Master's Thesis Presentation for the 143th Dissertation Evaluation

A Study on Effective Encoder-Decoder Neural Networks for Multiple Multi-Scale Disordered Masks in Vehicle Damage Segmentation

불규칙한 여러 개의 다중 스케일 마스크를 위한 효과적인 자동차 파손 분할 연구

Minkyung Sung

clary9752@gmail.com

25th Apr, 2025

중앙대학교
CHUNG-ANG UNIVERSITY

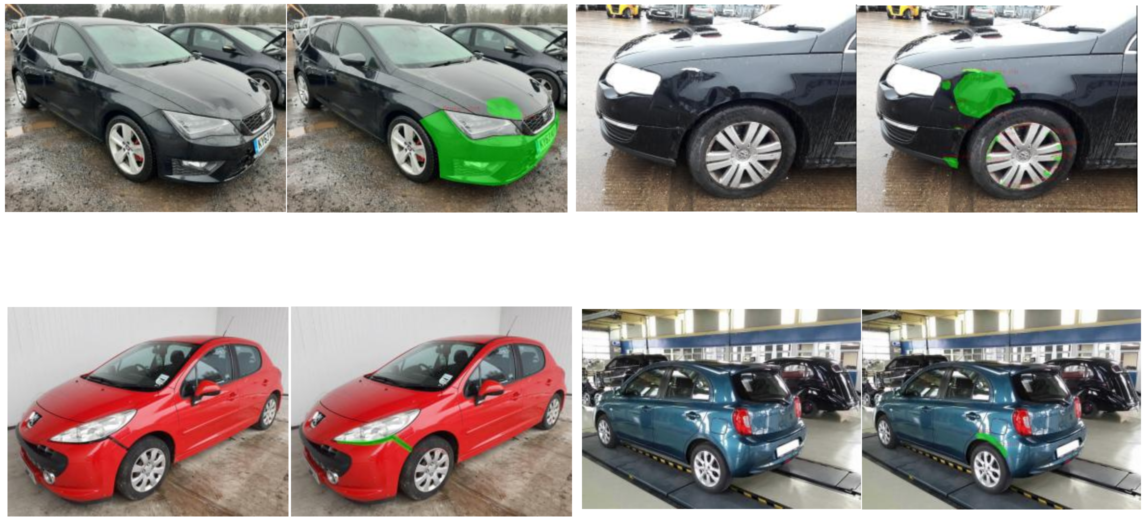
Contents

AutoML

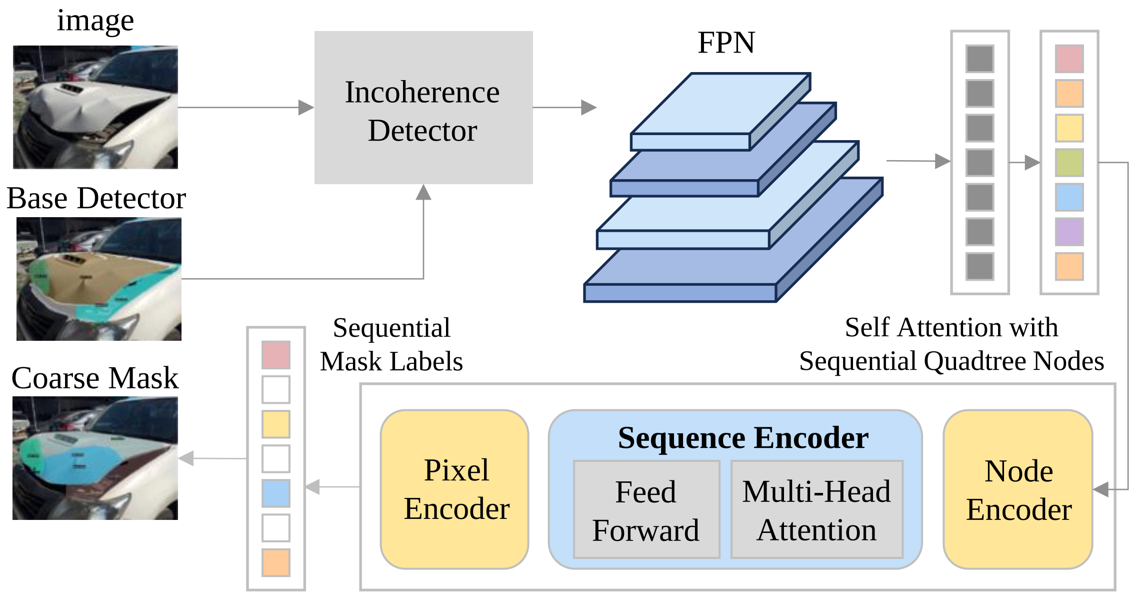
Introduction

Vehicle Damage Segmentation

- As car ownership and the incidence of car accidents increase, the efficiency and accuracy of insurance claims is becoming increasingly important, but the traditional manual process is time-consuming, costly, and prone to errors and fraud [1].
- Vehicle Damage Segmentation (VDS) is shifting toward automatic vehicle damage detection and evaluation using object detection and semantic segmentation, with recent research improving performance using transformer-based models and data augmentation [2].



[Figure 1] Small target area for VDS



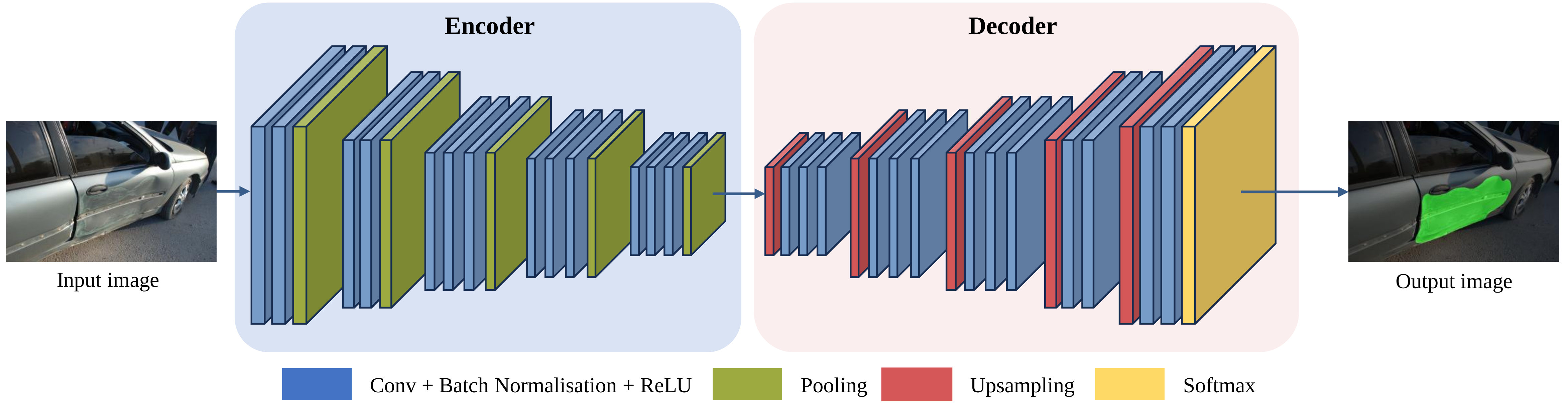
[Figure 2] One of the latest research trends in VDS [2]

