Improving Road Damage Detection Accuracy

Using Deep Learning Image Enhancement Models

Institute for Homeland Security

Sam Houston State University

Vung Pham

Abstract

It is important to accurately detect and classify road damage to ensure road human safety. Current road damage detection systems suffer from detection accuracy loss due to low-quality images. This work endeavors to elevate the efficacy of road damage detection systems by leveraging state-of-the-art deep learning techniques in image enhancement. Specifically, this project experiments with different image enhancement models (including Super Resolution Residual Network or SRResNet and Super Resolution Generative Adversarial Network or SRGAN) and evaluates how they help detection models (You Only Look Once versions 7 and 9 or YOLOv7 and YOLOv9) improve their detection accuracy. The experimental results show that SRResNet helps generate super-resolution images that subsequently improve the detection performances by approximately seven times, and SRGAN, though it produces good-looking images, does not help improve the performance as much as SRResNet due to generating unrealistic patterns.

Keywords: road damage detection, image enhancement, deep-learning vision models.

1 Introduction

Accurately detecting and classifying road damage is pivotal in ensuring road safety and maintenance. Current road damage detection systems often grapple with low-quality images acquired through smartphone-mounted devices on diverse vehicles, including cars, motorcycles, and bicycles. These devices, while convenient, often yield subpar image quality, consequently hindering detection accuracy. This work endeavors to elevate the efficacy of road damage detection systems by leveraging state-of-the-art deep learning techniques in image enhancement. The primary focus lies in exploring diverse deep neural network architectures tailored for image enhancement, specifically targeting low-quality images acquired through vehicular-mounted devices.

The overarching objective is to augment the quality of these images, subsequently amplifying the precision and recall rates of road damage detection. The proposed methodology comprises a multi-stage approach. Initially, a comprehensive analysis of prevalent image enhancement models, encompassing techniques such as super-resolution, denoising, and contrast enhancement, will be conducted. This analysis will guide the selection of the most apt models for enhancing low-quality road damage images. Subsequently, the chosen models will be fine-tuned and optimized to cater specifically to the idiosyncrasies of the low-quality images obtained from diverse vehicular platforms.

The adaptability of these models to varying environmental conditions, such as different lighting and weather scenarios, will be a paramount consideration during the optimization process. Furthermore, the enhanced images will undergo rigorous evaluation through established road damage detection frameworks. This evaluation will gauge the efficacy of the image enhancement models in improving the accuracy of damage detection and classification tasks. Metrics encompassing precision, recall, and overall accuracy will be employed to comprehensively quantify the performance enhancements achieved.

The outcomes of this research encompass a substantial advancement in road damage detection accuracy. By harnessing the capabilities of deep learning-based image enhancement, this project significantly mitigates the challenges posed by low-quality images obtained from vehicular-mounted devices. The resultant improvements in detection accuracy are poised to have far-reaching implications in enhancing road safety and optimizing maintenance efforts.

2 Gap assessment

2.1 Image quality

Despite rapid technological advancements, effectively managing low-quality images remains a significant challenge in the transportation sector. These images often result from unfavorable environmental conditions, aging infrastructure, equipment limitations, or restricted access to high-resolution data. Specifically, manually collecting road crack data is both labor-intensive and unsafe. Consequently, the emerging preferred method involves using recording devices (e.g., dash cams or smartphones) mounted on vehicles (e.g., cars, motorbikes, or drones) to capture digital photos and videos while moving along roads [PPD20]. However, due to the speed of these vehicles, particularly on highways, the resulting videos often suffer from low quality. For instance, Google Street View provides images captured by Google's advanced specialized devices, but even these images are limited to a maximum size of 640×640 [PND22]. These images are frequently low-resolution as they cover large areas, with much of the view including the sky and other objects above the roads. Additionally, low-quality images are inevitable due to varying lighting conditions, shadows, and weather changes.

