ASM Sc. J., 20(1), 2025
https://doi.org/10.32802/asmscj.2025.1955

Millimetre Wave Radar in Residual Neural Network with Multiple Large Model Parameters

C.H. Zhao and W.Y. Leong*

INTI International University, 71800 Nilai, Malaysia

In recent years, millimetre-wave radar has shown important application value in automatic target detection, autonomous driving perception, and multimodal data fusion tasks. After combining it with deep learning, especially with residual neural network (ResNet) and its derivative structures, millimetre-wave radar has further improved its performance in object detection. This study systematically compares the experimental performance of four deep learning models (ResNet50, ResNet101, Res2Net101 and Swin Transformer, referred to as Swin-T) on KITTI, nuScenes and BDD datasets. We evaluated key performance indicators including mAP, mAP0.5, mAP0.75, mAPS, mAPM, mAPL, mAR, mARS, mARM, mARL, FPS, and inference speed (FPS) in order to provide guidance for model selection for mmWave radar combined with deep learning.

Keywords: Millimetre-wave radar; residual neural networks; product innovation

I. INTRODUCTION

In the field of target detection and autonomous driving, millimetre-wave radar is valued for its stability and high penetration in adverse weather conditions (Venon, A *et al.*, 2022). The development of deep learning technology is also an important factor in the deeper application of automatic detection technology, especially the residual neural network (ResNet) and its derivative structures, which have shown strong feature extraction capabilities in the field of image and signal processing and have been widely used (Zhao, 2025a; Leong, 2024c).). This review aims to explore the application of millimetre wave radar and residual neural network in target detection tasks and comparatively analyses the performance of different deep learning models on multiple standard data sets.

In order to systematically evaluate the performance of different deep learning models on millimetre wave radar data, the authors selected four representative models: ResNet50, ResNet101, Res2Net101 and Swin Transformer (Swin-T for short), and conducted experiments on three widely recognised datasets: KITTI, nuScenes and BDD. The author's choices include average precision (mAP), precision at different IoU thresholds (mAP0.5, mAP0.75), precision at different scales (mAPS, mAPM, mAPL), average recall (mAR) and its variant (mARS, mARM, mARL) and frame rate (FPS) and other key performance indicators to evaluate model performance (Zhao, 2025b).

The goal of this review is to provide guidance for the selection of millimetre-wave radar combined with deep learning models. Through comparative analysis, the author hopes to reveal the advantages and challenges of different models in processing millimetre-wave radar data and provide a reference for future research and applications.

II. MODEL INTRODUCTION

ResNet is a residual network proposed by Microsoft Research. Since it won the CVPR Best Paper Award in 2015, it has been considered the gold standard architecture in the field of computer vision (Mehta, O *et al.*, 2022).

The ResNet50 network contains 50 layers of depth. Its overall structure is divided into stage0 to stage4 from input to output. Each stage is composed of multiple residual blocks (Bottleneck) (Zhang, C *et al.*, 2022). Specifically, ResNet50 has four groups of large residual blocks, each group contains 3, 4, 6, and 3 small residual blocks respectively, and each small residual block contains three

*Corresponding author's e-mail: waiyie.leong@newinti.edu.my

convolutional layers. In addition, the network starts with a single convolutional layer, so there are a total of 49 convolutional layers plus a fully connected layer, for a total of 50 layers.

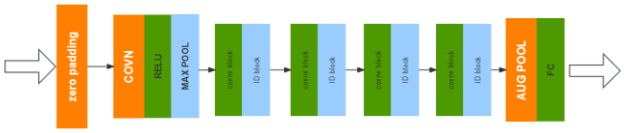

Figure 1. ResNet50 structure diagram

Researchers can use the pre-trained ResNet50 model to adapt it to specific classification tasks. The image to be classified is input into the trained ResNet50 model. Through its powerful feature extraction capabilities, it can accurately locate and identify targets in images (Chun, 2010).

ResNet101 is a deep residual network model that has revolutionary applications in deep learning and image recognition. The core features of ResNet101 are: ResNet101 has a 101-layer deep hierarchical structure designed to learn abstract features at different levels. The depth of the network enables it to capture low-level and high-level features (Vaishali, S *et al.*, 2024); ResNet101 can handle different input shapes. ResNet101 has the characteristics of depth and complexity, making it ideal for image classification tasks that require high accuracy. It has demonstrated excellent performance in a variety of challenging environments, including medical image analysis, satellite image interpretation, and advanced visual recognition tasks. As the network depth increases, the calculation amount and memory consumption of the model will also increase accordingly (Yao, P *et al.*, 2020). In practical applications, appropriate models need to be selected based on specific tasks and resource constraints.

