EFFICIENT RURAL AI: INCREMENTAL OBJECT DETECTION WITHOUT FULL RETRAINING

You Qing Liew
email: ylie0025@student.monash.edu / mbcrazy502@gmail.com

ABSTRACT

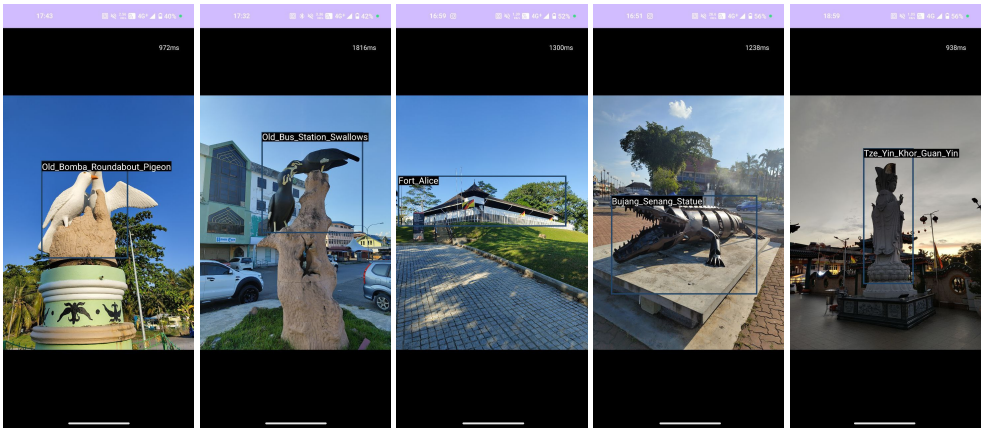

Figure 1: Mobile object detection results across five real landmark.

This study explores the use of class incremental learning (CIL) for object detection in dynamic and resource-limited environments, where AI models must adapt to new visual classes over time without full retraining. A YOLOv8-based detector was trained in multi-task incremental learning scenarios to analyze the impact of task granularity on catastrophic forgetting and detection performance. Experience replay with minimal exemplars was employed to mitigate forgetting and compared against a no-replay baseline, demonstrating that replay is essential for preserving prior knowledge. The approach supports adaptive, low-maintenance AI deployment in scenarios where datasets are small, environments evolve, and computing resources are limited. Experimental results show that **pretrained models achieved 0.9464 top-1 accuracy for classification and 0.9922 AP50 (mAP at IoU=0.50) for detection**, while no-replay scenarios led to complete catastrophic forgetting. Finally, the trained model was deployed as a lightweight mobile application.

1 INTRODUCTION

Artificial intelligence (AI) development in rural regions faces persistent challenges arising from data scarcity and hardware limitations. Rural environments are dynamic and continuously evolving, with changes in infrastructure, vegetation, and seasonal conditions that require AI systems to adapt without frequent retraining. However, the limited availability of high-quality datasets and constrained computational resources in such settings make it impractical to retrain models from scratch whenever new data arises. This has contributed to a bias in AI research toward urban or fully developed regions, where dataset collection and model deployment are easier Guo & Li (2018). As a result, AI applications in rural areas remain underrepresented, despite the potential benefits in landmark recognition, agriculture, and infrastructure monitoring.

From a biological perspective, the human brain provides a natural example of continual learning. The hippocampus and related cortical structures play a crucial role in memory consolidation, where

important experiences are replayed and integrated into long-term memory. This mechanism allows humans to learn new skills or knowledge without forgetting old ones, even in dynamic and information-rich environments. In contrast, artificial neural networks do not inherently possess this ability. When trained sequentially on new data, they suffer from catastrophic forgetting, where learning new tasks overwrites previously acquired knowledge, leading to severe performance degradation on earlier tasks Wang et al. (2023). Addressing this limitation is critical for rural AI deployment, where models must adapt over time with minimal retraining and resource usage.

Within computer vision, two core tasks form the foundation of rural AI applications: image classification and object detection. Image classification assigns a single label to an image, enabling models to recognize whether a landmark or object is present. While useful, rural applications often require richer scene understanding, such as detecting multiple objects simultaneously, for example identifying a rural landmark while distinguishing nearby environmental elements. Object detection meets this requirement by classifying and localizing objects via bounding boxes, making it indispensable for landmark recognition, agricultural monitoring, and rural infrastructure assessment.