2.2 Detection models

Analysis, understanding, and utilization of low-quality and latent image data in critical infrastructure sectors remains a prevalent yet often overlooked challenge. Implementing advanced deep learning architectures tailored for such images can drive significant advancements in various fields, including transportation infrastructure monitoring, crime scene investigations, healthcare, and more. To this end, we customized and trained Faster R-CNN models [PPD20], achieving state-of-the-art results in the Global Road Damage Detection Challenge 2020, part of the IEEE Big Data 2020 Big Data Cup Challenge [AMG+20]. More recently, we customized YOLOv7 for training road damage detection models [PND22], securing Rank 3 (Rank 1 among academic participants) in the Crowdsensing-based Road Damage Detection Challenge, a track in the IEEE Big Data 2022 Big Data Cup Challenge. In both competitions, the training images were of low quality. Our work training and optimizing these computer vision models with low-quality images in critical infrastructure sectors demonstrated that while cutting-edge architectures perform well on high-quality image datasets like ImageNet or Microsoft COCO (Common Objects in Context), their performance can be significantly compromised when trained with low-quality and latent images.

2.3 Image enhancement models

Various approaches exist to enhance images to super-resolutions, but they encounter two significant issues. First, these models may unintentionally amplify noise or other errors present in the original low-quality image, resulting in distortions in the enhanced output. Second, the generation of high-frequency artifacts is a common problem, as these models can create unrealistic textures or patterns that confuse subsequent prediction models [NJT+22]. Some methods attempt to address these issues; however, most focus on aesthetic improvements, such as enhancing facial beauty [WLZS21, ZCLL22]. For example, the fine-line patterns in road cracks might be removed similarly to how facial wrinkles are smoothed for aesthetic purposes, leading to reduced detection accuracy.

3 Topic discussions

3.1 Approach

The overview of the research approach for this project is shown in Figure 1. Specifically, during training, original images are resized to create low-resolution images. They (pairs of original images and low-resolution images) are then utilized to train image enhancement models that help reproduce original images from the low-resolution ones. Concomitantly, the original images and their annotations are utilized to train detection models. At the inference time, the low-resolution images are passed through the trained detection models to evaluate the prediction performance (*Score 1*). On the other hand, the low-resolution images are also passed through the trained enhancement models to produce super-resolution images. These images are then also passed through the trained detection models to evaluate prediction performance (*Score 2*). These two scores are then compared to determine if the enhancement models help improve the production performance (i.e., $Score1 < Score2$).

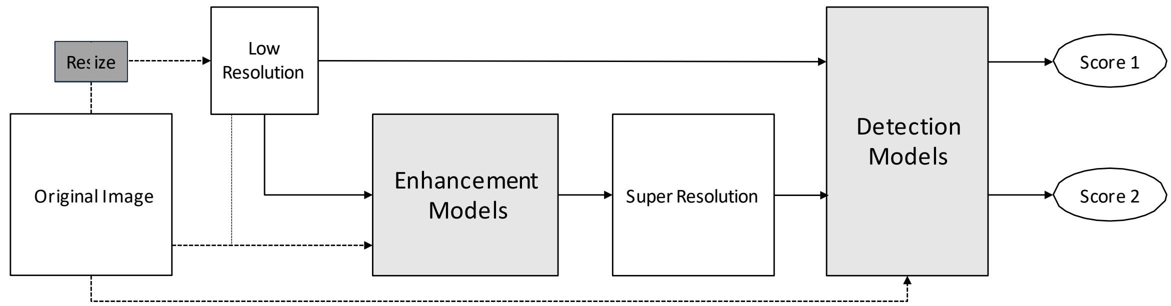

Figure 1: Research approach: Gray, filled boxes are the models, squares are images, and dashed arrows and solid arrows represent data flow during training and inference, respectively.

3.2 Experiments

3.3 Dataset

Recent advancements in smartphones, dashcams, and drones have facilitated automated roadway damage data collection. The smartphone and dashcam methods are simpler to set up, easier to implement, and more efficient than manually scanning for roadway damages. However, they still require human drivers to navigate the roads and record videos. In contrast, drones can be utilized to significantly reduce human workload.

Manually labeling road damage data in the collected images to create training datasets would be extremely labor-intensive. Therefore, we leverage existing datasets as benchmarks for training deep learning models to automatically detect road damage from collected videos. Several datasets are available, but the one provided in the IEEE 2020 Big Data Challenge Cup by Sekilab[1] is particularly practical [PND22].

