Comparative Study of Deep Learning Architectures for Thermal vs. Millimeter-Wave Concealed Weapon Detection

***Akinbohun, Sylvester., Arthur Isidaehomen. Edeoghon. Okosun Oduware**

Department of Computer Engineering, Faculty of Engineering, University of Benin, Benin City, Nigeria

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11070178

Received: 03 August 2026; Accepted: 08 August 2026; Published: 18 August 2026

ABSTRACT

CWD is becoming more and more essential for smart city security, while TIR and PMMW are emerging as promising sensing techniques that can be used in a privacy-respectful way to achieve the CWD objective. However, current studies focus on evaluating a single deep learning architecture with respect to a single sensor modality, raising the question of which architecture achieves the best performance on each sensor modality. This work aims to fill this gap by carrying out a systematic comparison among eight different deep learning architectures, a custom CNN, and seven transfer-learning architectures (namely, ResNet-18, ResNet-50, ResNet-152, DenseNet-121, DenseNet-201, and GoogleNet) in terms of their performance on balanced and curated sets of TIR and PMMW images, comprising respectively 686 samples for each category. The models were trained using the same training settings (Adam optimisation function, cross-entropy loss, learning rate 0.001, and batch size 32), while being evaluated with respect to accuracy, precision, recall, F1-score, inference latency, and size. The custom CNN was the only architecture to generalise well to both modalities, achieving 97.6% accuracy. On the other hand, TIR outperforms its nearest competitor on the same modality (ResNet-50, 78.5%) by 21.5 percentage points and also happens to be the most compact and computationally efficient network for this modality. This suggests that the choice of architecture for hidden weapon recognition should be modality-specific and not generic, and that specialized domain-trained architectures are worth considering for sensing modalities, which deviate significantly from natural images on which transfer learning architectures are pre-trained conventionally.

Keywords: concealed weapon detection, deep learning, transfer learning, thermal infrared imaging, passive millimetre-wave imaging

INTRODUCTION

However, due to the development of smart cities, the need has arisen for security measures that would be able to detect hidden weapons in real-time. Traditional screening techniques, such as metal detectors, walk-through X-ray scanning machines, and human-monitored closed-circuit television, are not very effective when it comes to detecting non-metallic objects and are known to generate false alarms. In addition to that, they do not have the ability to perform continuous pre-emptive surveillance that is required in smart city environments (Murugan et al., 2025). Thermal infrared (TIR) imaging and passive millimeter-wave (PMMW) imaging are suitable for privacy-preserving security measures, where TIR imaging works by detecting thermal radiation caused by a hidden object on clothes (Khor, Chen, Roberts, & Ciampa, 2024a).

Convolutional neural networks (CNNs) and transfer learning have resulted in remarkable advances in automated detection in both modalities (Bhardwaj et al., 2025; Dimanov et al., 2023; Muñoz et al., 2025). However, the scientific literature often focuses on a particular architecture in a particular modality, thus leaving two critical issues unanswered: first, whether a transfer learning pre-trained network or a specialized architecture for the domain achieves superior results for a particular modality; second, whether an architecture achieving high accuracy for one modality will do the same for another one (Cantero et al., 2022). This is crucially important since TIR and PMMW images vary significantly in resolution, level of noise, and degree of resemblance to the RGB images, on which the vast majority of pre-trained networks are trained.

This study addresses this gap through a systematic, controlled comparison of eight deep learning architectures, a custom CNN trained from scratch, and seven transfer-learning architectures, evaluated identically on balanced TIR and PMMW datasets. The rest of this paper is structured as follows: the Methodology section covers the preparation of datasets, architecture of models, and training setup; the Results section shows the performance comparison of the proposed approach on both modalities; the Discussion section analyses this comparison by considering the different statistical nature of the two modalities; and the Conclusion section presents practical conclusions of this work regarding architecture choice.