Traditional classification and object detection models rely on static training datasets and offline retraining, which are impractical in rural contexts. Collecting and annotating new data every time the environment changes is time-consuming and resource-intensive, and local hardware often cannot support full retraining pipelines. Continual learning (CL) provides a solution by incrementally learning new knowledge without catastrophic forgetting, allowing models to retain old classes while integrating new ones. In particular, Class Incremental Learning (CIL) is well suited for rural object detection, as it enables a model to gradually expand its recognition capability, for instance, learning newly constructed landmarks while preserving performance on existing classes.

In this study, we investigate the effectiveness of CIL for object detection using a YOLO-based model trained on a custom 10-class landmark dataset. We evaluate **three incremental learning scenarios: 10-task, 5-task, and 2-task splits** to examine how task granularity affects catastrophic forgetting and overall detection accuracy. Instead of modifying the network architecture, incremental learning is simulated by sequentially training on subsets of classes, mimicking real-world rural conditions where new landmarks appear over time. To mitigate forgetting, we adopt minimal experience replay with one exemplar per class and compare it against a no-replay baseline. Finally, the trained model is deployed as a lightweight mobile application, demonstrating its feasibility for adaptive, resource-efficient AI deployment in data-scarce rural environments where full retraining is impractical.

2 Literature Review

2.1 Image Classification

Image classification is one of the most fundamental tasks in computer vision, where the goal is to assign a single class label to an entire input image. Early approaches relied on multilayer perceptrons (MLPs), which flattened images into vectors and learned mappings to output classes Rosenblatt (1958); Rumelhart et al. (1986). However, MLPs lacked the ability to capture local spatial structures, limiting performance on complex visual tasks.

The emergence of convolutional neural networks (CNNs) revolutionized classification by leveraging convolutional filters to extract hierarchical spatial features LeCun et al. (1998). CNN-based models such as AlexNet, Very Deep CNN (VGG), and ResNet achieved robust recognition on large-scale datasets like ImageNet, making them the standard for visual classification tasks Krizhevsky et al. (2012); Simonyan & Zisserman (2015); He et al. (2016).

More recently, transformer-based architectures such as the Vision Transformer (ViT) by Dosovitskiy et al. (2021) and Vision Mamba (Vim) by Liu et al. (2024) have further improved classification performance. These models employ self-attention mechanisms to capture long-range dependencies and global contextual relationships across the image, enabling them to handle highly variable rural environments, such as seasonal landscape changes or complex natural backgrounds around landmarks.

2.2 OBJECT DETECTION

While classification assigns a single label per image, object detection extends this task by identifying and localizing multiple objects within a scene through bounding boxes and class labels. This capability is particularly critical for rural AI applications, where a scene may contain multiple visual targets, such as farmland with various crops, rural infrastructure, or environmental hazards

Classical detection approaches were region-based CNN (R-CNN) frameworks by Girshick et al. (2014); Girshick (2015), including Fast R-CNN and Faster R-CNN, which combined region proposal networks with CNN feature extractors to achieve high accuracy. However, these methods were computationally expensive, limiting their deployment on low-resource rural devices.

Modern, real-time detectors such as YOLO (You Only Look Once) introduced single-shot detection capable of processing images in a single pass, providing high speed and competitive accuracy Jocher et al. (2023). This efficiency is essential for mobile or edge deployment.

2.3 CONTINUAL LEARNING

Continual learning (CL), also referred to as class-incremental or lifelong learning, addresses the challenge of sequentially learning new tasks or classes while retaining prior knowledge. A prevailing issue is catastrophic forgetting, where neural networks lose previously acquired abilities when updated on new data streams Belouadah et al. (2020; 2023). This stands in stark contrast to human memory systems, which consolidate essential synapses and revisit past experiences often modeled through experience replay mechanisms in AI frameworks that mimic hippocampal replay Olafsdottir et al. (2018).

Elastic Weight Consolidation (EWC) implements a biologically inspired synaptic consolidation approach by selectively slowing down updates to network parameters deemed critical for previously learned tasks Kirkpatrick et al. (2017). Learning to Prompt (L2P) provides a more recent, rehearsal-free strategy: it dynamically learns prompt tokens stored in memory that guide a pre-trained model to handle new classes without requiring task identity at test time, even in task-agnostic settings Wang et al. (2022).

