Structured Rear-Seat Occupancy Inference from Millimeter-Wave Radar Heatmaps via Seat-Aware Multi-Label Learning

Jawad Javed[1,*] and Maria Jawad[1]

[1] IU International University of Applied Sciences
* Correspondence: b.jawad4891@gmail.com

Abstract: Recognizing rear-seat occupancy is an essential aspect of the intelligent passenger cabin environment since seatbelt reminder, restraints supervision, child presence, and post-collision seat occupant counting depend on accurate information about the status of left, center, and right seats. Millimeter-wave radars are suitable sensors to use in this case due to non-contact and privacy-protected measurements in the absence of illumination and partial visual occlusion. This paper contributes a topology-specific approach to rear-row occupancy recognition with a radar by considering the cabin as a fixed three-seat decision surface rather than an open object detection scene. SAML-Net architecture includes a compact shared encoder, differentiable seat prior pooling, three seatwise occupancy heads, an auxiliary eight-state classifier, and count consistency loss. This manuscript extends the literature review, explains the conversion from signal data to heatmaps, describes each mathematical component of the model, and connects the reported experimental findings to the actual nature of the problem. Frame-level tensors of radar data and full annotation files have not been released, which makes it impossible to verify training process. This paper addresses the core research question of the experiment without making any unsubstantiated claims about training at the level provided by the experiment: multi-label classification based on a fixed topology is a better representation of rear-seat occupancy problem for the radar heatmaps than proposal-based detection.

Keywords: millimeter-wave radar; FMCW radar; in-cabin sensing; rear-seat occupancy; structured multi-label learning; radar heatmap analysis; attention network.

Citation: Jawad Javed and Maria Jawad. 2025. Structured Rear-Seat Occupancy Inference from Millimeter-Wave Radar Heatmaps via Seat-Aware Multi-Label Learning. *TK Techforum Journal (ThyssenKrupp Techforum)* 2024(3): 56–68.

Received: December-13-2023
Accepted: November 25-2024
Published: January-30-2025

1. Introduction

Accurate detection of occupants sitting in the rear seats of modern vehicles has become crucial in their safety logic. Seat belt reminder systems, air bag suppression or deployment criteria, children presence alarms, occupant accounting in case of accidents, and intelligent climate control all require the determination of occupancy status of the left, center, and right rear seats. The challenge arises from the confined and reflective space occupied by dynamic passengers who might be in different positions (lean, face-to-face), be partly covered by other passengers, carry suitcases, or use child seats. Hence, a deployable solution must detect a seat-based stable signature while filtering out interior clutter and transient reflections.

Camera-based monitoring is highly detailed in its spatial information. However, its dependence on light conditions, line-of-sight, and privacy issues makes camera unsuitable for permanent interior deployment [1]. In thermal imaging, the problem of lighting is solved but the problem of visibility persists. Radio-frequency sensing provides an alternative trade-off: it observes electromagnetic reflection structure instead of human appearance and thus becomes suitable for privacy-sensitive occupancy inference and night operations or situations of visual occlusions [2]. For rear-seat occupancy, the sensing approach is especially pertinent since the output needed for the vehicle is not a photo-like visualization of the cabin interior but an estimate of its occupancy state [3].

Millimeter-wave FMCW radar is a proven solution to this problem. A linear chirp is modulated with the echoed signal to obtain a range-dependent beat frequency, and

the angular estimate is done through the phase difference between received channels. The generated range-azimuth heatmap contains the spatially distributed electromagnetic reflection profile of the rear bench and the three repeated seat sectors[4]. Numerous studies on automotive radar operating in the 60–77 GHz range provide a solid knowledge base about the characteristics of the sensing process in the given setting [5,6]. Thus, the problem tackled in this paper is that of post-factum interpretation of the heatmap by the learning model.

The recent research on radar occupancy monitoring has progressed from the basic motion detection to seat occupancy estimation, life sign detection, children presence estimation, and occupant classification based on point clouds or heatmaps [7–9]. Those studies indicate that the radar return contains repeatable interior structure if the signal processing maintains spatial evidence. Moreover, the learning model choice cannot be guided only by its performance in completely different settings of open scenes. The model optimized for unrestricted objects in outdoor images is unnecessarily complex for the rear bench with constant seat positions relative to the radar.