METHODOLOGY

Dataset Acquisition and Preprocessing

Two data sets have been collected from the secondary sources to facilitate a comparative and cross-modal study. The TIR data set has been collected from the public data set containing 358 images captured using a FLIR Breach PTQ136 thermal imaging camera of 568 × 351 pixels resolution, having images of people carrying both hidden and non-hidden pistols. The PMMW data set has been collected from a publicly available data repository having 1,216 images captured using a Ka-band (34 GHz) passive radiometer of 2.5–5 m standoff distance and a field of view of 40° × 22° showing an extreme class imbalance towards the object class.

Both data sets went through a similar four-stage pipeline to create a truly comparable basis for cross-modal comparisons: labelling all images as either weapon or no weapon; balancing and augmenting the images using stratified sampling on the PMMW data set and flipping and rotating the images of the TIR data set to have 686 balanced images per data set; resizing the images to a common size of 224 x 224 pixels; and dividing the images into training, testing, and validation data sets of 70%, 20%, and 10% respectively (480, 137, and 69).

Model Architectures

There were two approaches to modeling. First, a transfer learning approach using seven models pretrained on ImageNet (Deng et al., 2009) was tested: ResNet-18, ResNet-50, and ResNet-152 (He et al., 2016); DenseNet-121 and DenseNet-201 (Huang et al., 2017); and GoogleNet (Szegedy et al., 2015). In all cases, the classification layer was substituted for a two-class one, and then independently fine-tuned on TIR and PMMW datasets.

The second approach is to train our own CNN from scratch, based on the idea that features obtained in the specific sensor domain can be more generic than those transferred from the natural RGB image domain (especially for the low-resolution, noisy PMMW images). The model consists of four convolutional layers with 32, 64, 128, and 256 filters, respectively (3×3 kernels, ReLU activation, 2×2 max pooling), after which there is flattening, then a fully connected layer with 512 units and dropout (Srivastava et al., 2014), and finally a two-unit output layer using SoftMax activation function.

Training Configuration

All eight models had the same baseline setup for training; this comprised Adam optimiser (Kingma & Ba, 2015), cross-entropy loss, learning rate of 0.001, and batch size of 32. The optimality of the above parameters was tested using a grid search for learning rate, batch size, dropout, weight decay, and optimiser, resulting in the conclusion that the above were the best parameters.

Evaluation Metrics

The models were assessed on the withheld testing data set by evaluating accuracy, precision, recall, and F1-score, following the recommendation that performance of a binary classifier be presented by more than one metric rather than just accuracy (Chicco & Jurman, 2020). The training duration, inference speed, and model size were also considered in order to assess the deployment feasibility, in agreement with the recommendation that detection accuracy be evaluated alongside the computational efficiency (Cantero et al., 2022).

RESULTS

Test-set results from all eight architecture models are tabulated in Table 1. All pre-trained transfer learning models performed nearly perfectly in TIR images, where ResNet-18, GoogleNet, DenseNet-121, and ResNet-152 attained 100.00% accuracy and an F1-score of 1.000. However, all four models exhibited poor performance in PMMW images, where GoogleNet, DenseNet-121, and DenseNet-201 performed at 49.8-50.2% accuracy, which was almost chance-level accuracy for a two-class problem, while ResNet-18 and ResNet-50 performed at 62.4% and 78.5%, respectively. The custom CNN architecture was the only model that performed well in both classes of images.

Table 1: Comprehensive Model Performance by Modality

Model	Modality	Accuracy	Precision	Recall	F1-Score
ResNet-18	TIR	100.00%	1.000	1.000	1.000
ResNet-18	PMMW	62.44%	0.785	0.624	0.565
GoogleNet	TIR	100.00%	1.000	1.000	1.000
GoogleNet	PMMW	50.24%	0.251	0.502	0.334
DenseNet-121	TIR	100.00%	1.000	1.000	1.000
DenseNet-121	PMMW	49.76%	0.249	0.498	0.332
DenseNet-201	PMMW	50.24%	0.251	0.502	0.334
ResNet-50	PMMW	78.54%	0.816	0.785	0.780
ResNet-152	TIR	100.00%	1.000	1.000	1.000
Custom CNN	TIR	97.57%	0.979	0.976	0.956
Custom CNN	**PMMW**	**96.60%**	**0.989**	**0.966**	**0.967**

Figure 1 presents PMMW performance in the six models considered here. The custom CNN is better than the best pre-trained model, ResNet-50, by 18.1 percentage points and better than the median pre-trained model by about 34 percentage points.