[1] A. Mateen, et al., "Challenges, issues, and recommendations for Blockchain-and cloud-based automotive insurance systems," Appl. Sci., vol. 13, no. 6, p. 3561, 2023.
[2] T. Panboonyuen, et al., "Mars: Mask attention refinement with sequential quadtree nodes for car damage instance segmentation," in Proc. Int. Conf. Image Anal. Process. (ICIAP), Cham: Springer Nature Switzerland, 2023.

Introduction

Objective and Standard Procedure

- VDS research aims to analyze images of vehicle damage to assess the location and extent of damage with pixel-perfect precision and improve performance to better identify damage and minimize misjudgment [3].
- VDS detects the vehicles in images, identifies damage areas, and then segments damage pixel-by-pixel to analyze damage type, severity, and location for insurance, vehicle maintenance, and vehicle condition management [4].



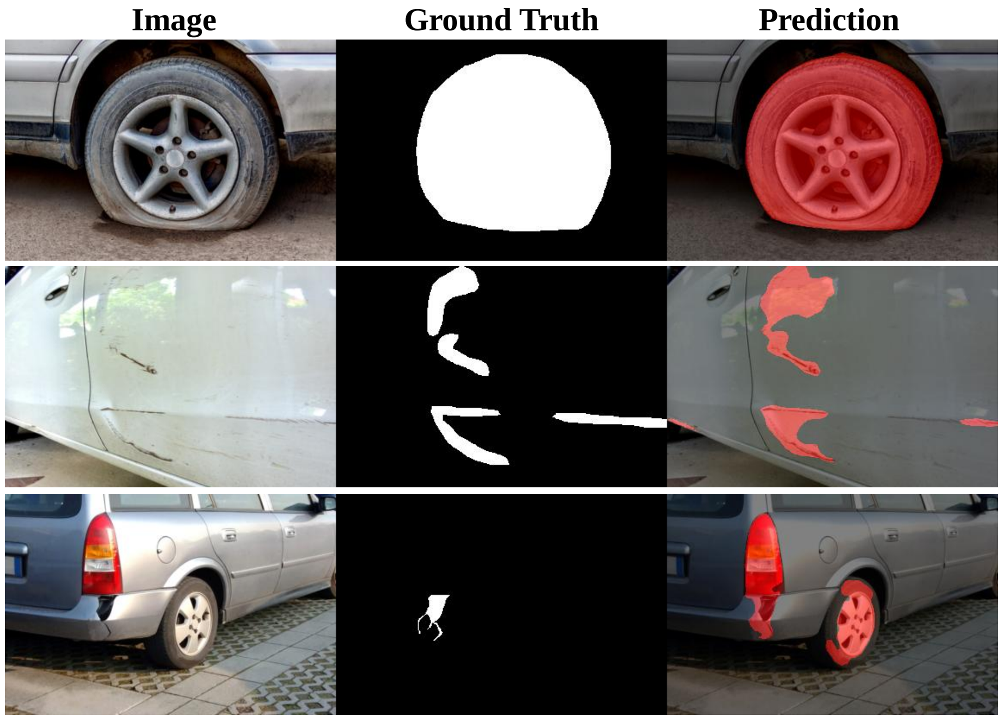
[Figure 3] VDS operation process

[3] Q. Zhang, X. Chang, and S. B. Bian, "Vehicle-damage-detection segmentation algorithm based on improved mask RCNN," IEEE Access, vol. 8, pp. 6997–7004, 2020.
[4] J. Qaddour and S. A. Siddiqa, "Automatic damaged vehicle estimator using enhanced deep learning algorithm," Intell. Syst. Appl., vol. 18, p. 200192, 2023.