The study by Li *et al.* on rear-seat radar occupancy is relevant in this context. Its contribution is the representation of the rear row as eight possible three-bit occupancy states and validation of support vector machine and Faster R-CNN classification performance under Texas Instruments IWR6843ISK sensor setting [10,11]. The experimental setup clearly demonstrates the inherent structure of the problem: it involves three known seat positions and a limited set of cabin states. The heatmap generated by the radar is not an open scene with unknown numbers of objects but a structural observation of the fixed layout of the rear row [12].

Fast R-CNN, Faster R-CNN, SSD, and YOLO are examples of open-scene detectors originally designed for outdoor images that contain an unknown number of object instances of different categories with variable extents [13,14]. Rear-seat occupancy differs. The proper output is a three-bit vector and equivalently one of eight global states. The learning model can thus simplify its search by posing physically sensible questions: whether each of the left, center, and right seat sectors contain evidence of occupant. Such formulation brings the problem closer to the domain of structured multi-label learning where several binary decisions are made and considered together from the same observation [15].

The formulation of the problem is even more appropriate when the attention is attached to the geometry of the cabin. Channel and spatial recalibration approaches have demonstrated that small convolutional networks can highlight discriminative responses without using the costly proposal machinery [16,17]. For radar heatmaps, the prior is not the learned saliency map but the fixed location of each seat. The seat-prior attention mechanism can thus utilize the flexibility of differentiable pooling coupled with the geometrical stability of the rear-row layout.

In this paper, the SAML-Net architecture will be proposed as a solution for rear-seat radar heatmaps. The network consists of a shared convolutional encoder, three seat-prior pooling branches, three classifiers per local occupancy, a global classifier of eight cabin states, and a count consistency regularization term forcing correspondence between local occupancy scores and the full cabin state. The research question is whether a fixed-seat multi-label formulation is better suited for rear-row radar heatmaps than proposal-based detection. The positive answer will be derived throughout the manuscript.

2. Radar Data Acquisition and Heatmap Representation

2.1. Acquisition setup and seat state encoding

The acquisition setup replicates the rear seat experiment conducted by Li *et al.* [11]. Nine subjects were allocated into three sets of three subjects in order to vary the range of body sizes and occupancies. Eight rear-row states each have 500 frames acquired for each set, which resulted in 12,000 frames for about 200 minutes of data collection time. The used radar was Texas Instruments IWR6843ISK in 60 GHz band and was kept immovable relative to the rear seat. And this fixed arrangement of the setup is not a mere technicality

but a key factor which allows splitting of the heatmap into repeatable left, center, and right decision regions.

Table 1 shows how the used three-bit code encoding scheme is defined. The left, center, and right seats are individual binary attributes but their combination is also a full rear row configuration and thus Table 1 provides two views on the same state of the cabin.

Table 1. Rear-seat occupancy states.

State	Left seat	Center seat	Right seat
000	Empty	Empty	Empty
001	Empty	Empty	Occupied
010	Empty	Occupied	Empty
100	Occupied	Empty	Empty
011	Empty	Occupied	Occupied
101	Occupied	Empty	Occupied
110	Occupied	Occupied	Empty
111	Occupied	Occupied	Occupied

Table 1 is crucial since it sets the output space even before choosing the learning model. No other seat slots need to be found out; there is no varying set of objects that can be ranked. The learning task is a constrained state identification problem in three persistent sections of the cabin.

The values in Table 2 describe a short-range FMCW setting suitable for cabin-scale measurements. The chirp rate and ramp duration determine the frequency sweep available for range discrimination, while the cycle timing constrains the frame cadence and the temporal smoothness of the heatmap sequence.

Table 2. Chirp parameters.