Res2Net101 is a new deep residual network structure that achieves multi-scale feature representation by constructing hierarchical residual connections within a single residual block (Qu, Z *et al.*, 2021). The core idea of Res2Net101 is to construct multi-scale feature representations within a single residual block. This design enables the network to capture features of different scales at a finer granularity and increase the receptive field range of each layer. In this way, Res2Net101 enhances the multi-scale representation ability of the network, thus achieving performance improvements on a variety of visual tasks.

For performance comparison, this article also introduces the most cutting-edge Transformer architecture model for performance comparison. Swin Transformer is a layered Transformer structure specially designed for computer vision tasks. Swin Transformer adopts the technology of moving window segmentation between consecutive Transformer blocks. This movement ensures that image areas processed separately in the previous layer can interact with each other in the next layer, promoting better local and global context information. Integrate (Liu, Z *et al.*, 2021). As a general backbone network for computer vision, Swin Transformer has achieved excellent performance and results in tasks such as object classification, target detection, semantic and instance segmentation, and target tracking (Leong, 2002; 2003).

III. DATASET INTRODUCTION

This article uses relevant experimental data on KITTI, nuScenes and BDD data sets for analysis. These datasets cover a variety of driving scenarios, with different object types and density distributions.

The KITTI dataset is a dataset for research in the field of autonomous driving, jointly created by the Karlsruhe Institute of Technology (KIT) in Germany and the Toyota Technological Institute of Chicago (TTI-C) in the United States. The KITTI data set contains data collected by a variety of on-board sensors. Unlike data sets generated through computer graphics technology, the KITTI data set is closer to the real situation (Ristić-Durrant *et al.*, 2021). The KITTI data set covers stereo vision, optical flow estimation, visual odometry, 3D object detection, object tracking, road surface and lane detection and many other aspects. The KITTI dataset has become an important resource for research and algorithm evaluation in the field of autonomous driving due to its characteristics of multi-modal data, real-world scenarios and multi-task coverage (Leong, 2025a). The nuScenes dataset is a large-scale public dataset designed for autonomous driving research and developed by Motional (formerly nuTonomy). The nuScenes data set is inspired by the KITTI data set and has a wide range of application scenarios in the field of autonomous driving

(Pham, Q *et al.*, 2020). It mainly includes using camera images and lidar data to detect and track vehicles, pedestrians and other targets on the road. As an evaluation benchmark for autonomous driving algorithms, it helps researchers evaluate the performance and generalisation ability of the algorithms (Leong, 2024a; 2024b).

The BDD dataset, full name BDD100K (Berkeley DeepDrive 100K), is a large-scale and diverse autonomous driving video database released by the University of California, Berkeley AI Laboratory (BAIR) (Chen, Y *et al.*, 2022). The BDD100K data set supports a variety of tasks related to autonomous driving, including but not limited to object detection, semantic segmentation, lane line detection, drivable area segmentation and scene understanding (Leong, 2025b).

IV. STATISTICAL ANALYSIS OF PERFORMANCE DATA OF RESNET50, RESNET101, RES2NET101 AND SWIN TRANSFORMER ON KITTI, NUSCENES AND BDD DATASETS

Because different hardware platforms will result in different parameter data, this article uses the original test performance data analysis of KITTI, nuScenes and BDD datasets. Because the data involves parameters such as mAR, mARS, mARM, mARL, mAP, mAP0.5, mAP0.75, mAPS, mAPM, mAPL, and FPS, this article explains the above parameters.