Figure 1: PMMW test accuracy by architecture. The custom CNN (orange) substantially outperforms all pre-trained transfer-learning alternatives.

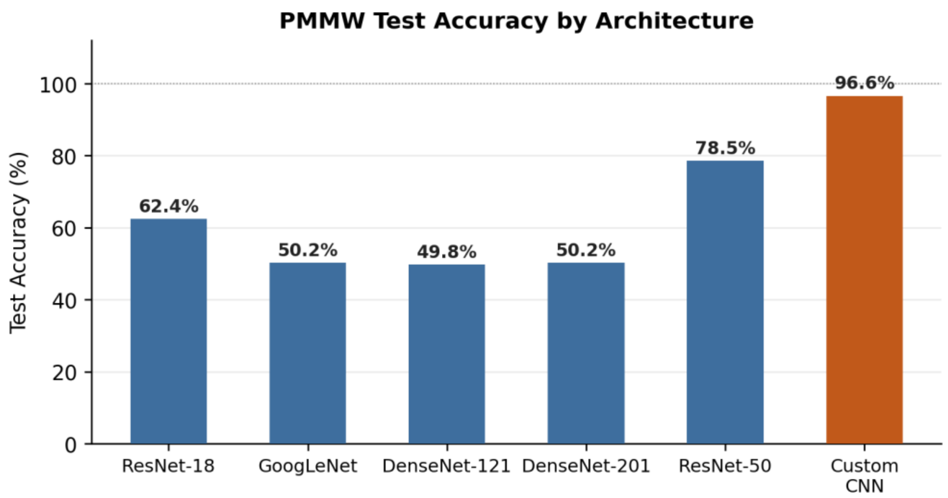

Table 2 provides performance results for inference of the best-performing model in each modality, after containerized deployment on local edge hardware. The custom CNN is the most accurate, smallest, and fastest performing model in the PMMW modality.

Table 2: Measured Edge Inference Performance (Best Model per Modality)

Model	**Modality**	**Inference Time (ms)**	**Model Size (MB)**
GoogLeNet	TIR	23.4	45.2
Custom CNN	PMMW	18.7	12.8
ResNet-50	PMMW (alt.)	41.2	97.8

DISCUSSION

The difference in Table 1 can be attributed to the different nature of the statistics of the two imaging techniques. TIR images, even though different from RGB images, share enough texture and contrast information with ImageNet statistics for transferred filters to still retain some meaning, which has been demonstrated previously by successful transfer learning results on TIR images containing concealed weapons (Khor et al., 2024a; Muñoz et al., 2025; Veranyurt & Sakar, 2023). PMMW images, which are relatively low resolution and noisier (Pang et al., 2020; Wang et al., 2019) than the other images, are farther away from the ImageNet distribution. Therefore, ImageNet biases embedded into the filters do not transfer well into the new problem domain. The CNN tailored for PMMW images was able to learn appropriate features.

Another observation is that the ResNet-50 model performed significantly better than other pre-trained models when used for PMMW (78.5% compared to 49.8-62.4%). This implies a non-linear association between network depth and PMMW performance, which cannot be attributed solely to the family of architectures (since ResNet-18 and ResNet-152 performed poorly compared to the mid-depth ResNet-50). This calls for further study, but not within the scope of this current study, and indicates that model capacity might play a more critical role than architecture family for PMMW transfer learning.

Indeed, on a practical note, this finding reveals that efficiency and accuracy do not inherently have to come at each other's expense: in PMMW classification, the custom-designed CNN was both the most efficient and the most accurate (Table 2), which is an optimal but not an inevitable result and needs to be checked in other sensing modalities. Another consequence of the comparison directly relates to design: instead of choosing one architecture for a multi-modal concealed weapon detection system, architecture selection must take place on a modality-by-modality basis, as the best-performing model in TIR (pre-trained GoogleNet) differs substantially, not just parameter-wise, from the best-performing model in PMMW (custom-made CNN).