This benchmark dataset includes one training set (*train*) and two test sets (*test1* and *test2*). The training set contains 21,041 images (2,829 from Czechia, 7,706 from India, and 10,506 from Japan). The two test sets contain 2,631 and 2,664 images, respectively. The training set includes 34,702 ground-truth labels (bounding boxes and damage types), covering four types of damage: longitudinal cracks (D00), transverse cracks (D10), alligator cracks (D20), and potholes (D40).

Our experiments with this benchmark dataset revealed that models trained on data from one country do not perform well in other countries due to differences in road types. Thus, training deep learning models on this dataset does not scale well to roads in the United States. As a result, we collected a separate dataset specifically for the United States. We utilized the Google Street View API[2] to download images from Google Street View.

Google Street View is an excellent resource, offering a vast number of road images worldwide. In this project, we focused on states in the US, though theoretically, we could collect road damage data from any city supported by Google Street View. Furthermore, Google frequently updates its images; most images in this dataset were captured in 2020 or later. The highest downloadable image resolution is 640 by 640, which is suitable for training deep learning models.

In this project, we experimented with 6,005 images collected from Google Street View. As shown in Figure 2, there are 8,303; 4,121; 1,068; and 175 damages for longitudinal (D00), traverse (D10), alligator (D20), and pothole (D40) damages, correspondingly. With this dataset, we were awarded the silver prize for data contributor in the Crowdsensing-based Road Damage Detection Challenge

[1]https://rdd2020.sekilab.global/
[2]https://developers.google.com/maps/documentation/streetview/overview

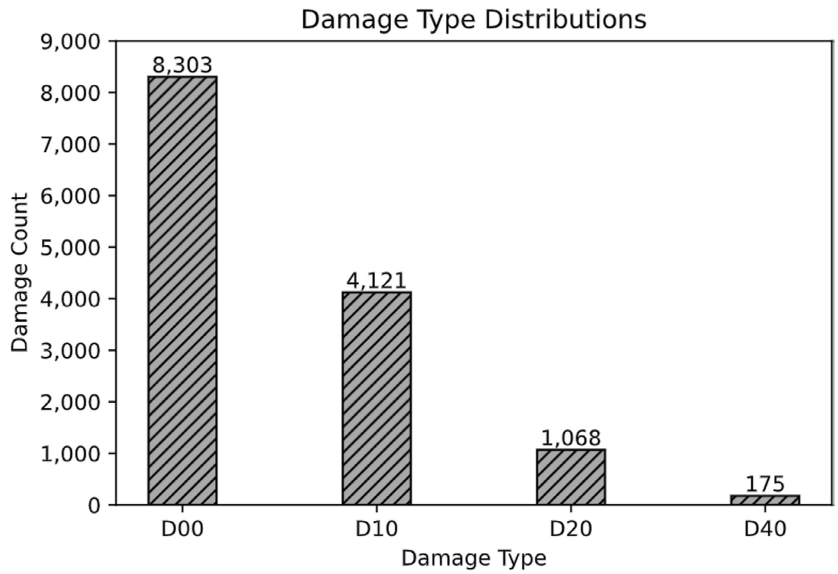

Figure 2: Distribution of damage types in the newly collected set of images from the United States using Google Street View. There are four types of damages considered: longitudinal (D00), traverse (D10), alligator (D20), and pothole (D40) damages, respectively.

(CRDDC2022), IEEE BigData 2022 [PND22]. These images are used as original resolution images. They are resized by one-fourth (1/4) using the bi-cubic algorithm to generate the low-resolution images. Out of the 6,005, 1200 of them were reserved for the CRDDC2022 competition, so we are not using them in this project. The remaining images and their annotations (4,805) of this dataset are divided into two sets (*train* and *val*) for training and validation, with 4,324 and 481 images, respectively.

3.3.1 Detection Models

Computer vision is a crucial component of many contemporary businesses, including automobiles, robotics, and manufacturing, and its market is growing rapidly. Modern computer vision tasks include object detection, instance segmentation, key-point detection, semantic detection, and panoptic segmentation [Pha23]. These modern computer vision tasks are enabled thanks to the development of modern deep learning architectures. The common deep learning methods in this area include Faster Region-Based Convolutional Neural Networks (Faster R-CNN) [RHGS15] and You Only Look Once (YOLO) [BWL20].