mAR (Mean Average Recall) refers to the mean of the average recall rates of all categories in the object detection task (Hosang, J *et al.*, 2015). Recall is a measure of the model's ability to detect all positive samples (real objects). It is defined as the number of successfully detected positive samples divided by the number of all true positive samples. For each category, the recall rate is the proportion of all positive samples in that category that are correctly detected by the model. For example, for the category, the recall rate Rc can be expressed as:

$$Rc = \frac{TPc}{TPc+FNc} \quad (1)$$

In the above formula, *TPc* is the number of true positives, that is, the number of samples of this category that the model correctly detects. *FNc* is the number of false negatives, which is the number of samples of this category that the model fails to detect. mAR measures the overall recall of the model on different categories by calculating the recall rate of each category and then averaging these recall rate values. mAR is the mean of the average recall rates of all categories. If there are *C* categories, *mAR* can be expressed as

$$mAR = \frac{1}{C}\sum_{c=1}^{R} Rc \quad (2)$$

mAR is an important performance metric that reflects the model's ability to detect objects on average across all categories (Zheng, L *et al.*, 2016). mARS specifically measure the model's recall performance when detecting small objects. mARM refers to recall performance on medium-size objects. mARL measures the model's recall performance when detecting large objects.

mAP is the abbreviation of Mean Average Precision, which measures the overall performance of the target detection model on all categories (Sun, M *et al.*, 2022). mAP is calculated by calculating the AP (Average Precision) value for each category and then averaging these AP values. AP (Average Precision) means that in the target detection task, for a single category, by changing the confidence threshold, the precision (Precision) at different recall levels is calculated, and then the average of these precision values is calculated. The calculation of mAP involves calculating AP separately for each category and then averaging the AP values across all categories. If there are *K* categories in the dataset, the *mAP* calculation formula is:

$$mAP = \sum_{c=1}^{k} APc \quad (3)$$

In the formula, APc refers to the AP value of the cth category. mAP is mainly used to evaluate the performance of the target detection model. The higher the mAP, the better the overall performance of the model (An, Q *et al.*, 2019). mAP0.5 is the mAP when the Intersection over Union (IoU) threshold is 0.5. IoU is an indicator that measures the degree of overlap between the predicted bounding box and the true bounding box. mAP0.75 is the mAP when the IoU threshold is 0.75. mAPS refer to the mAP on small objects. mAPS specifically measure the performance of the model in detecting small objects. mAPM refers to the mAP on medium-sized objects. This indicator focuses on the performance of the model in detecting medium-sized objects. mAPL refers to the mAP on large objects. This

indicator measures the performance of the model in detecting large-sized objects.

In target detection and computer vision tasks, FPS is often used to measure the inference speed of the model, that is, the number of frames the model can process per second.

Next, we explore the performance test parameters of the original datasets, which are all from the official websites of KITTI, nuScenes, and BDD datasets.