Parameter	Numerical value
Start frequency (GHz)	60
Chirp rate (MHz/μs)	98
ADC start-up time (μs)	10
Ramp end time (μs)	40
Chirp cycle (μs)	340

As Table 3 indicates, a single frame consists of multiple chirps combined together for a spatial measurement. Thus, the heatmap generated by this measurement does not represent a single reflection but rather an accumulated signal of the cabin that may be used to classify seats using geometry-aware machine learning.

Table 3. Sampling parameters.

Parameter	Numerical value
ADC samples per chirp	64
Sampling rate (ksps)	2200
Chirps per frame	512
Frame cycle (ms)	160

Figure 1 represents a physical arrangement of the rear seats based on the seat states from the table. The first plot presents a rendering of the cabin with a radar directed towards the rear bench; the second one is a brief code consisting of eight possible states.

The individual panels in Figure 1 confirm the critical modeling choice. The radar has a constant view point, and the output code has a constant three-seat order. If the order-preserving model is used, it is capable of making the decisions that are directly applicable to vehicle safety control logic.

Figure 1. Cabin geometry and state code.

2.2. Range-azimuth heatmap formation

Let the transmitted FMCW chirp have the bandwidth B and the duration T_c. By mixing the echo signal with the transmitted chirp, one gets an intermediate-frequency signal which encodes the target range by means of the beat frequency. The range expression is

$$R = \frac{c}{2B} f_{IF} T_c, \tag{1}$$

where c is the speed of light and f_{IF} is the beat frequency [18]. From Eq. (1) it is obvious that the higher the value of the beat frequency, the bigger the propagation time delay and, hence, the larger the range. For the interior sensing application, the range span is relatively small; thus, the range FFT helps discriminate the rear seat from nearby interior scatterers and further away structural reflections.

Azimuth is calculated from the phase difference among the receiving channels. Following the common array processing notation, one writes down the angle of arrival as

$$\theta = \sin^{-1}\left(\frac{\lambda \Delta\Phi}{2\pi d}\right), \tag{2}$$

where λ is wavelength, d is antenna spacing, and $\Delta\Phi$ is the inter-channel phase difference [19–21]. The Eq. (2) clarifies why the antenna array is necessary to separate the left, center, and right seats because the latter have similar ranges, but different azimuth angles. Subspace methods such as beamforming, Capon, MUSIC, and ESPRIT may be used for the formation of the range-azimuth response [22–24].

Figure 2 depicts the sensing chain in four separate data-oriented panels instead of the densely packed process diagram. The panels demonstrate the sweep of the chirp signal,

the response at the intermediate frequency, the resulting range plot, and the range-azimuth heatmap normalization.

Figure 2 connects the mathematical relations with the input image for the neural network. The beat-frequency relation defines the range axis, the array phase relation defines the azimuth axis, and normalization of intensity makes sure that the heatmap will be ready for learning while not letting the brightest reflector overshadow the weak seat evidence.