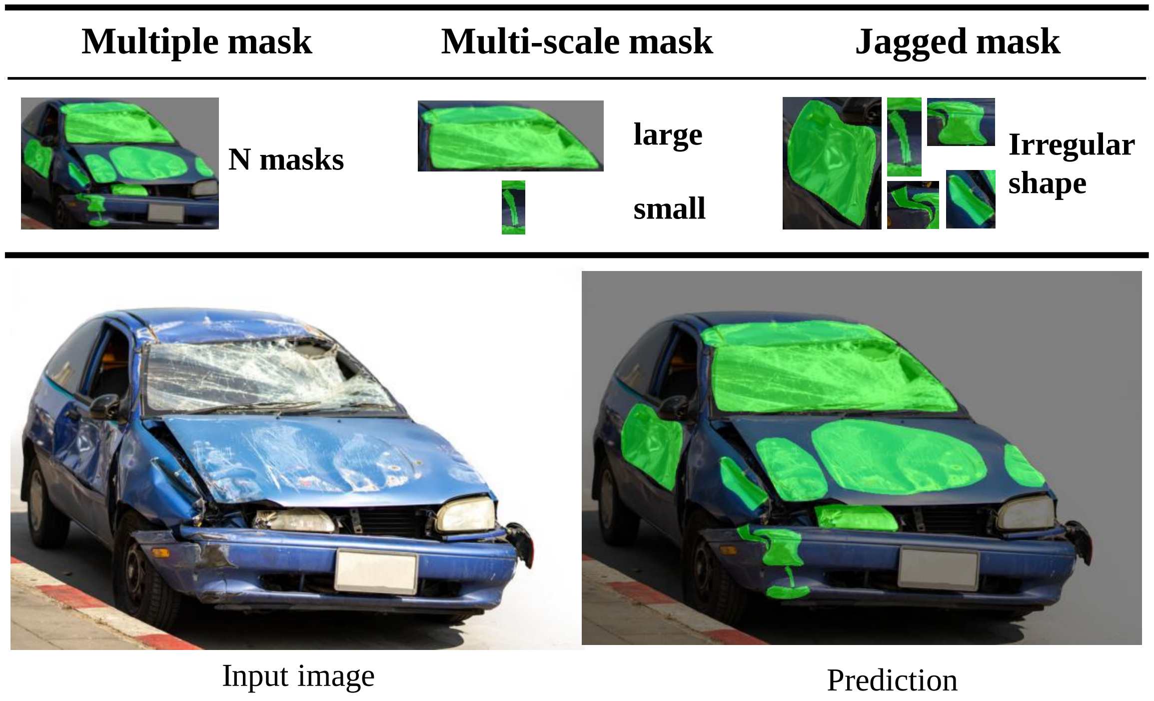
Introduction

Problem and Strategy

- Existing segmentation models have been proposed to solve a single problem, and many previously studied general models have been used because no suitable model has been proposed for multiple multi-scale disordered masks.
- A proposed model with a pre-trained model to eliminate noise, a multi-channel structure to enhance contextual information, ASPP to improve the separation performance of small objects, and a multi-layer convolutional decoder to separate multiple masks.



[Figure 4] Qaddour et al., (2023) output



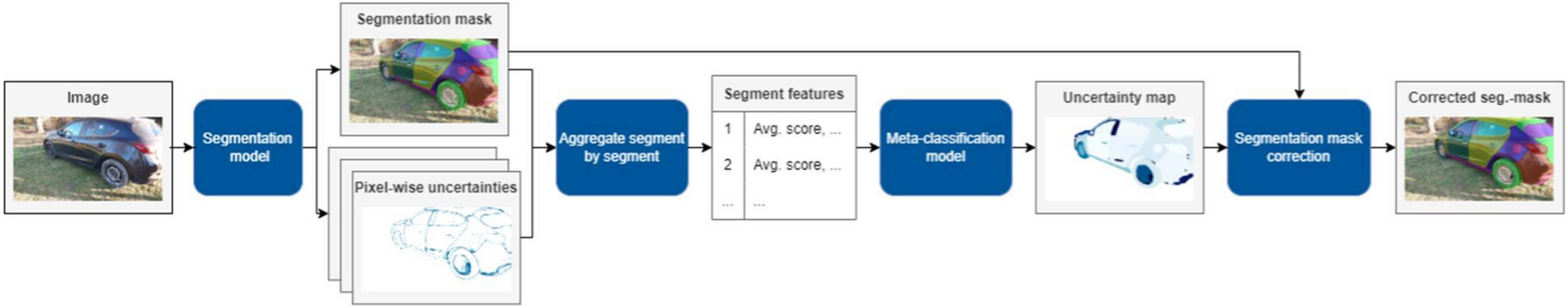
[Figure 5] Replace Attention U-Net with backbone in DeepLabV3+

[4] Qaddour, Jihad, and Syeda Ayesha Siddiqa. "Automatic damaged vehicle estimator using enhanced deep learning algorithm." *Intelligent Systems with Applications* 18 (2023): 200192.

Related Work

Uncertainty estimates for semantic segmentation (2024) [5]

- In the context of vehicle image analysis, a segmentation mask is generated for each component to quantify pixel-level uncertainty, with the calculated feature vector for each segment then being input into a meta-classification model to enhance the accuracy.
- Despite assertions of enhanced performance through the elimination of high-uncertainty regions, the system demonstrates deficiencies in tasks involving reflective backgrounds on vehicle surfaces or windows, indistinct tire boundaries, and minor scratches or dents.



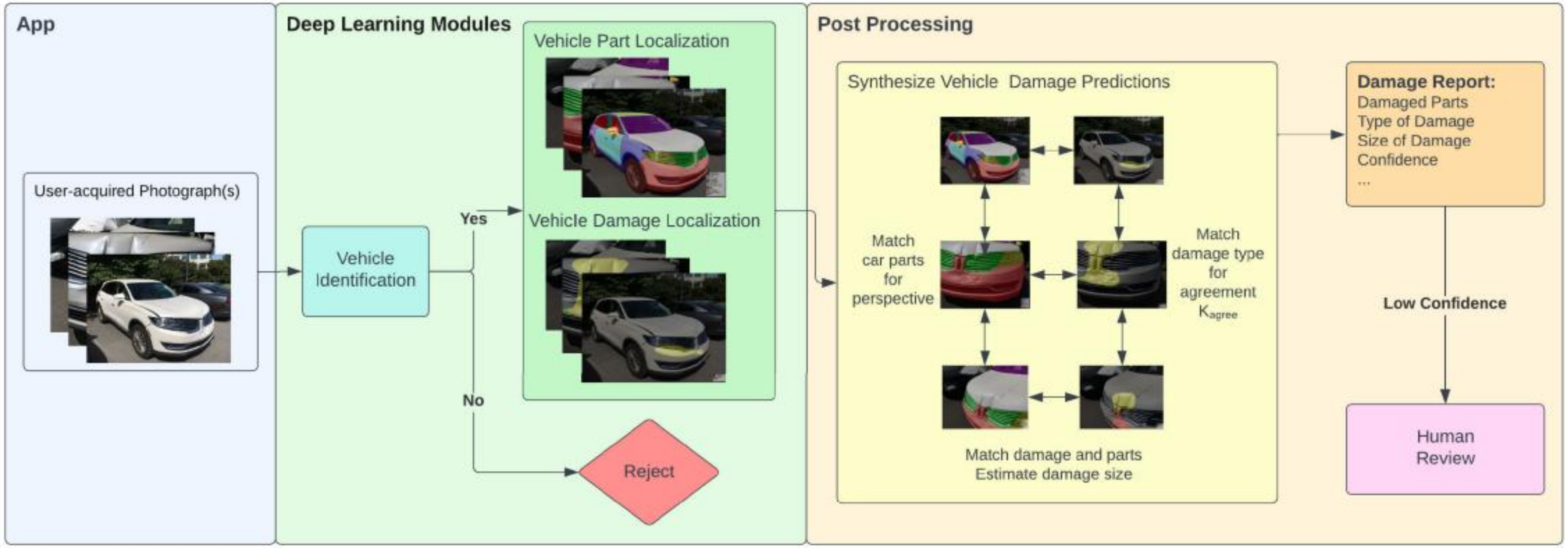
[Figure 6] Model architecture using uncertainty regions

[5] J. Küchler, et al., "Uncertainty estimates for semantic segmentation: Providing enhanced reliability for automated motor claims handling," Mach. Vis. Appl., vol. 35, no. 4, p. 66, 2024.