Ross Girshick et al. [GDDM14] propose R-CNN, an approach with three main modules. R-CNN takes a long time to train and predict. Therefore, Girshick proposed Fast R-CNN [Gir15] to tackle these issues. This architecture takes the images and proposes candidate regions, then passes them through a popular, pre-trained image classification model (e.g., ResNet [HZRS16], VGG-16 [SZ14]) to extract features from the candidates. Though Fast R-CNN improves the training and predicting time, it requires the region proposal as the input. Therefore, Ren et al. propose Faster R-CNN [RHGS15] to tackle this issue. The PI's recently published book [Pha23] also provides great visualizations and codes for exploring and explaining Faster R-CNN architecture.

Another famous family of neural network architectures for modern computer vision tasks is YOLO, with different versions such as YOLO [RDGF16], YOLOv2 [RF17], YOLOv3 [RF18], YOLOv4 [BWL20], and many more. Different YOLO versions may differ in terms of architectures and techniques used. However, generally, it divides an image into a grid of cells. Features extracted from each cell are used to predict objects with centers of the bounding boxes that fall into the cell. The advantage of this method is that it is faster to train and predict. However, the benefit comes with a slightly lower accuracy than Faster R-CNNs. Still, YOLO is an active development architecture, and recent versions have made improvements to increase accuracy while maintaining their speed advantage.

Toward this end, we customized and trained Faster R-CNN models [PPD20] and achieved state-of-the-art results [AMG+20] in the Global Road Damage Detection Challenge 2020, a track in the

IEEE Big Data 2020 Cup Challenge. Specifically, Faster R-CNN models were optimized and trained on low-quality road images to detect road cracks. Recently, we also trained road damage detection models [PND22] using YOLOv7 [WBL22] and achieved Rank 3 (Rank 1 among academic participants) in the Crowdsensing-based Road Damage Detection Challenge (CRDDC2022), a track in IEEE Big Data 2022 Cup Challenge. Specifically, YOLOv7 has several optimization techniques, such as adding coordinate attention layers [HZF21] or utilizing the label smoothing technique [SVI+16] to overcome poor labeling quality in low-quality images of road damage.

Therefore, this project utilized this optimized YOLOv7 model [PND22] and also experimented with the latest YOLO release (by the time of this writing, it is YOLOv9 [WL24]) as a detection model. Specifically, these models are trained on the original images and their annotations, as described in Section 3.3.

3.3.2 Enhancement Models

Super-resolution approaches are techniques to enhance image quality. These approaches can be broadly classified into two categories: traditional image processing techniques and deep learning techniques. The former category has a long development history with various interpolation techniques. These techniques perform poorly if the input image is of low quality or altered (e.g., blurred or has noise). Therefore, our proposed work focuses on deep-learning alternatives. Specifically, Super-Resolution Convolutional Neural Network or SRCNN [DLHT15] is one of the first approaches that utilize deep learning to enhance images. This technique then became a benchmark for several deep-learning approaches for this task. One variety of this model is the Fast Super-Resolution Convolutional Neural Network (FSRCNN) [DLT16].

At the same time, popular deep neural network architectures developed and succeeded in general image classification tasks. These successful architectures include VGG (Visual Geometry Group) Neural Networks, ResNet (Residual Neural Network), GAN (Generative Adversarial Networks), and Transformers. Thus, these popular architectures have been adopted for super-resolution image enhancement tasks, too. Specifically, Very Deep Super-Resolution Convolutional Neural Networks (VDSR) [KLL16] adopts the VGG style of developing a deep neural network with a skip connection (ResNet) to learn the residual of the high-resolution image out of the original input. Similarly, Ledig et al. [LTH+17] utilize ResNet for this task and create SRResNet (Super-Resolution ResNet). These authors also use the SRResNet as the base and propose SRGAN, a generative adversarial network (GAN), for the image super-resolution (SR) task.

In the same direction, ESRGAN (Enhanced SRGAN) [WYW+18] enhances SRGAN. These authors recently introduced Real-ESRGAN [WXDS21] attempts at "blind" (i.e., utilizing different degradation techniques such as noise, blur, down-sampled, and image compression) super-resolution image enhancement. Another notable work in the GAN direction for image enhancement tasks is Generative Facial Prior GAN or GFPGAN [WLZS21], which focuses on enhancing facial images with the blind approach. Similarly, some works utilize Transformers for image-enhancement tasks. A typical project in this direction should be Codebook Lookup Transformer (CodeFormer) [ZCLL22]. This approach aims at the blind face restoration task, and it adds Generative Facial Prior (GFP) to the face restoration process.