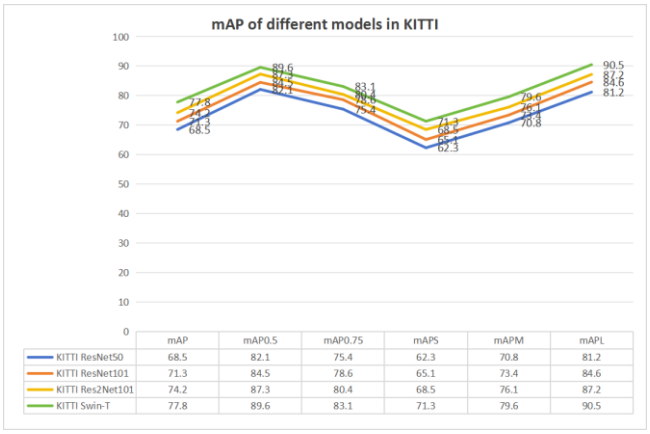

Figure 2. KITTI performance data diagram

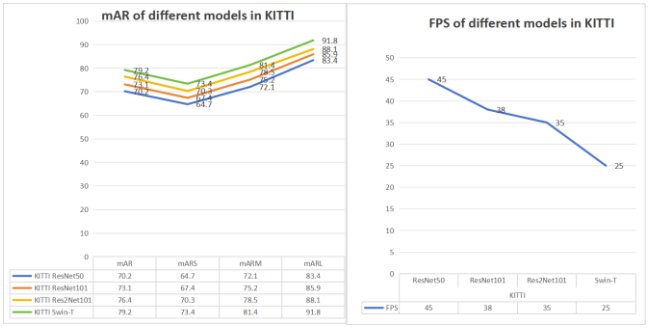

Figure 3. KITTI performance data diagram

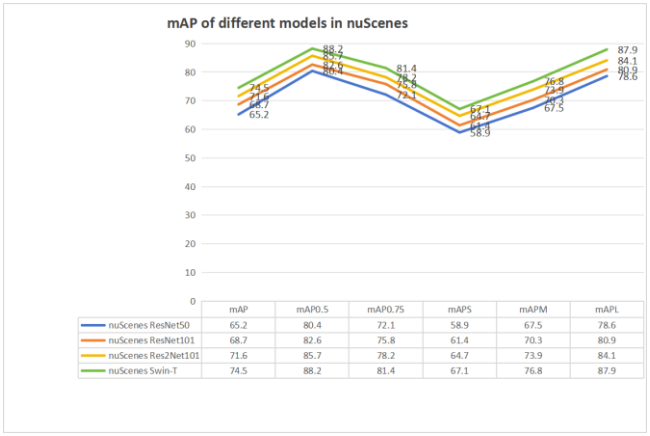

Figure 4. nuScenes performance data diagram

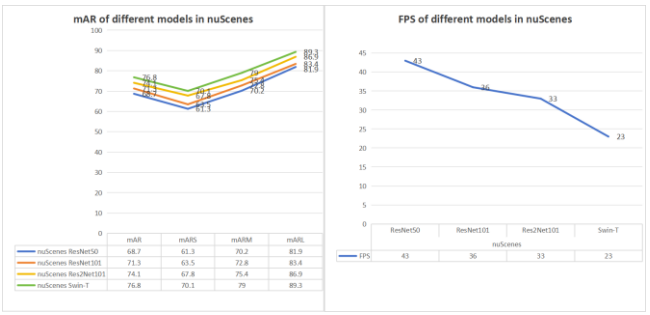

Figure 5. nuScenes performance data diagram

Figure 6. BDD performance data diagram

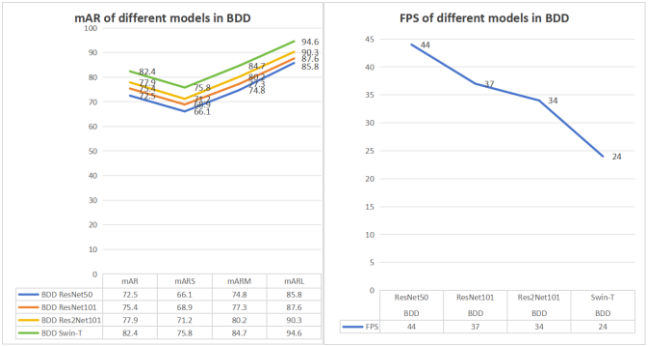

Figure 7. BDD performance data diagram

Judging from the results, the Swin-T model performs most prominently in various indicators, especially in large target detection mAPL, mARL and overall mAP, which is significantly better than the ResNet series models. However, its inference speed is low, which limits its applicability in real-time application scenarios. By analysing the advantages and disadvantages of different models, this paper believes that ResNet50/101 is simple and efficient, suitable for small and medium-sized model deployment scenarios. Res2Net101 benefits from its multi-scale feature capabilities and performs better than ResNet101 on small and medium-

sized targets. Although Swin-T has the best performance, it has higher hardware requirements and is suitable for scenarios with high precision requirements.

V. RESULTS AND DISCUSSION

This review provides guidance for the selection of deep learning models for millimetre-wave radar by comparing and analysing the performance of four deep learning models on KITTI, nuScenes, and BDD datasets. Future work will focus on further optimising these models, improving their performance on millimetre-wave radar data, and exploring new model structures and training strategies to meet the growing market demand.

VI. REFERENCES

An, Q, Pan, Z, Liu, L & You, H 2019, 'DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images', IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 8333-8349.

Chen, Y, Ma, M, Yu, Q, Du, Z & Ding, W 2022, 'Road Bump Outlier Detection of Moving Videos Based on Domestic Kylin Operating System', in Proceedings of the 6th International Conference on High Performance Compilation, Computing and Communications, pp. 137-143.

Chun, WX & Leong, WY 2010, 'Composite defects diagnosis using parameter optimization-based support vector machine', in 2010 5th IEEE Conference on Industrial Electronics and Applications, pp. 2300-2305.

Hosang, J, Benenson, R, Dollár, P & Schiele, B 2015, 'What makes for effective detection proposals?', IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 4, pp. 814-830.

