applied system innovation

MDPI

Article

Detection of Road Surface Ice and Snow Cover Using the HPC-D Segmentation Algorithm Combined with the SCI Illumination Enhancement Model

Yuming Mu and Sen Yang *

School of Control and Information Engineering, Northeast Forestry University, Harbin 150040, China; w05478040@126.com
* Correspondence: yangsennefu@163.com

Abstract

Road-surface ice, snow and standing water can impair winter driving, while irregular boundaries and changing illumination complicate image-based assessment. This study investigates pixel-wise semantic segmentation using HPC-D, a DeepLabv3+ architecture incorporating convolutional block attention (CBAM), parallel axial attention (PAA) and hierarchical feature fusion (HFF), together with self-calibrated illumination (SCI) enhancement for low-light images. The reported dataset contains 5000 image entries from a campus road environment, with a reported training/validation/test allocation of 3000/1000/1000. The original experimental tables report a mean intersection-over-union (mIoU) of 0.96, a reported mean average precision (mAP) of 0.95 and an accuracy of 0.98 for HPC-D. Relative to the other rows in the segmentation comparison, the numerical differences are 2–21 percentage points in mIoU, 0–10 percentage points in mAP and 3–14 percentage points in accuracy. In the illumination comparison, the SCI row exceeds the no-enhancement row by 16 percentage points in mIoU and 17 percentage points in mAP. These are descriptive differences in the reported results. Incomplete documentation of image lineage, acquisition-session separation, metric implementation and comparator configurations limits their interpretation as evidence of independent generalization. The contribution is therefore an architectural integration and a preliminary evaluation for visually annotated road-surface segmentation; cross-camera robustness, shadow rejection and real-time vehicle operation remain unvalidated.

Keywords: winter driving safety; semantic segmentation; illumination enhancement; convolutional attention; feature fusion; downsampling

Check for updates

Academic Editor: Abdelkader Sbihi

Received: 28 August 2026
Revised: 20 October 2026
Accepted: 27 October 2026
Published: 29 October 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cold weather can produce water, ice and snow on road surfaces and alter vehicle braking and stability. Research on pavement friction, winter maintenance and weather-related traffic conditions motivates reliable road-surface assessment [1–6]. The magnitude of these risks depends on the road, weather regime and traffic conditions. The present work focuses on locating visually identifiable surface-cover regions in images, rather than predicting accident frequency, measuring friction or controlling a vehicle.

Existing approaches address several related but distinct tasks. Yang et al. developed an asphalt-pavement icing prediction model using support vector regression and Bayesian optimization [7]. Li et al. studied spatiotemporal road-scene reconstruction using

a superpixel-based Markov random field [8]. Wang et al. investigated autonomous emergency braking on snow–asphalt joint pavement [9]. Musci et al. investigated ice detection on aircraft surfaces using hyperspectral and multispectral images [10]. These studies provide context for sensing, scene reconstruction and safety control; they are not interchangeable examples of pixel-wise road-ice segmentation.

Broader transportation and perception studies discuss intelligent transport systems [11], image recognition in difficult weather [12], millimeter-wave (mmWave) backscattering with a one-dimensional convolutional neural network (CNN) for black-ice detection [13], road-object detection [14], autonomous-driving system design [15], road-damage detection [16] and parallel driving in cyber–physical–social systems [17]. Their outputs and sensing modalities differ from the semantic masks considered here. More directly related work includes dual-spectrum roadway snow detection by He et al. [18]. Li et al.'s ParallelEye pipeline [19] concerns synthetic image generation for vehicle perception. Winter road-surface detection was investigated by Moroto et al. [20].

Building on the DeepLab family [21–24], this study combines HPC-D (HFF+PAA+CBAM–DeepLabv3+) with self-calibrated illumination (SCI) enhancement. The task is semantic segmentation: assigning a surface-category label to each evaluated pixel. Here, detection denotes identifying surface-cover regions through segmentation, not bounding-box object detection. SCI is an image-enhancement stage; it is distinguished from sample-expanding training augmentation. The work examines whether the integrated architecture is associated with higher reported segmentation scores within the collected road-image dataset. Its evidential scope is limited by the protocol details discussed in Sections 2, 3.3 and 4.4.

Section 2 describes acquisition, annotation and the limitations of the dataset documentation. Section 3 presents the unchanged HPC-D architecture and SCI processing, clarifies loss notation and defines the evaluation quantities. Section 4 discusses the original ablation, segmentation and illumination-comparison tables and their interpretation. Section 5 states the conclusions and the validation needed before broader deployment.

2. Dataset and Preprocessing

2.1. Image Acquisition

The target categories are ice, snow, dry road and standing water. Images were acquired using an EOS 850D camera (Canon Inc., Tokyo, Japan) with the lens aperture reported as f/5.6, and a fixed tripod was used to set the viewing angle. A UT383B mini light meter was used to measure illuminance. Illuminance is a lighting measurement; it does not determine whether a road surface is physically frozen. The acquisition description does not establish a measured road-surface temperature, a friction measurement or a synchronized meteorological reference for every image.