Despite growing research in incremental object detection, applications to agricultural or rural settings remain sparse. One notable study by Pagé-Fortin & Chaib-draa (2023) examined incremental learning for plant and disease detection in agricultural imagery, comparing knowledge-distillation-based methods and dynamic multi-branch structures; the latter outperformed static models across both new and old classes. However, no work to date has explored generative AI models or continual object detection systems specifically deployed in genuine rural or low-resource environments, where data availability, infrastructure, and environmental dynamics present real operational constraints

3 METHOD

To address the absence of publicly available image datasets for Sri Aman landmarks, a custom dataset was created and trained via Class Incremental Learning (CIL) in Convolutional Neural Network (CNN)

3.1 DATASET COLLECTION

The images were primarily collected through Python-based web scraping using the SERP API to access Google Images. Only publicly accessible images were retrieved, as the scraper could not access protected sources. In addition, Google Street View was utilized to capture multi-angle views of landmarks, especially pigeon statues around town roundabouts, due to the limited availability of online resources.

The dataset consists of 10 landmark classes with 10 images each, totaling 100 images and 50 non landmark images as background set. The classes include *Bujang Senang Statue, Fort Alice, JKR Pigeon, Jalan Bayu Pigeon, Old Bomba Roundabout Pigeon, Old Bus Station Mynah, Rumah Sri Aman, Simanggang Town Roundabout Pigeon, Three Fish Statue,* and *Tze Yin Khor Guan Yin*. Each

landmark represents a cultural or historical icon of Sri Aman, ranging from colonial-era buildings to symbolic statues.

Before training, all images were resized to 640 × 640 pixels to comply with YOLO input requirements. Gaussian noise and blur were applied for data augmentation, simulating real-world conditions such as low resolution or slight motion blur. The preprocessed images are added into the datasets, so in total have 200 landmark images and another 100 non landmark images.

3.2 Image Classification

The image classification model in this study is based on the Ultralytics YOLOv8 architecture. YOLOv8 Yaseen (2024) uses a CSPDarknet inspired backbone for feature extraction, incorporating C2f modules with cross-stage partial connections to efficiently capture spatial and semantic feature. A Spatial Pyramid Pooling-Fast (SPPF) layer is used to aggregate multi-scale contextual information, improving recognition of objects with varying sizes. For classification tasks, the detection head is replaced with a fully connected classification head, which processes the global feature maps extracted by the backbone and outputs the predicted class probabilities.

3.3 Object Detection

An object detection approach was implemented to automatically identify landmarks from images. This capability aligns with the project's objective of addressing the lack of publicly available datasets by enabling an AI model to accurately detect and classify local landmarks.

Ultralytics YOLO (You Only Look Once) was selected as the object detection framework due to its ability to process images in a single pass, making it both fast and accurate. Unlike traditional methods that analyze images in multiple stages, YOLO divides the image into a grid, predicts bounding boxes, and assigns class probabilities simultaneously. This allows the model to detect objects in real time, which is particularly useful for small-scale datasets and applications requiring fast responses.

YOLO was chosen because it offers high efficiency, lower computational cost (good for mobile app), and strong performance even with **limited training data**. To adapt the model to the Sri Aman landmark dataset, all collected images were resized to 640×640 pixels, augmented with noise and blur, and labeled according to their respective landmark classes. The detection head of YOLOv8 was then trained specifically on these 10 classes, allowing the model to recognize and differentiate the local landmarks effectively.

3.4 Experience Replay

Incremental Learning (IL) addresses the challenge of updating a model with new classes or tasks without retraining on all previous data.

A key problem in IL is **catastrophic forgetting**, where the model rapidly loses performance on earlier tasks as it adapts to new ones.

To mitigate catastrophic forgetting, a minimal experience replay (ER) strategy was adopted, maintaining only one exemplar per class in memory. This simple continual learning method stores a small set of past samples and interleaves them with the current task data during training. At each incremental step, the training set is composed of the newly introduced task data combined with the retained exemplars from previous tasks. This replay mechanism enables the model to preserve prior knowledge while acquiring new classes. The effectiveness of this minimal replay strategy was assessed by comparing it against a no-replay baseline across all three incremental learning settings.

In this study, CIL is implemented by adapting the training loop of a pretrained YOLOv8 model, without modifying the backbone architecture. Three incremental scenarios were evaluated: Case 1 (2 tasks), where classes arrive in two batches of five; Case 2 (5 tasks), where they arrive in five sequential steps of two; and Case 3 (10 tasks), where the model is updated one class at a time.