In this project, the 'blind' approaches are not required, and we are not focusing on improving the aesthetic look of the images (such as for the faces) because these approaches may bring side effects (such as erasing the road cracks like clearing face wrinkles for aesthetic purposes). Therefore, we decided to experiment with SRResNet and SRGAN models [LTH+17] for image enhancement purposes. Specifically, the pairs of original images and the corresponding low-resolution images (described in Section 3.3) are used to train these image enhancement models, and the scale factor utilized is four (i.e., with and height are scaled by 4 times each).

3.4 Results and Discussions

Performance metrics are essential for evaluating the accuracy and efficiency of object detection models. They provide insights into how effectively a model can identify and localize objects within images. Key concepts for assessing detection performance include: Intersection over Union (IoU), Precision and Recall, Average Precision (AP), Mean Average Precision (mAP), and F1 Score. IoU quantifies the overlap between a predicted bounding box and a ground truth bounding box, playing a crucial role in

Table 1: Summary of the experiment results

Detection Model	**YOLOv7**			**YOLOv9**		
Image resolution	**Precision**	**Recall**	**mAP50**	**Precision**	**Recall**	**mAP50**
Original	0.609	0.698	0.670	0.712	0.557	0.624
Low-resolution	0.191	0.085	0.046	0.095	0.098	0.049
SRResNet	0.456	0.352	0.325	0.507	0.321	0.349
SRGAN	0.347	0.389	0.267	0.340	0.288	0.248

SRResNet: Images generated by SRResNet, SRGAN: Images generated by SRGAN, mAP50: mean average precision at an intersection over union (IoU) threshold of 0.5

evaluating object localization accuracy. Precision measures the proportion of true positives among all positive predictions, indicating the model's ability to avoid false positives. Recall, on the other hand, measures the proportion of true positives among all actual positives, reflecting the model's capacity to detect all instances of a class. AP computes the area under the precision-recall curve, providing a single value that encapsulates the model's precision and recall performance. The mAP metric extends AP by averaging the AP values across multiple object classes, making it useful for multi-class object detection scenarios to offer a comprehensive evaluation of the model's performance. The F1 Score is the harmonic mean of precision and recall, offering a balanced assessment of a model's performance by considering both false positives and false negatives.

Table 1 shows the summary of the experiment results. Notably, the F1 Score metric is normally a good indicator of prediction performance for a specific dataset, but it is not a good indicator of how well the model can generalize to future and unseen datasets [Pha23]. Therefore, we do not report this score in this project. Additionally, we can utilize different approaches, such as an ensemble of models, and increase the overall performance as described in our previous work [PND22]. However, the main purpose of this project is to evaluate how well image enhancement can help to improve prediction performance. Thus, we do not attempt these approaches.

This table shows that the performances of the trained detection models (both YOLOv7 and YOLOv9) degrade severely while predicting road damages from the low-resolution images (mAP50 at 0.046 and 0.049, respectively). These models perform relatively well on enhanced images (generated from low-resolution images using SRResNet). Specifically, the mAP50 performance metric improved by approximately seven times for both trained detection models. Another interesting note is that SRGAN does not improve prediction performances than SRResNet even though SRGAN claims to generate more aesthetic super-resolution images. One indication of this is that SRGAN may generate images that are good-looking for human eyes but may (1) inadvertently amplify noise or other errors in the original, low-quality image, causing distortions in the enhanced version, and (2) generate high-frequency artifacts. This model can produce unrealistic textures or patterns, confusing subsequent prediction models.