Related Work

A System for Automated Vehicle Damage Localization and Severity Estimation (2023) [6]

- The proposed an end-to-end system that automatically identifies vehicle damage using user-photos as input, the model is trained using pre-trained weights and complex scaling techniques with large datasets, with the objective of enhance model performance.
- The squeeze-and-excitation block in EfficientNet enhances the big damaged area through weighting techniques, but this approach is limited by recognizing small damaged areas, which can be hindered by distractions like window reflections, lighting changes, etc.



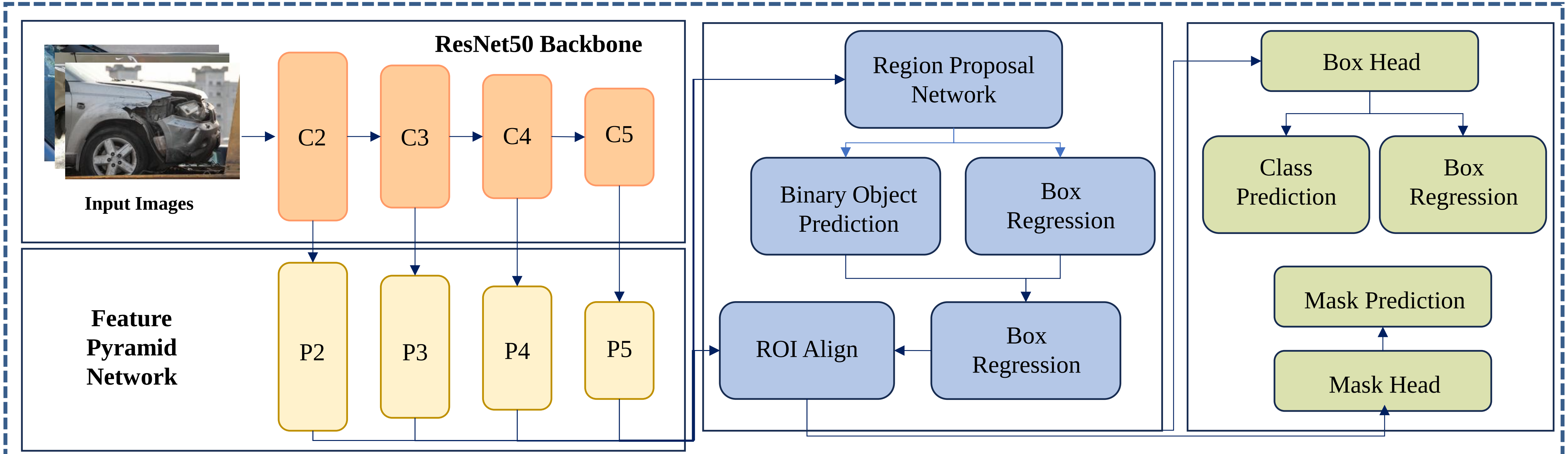
[Figure 7] Model architecture using DeepLabV3+ with EfficientNet

[6] Y. Ma, et al., "A system for automated vehicle damage localization and severity estimation using deep learning," IEEE Trans. Intell. Transp. Syst., 2023.

Related Work

Vehicle Damage Severity Estimation for Insurance Operations (2023) [7]

- The study proposes a smartphone-based image a process to automatically detect vehicle damage, assess its severity, and calculate repair costs, aiming to provide faster and more reliable vehicle damage analysis compared to manual assessment.
- Deformable convolution has proven effective at detecting damage patterns of various sizes and shapes, but its performance has been limited by its inability to detect damage patterns that include distorted patterns such as shapes or light reflections.



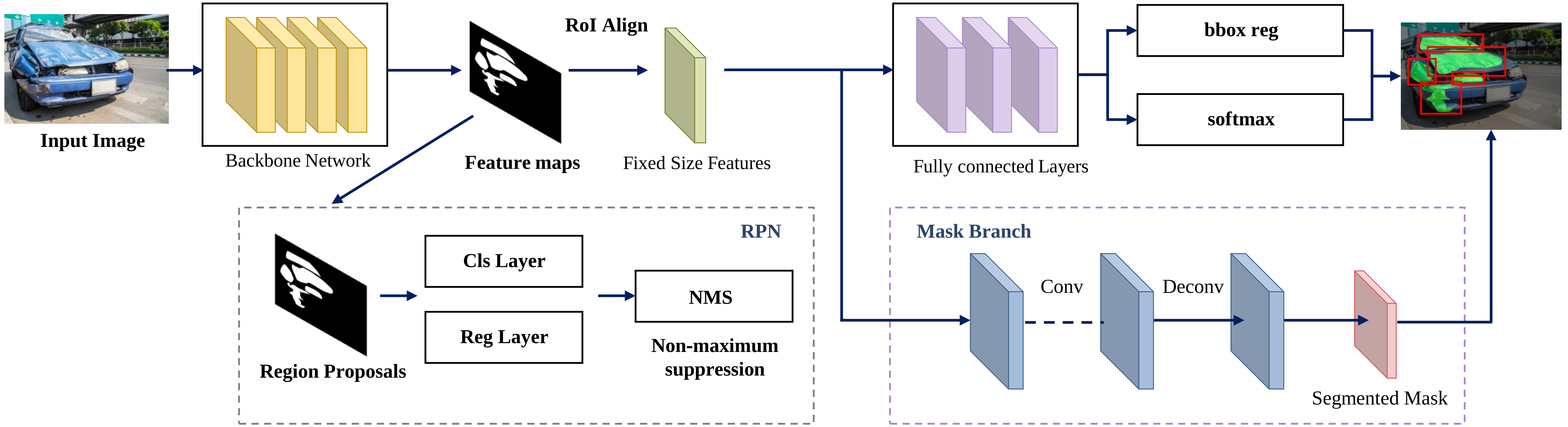
[Figure 8] Model architecture using MaskRDCN

[7] D. Mallios, et al., "Vehicle damage severity estimation for insurance operations using in-the-wild mobile images," IEEE Access, 2023.

Related Work

Automatic damaged vehicle estimator (2023) [4]

- The model is a two-step approach that integrates object detection and instance segmentation for the proposed bounding box by predicting the bounding box to detect objects and then generating a pixel-by-pixel mask within the same region.
- Despite its efficacy in detecting vehicle damage of various sizes through the utilization of a Feature Pyramid Network, the system still has limitations include the inability to recognize certain types of damage or to detect an area larger than the damage site itself.