Data were collected on asphalt road surfaces on the Northeast Forestry University campus in Harbin over a reported period of 14 months. The collection includes different viewing angles, times, weather, illumination and traffic conditions. The image resolution is 2400 × 1600 pixels. This acquisition diversity is confined to the documented site and camera; it does not by itself establish independence between subsets or representativeness of other roads and imaging devices.

2.2. Annotation and Physical-State Interpretation

Images were annotated manually with LabelMe (version not preserved in the original annotation records) for the four surface-cover categories, and the background was also annotated. These masks provide visual reference labels. The available annotation description does not document a separate physical-state measurement, the number of annotators, an adjudication procedure or an agreement statistic. Consequently, the labels must not be

interpreted as independently instrument-verified ice or water states. Transparent ice, wet asphalt, shadows, slush and mixed boundaries can be visually ambiguous. The treatment of such ambiguity and the inclusion or exclusion of background in the reported scores are not sufficiently specified to support a stronger claim.

2.3. Dataset Accounting and Partition Independence

The reported dataset contains 5000 image entries, allocated in a 6:2:2 ratio: 3000 training entries, 1000 validation entries and 1000 test entries. Random cropping, rotation, scaling and color transformations are described as augmentation operations. However, the dataset summary does not distinguish the number of original captures from the number of derived images, provide per-class image or pixel counts, or establish whether partitioning preceded augmentation. These counts are retained as reported and are not relabeled as counts of independent original photographs.

For a leakage-controlled protocol, the split unit must be the original capture or, more conservatively, an acquisition-session group. All crops and transformed descendants of one source, and related captures within a group, must remain in the same subset. Random augmentation should be applied only to training data after the groups are assigned; validation and test processing should be deterministic. Validation results may guide model selection; test images, labels and scores must be excluded from training and selection of model or enhancement settings. This paragraph states the required protocol. The present documentation does not establish that the original experiment satisfied it.

The number of original images, augmented derivatives, acquisition sessions and class-specific pixels cannot be recovered from the aggregate total of 5000. An image can contain several surface categories, so image-level class presence and pixel-level class proportions are different quantities. No equal class distribution, deduplication outcome or absence of cross-subset derivatives is asserted. This incompleteness limits the interpretation of the original results, especially the high reported mIoU. Representative road-image examples are presented in Figure 1.



Figure 1. Examples from the reported road-image dataset.

3. Principles and Methods

3.1. Principles of the HPC-D Segmentation Algorithm

HPC-D retains DeepLabv3+ as its segmentation framework and integrates convolutional block attention (CBAM), parallel axial attention (PAA), and hierarchical feature fusion (HFF) into the feature-processing and decoder paths shown in Figure 2. DeepLabv3+

uses atrous spatial pyramid pooling and decoder refinement to combine contextual and spatial information [24]. In this application, indistinct surface boundaries and heterogeneous coverage motivate additional feature selection and fusion. The following descriptions explain the intended roles of the existing modules without introducing a new architecture or changing their connectivity.

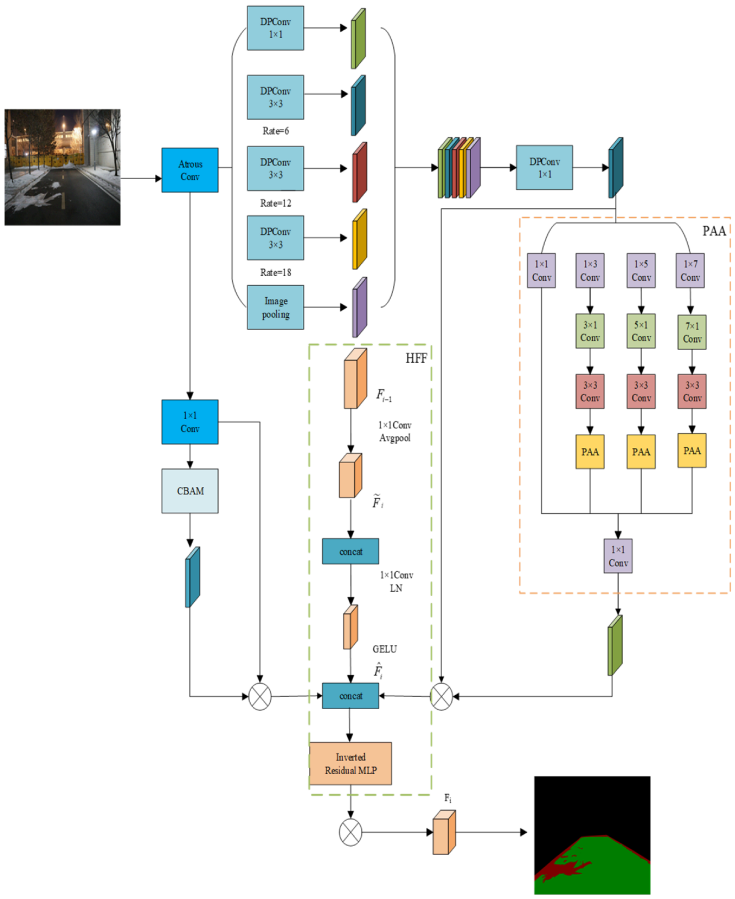


Figure 2. HPC-D architecture with convolutional block attention (CBAM), parallel axial attention (PAA), and hierarchical feature fusion (HFF) paths. Red and green in the example output denote distinct color-coded prediction regions.

