.
- [48] P. Ganesh Pawar and V. Devendran, 'Scene Understanding: A Survey to See the World at a Single Glance', 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), 2019.
- [49] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, 'Transferable Representation Learning with Deep Adaptation Networks', *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 12, pp. 3071–3085, Dec. 2019, doi: 10.1109/TPAMI.2018.2868685.
- [50] A. Chakraborty, Cosmin Anitescu, Xiaoying Zhuang, and T. Rabczuk, 'Domain adaptation-based transfer learning approach for solving PDEs on complex geometries', vol. 1, p. 3, doi: 10.1007/s00366-022-01661-2.
- [51] S. J. Pan and Q. Yang, 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. doi: 10.1109/TKDE.2009.191.
- [52] D. P. Thesis, 'Transfer Learning for Object Category Detection'.
- [53] M. Afif, R. Ayachi, Y. Said, and M. Atri, 'Deep Learning Based Application for Indoor Scene Recognition', *Neural Process Lett*, vol. 51, no. 3, pp. 2827–2837, Jun. 2020, doi: 10.1007/s11063-020-10231-w.
- [54] R. Xiao, Y. Wang, and C. Tao, 'Fine-Grained Road Scene Understanding from Aerial Images Based on Semisupervised Semantic Segmentation Networks', *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2022, doi: 10.1109/LGRS.2021.3059708.
- [55] Ge, F., Zhang, Y., Wang, L., Tan, Y., Coleman, S., & Kerr, D. (2025). MDAG-Net: Multi-Domain Association-Guided Network for Image-based Long-term Visual Localization. *IEEE Transactions on Cognitive and Developmental Systems*, 1–15. <https://doi.org/10.1109/TCDS.2024.3525180>.
- [56] J. Lu, N. Verma, and N. K. Jha, 'Convolutional Autoencoder-Based Transfer Learning for Multi-Task Image Inferences', *IEEE Trans Emerg Top Comput*, vol. 10, no. 2, pp. 1045–1057, 2022, doi: 10.1109/TETC.2021.3068063.
- [57] F. Husain, B. Dellen, and C. Torras, 'Scene Understanding Using Deep Learning', *Handbook of Neural Computation*, pp. 373–382, Jan. 2017, doi: 10.1016/B978-0-12-811318-9.00020-X.
- [58] S. Safavi and M. Jalali, 'DeePOF: A hybrid approach of deep convolutional neural network and friendship to Point-of-Interest (POI) recommendation system in location-based social networks', *Concurr Comput*, vol. 34, no. 15, Jul. 2022.
- [59] Mane D, Arora S and Shelke S (2026) IndiaScene365: a transfer learning dataset for Indian scene understanding in diverse weather condition. *Front. Artif. Intell.* 8:1669512. doi: 10.3389/frai.2025.1669512

- [60] Mane, D., Arora, S., Shelke, S. (2025). Target-Aware Generative Adversarial Network for Domain Adaptive Yolo in Cross Weather. International conf. Integration of Artificial Intelligence in IoT: Opportunities and Challenges.
- [61] BDD Dataset Link: <https://www.kaggle.com/datasets/awsaf49/bdd100k-dataset/data>
- [62] IndiaScene365 Dataset Link: <https://data.mendeley.com/datasets/pwffhg6nhz/2>

Authors' Profiles



Deepa Mane is Research Scholar in Computer Engineering at SPPU. Received M.E. (Information Technology) Pune Institute of computer technology, Pune, India. B.E. (Computer Engineering) from Mumbai University. Her Research interest are in Machine Learning, deep learning and Data Science. Working as assistant Professor in SPPU. Has Received Best Paper Presented in National Conference on Technical Revolution 2016.



Sandhya Arora is working as professor in Department of Computer Engineering, Cummins college of Engineering for women Pune. She has completed Ph.D. (Computer Science & Engineering) from Jadavpur University Kolkata in 2012, M. Tech from Banasthali Vidyapith, Rajasthan, India and B.E. (Computer Engineering) from University of Rajasthan, India. She has teaching experience of 26+ years. She is a life member of ISTE, CSI, ACM. She has more than 60 research publications. Also, she has written 7 books under university press, India and CRC press, USA.

How to cite this paper

Deepa Mane, Sandhya Arora, "Cross-Domain Outdoor Scene Understanding Using PC-GAN and TL-SZKSE-DenseNet in Adverse Weather Conditions", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.18, No.5, pp. 140-162, 2026. DOI:10.5815/ijieeb.2026.05.09