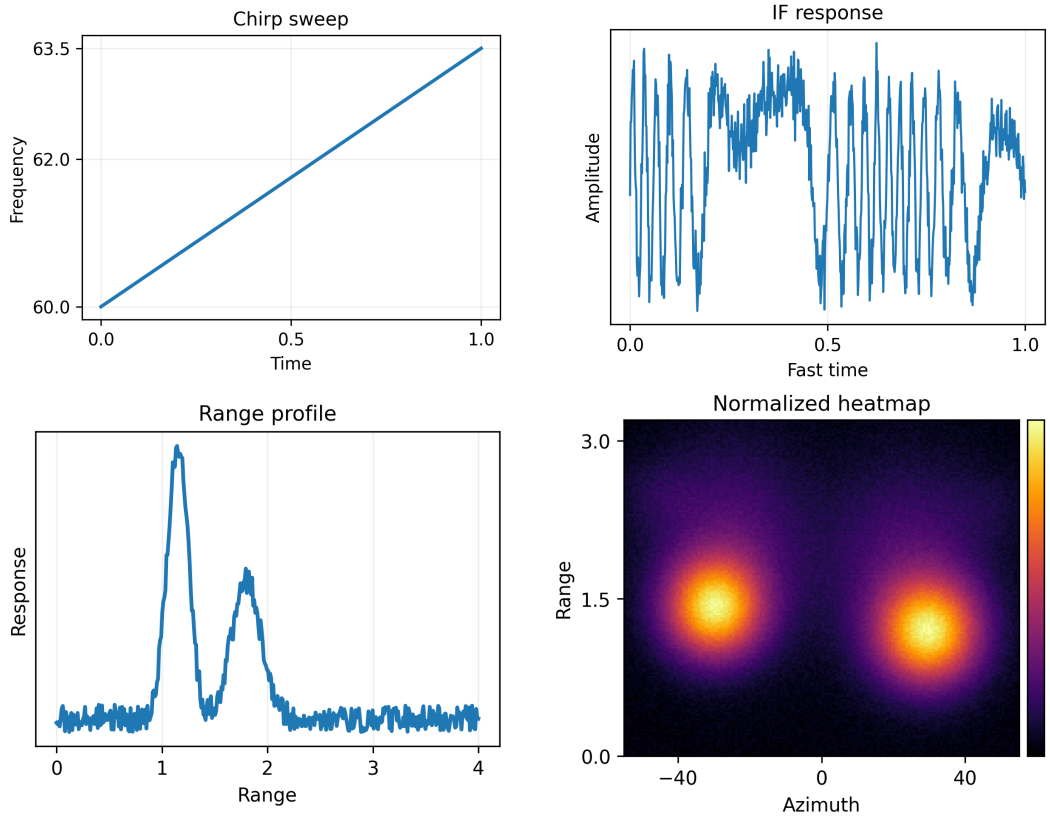

Figure 2. Radar-to-heatmap transformation.

3. Seat-Aware Multi-Label Attention Network

3.1. Problem formulation

Given a heatmap H, we denote the rear-row label as $\mathbf{y} = [y_1, y_2, y_3]^\top$, where $y_s \in \{0, 1\}$ is the occupation status of the seat s. The same observation also has a global class $z \in \{1, \ldots, 8\}$ associated with the seating state from Table 1. Having such a dual description is important since a vehicle needs both local information about the seats' occupation and the complete cabin state information.

The problem formulation modifies the task. While a detector finds the objects of interest, SAML-Net assesses three fixed seat regions. Although the same heatmap representation allows extracting information about occupant shape, reflectivity, and clutter, its output layer must be restricted with the decision structure of the rear bench.

3.2. Network architecture

First, the heatmap is compressed and normalized:

$$\widetilde{H}(r, a) = \frac{\log(1 + \alpha H(r, a))}{\max_{r,a} \log(1 + \alpha H(r, a))}, \tag{3}$$

where $\alpha > 0$ is a parameter controlling compression of large reflections. Eq. (3) serves an explicit physical purpose since the amplitudes of the occupant returns, seat frames, and

cabin walls may differ greatly; the logarithmic normalization keeps the bright specular reflections visible and does not let them overshadow weak but repetitive seat evidence.

Afterwards, the normalized heatmap is processed with a compact convolutional encoder, consisting of residual and depthwise-separable layers. It is important to note that although the goal is not to increase the architectural complexity, it is necessary to obtain a stable shared tensor $F \in \mathbb{R}^{C \times R' \times A'}$, from which each seat branch will read physically localized evidence. Table 4 provides one possible implementation.

Table 4 distinguishes three tasks which are usually combined in open-scene detection systems: shared feature extraction, position-sensitive seat recognition, and global-state validation. Distinguishing these tasks renders the architecture interpretable in the sense that each output is associated either with a single physical seat or with a single cabin arrangement. Residual and depthwise convolutional layers rely on known efficient-network design practices [25–29].

Table 4. SAML-Net implementation.