CBAM applies channel and spatial attention to refine feature responses [25]. The existing HPC-D configuration uses this mechanism in its feature-processing path. Average and maximum pooling provide complementary summaries for channel attention. The intended role is to emphasize informative feature responses, but the aggregate scores in Table 1 do not independently establish which individual boundaries or channels account for the observed differences.

The PAA branch uses parallel axial attention to aggregate information along the horizontal and vertical axes. The PAA encoder design described in UACANet provides the relevant method source [26], including multi-scale receptive-field processing and channel reduction. In HPC-D, PAA is retained at the location shown in Figure 2. Channel reduction and spatial downsampling are distinct operations; the description does not imply that an attention operation alone reduces spatial resolution. Preserving every image detail, reducing the complete model's parameter count and preventing overfitting are not established by the reported aggregate scores.

Table 1. Reported ablation results.

Model	mIoU	mAP*	Accuracy
DeepLabv3+	0.81	0.79	0.87
CBAM-D	0.92	0.87	0.94
PAA-D	0.91	0.86	0.91
HFF-D	0.89	0.93	0.88
HPC-D	0.96	0.95	0.98

All numeric entries are unchanged. CBAM-D, PAA-D and HFF-D are the separately named variants in the original table. mAP* denotes the originally reported mAP with the implementation limitation explained in Section 3.3. The table does not establish a cumulative ablation or repeated-run significance.

The HFF block integrates features from different levels through the existing cascaded fusion path. The hierarchical fusion design in HiFuse is cited as the relevant source for the HFF concept and its inverted residual multilayer perceptron component [27]. Figure 2 specifies the arrangement used here. This attribution does not equate HPC-D with the complete HiFuse classification network, nor does it establish that every source-code detail is identical. The architectural contribution examined in this study is the integration of the modules into the road-surface segmentation framework.

The fused features are passed to the decoder for pixel-wise prediction. The originally reported training objective combines a cross-entropy term with an intersection-over-union term. The positive-label expression printed in the original manuscript was incorrectly named binary cross-entropy. Equations (1)–(3) retain their additive structure and algebraic terms, using the neutral symbol L_{CE} in place of the misleading L_{BCE} label. This is a correction of mathematical notation, not a change to the trained model or a retrospective verification of the training implementation.

$$L = L_{CE} + L_{IoU} \tag{1}$$

$$L_{CE} = -\sum_i y_i \, log(p_i) \tag{2}$$

$$L_{IoU} = 1 - IoU \tag{3}$$

In Equation (2), y_i and p_i denote a target indicator and the corresponding predicted probability. With one-hot, mutually exclusive class targets and summation over classes, this positive-label expression has the categorical cross-entropy form for a pixel. A binary cross-entropy expression also includes the negative-target contribution $-(1 - y_i) \log(1 - p_i)$. The original description does not identify the activation, tensor layout, class reduction or implementation sufficiently to establish which objective produced the trained weights. Likewise, Equation (3) does not specify the soft-IoU aggregation, smoothing or background policy used in training. These implementation questions remain distinct from correcting the printed formulas.

Combining a label-fitting term with an overlap term is intended to account for both category assignment and spatial overlap. The available experimental tables do not isolate the effect of the loss terms. Figure 2 reproduces the existing HPC-D architecture; no layer, branch or feature-fusion connection is changed in this revision.

3.2. SCI Enhancement Algorithm

SCI is the self-calibrated illumination method introduced by Ma et al. [28]. Under the Retinex image-formation approximation, the observed low-light image is represented by an element-wise product of reflectance and illumination, rather than a vector dot product. SCI estimates illumination through a weight-sharing learning process. Its reported update

structure is retained in Equation (4). This description attributes SCI to its original authors and does not present it as a newly proposed component of HPC-D.

$$u^t = H_\theta(x^t),\ x^0 = y;\ x^{t+1} = x^t + u^t \tag{4}$$

In Equation (4), u^t and x^t are the residual and illumination estimate at stage t, and y denotes the input low-light image. The network parameters θ are shared across stages; the residuals and illumination estimates themselves can vary with t. SCI introduces a self-calibration operation to encourage consistent stage outputs during training [28]. Its reported expression is retained in Equation (5). Parameter sharing should not be confused with equality of all intermediate image estimates.

$$z^t = y \oslash x^t;\ s^t = k_\theta(z^t);\ v^t = y + s^t \tag{5}$$

Here, z^t is formed through element-wise division of y by the illumination estimate x^t; s^t is the estimated calibration correction; and v^t is the calibrated input. The mathematical notation describes SCI's published mechanism. The road-image experiment does not document the enhancement checkpoint, its training data, numerical safeguards or adaptation settings in sufficient detail for an implementation-level reconstruction.

The original SCI study separates its training procedure from the lightweight enhancement path used for inference [28]. That efficiency result is attributable to the original SCI implementation and its test conditions; it is not a measured latency for the combined SCI–HPC-D pipeline in this work. Here, SCI is treated as illumination preprocessing. Section 4.3 reports its association with downstream segmentation scores and distinguishes that observation from generalization, physical-state verification and deployment speed.