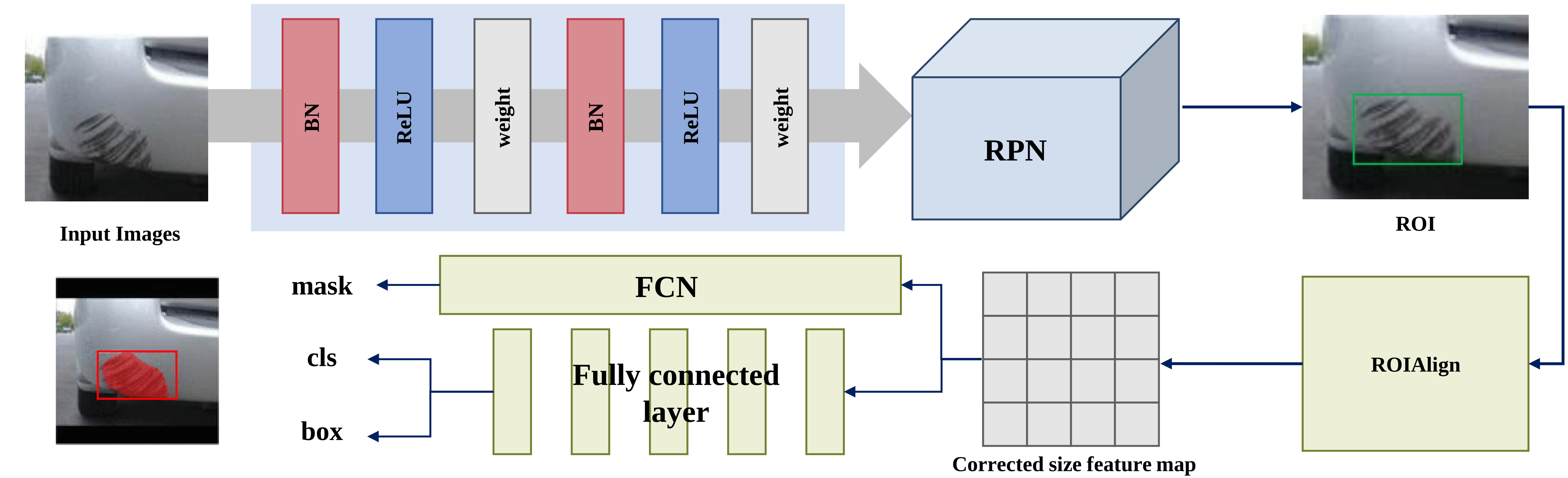
[Figure 9] Model architecture using MaskRCNN

[4] J. Qaddour and S. A. Siddiqa, "Automatic damaged vehicle estimator using enhanced deep learning algorithm," Intell. Syst. Appl., vol. 18, p. 200192, 2023.

Related Work

Vehicle-Damage-Detection Segmentation Algorithm (2020) [3]

- The process of generating and predicting a pixel-by-pixel mask from a proposed bounding box, and applying batch normalization before the active function, ensures high accuracy and efficiency under varying lighting, angles, and color variations.
- In a typical Mask R-CNN model, adopting ResNetV2 as the underlying framework enhances the efficacy of detecting minor damage, however it remains constrained in accurately identifying damage of diverse dimensions.



[Figure 10] Model architecture using MaskRCNN with ResNetV2

[3] Q. Zhang, X. Chang, and S. B. Bian, "Vehicle-damage-detection segmentation algorithm based on improved mask RCNN," IEEE Access, vol. 8, pp. 6997–7004, 2020.

Related Work

Common Drawback and Conventional Methods

- Vehicles that have been subject to damage in automobile accidents, such as scratches, frequently exhibit mask information that is irregular in shape and size, often presenting multiple masks of varying dimensions within a single imagery.
- This characteristic results in the loss of information from small damage area during the feature extraction phase of the encoder, and decoders with a high degree of simplicity are incapable of recuperating adequate data, thereby inducing performance degradation.

[Table 1] Counterpart Summary

Method	Venue	Summary	Drawback
Küchler et al., (2024) [5]	Machine Vision and Applications	The objective of this initiative is to enhance the reliability of the insurance claims system by leveraging uncertainty estimation techniques to optimize vehicle damage assessments.	Minor damage detection limited
Ma et al., (2023) [6]	Transactions on Intelligent Transportation Systems	The objective of this technology is to enhance the efficiency of the insurance claims process by automating the detection and severity assessment of automobile damage.	Reduced performance due to noise (such as window reflections)
Mallios et al., (2023) [7]	IEEE Access	A proposal has been made for a damage severity analysis system that considers the variability of vehicle damage images taken in a mobile environment.	Limited detection of specific damage types
Qaddour et al., (2023) [4]	Intelligent Systems with Applications	The objective is to enhance the process of auto insurance ratings by leveraging deep learning-based vehicle damage detection to achieve greater precision than that of conventional methods.	Missing some damage recognition
Zhang et al., (2020) [3]	IEEE Access	The objective is to develop an algorithm that performs more accurate damage segmentation by compensating for the limitations of the existing Mask R-CNN based vehicle damage detection.	Limited detection of damage of different sizes

Proposed Method

Notation and Definition

- The input image (X) consists of a height (H), width (W), and number of channels (C), and each pixel that exists as an integer between 0 and 255 is normalized to a real number between 0 and 1 by dividing by 255 to use as model input.
- The model f_θ is a function that generates a segmentation mask $\hat{S}$ by predicting the damage region pixel by pixel in the input image X, with the goal of making the predicted mask $\hat{S}$ as close as possible to the ground truth mask S.