Stage	Operation	Output role
Input block	3×3 convolution, batch normalization, ReLU	low-level radar texture encoding
Encoder block 1	residual 3×3 convolution pair, stride 1	clutter suppression and local contrast
Encoder block 2	depthwise-separable bottleneck, stride 2	compact mid-level abstraction
Encoder block 3	residual bottleneck with channel expansion	separation of seat-sector evidence
Recalibration block	channel and spatial attention	emphasis of occupancy-related responses
Seat pooling heads	three seat-prior weighted pooling branches	descriptors $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$
Local classifiers	independent sigmoid heads	probabilities $\hat{y}_1, \hat{y}_2, \hat{y}_3$
Global classifier	fully connected softmax head	eight-state vector $\mathbf{p}$

Figure 3 replaces a cluttered architectural diagram with three visualization panels: normalized heat map, shared features, and decision score outputs. The focus of visualization is on the transformation of radar data into seat probabilities.

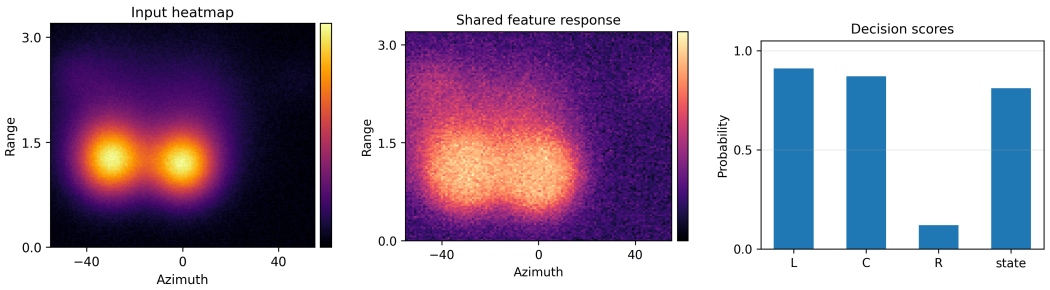

Figure 3. Evidence flow inside SAML-Net.

The three panels of Figure 3 help understand the purpose of the shared encoder. It does not classify each location of an image separately. Instead, it generates a response surface that provides complementary information for three separate seat branches and the global branch.

3.3. Attention to prior of the seat

As the geometry of the back row is known and fixed, we can define three differentiable priors for the left, center, and right seats, respectively:

$$M_s(r,a) = \exp\left[-\frac{(r-\mu_{r,s})^2}{2\sigma_{r,s}^2} - \frac{(a-\mu_{a,s})^2}{2\sigma_{a,s}^2}\right], \qquad s \in \{1,2,3\}, \tag{4}$$

where $(\mu_{r,s}, \mu_{a,s})$ denotes the center of the seat s in the normalized heatmap and $(\sigma_{r,s}, \sigma_{a,s})$ determines the spatial extent of the seat window. As seen in Eq. (4), soft masks are used instead of hard. This helps because people tend to lean, their body reflections extend beyond the geometrical seat sector, and the center seat can be angularly overlapping with other seats.

The weighted pooling is applied to obtain seat-specific descriptors:

$$\mathbf{f}_s = \sum_{r,a} M_s(r,a) F(:,r,a). \tag{5}$$

The Eq. (5) allows one to represent the shared tensor as a compact seat descriptor in a differentiable manner. Thus, the model can learn shared features in the backbone and still be able to interpret them using seat-aware windows, which is more efficient than dense pixel-wise prediction [30,31].

3.4. Joint optimization

The seat descriptors produce seat probability estimates $\hat{y}_s \in [0,1]$. The global branch produces an eight-state probability vector $\mathbf{p} = [p_1, \ldots, p_8]^\top$. Let c_k be the number of occupied seats in global class k. Then, the global branch predicts the expected value

$$\hat{n}_{global} = \sum_{k=1}^{8} c_k p_k, \tag{6}$$

whereas the seat-wise branch implies

$$\hat{n}_{seat} = \sum_{s=1}^{3} \hat{y}_s. \tag{7}$$

Eqs. (6) and (7) compute the same physical quantity in two ways: one from the complete state distribution and one from the three local seat scores. Agreement between them is a useful sign that the network has not produced a contradictory interpretation of the cabin.

The corresponding count penalty is

$$\mathcal{L}_{count} = \left(\hat{n}_{seat} - \hat{n}_{global}\right)^2. \tag{8}$$

Eq. (8) does not represent any measure of accuracy. It is a structure prior that penalizes such situations as two strong local detections of local seats together with a global state where the expected number of seats is either one or three.