3.3. Evaluation Quantities and Implementation Limitations

Let C denote the set of evaluated classes and Ω the evaluated pixels after any explicit ignore mask. For class c, TP_c, FP_c and FN_c are pixel-level true-positive, false-positive and false-negative counts. The standard overlap and pixel-accuracy quantities are defined below. These definitions make the intended terminology explicit; they do not independently verify the class set, background policy or aggregation code behind the original scores.

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c} \tag{6}$$

$$mIoU = \frac{1}{|C|}\sum\nolimits_{c \in C} IoU_c \tag{7}$$

$$Accuracy = \frac{\sum_c TP_c}{N},\ N = \sum\nolimits_c N_c \tag{8}$$

In Equations (6)–(8), N_c is the number of evaluated pixels whose reference label is c, and N is the total number of evaluated pixels. The sums for accuracy include every evaluated reference class, even when a selected class subset is used for macro-averaged IoU. Background inclusion, ignored labels, zero-union classes and pooling across images must be stated for numerical reproducibility. The original score tables do not resolve these choices.

For pixel-wise average precision, a defensible definition treats each class as a one-versus-rest ranking problem using continuous per-pixel scores. At threshold τ, precision $P_c(\tau)$ and recall $R_c(\tau)$ are formed from the corresponding pixel counts. With thresholds ordered so that recall increases, non-interpolated average precision and macro-averaging can be written as follows [29].

$$AP_c = \sum\nolimits_k (R_{c,k} - R_{c,k-1})P_{c,k} \tag{9}$$

$$mAP = \frac{1}{|C|} \sum_{c \in C} AP_c \tag{10}$$

Here, k indexes successive distinct score thresholds, with $R_{c,0} = 0$. An implementation collects the reference indicator and the continuous score for each evaluated pixel, groups tied scores, sorts by descending score, accumulates true and false positives, and sums precision weighted by recall increments. AP must be computed before reducing probabilities to hard class labels. The handling of classes with no positive reference pixels must be disclosed. This is a transparent evaluation specification, not a claim that a newly implemented evaluator reproduced the original mAP values.

The original documentation does not identify whether the reported mAP was computed by this pixel-ranking procedure, a different interpolation convention, or an average of fixed-threshold precision values. These quantities are not interchangeable. The mAP column is retained and marked with an asterisk in Tables 1–3 to indicate unresolved implementation provenance. No detection-style mAP at bounding-box or instance-mask IoU thresholds is inferred. The reported mIoU and accuracy also remain conditional on verification of their evaluation masks and class policies.

4. Experiments and Analysis

Training was reported on a Colab server using a Tesla T4 graphics processing unit (GPU) with 16 GB of GPU memory. The reported HPC-D settings are batch size 8, stochastic gradient descent, a step-wise learning-rate schedule, initial learning rate 7×10^{-3}, minimum learning rate equal to 0.01 of the maximum, and 200 epochs with evaluation every two epochs. Pretrained weights were used. These settings are retained. The backbone/checkpoint identifier, network input size after preprocessing, random seed, software versions and checkpoint-selection rule are not fully specified. The acquisition resolution must therefore not be assumed to be the training or inference tensor resolution. Figure 3 retains the original illustrative segmentation outputs.

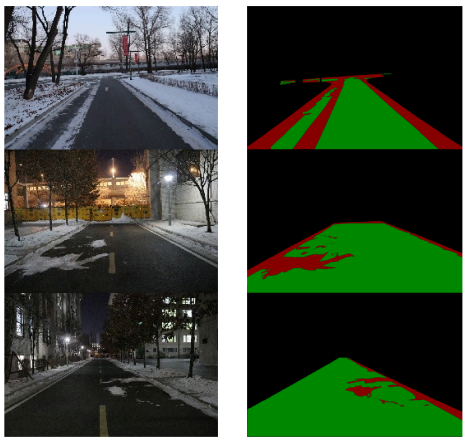


Figure 3. Original illustrative HPC-D segmentation outputs. The red and green regions visually distinguish predicted segmentation regions, rather than calibrated physical-state measurements.

4.1. Ablation Studies

Table 1 reports DeepLabv3+, CBAM-D, PAA-D, HFF-D and the combined HPC-D configuration. The row labels describe separate module variants and a combined configuration; they do not document a cumulative sequence of CBAM, then CBAM+PAA, and then CBAM+PAA+HFF. HPC is not a fourth added module. The interpretation below follows the rows actually reported. A complete factorial ablation and repeated-run uncertainty

estimates are not available, so module interactions and statistical significance cannot be quantified from this table.

Table 1 retains all original scores. The following differences are arithmetic comparisons of the reported rows, not new experimental results.

Relative to DeepLabv3+ (0.81 mIoU, 0.79 mAP and 0.87 accuracy), HPC-D reports 0.96, 0.95 and 0.98, respectively. The corresponding differences are 0.15, 0.16 and 0.11, or 15, 16 and 11 percentage points. The mAP label is subject to the implementation limitation in Section 3.3.