CONCLUSION

This work presents a comparative analysis of eight deep learning architectures, one custom CNN, and seven transfer-learning networks on balanced TIR and PMMW concealed weapon detection datasets using the same training and testing protocol. Pre-trained transfer-learning architectures performed near perfection on the TIR dataset but failed catastrophically on the PMMW dataset, while the custom CNN is the only network that performed well on both datasets and performed significantly better than the best transfer-learning approach on the PMMW dataset by 18.1% while being the smallest and fastest model in the set. The results of this work validate two concrete suggestions: first, that the choice of architecture should be modality-dependent when it comes to concealed weapon detection, and second, that a domain-custom architecture can be seriously considered for sensing modalities that are different from natural images. This research work needs to be extended further by analysing other architectures and other categories of concealed threats aside from the handgun, and the investigation of a principled way of combining the best modality-specific architectures into a single detection pipeline.

ACKNOWLEDGEMENTS

In appreciation of the academic supervision provided by Dr. I.A. Edeoghon of the Department of Computer Engineering, University of Benin, for the Ph.D. research work from which this study emanates.

REFERENCES

1. Bhardwaj, D., Ramamoorthy, A., & Goyal, P. (2025). DEF-YOLO: Leveraging YOLO for concealed weapon detection in thermal imaging. arXiv. https://arxiv.org/abs/2510.13326
2. Cantero, D., Esnaola-González, I., Miguel-Alonso, J., & Jauregi, E. (2022). Benchmarking object detection deep learning models in embedded devices. Sensors, 22(11), Article 4205. https://doi.org/10.3390/s22114205
3. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over the F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, Article 6. https://doi.org/10.1186/s12864-019-6413-7
4. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848
5. Dimanov, D., Singleton, C., Rostami, S., & Balaguer-Ballester, E. (2023). MEOW: Automatic evolutionary multi-objective concealed weapon detection. Proceedings of the Second International Conference on Automated Machine Learning, 224. https://proceedings.mlr.press/v224/dimanov23a.html
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 770–778. https://doi.org/10.1109/CVPR.2016.90
7. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708. https://doi.org/10.1109/CVPR.2017.243
8. Khor, W., Chen, Y. K., Roberts, M., & Ciampa, F. (2024a). Automated detection and classification of concealed objects using infrared thermography and convolutional neural networks. Scientific Reports, 14, Article 8353. https://doi.org/10.1038/s41598-024-56636-8
9. Khor, W., Chen, Y. K., Roberts, M., & Ciampa, F. (2024b). Non-contact, portable, and stand-off infrared thermal imager for security scanning applications. AIP Advances, 14(4), Article 045314. https://doi.org/10.1063/5.0188862
10. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations. https://arxiv.org/abs/1412.6980
11. Muñoz, J. D., Ruiz-Santaquiteria, J., Deniz, O., & Bueno, G. (2025). Concealed weapon detection using thermal cameras. Journal of Imaging, 11(3), Article 72. https://doi.org/10.3390/jimaging11030072
12. Murugan, T., Badusha, N. A. N. M., Semaihi, A. R. O. A., Alkindi, M. M. R., Alnaqbi, E. M. R., & Alketbi, G. H. T. (2025). AI-based weapon detection for security surveillance: Recent research advances (2016–2025). Electronics, 14(23), Article 4609. https://doi.org/10.3390/electronics14234609
13. Pang, L., Liu, H., Chen, Y., & Miao, J. (2020). Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm. Sensors, 20(6), Article 1678. https://doi.org/10.3390/s20061678
14. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.
15. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, 1–9. https://doi.org/10.1109/CVPR.2015.7298594
16. Veranyurt, O., & Sakar, C. O. (2023). Concealed pistol detection from thermal images with deep neural networks. Multimedia Tools and Applications, 82, 44259–44275. https://doi.org/10.1007/s11042-023-15358-1

17. Wang, Z., Chang, T., & Cui, H.-L. (2019). Review of active millimeter wave imaging techniques for personnel security screening. IEEE Access, 7, 148336–148350. https://doi.org/10.1109/ACCESS.2019.2946736