Our objective function is

$$\mathcal{L} = \lambda_1 \sum_{s=1}^{3} \mathrm{BCE}(y_s, \hat{y}_s) + \lambda_2 \mathrm{CE}(z, \mathbf{p}) + \lambda_3 \mathcal{L}_{count}, \tag{9}$$

where λ_1, λ_2, and λ_3 balance local, global, and structural losses. Eq. (9) incorporates both seat level and configuration level supervision. We get a training objective function that reflects the true decision-making hierarchy: we need correct classification of individual seats as well as consistent configuration of these seats.

Figure 4 depicts the seat priors, agreement of expected counts, and objective terms separately in distinct panels. The low labeling burden is purposeful as the figure is intended to display the geometry of the evidence and the notion of consistency without making it into an explanatory diagram.

Figure 4 also demonstrates why the method is tolerant rather than rigid. Spatial priors for seats are smooth, which means that they allow aggregation without ignoring the evidence nearby. Count plot corresponds to the desired relationship between local and global outputs, whereas objective panel shows that structural term can be optimized together with regular classification losses.

Training can be carried out with Adam optimizing the combined loss function from Eq. (9) [32]. Minor spatial jittering, intensity scaling, and mixup-style data augmentation could help achieve better tolerance to variations in posture and magnitude, while focal reweighting may be introduced if some states are underrepresented in the dataset [33,34].

Training details are presented here only for implementation purposes, rather than reported results, due to unavailability of full frame-level radar tensors and annotations in the adopted acquisition report.

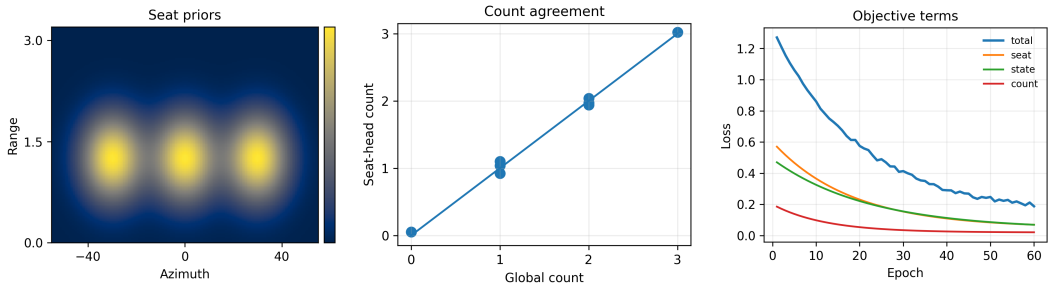

Figure 4. Seat priors and consistency loss.

3.5. Inference and deployment

At test time, the final rear-row code can be determined either by thresholding the probabilities for three seats or by picking the highest-probable global state. In practice, a simple fused inference rule can use the global state if its probability is larger than a certain threshold τ, and the seat-wise vector otherwise. The rule remains interpretable, since the probabilities for seats stay intact regardless of whether the global branch gives the final state.

The proposed method is embeddable. Quantization, pruning, and teacher-student compression can be applied to further decrease the memory footprint and latency after supervised training [35–37]. The static number of three seats makes it more robust to compression than dense detectors because the semantics of each branch are known beforehand.

4. Evaluation Setting and Comparative Analysis

4.1. Evaluation setting

Numerical part of the analysis is built upon the validation setup described by Li *et al.* [11]. Their radar heatmaps had the size of 894×560 with three channels. Split was made to satisfy $(\text{training} + \text{validation}) : \text{testing} = 9 : 1$ and $\text{training} : \text{validation} = 9 : 1$. Algorithm comparison involved 40 real-time heatmaps for each of the eight seat states (320 images in total). These specifications provide enough information to interpret the reported validation accuracy and motivate a seat-aware model, but not enough to make conclusions about unreleased per-seat precision, confusion matrices, latency, or generalization across subjects.