Leong, WY & Homer, J 2002, 'Hop selection in peer-to-peer WPAN networks', in CCS 2002: 8th International Conference on Communications Systems (volume 1 and 2), IEEE, vol. 2, pp. 870-872.

Leong, WY & Homer, J 2003, 'Enhancing interference mitigation in communication', Fourth International Conference on Information, Communications and Signal Processing 2003 and the Fourth Pacific Rim Conference on Multimedia, Singapore, vol. 1, pp. 587-591.

Leong, WY, Leong, YZ & Leong, WS 2024a, 'Miniature THz Antenna Design', 2024 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM), Taiwan, pp. 1-2.

Leong, WY, Leong, YZ & Leong, WS 2024b, 'Nuclear Technology in Electronic Communications', 2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB), Taiwan, pp. 684-689.

Leong, WY 2024c, 'Industry 5.0: Design, standards, techniques and applications for manufacturing. Institution of Engineering and Technology.

Leong, WY 2025a, 'Cognitive Spectrum Management and Coexistence Strategies for 6G Wireless Ecosystems', 2025 IEEE International Conference on Geoinformation Science and Communication Technology (GSCT), 16-18 October, Shanghai.

Leong, WY 2025b, 'AI-Native 6G Networks: Architectures and Protocols', 2025 IEEE International Conference on Geoinformation Science and Communication Technology (GSCT), 16-18 October, Shanghai.

Liu, Z, Lin, Y, Cao, Y, Hu, H, Wei, Y, Zhang, Z ... & Guo, B 2021, 'Swin transformer: Hierarchical vision transformer using shifted windows', in Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012-10022.

Mehta, O, Liao, Z, Jenkinson, M, Carneiro, G & Verjans, J 2022, 'Machine learning in medical imaging–clinical applications and challenges in computer vision', Artificial Intelligence in Medicine: Applications, Limitations and Future Directions, pp. 79-99.

Pham, QH, Sevestre, P, Pahwa, RS, Zhan, H, Pang, CH, Chen, Y ... & Lin, J 2020, 'A* 3d dataset: Towards autonomous driving in challenging environments', in 2020 IEEE International conference on Robotics and Automation (ICRA), IEEE, pp. 2267-2273.

Qu, Z, Chen, W, Wang, SY, Yi, TM & Liu, L 2021, 'A crack detection algorithm for concrete pavement based on attention mechanism and multi-features fusion', IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11710-11719.

Ristić-Durrant, D, Franke, M & Michels, K 2021, 'A review of vision-based on-board obstacle detection and distance estimation in railways', Sensors, vol. 21, no. 10, p. 3452.

Sun, M, Zhang, H, Huang, Z, Luo, Y & Li, Y 2022, 'Road infrared target detection with I-YOLO', IET Image Processing, vol. 16, no. 1, pp. 92-101.

Vaishali, S & Neetu, S 2024, 'Enhanced copy-move forgery detection using deep convolutional neural network (DCNN) employing the ResNet-101 transfer learning model', Multimedia Tools and Applications, vol. 83, no. 4, pp. 10839-10863.

Venon, A, Dupuis, Y, Vasseur, P & Merriaux, P 2022, 'Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey', IEEE Transactions on Intelligent Vehicles, vol. 7, no. 3, pp. 533-555.

Yao, P, Wu, H, Gao, B, Tang, J, Zhang, Q, Zhang, W, ... & Qian, H 2020, 'Fully hardware-implemented memristor convolutional neural network', Nature, vol. 577, no. 7792, pp. 641-646.

Zhang, C, Bengio, S & Singer, Y 2022, 'Are all layers created equal?', Journal of Machine Learning Research, vol. 23, no. 67, pp. 1-28.

Zhao, CH & Leong, WY 2025a, 'Application of Millimeter Wave Radar in Residual NeuralNetwork: Review of Performance of Multiple Large Model Parameters', Artificial Intelligence Technology Research, vol. 2, no. 8.

Zhao, CH, Leong, WY 2025b, 'Optimization solutions and simple innovative solution research on ResNet50 model', ASM Science Journal, vol. 20, no. 1.

Zheng, L, Bie, Z, Sun, Y, Wang, J, Su, C, Wang, S & Tian, Q 2016, 'Mars: A video benchmark for large-scale person re-identification', in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Part VI 14, Springer International Publishing, pp. 868-884.