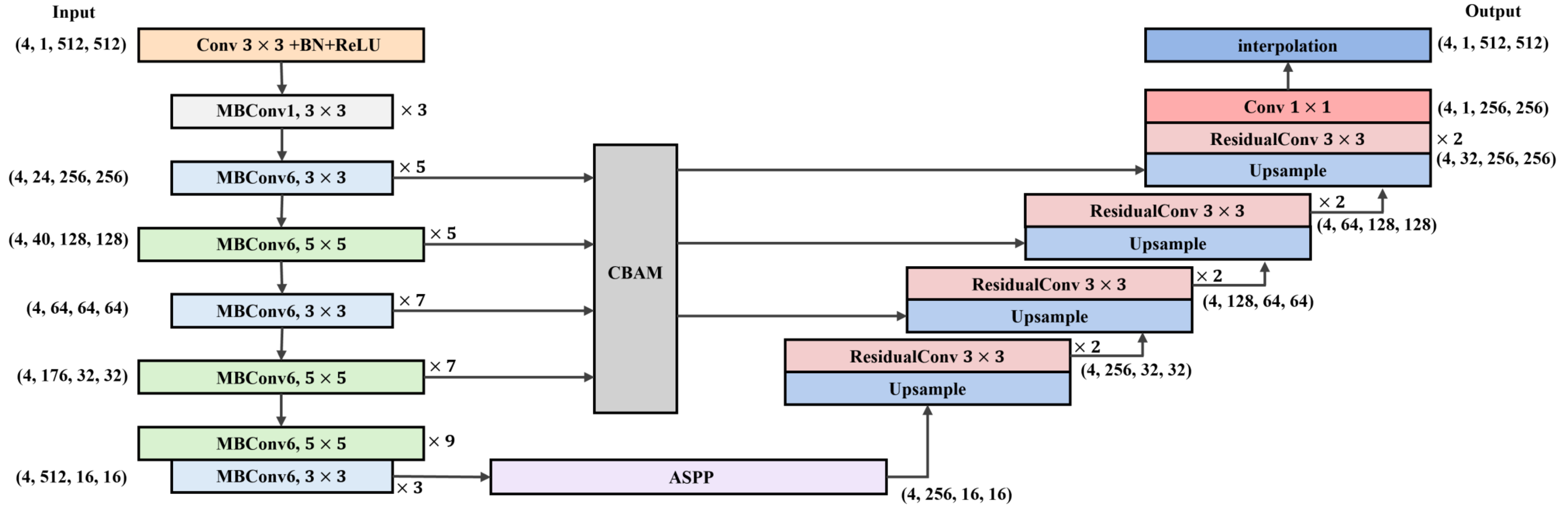
[Table 2] Notation Symbol Definition

Notation	Name	Definition
$X = \{X_i\}_{i=1}^{m}, \quad X_i \in \mathbb{R}^{H\times W\times C}$	Input Image	Vehicle images (RGB) used as input for the segmentation model
$S = \{S_i\}_{i=1}^{m}, \quad S \in \{0, 1, \ldots, C\}^{H\times W}$	Ground Truth Masks	Pixel-wise labeled masks for vehicle damage, where each pixel belongs to a specific damage category
$f_{MBConv(e)}(\boldsymbol{X}) = P\left(SE\left(D_{k,s}\left(E_e(\boldsymbol{X})\right)\right)\right) + 1_{(s=1)\wedge(C_{in}=C_{out})}\boldsymbol{X}$	MBConv	MBConv1 automatically becomes $E_e = Id$ when set to e=1, so the expression works as above. MBConv6 only needs to substitute $e = 6$.
$f_{ASPP}(\boldsymbol{X}) = C_{1\times1}(\left[DIL_{r_1}(\boldsymbol{X}), \ldots, DIL_{r_m}(\boldsymbol{X}), G(\boldsymbol{X})\right])$	ASPP	$DIL_{r_i}(\boldsymbol{X})$ = conv $k \times k$ with dilatioin $r_i \longrightarrow BN \longrightarrow \phi$, $G(\boldsymbol{X}) = \phi\left(C_{1\times1}\left(UP\left(GAP(\boldsymbol{X})\right)\right)\right)$
$f_{CBAM}(\boldsymbol{X}) = \left(\boldsymbol{X}\odot\mathcal{C}(\boldsymbol{X})\right)\odot\mathcal{S}\left(\boldsymbol{X}\odot\mathcal{C}(\boldsymbol{X})\right) + \boldsymbol{X}$	CBAM	Channel Attention $\mathcal{C}(\boldsymbol{X})$, Spatial Attention $\mathcal{S}(\boldsymbol{X})$,
$f_{RCB}(\boldsymbol{X}) = \phi\left(SE\left(B_2 \circ C_{3\times3}^{(2)} \circ \phi \circ B_1 \circ C_{3\times3}^{(1)}(\boldsymbol{X})\right) + \left[1_{C_{in}=C_{out}}\boldsymbol{I} + 1_{C_{in}\neq C_{out}}B_r \circ C_{1\times1}^{r}\right](\boldsymbol{X})\right)$	Residual Block	Conv-BN-ReLU-Conv-BN-SE in a single line, followed by the conditional residual, adding the two values together, and then relu. For the conditional residual, it depends on the channel and resolution. $\phi(\cdot)$ is ReLU

Proposed Method

Procedural Overview

- Pre-trained models can capture sophisticated features, identify subtle changes, and apply ASPP to identify patterns at various scales, resolutions, and feature spaces, reflecting complex contextual information.
- After ASPP, a decoder with multiple convolutional blocks processes features to restore resolution, preserving fine structure and boundary information for enhanced recognition and skip-connections are used to accurately restore the structure of small objects.

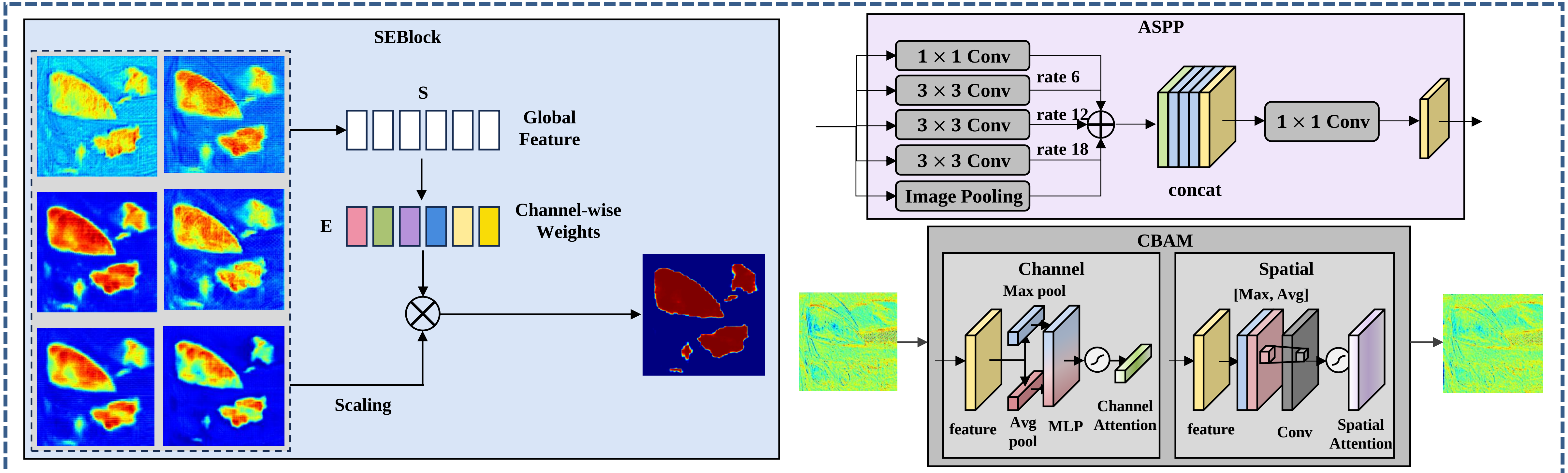


[Figure 11] Proposed Model architecture

Proposed Method

Details

- To effectively recognize multiple objects in a single image, SEBlock analyzes information from multiple channels to learn their importance and reflects the generated weights in a feature map to effectively recognize objects.
- The ASPP extracts multi-scale information from a single feature map to recognize objects of different sizes, the CBAM uses channel-space attentions to reflect visual saliency and location information to effectively segment irregular masks.