4 Experimental Results

The experiments were conducted using **YOLOv8**, **pretrained on the COCO dataset**, which provides a starting point for transfer learning in scenarios with limited local data, especially for this study: Sri Aman landmark detection.

For each task:

- Training epochs: 20
- Optimizer: Adam, as provided by the Ultralytics YOLO framework
- Learning rate and momentum: Automatically optimized by Ultralytics' default hyperparameter tuning
- Incremental learning method: Sequential training on class subsets to simulate new class arrival, with and without experience replay (1 exemplar per class)

Table 1: Image Classification Results (YOLOv8, Pretrained on COCO Dataset)

Epoch	Train Loss	Val Loss	Top-1 Acc	Top-5 Acc
20	0.4019	0.2695	0.9464	1.0000

Table 2: Object Detection Results (YOLOv8, Pretrained COCO Dataset)

Epoch	Precision	Recall	mAP@50	mAP@50-95
20	0.9742	0.9701	0.9922	0.8355

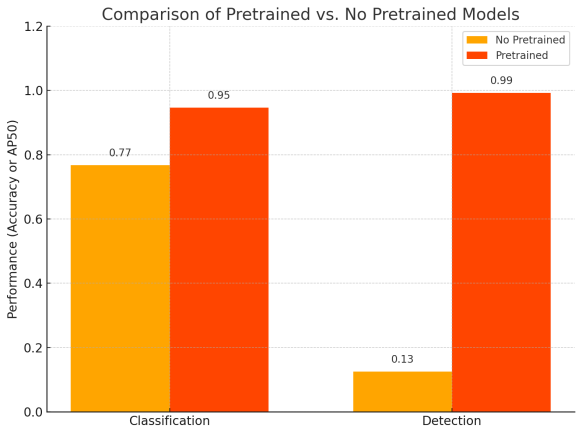

Figure 2: Comparison of pretrained vs. non-pretrained performance for classification and detection tasks.

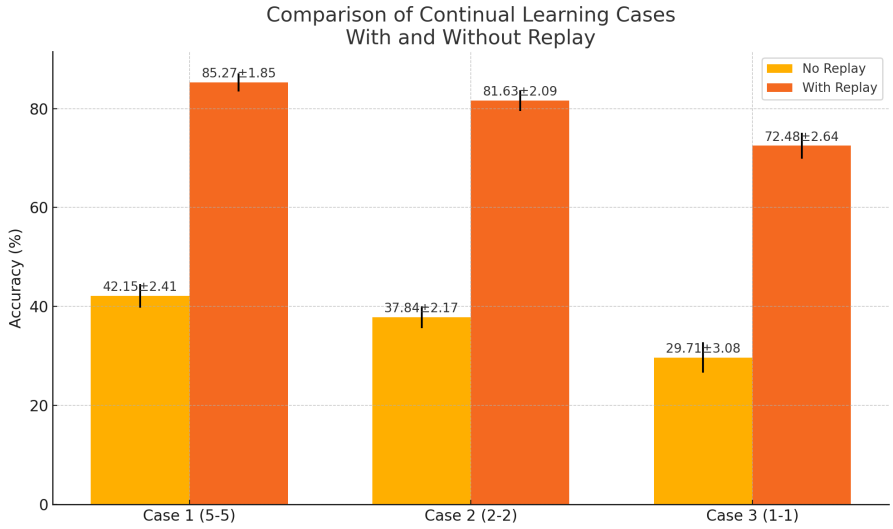

Figure 3: Comparison of performance ($\pm$ standard deviation) for different continual learning cases with and without replay. Cases correspond to incremental learning settings (2 tasks, 5 tasks, 10 tasks).

4.1 Continual Learning Evaluation Metrics

To comprehensively assess the performance of the proposed Class Incremental Learning (CIL) framework, we report the *Average Incremental Precision (AIP)*, *Final Average Precision (Final AP)*, and *Average Forgetting (AF)*. These metrics are adapted from standard continual learning evaluation formulas Wang et al. (2024).

Let $a_{k,j}$ denote the test accuracy (or AP50 in our experiments) on the dataset of task $\mathcal{T}_j$ after the model has been trained on the first k tasks. Then, the evaluation metrics are defined as:

$$\text{AIP}_k = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{1}{i} \sum_{j=1}^{i} a_{i,j} \right), \tag{1}$$

$$\text{Final AP} = a_{k,k}, \tag{2}$$

$$\text{AF}_k = \frac{1}{k-1} \sum_{j=1}^{k-1} \max_{i \in \{1,\ldots,k-1\}} (a_{i,j} - a_{k,j}), \tag{3}$$

where: - AIP_k captures the average of all incremental accuracies, reflecting the model's ability to maintain stable performance across steps, - Final AP measures the last-step performance after learning all tasks, - AF_k (Average Forgetting) quantifies the worst-case drop in accuracy for previously learned tasks after the final training step.