The CBAM-D, PAA-D and HFF-D rows exceed the baseline mIoU by 11, 10 and 8 percentage points, respectively. Their mAP differences are 8, 7 and 14 points, and their accuracy differences are 7, 4 and 1 point. HFF-D therefore has the largest mAP difference among the individual-module rows, but smaller differences in mIoU and accuracy. The different metric ordering prevents a single universal ranking of individual-module contributions.

HPC-D has the highest reported values in Table 1. This pattern is consistent with a potentially useful combination of feature attention and fusion, but it does not prove additive gains, a particular mechanism of error reduction or statistical superiority. The table does not establish performance in bounding-box detection, image-level classification or independent deployment domains.

The ablation results support a descriptive comparison of the recorded configurations, conditional on the unresolved split, evaluation and implementation details. They are not evidence that overfitting or data leakage has been excluded.

4.2. Performance Comparison Experiment of Segmentation Algorithms

Related U-Net-based methods include Unet3+ [30], RUC-Net [31] and MCA-UNet [32]. MCA-UNet uses multi-scale cross co-attention for medical image segmentation and is cited here as related work only, not as an evaluated comparator. Table 2 reports comparisons with Unet3+, RUC-Net, GMD-Net, MFCA-Net, DA-TransUNet, SeNet, DHF-Net and GCIFF-Net [30,31,33–38]. These methods originate in different segmentation domains. The DHFNet source [37], for example, uses RGB–thermal inputs, whereas the present acquisition description documents RGB camera images. The original comparison does not specify the input adaptation, pretrained weights and training configuration for every baseline. The listed scores therefore cannot establish a fully controlled benchmark.

Table 2. Reported segmentation comparisons with provenance limitations.

Model	mIoU	mAP*	Accuracy
Unet3+ [30]	0.75	0.86	0.88
RUC-Net [31]	0.81	0.85	0.91
GMD-Net [33]	0.83	0.89	0.93
MFCA-Net [34]	0.89	0.93	0.94
DA-TransUNet [35]	0.86	0.88	0.84
SeNet [36]	0.92	0.91	0.94
DHF-Net [37]	0.94	0.95	0.95
GCIFF-Net [38]	0.90	0.89	0.93
HPC-D	0.96	0.95	0.98

The numeric entries for all remaining comparator rows are unchanged. mAP* is subject to Section 3.3. MCA-UNet [32] is cited as related work only; no performance values are reported for it. Corrected bibliographic identities do not verify the code or input adaptation used for the original comparator runs.

HPC-D reports mIoU 0.96, mAP 0.95 and accuracy 0.98. Across the other Table 2 rows, the numerical differences are 0.02–0.21, 0.00–0.10 and 0.03–0.14, respectively, corresponding

to 2–21, 0–10 and 3–14 percentage points. The mAP value is tied with DHF-Net, rather than exceeding every comparator. These arithmetic observations do not establish causal contributions of particular modules or show that PAA prevents overfitting. The original Figure 3 illustrates segmentation outputs; Figure 4 illustrates low-light inputs and is not a module-isolation experiment.

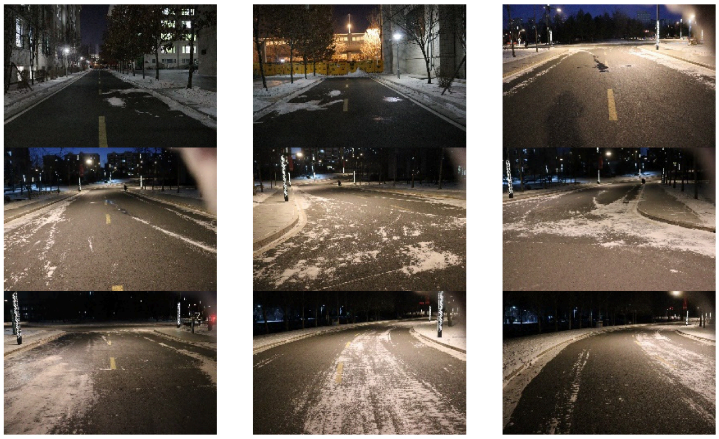


Figure 4. Examples of nighttime low-light road images.

4.3. Comparative Experiments on Illumination Enhancement Methods

The image collection includes nighttime low-light road scenes. Figure 4 retains examples from this part of the collection. These examples illustrate the appearance of the inputs; they do not quantify the frequency of difficult cases or constitute a separate shadow-confounder test.

The original illumination experiment describes Group A as images with measured illuminance above 500 lx and Group B as images below 500 lx. The 500 lx value is an operational grouping threshold used in this study, not a universal physical definition of low light. The group sizes, treatment of measurements exactly equal to 500 lx, measurement geometry and exposure settings are not specified. The summary also does not establish whether identical held-out images, fixed segmentation weights and unchanged masks were used across every enhancement method.

The illumination-comparison table retains the labels RetinexNet, FIDE, DRBN, KinD, EnGAN, ZeroDCE, RUAS and SCI. References [28,39–45] have been aligned with the intended low-light methods: FIDE with decomposition-and-enhancement [40], DRBN with the semi-supervised recursive-band method [41], KinD with Kindling the Darkness [42], and the EnGAN label with the intended EnlightenGAN method [43]. These corrected method citations do not independently identify the code or checkpoint used for each original row. Figure 5 retains the original enhancement examples, with panels (a)–(h) in the same method order.