[Figure 11] Proposed Model architecture

Experimental Settings

Dataset

- The dataset under consideration employed a public dataset that developed its own data collection and labeling process for images of damaged vehicles, utilizing a training and evaluation framework for assessing vehicle daamage.
- The data collected is precisely labeled according to damage type, location, and more, and consists of images taken from different angles and lighting conditions to help improve the model's generalization performance.

[Table 3] Datasets Information

Dataset	Class	Image	Remark
CarDD [8]	7	4,000	A comprehensive dataset designed for developing and testing machine learning models to detect and assess vehicle damage.

Cross Validation

- Number of Iterations: 10
- Data Split Ratio (Train: Validation: Test) = 0.6 : 0.2: 0.2

Experimental Settings

Hyperparameter Settings

- All experiments were conducted using 1 or 2 Geforce RTX 3090 GPUs

[Table 4] Hyperparameter Settings

Method	Batch size	Iterations	Loss function	Learning Rate
Küchler et al., (2024) [5]			Cross entropy	
Ma et al., (2023) [6]			Dice BCE Loss	
Mallios et al., (2023) [7]	8	100		1e-4
Qaddour et al., (2023) [4]			BCE Loss	
Zhang et al., (2020) [3]				

Experimental Settings

Evaluation Measures

[Table 5] Evaluation Metrics

Evaluation Metrics	notation	description
Intersection over Union	$IoU = \frac{\lvert S \cap \hat{S} \rvert}{\lvert S \cup \hat{S} \rvert}$	• $\lvert S \cap \hat{S} \rvert$ is the number of intersection pixels between the actual and predicted damaged areas, and $\lvert S \cup \hat{S} \rvert$ is the number of union pixels. • Consequently, IoU is a ratio that quantifies the extent of overlap of the damaged areas.
F1-Score	$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$	• The F1 score is the harmonic mean of precision and recall, a metric that balances the two values.
Precision	$Precision = \frac{TP}{TP + FP}$	• TP means True Positive, which is the number of pixels that correctly detected the actual damaged area. • FP means False Positive, which is the number of pixels that are falsely predicted to be damaged when they are not.
Recall	$Recall = \frac{TP}{TP + FN}$	• FN means False Negative, which is the number of pixels that have damage but are not detected. • Thus, Recall is the percentage of actual corruption areas that the model correctly detects.

Experimental Results

Performance Comparison

- The following comparison is presented of the results of the counterpart algorithm experiments on the CarDD dataset, with all experiments undergoing 10 iterations.
- The metrics employed in this study included IoU, precision, recall, and F1-score and the experiments were conducted under the same experimental conditions and in the same environment.

[Table 6] CarDD dataset Performance

Model	IoU	F1-Score	Precision	Recall
Proposed	**0.6772 ± 0.0106**	**0.7732 ± 0.0100**	**0.8264 ± 0.0122**	0.7770 ± 0.0158
Küchler et al., (2024) [5]	0.3462 ± 0.0077	0.5151 ± 0.0076	0.6003 ± 0.0191	0.4651 ± 0.0090
Ma et al., (2023) [6]	0.6627 ± 0.0114	0.7621 ± 0.0096	0.8168 ± 0.0114	0.7648 ± 0.0155
Mallios et al., (2023) [7]	0.2876 ± 0.0086	0.4595 ± 0.0169	0.3453 ± 0.0173	0.9563 ± 0.0110
Qaddour et al., (2023) [4]	0.2839 ± 0.0080	0.4440 ± 0.0116	0.3312 ± 0.0110	0.9690 ± 0.0025
Zhang et al., (2020) [3]	0.2837 ± 0.0081	0.4439 ± 0.0121	0.3306 ± 0.0113	**0.9691 ± 0.0028**

Experimental Results

Ablation study

- In order to validate the performance stability of the proposed model, an experimental analysis was conducted to assess the performance change when each of the three core modules was systematically removed one by one.
- The experimental findings indicated that the three modules exhibited a substantial influence on the performance of the model, thus substantiating the significance of each component.

[Table 7] Ablation study Performance comparison

Model	IoU	F1-Score	Precision	Recall
Proposed	**0.6772 ± 0.0106**	**0.7732 ± 0.0100**	0.8264 ± 0.0122	**0.7770 ± 0.0158**
Proposed (- CBAM)	0.6741 ± 0.0088	0.7701 ± 0.0079	0.8255 ± 0.0090	0.7725 ± 0.0153
Proposed (- ASPP)	0.6755 ± 0.0150	0.7720 ± 0.0147	**0.8356 ± 0.0076**	0.7680 ± 0.0227