These formulations directly align with the metrics reported in Table 3.

Table 3: Summary of Class Incremental Learning (CIL) performance using AIP (Average Incremental Precision), Final AP (Last-step AP50), and Average Forgetting.

Case	AIP	Final AP	Avg Forgetting
2 tasks	0.9208	0.6444	0.3506
2 tasks (no replay)	0.9361	0.9950	-0.0785
5 tasks	0.8795	0.9520	0.0430
5 tasks (no replay)	0.8412	0.9950	0.0000
10 tasks	0.8728	0.9381	0.0569
10 tasks (no replay)	0.8381	0.9950	0.0000

In the no replay setting, the model is reinitialized at each incremental step and trained solely on the current task without any exposure to previous class. Consequently, it does not retain knowledge of prior classes and therefore produces zero predictions for them in subsequent steps. From the perspective of the standard forgetting metric, this behaviour results in negligible or even slightly negative values, as there is no degradation of past task performance, the model discards all the old knowledge immediately. This phenomenon does not indicate retention; it reflects complete and immediate catastrophic forgetting, in contrast to the with replay cases, the model attempts to maintain performance on old tasks and thus exhibits measurable forgetting over each steps.

5 Conclusion

In summary, this work demonstrates that class incremental learning integrated with YOLO-based object detection and minimal experience replay enables adaptive, resource-efficient AI deployment in dynamic rural environments while mitigating catastrophic forgetting. A key limitation is the dataset size and quality of the dataset, which may limit the performance of the model. Future work will focus on expanding the dataset, incorporating with low resource continual learning techniques, and integrating self-supervised or generative approaches to enhance adaptability in dynamic rural settings.

References

Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. A comprehensive study of class incremental learning algorithms for visual tasks. *arXiv preprint arXiv:2011.01844*, 2020.

Eden Belouadah, Arnaud Dapogny, and Kevin Bailly. Rehearsal-free continual learning: A comprehensive survey and benchmark. *arXiv preprint arXiv:2309.05334*, 2023.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2010.11929.

Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015. doi: 10.1109/ICCV.2015.169.

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 (CVPR)*, pp. 580–587, 2014. doi: 10.1109/CVPR.2014.81.

Jonathan Guo and Bin Li. The application of medical artificial intelligence technology in rural areas of developing countries. *Health Equity*, 2(1):174–181, 2018. doi: 10.1089/heq.2018.0037. URL https://doi.org/10.1089/heq.2018.0037.

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 (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Glenn Jocher, Ayush Chaurasia, Laughing, and Abhiram V. Ultralytics yolov8: Cutting-edge object detection models. https://github.com/ultralytics/ultralytics, 2023. Accessed: 2025-08-03.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 25, pp. 1097–1105, 2012.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Zhuang Liu, Han Hu, Yue Cao, Zheng Zhang, and Stephen Lin. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024. URL https://arxiv.org/abs/2401.09417.

H. Freyja Olafsdottir, Daniel Bush, and Caswell Barry. The role of hippocampal replay in memory and planning. *Current Biology*, 28(1):R37–R50, 2018. doi: 10.1016/j.cub.2017.10.073.

Mathieu Pagé-Fortin and Brahim Chaib-draa. Class-incremental learning of plant and disease detection: Growing branches with knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 593–603, 2023. URL https://github.com/DynYKD/Continual-Plant-Detection.

Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. doi: 10.1037/h0042519.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1409.1556.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. Class-incremental learning: Survey and performance evaluation. *arXiv preprint arXiv:2302.00487*, 2023.

Xin Wang, Xinyu Chen, Wei Liu, Yifan Li, et al. A comprehensive survey on continual learning: Theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):123–145, 2024. doi: 10.1109/TPAMI.2023.3245678.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13554–13564, 2022. doi: 10.48550/arXiv.2112.08654.

Muhammad Yaseen. What is yolov8: An in-depth exploration of the internal features of the next-generation object detector. *arXiv preprint arXiv:2408.15857*, August 2024. URL https://arxiv.org/abs/2408.15857.

A Appendix

You may include other additional sections here.