In Table 3, SCI reports mIoU 0.97, mAP 0.95 and accuracy 0.96, compared with 0.81, 0.78 and 0.82 without enhancement. The differences are therefore 16, 17 and 14 percentage points, respectively. Relative to the other enhancement rows, the differences are 4–9 points in mIoU, 3–8 points in mAP and 3–7 points in accuracy. These values correct the inconsistent ranges and reversed metric assignments in the original text without changing a table entry. Because Table 3 does not present a separate, matched Group A reference, equality with normal-light performance is not established. The results are descriptive until the image pairing, model settings and metric implementation are verified.

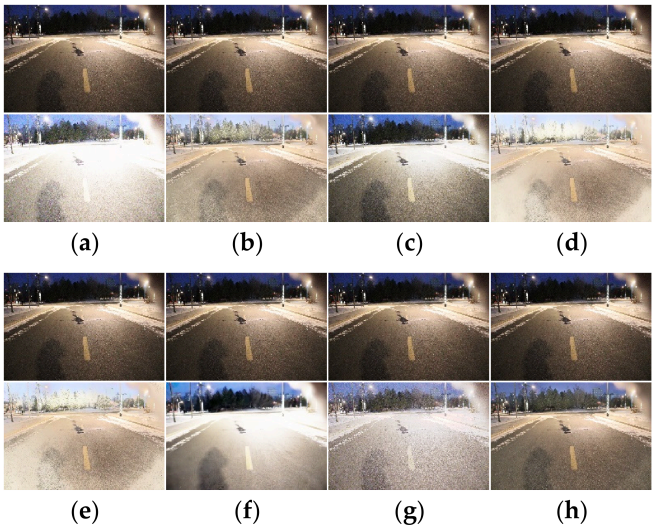


Figure 5. Original enhancement examples for RetinexNet, FIDE, DRBN, KinD, EnGAN, ZeroDCE, RUAS, and SCI in panels (**a**–**h**).

Table 3. Reported illumination-enhancement comparisons.

Enhancement Method	mIoU	mAP*	Accuracy
No enhancement	0.81	0.78	0.82
RetinexNet	0.91	0.91	0.93
FIDE	0.91	0.89	0.91
DRBN	0.93	0.92	0.89
KinD	0.89	0.87	0.90
EnGAN	0.88	0.88	0.89
ZeroDCE	0.91	0.89	0.92
RUAS	0.92	0.92	0.93
SCI	0.97	0.95	0.96

All numeric entries are unchanged. mAP* is subject to Section 3.3. Group sizes, identical-image pairing, enhancement checkpoints and segmentation-training controls are not established by this summary. No separate matched normal-light reference is provided in the table.

4.4. Scope and Limitations of the Evidence

The three tables contain aggregate results from the reported experiment. They provide neither repeated-run variability nor per-image or per-class measurements from which uncertainty, class balance or paired significance tests can be reconstructed. A higher mean score is not sufficient evidence that all surface categories improve. Likewise, the large number of pixels in a segmentation dataset is not the number of statistically independent road scenes.

Dataset independence is the main unresolved issue. Original captures, their augmented descendants and closely related images from the same acquisition session may share appearance information. A reported 6:2:2 allocation alone cannot rule out this dependence. Until the original image-to-subset and session records are verified, the table values must not be interpreted as leakage-free estimates of generalization to new roads or sessions.

Camera transfer has not been evaluated. Differences in sensor noise, spectral response, exposure, white balance, optics, resolution and image processing can shift the input distribution. Resizing alone does not remove these effects. The use of one camera and one campus road environment confines the present evidence to that acquisition setting.

Shadows and other visual confounders remain a specific failure mode. Dark dry asphalt may resemble wet or icy pavement; bright reflections, markings, slush and partially

transparent ice can also confuse visual labels. Global brightening does not prove that these ambiguities are resolved, and illumination enhancement can alter contrast or amplify noise. No dedicated hard-negative test is reported, so robustness to shadow look-alikes is not claimed.

The Tesla T4 and 16 GB specification describes the reported training platform, not a minimum deployment requirement. Parameter count, multiply–accumulate operations, end-to-end latency, throughput, inference memory and power were not measured in the reported study. Computational demand depends on actual input size, channel dimensions, attention implementation, enhancement settings and hardware. A real-time claim therefore cannot be inferred from the architecture or from the efficiency of SCI in its original publication.

A typical dense convolution with output height H, output width W, kernel size k and input/output channels C_{in} and C_{out} requires approximately $H \times W \times C_{in} \times C_{out} \times k^2$ multiply–accumulate operations, ignoring bias and boundary details. Axial affinity computation may scale as $H \times W \times (H + W)$ times its feature dimension when attention is formed separately along both axes. These are explanatory operation-count relationships, not computed complexity measurements for HPC-D. A deployed pipeline would also include acquisition, preprocessing, enhancement, data transfer and postprocessing.

The practical value of the study is to identify an integration worth examining under better-documented evaluation. It does not establish vehicle control performance, stopping distance reduction or a validated safety system. The present revision makes these limits explicit while retaining the original architecture and the remaining reported numerical results.