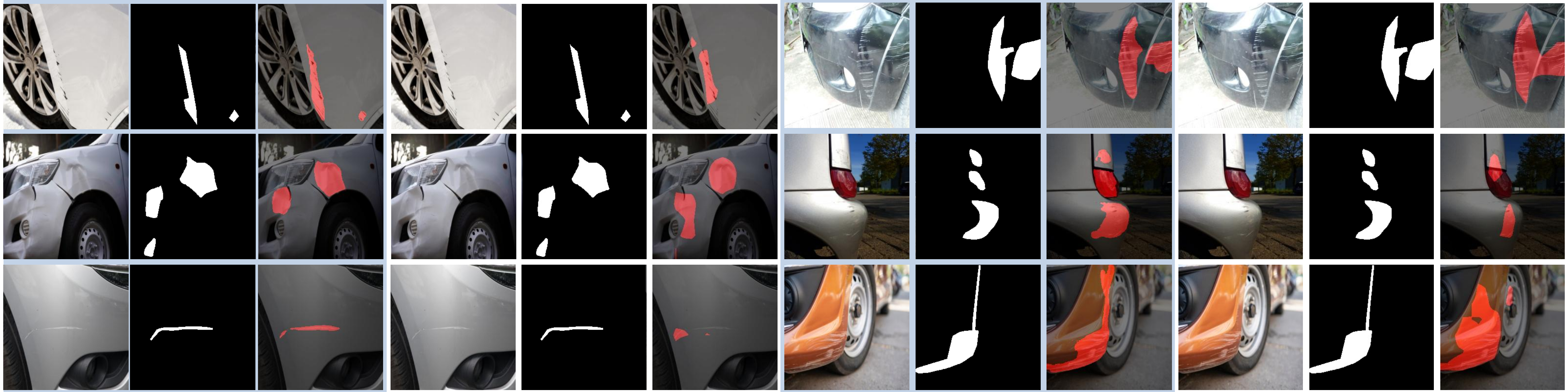
Experimental Results

Qualitative Analysis

- A visual comparison of the prediction results of the proposed model and the existing model was conducted to ascertain the superiority of the proposed model in terms of segmentation and boundary representation of the damaged area.
- The proposed model segmentation performance is quantitatively verified by an ablation study and a visual comparison analysis, and the results show that the proposed model exhibits superior accuracy and boundary discrimination compared to existing techniques.

[Table 8] Compare predict images   Proposed method   J. Küchler, et al (2024) [5]

원본	GT	예측	원본	GT	예측	원본	GT	예측	원본	GT	예측



Conclusion and Contribution

Conclusion

- The objective of VDS research is to automate the detection and severity assessment of vehicle damage, thereby eliminating the drawbacks associated with inconsistent manual processing.
- The proposed model is designed to learn different resolutions with a multi-channel structure, and effectively segment multiple objects and irregular masks through residual blocks and skip connections with CBAM.
- The performance of proposed model is demonstrated in the Comparison to Conventional Methods section, where it is compared to existing VDS techniques, and the statistical tests that support it show a significant improvement.

Achievement

- Journal
  - Real-Time On-Device Continual Learning Based on Combined Nearest Class Mean and Replay Method for Smartphone Gesture Recognition, **Sensors**, 2025
- Conferences
  - Recent Research Trends In Open Vocabulary Segmentation Based-on Vision-Language Pre-trained Model , ***IEIE***, 2024
  - Automatic Playlist Generation: A Comprehensive Crux of a Methodology for Technology Trend, ***ICCE***, 2025
- Others
  - ***2023 Singapore K-Tech Festival - AI/ML Innovation Research Center Booth***, Singapore, Singapore, 2023
  - Automatically generate playlists and cover images that reflect your music taste, *AI Technology-based Employment and Start-up Workshop*, 2024, 2^{nd} prize

Thank you

Appendix

Performance Comparison

- Performance comparison table when experimenting with the general model.

[Table 6] CarDD Dataset Result

Model	IoU	F1-Score	Precision	Recall
Proposed	**0.6772 ± 0.0106**	**0.7732 ± 0.0100**	**0.8264 ± 0.0122**	**0.7770 ± 0.0158**
SegNet	0.5114 ± 0.0133	0.6195 ± 0.0128	0.7411 ± 0.0155	0.6099 ± 0.0185
U-Net	0.5397 ± 0.0105	0.6506 ± 0.0110	0.7456 ± 0.0149	0.6492 ± 0.0205
DeepLabV3+	0.6608 ± 0.0176	0.7605 ± 0.0160	0.8188 ± 0.0069	0.7605 ± 0.0160
DS-TransUnet	0.4172 ± 0.0191	0.5319 ± 0.0200	0.6659 ± 0.0109	0.5334 ± 0.0273
CaraNet	0.4888 ± 0.0212	0.5935 ± 0.0234	0.7183 ± 0.0069	0.6002 ± 0.0275
OTSNet	0.5696 ± 0.0113	0.6744 ± 0.0109	0.7792 ± 0.0190	0.6694 ± 0.0193
DUCKNet	0.5052 ± 0.0122	0.6188 ± 0.0124	0.7249 ± 0.0201	0.6465 ± 0.0247
DCSAU-Net	0.4935 ± 0.0095	0.6035 ± 0.0101	0.7258 ± 0.0111	0.5998 ± 0.0139
MEGANet	0.5141 ± 0.0204	0.6231 ± 0.0201	0.7258 ± 0.0105	0.6253 ± 0.0246

Reference

[1] A. Mateen, et al., "Challenges, issues, and recommendations for Blockchain-and cloud-based automotive insurance systems," Appl. Sci., vol. 13, no. 6, p. 3561, 2023.

[2] J. Parslov, E. Riise, and D. P. Papadopoulos, "CrashCar101: Procedural generation for damage assessment," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2024.

[3] Q. Zhang, X. Chang, and S. B. Bian, "Vehicle-damage-detection segmentation algorithm based on improved mask RCNN," IEEE Access, vol. 8, pp. 6997–7004, 2020.

[4] J. Qaddour and S. A. Siddiqa, "Automatic damaged vehicle estimator using enhanced deep learning algorithm," Intell. Syst. Appl., vol. 18, p. 200192, 2023.

[5] J. Küchler, et al., "Uncertainty estimates for semantic segmentation: Providing enhanced reliability for automated motor claims handling," Mach. Vis. Appl., vol. 35, no. 4, p. 66, 2024.

[6] Y. Ma, et al., "A system for automated vehicle damage localization and severity estimation using deep learning," IEEE Trans. Intell. Transp. Syst., 2023.

[7] D. Mallios, et al., "Vehicle damage severity estimation for insurance operations using in-the-wild mobile images," IEEE Access, 2023.

[8] X. Wang, W. Li, and Z. Wu, "Cardd: A new dataset for vision-based car damage detection," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 7, pp. 7202–7214, 2023.

[9] V.-D. Hoang, et al., "Powering AI-driven car damage identification based on VeHIDE dataset," J. Inf. Telecommun., pp. 1–19, 2024.

[10] Humans in the Loop, Car Parts and Car Damages [Data set], Kaggle, 2023. [Online]. Available: https://doi.org/10.34740/KAGGLE/DSV/5884733

[11] K. H. Park, H. Chung, and S. Kwon, "Towards an efficient ML system: Unveiling a trade-off between task accuracy and engineering efficiency in a large-scale car sharing platform," arXiv preprint arXiv:2210.06585, 2022.