4 Way forward

The experiment results (Table 1) indicate that though the image enhancement models help to improve the subsequent prediction performances, the improved performances are still relatively low compared to the original images (about a half). Therefore, there is still room for improvement in performance by optimizing image enhancement models. The difference in performance improvements from the results of using SRResNet and SSRGAN suggests one direction that is worth exploring, which should be creating models that can enhance image quality while keeping fine-detail patterns, such as lines for the road cracks. Additionally, as the ultimate goal is to predict the road damage (rather than improve image quality), we can incorporate the two deep-learning architectures (one for image enhancement and one for detection) together and train a single neural network with the loss function that ensures prediction performance in the direction that first improves the image quality (and not any other random directions), because we know that improving image quality helps increase subsequent prediction accuracy.

5 Conclusion

This project utilizes images collected from Google Street View with road damage labels to train road damage detection models using the popular YOLOv7 and the recently released YOLOv9. Additionally, this project resizes the original images using the bi-cubic algorithm and generates pairs of low-and-high-resolution images to train image enhancement models using the SRResNet and SRGAN architectures. Experimental results show that the image enhancement models help to increase subsequent prediction performances by approximately seven times. Additionally, the trained SRGAN model can create aesthetic and good-looking images for human eyes, but it does not help improve performance compared to SRResNet. This discrepancy indicates an important result is that some advanced image super-resolution models may create aesthetic images, but they may also generate unrealistic patterns and features. Therefore, it is important to select enhancement models that can improve image quality while keeping the critical fine details from the original images.

References

[AMG+20] Deeksha Arya, Hiroya Maeda, Sanjay Kumar Ghosh, Durga Toshniwal, Hiroshi Omata, Takehiro Kashiyama, and Yoshihide Sekimoto. Global road damage detection: State-of-the-art solutions, 2020.

[BWL20] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[DLHT15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[DLT16] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016.

[GDDM14] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[Gir15] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[HZF21] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13713–13722, 2021.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[KLL16] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[LTH+17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[NJT+22] Fang Nan, Wei Jing, Feng Tian, Jizhong Zhang, Kuo-Ming Chao, Zhenxin Hong, and Qinghua Zheng. Feature super-resolution based facial expression recognition for multiscale low-resolution images. *Knowledge-Based Systems*, 236:107678, 2022.

[Pha23] Vung Pham. *Hands-On Computer Vision with Detectron2*. Packt, Birmingham-Mumbai, 2023.

[PND22] Vung Pham, Du Nguyen, and Christopher Donan. Road damage detection and classification with yolov7. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6416–6423, 2022.

[PPD20] Vung Pham, Chau Pham, and Tommy Dang. Road damage detection and classification with Detectron2 and Faster R-CNN. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 5592–5601. IEEE, 2020.

[RDGF16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[RF17] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[RF18] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[RHGS15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[SVI+16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[WBL22] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[WL24] Chien-Yao Wang and Hong-Yuan Mark Liao. YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv*, 2024.

[WLZS21] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9168–9178, 2021.

[WXDS21] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.

[WYW+18] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[ZCLL22] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022.

Vung Pham is an assistant professor at the Computer Science Department, Sam Houston State University, Texas, US. He is a passionate research scientist in machine learning, deep learning, data science, and data visualization. He has years of experience and numerous publications in these areas. His areas of expertise centered around data visualizations, data analytics, machine learning, and deep learning. He has published one patent, various peer-reviewed journals, and conference papers in these areas. He is currently working on projects that use deep neural networks (e.g., YOLO, R-CNN) to solve computer vision tasks (e.g., road damage detection or brain tumor detection). He is also the author of the book "Hands-On Computer Vision with Detectron2" by Packt Publishing. He also works on projects to analyze climate change and health information.

INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University

The Institute for Homeland Security at Sam Houston State University is focused on building strategic partnerships between public and private organizations through education and applied research ventures in the critical infrastructure sectors of Transportation, Energy, Chemical, Healthcare, and Public Health.

The Institute is a center for strategic thought with the goal of contributing to the security, resilience, and business continuity of these sectors from a Texas Homeland Security perspective. This is accomplished by facilitating collaboration activities, offering education programs, and conducting research to enhance the skills of practitioners specific to natural and human caused Homeland Security events.

Institute for Homeland Security
Sam Houston State University

Pham, Vung (2024) Improving Road Damage Detection Accuracy Using Deep Learning Image Enhancement Models. (Report No. IHS/CR-2024-1018). The Sam Houston State University Institute for Homeland Security.
https://doi.org/10.17605/OSF.IO/ZDKJN