5. Conclusions

This study investigates an existing integration of CBAM, PAA and HFF within DeepLabv3+ for semantic segmentation of visually annotated road-surface conditions. SCI provides an illumination-enhancement stage for low-light imagery. The architecture and image results are retained, and the values in the remaining table entries are unchanged; the revised interpretation separates the recorded scores from conclusions that require additional protocol evidence.

The reported HPC-D scores are mIoU 0.96, mAP 0.95 and accuracy 0.98. Relative to DeepLabv3+ in Table 1, the numerical differences are 15, 16 and 11 percentage points. Relative to the other entries in Table 2, the ranges are 2–21, 0–10 and 3–14 points. These observations indicate a promising configuration within the reported experiment, but do not establish statistically verified or independent-domain superiority. They do not support conclusions about object-detection AP or vehicle safety outcomes.

For the illumination experiment, the SCI row exceeds the no-enhancement row by 16 percentage points in mIoU, 17 in mAP and 14 in accuracy. This association motivates further examination of illumination preprocessing for segmentation. It does not demonstrate equivalence to normal-light performance, physical ice identification in ambiguous scenes, or real-time operation of the complete pipeline.

The principal limitations concern the undocumented original-image and session lineage, class balance, label verification, metric implementation and baseline adaptations. The results should therefore be treated as preliminary evidence for a site- and camera-specific segmentation study. Further work should establish independent acquisition-session evaluation, calibrated assessment of ambiguous ice/water/shadow cases, cross-camera transfer and measured end-to-end latency before considering integration with vehicle decision systems. Image-based surface segmentation alone does not provide a friction estimate or certify safe driving.

Author Contributions: Investigation, Y.M.; writing—original draft preparation, Y.M.; supervision, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets supporting the findings of this study are not publicly available. The original road images and annotation masks have not been released for public distribution by the authors.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-6) to assist with language editing and proofreading. The authors reviewed the suggested changes and take responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, X.; Jiang, Y.; Zheng, B.; Zhao, R. Theory and methodology on safety braking of autonomous vehicles based on the friction characteristic of road surface. *Chin. Sci. Bull.* **2020**, *65*, 3328–3340. [CrossRef]
2. Hääl, M.L.; Sürje, P. Environmental problems related to winter traffic safety conditions. *Balt. J. Road Bridge Eng.* **2006**, *1*, 45–53.
3. Theofilatos, A.; Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **2014**, *72*, 244–256. [CrossRef]
4. Usman, T.; Fu, L.; Miranda-Moreno, L.F. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accid. Anal. Prev.* **2010**, *42*, 1878–1887. [CrossRef]
5. Mashros, N.; Ben-Edigbe, J.; Hassan, S.A.; Hassan, N.A.; Yunus, N.Z.M. Impact of rainfall conditions on traffic flow and speed: A case study in Johor and Terengganu. *J. Teknol.* **2014**, *70*, 65–69. [CrossRef]
6. Abohassan, A.; El-Basyouny, K.; Kwon, T.J. Exploring the associations between winter maintenance operations, weather variables, surface condition, and road safety: A path analysis approach. *Accid. Anal. Prev.* **2021**, *163*, 106448. [CrossRef]
7. Yang, E.; Yang, Q.; Li, J.; Zhang, H.; Di, H.; Qiu, Y. Establishment of icing prediction model of asphalt pavement based on support vector regression algorithm and Bayesian optimization. *Constr. Build. Mater.* **2022**, *351*, 128955. [CrossRef]
8. Li, Y.; Liu, Y.; Zhu, J.; Ma, S.; Niu, Z.; Guo, R. Spatiotemporal road scene reconstruction using superpixel-based Markov random field. *Inf. Sci.* **2020**, *507*, 124–142. [CrossRef]
9. Wang, X.; Wang, J.; Sun, W.; Wang, Y.; Xie, F.; Guo, D. Development of AEB control strategy for autonomous vehicles on snow-asphalt joint pavement. *Int. J. Crashworthiness* **2022**, *27*, 1601–1621. [CrossRef]
10. Musci, M.A.; Mazzara, L.; Lingua, A.M. Ice detection on aircraft surface using machine learning approaches based on hyperspectral and multispectral images. *Drones* **2020**, *4*, 45. [CrossRef]
11. Wang, F.Y.; Lin, Y.; Ioannou, P.A.; Vlacic, L.; Liu, X.; Eskandarian, A.; Lv, Y.; Na, X.; Cebon, D.; Ma, J.; et al. Transportation 5.0: The DAO to safe, secure, and sustainable intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 10262–10278. [CrossRef]
12. Osipov, A.; Pleshakova, E.; Gataullin, S.; Korchagin, S.; Ivanov, M.; Finogeev, A.; Yadav, V. Deep learning method for recognition and classification of images from video recorders in difficult weather conditions. *Sustainability* **2022**, *14*, 2420. [CrossRef]
13. Kim, J.; Kim, E.; Kim, D. A black ice detection method based on 1-dimensional CNN using mmWave sensor backscattering. *Remote Sens.* **2022**, *14*, 5252. [CrossRef]
14. Mahaur, B.; Singh, N.; Mishra, K.K. Road object detection: A comparative study of deep learning-based algorithms. *Multimed. Tools Appl.* **2022**, *81*, 14247–14282. [CrossRef]
15. Zhao, J.; Zhao, W.; Deng, B.; Wang, Z.; Zhang, F.; Zheng, W.; Cao, W.; Nan, J.; Lian, Y.; Burke, A.F. Autonomous driving system: A comprehensive survey. *Expert Syst. Appl.* **2024**, *242*, 122836. [CrossRef]
16. Xiang, W.; Wang, H.; Xu, Y.; Zhao, Y.; Zhang, L.; Duan, Y. Road disease detection algorithm based on YOLOv5s-DSG. *J. Real-Time Image Process.* **2023**, *20*, 56. [CrossRef]
17. Wang, F.Y.; Zheng, N.N.; Cao, D.; Martinez, C.M.; Li, L.; Liu, T. Parallel driving in CPSS: A unified approach for transport automation and vehicle intelligence. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 577–587. [CrossRef]
18. He, X.; Wu, Y.; Zhang, K.; Zhu, X.; Yang, X. Roadway snow detection using dual-spectrum camera images and computer vision. *Transp. Res. Rec.* **2023**, *2677*, 220–232. [CrossRef]
19. Li, X.; Wang, K.; Gu, X.; Deng, F.; Wang, F.-Y. ParallelEye pipeline: An effective method to synthesize images for improving the visual intelligence of intelligent vehicles. *IEEE Trans. Syst. Man. Cybern. Syst.* **2023**, *53*, 5545–5556. [CrossRef]
20. Moroto, Y.; Maeda, K.; Ogawa, T.; Haseyama, M. Snow- or ice-covered road detection in winter road surface conditions using deep neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 2935–2950. [CrossRef]

21. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv* **2015**, arXiv:1412.7062. Available online: https://arxiv.org/abs/1412.7062 (accessed on 26 October 2026).
22. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
23. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. Available online: https://arxiv.org/abs/1706.05587 (accessed on 26 October 2026).
24. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
25. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19. [CrossRef]
26. Kim, T.; Lee, H.; Kim, D. UACANet: Uncertainty augmented context attention for polyp segmentation. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, 20–24 October 2021; pp. 2167–2175. [CrossRef]
27. Huo, X.; Sun, G.; Tian, S.; Wang, Y.; Yu, L.; Long, J.; Zhang, W.; Li, A. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification. *Biomed. Signal Process. Control* **2024**, *87*, 105534. [CrossRef]
28. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
29. Scikit-Learn Developers. average_precision_score Documentation. Reference Definition of Non-Interpolated Average Precision. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html (accessed on 19 October 2026).
30. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1055–1059.
31. Yu, G.; Dong, J.; Wang, Y.; Zhou, X. RUC-Net: A Residual-Unet-Based Convolutional Neural Network for Pixel-Level Pavement Crack Segmentation. *Sensors* **2023**, *23*, 53. [CrossRef]
32. Wang, H.; Cao, P.; Yang, J.; Zaiane, O. MCA-UNet: Multi-scale cross co-attentional U-Net for automatic medical image segmentation. *Health Inf. Sci. Syst.* **2023**, *11*, 10. [CrossRef]
33. Qi, Y.; Wan, F.; Lei, G.; Liu, W.; Xu, L.; Ye, Z.; Zhou, W. GMDNet: An irregular pavement crack segmentation method based on multi-scale convolutional attention aggregation. *Electronics* **2023**, *12*, 3348. [CrossRef]
34. Li, X.; Li, J. MFCA-Net: A deep learning method for semantic segmentation of remote sensing images. *Sci. Rep.* **2024**, *14*, 5745. [CrossRef]
35. Sun, G.; Pan, Y.; Kong, W.; Xu, Z.; Ma, J.; Racharak, T.; Nguyen, L.-M.; Xin, J. DA-TransUNet: Integrating spatial and channel dual attention with transformer U-net for medical image segmentation. *Front. Bioeng. Biotechnol.* **2024**, *12*, 1398237. [CrossRef]
36. Huang, Y.; Shi, P.; He, H.; He, H.; Zhao, B. Senet: Spatial information enhancement for semantic segmentation neural networks. *Vis. Comput.* **2024**, *40*, 3427–3440. [CrossRef]
37. Cai, Y.; Zhou, W.; Zhang, L.; Yu, L.; Luo, T. DHFNet: Dual-decoding hierarchical fusion network for RGB-thermal semantic segmentation. *Vis. Comput.* **2024**, *40*, 169–179. [CrossRef]
38. Chen, L.; Qu, Z.; Zhang, Y.; Liu, J.; Wang, R.; Zhang, D. Edge-enhanced GCIFFNet: A multiclass semantic segmentation network based on edge enhancement and multiscale attention mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4450–4465. [CrossRef]
39. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560. Available online: https://arxiv.org/abs/1808.04560 (accessed on 26 October 2026).
40. Xu, K.; Yang, X.; Yin, B.; Lau, R.W.H. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020. [CrossRef]
41. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020. [CrossRef]
42. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640. [CrossRef]
43. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [CrossRef]

44. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
45. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.