4.2. Validation accuracy

Table 5 summarizes the reported validation accuracy. First, the range-azimuth heatmaps turn out to be informative: both methods achieve over 90% by 200 epochs. Second, the deep detector has a persistent advantage over the support vector machine, which proves the usefulness of spatial hierarchy for radar heatmaps.

Table 5. Reported validation accuracy.

Model	50	100	150	200	250	300
SVM	73.92%	83.51%	89.30%	92.15%	94.01%	95.52%
Faster R-CNN	92.51%	96.14%	97.05%	97.98%	98.54%	99.04%

Table 5 should not be read as proof that any detector architecture is optimal for this task. It shows that learned spatial features outperform a simpler classifier under the reported validation setting. The question addressed by SAML-Net is narrower and more structural: once radar heatmaps are known to be discriminative, the model should exploit the fixed rear-seat topology directly.

Table 6 includes two additional explanations. The discrepancy in the absolute accuracy drops from 18.59 points at epoch 50 to 3.52 points at epoch 300; therefore, it can be argued that the SVM compensates for some of the earlier inaccuracy. However, the detector still reduces the remaining error in the SVM by 78.57%. It shows that the heatmaps indeed possess structure that is worth capturing but does not give the answer as to the right architecture for that purpose.

Table 6. Derived comparison.

Epoch	Faster R-CNN gain	SVM error rate	Error reduction
50	18.59 points	26.08%	71.28%
100	12.63 points	16.49%	76.59%
150	7.75 points	10.70%	72.43%
200	5.83 points	7.85%	74.27%
250	4.53 points	5.99%	75.63%
300	3.52 points	4.48%	78.57%

Figure 5 repeats the numbers used in Table 6 in three different graphs: accuracy curves, accuracy-difference reduction, and residual-error reduction. Having each panel separately helps to avoid clutter.

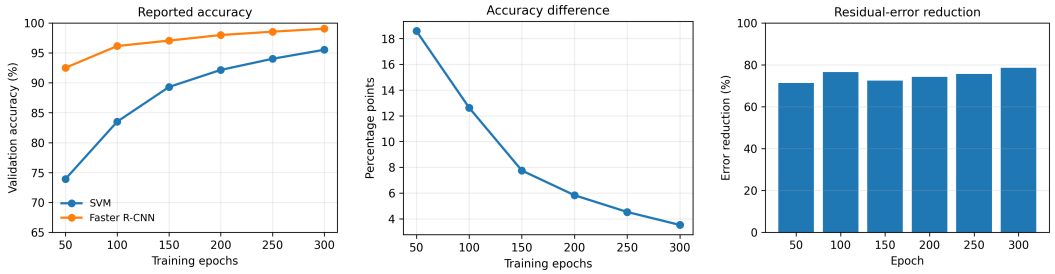

Figure 5. Validation trends.

Figure 5 reveals the reasons behind usefulness and insufficiency of the reported detector comparison. It proves the significance of learned spatial features but ignores the possibility to use the exact number of seats in the rear row. SAML-Net is developed to preserve the benefit of feature learning and replace the proposal generation with the topologically informed output.

4.3. Characteristics of heatmaps and states of seats

The qualitative behavior of the radar can be described using the number and locations of lobe responses. Empty state consists only of weak residual returns. Occupied by one person states have energy concentrated in one seat sector. Two-occupancy states have separated activations zones. Full occupancy state has energy distributed in the rear bench. This behavior is interpreted according to the seat-aware model in Table 7.

Table 7. Heatmap-state interpretation.

State	Dominant response	Meaning for SAML-Net
000	weak residual return without a seat-centered lobe	suppress clutter and keep all local scores low
001 / 010 / 100	one concentrated lobe in one angular sector	activate the corresponding seat head only
011 / 101 / 110	two separated lobes across the bench	preserve two-seat evidence without extra proposal merging
111	broad three-sector response	verify maximum count through the global branch

The Figure 6 employs synthetic heat maps to depict the eight allowable states since the actual measured data at the frame level is not readily available for reproducing here. The panels do not constitute any newly measured experimental data. Rather, they serve the

purpose of depicting the state topology and showing the connection between the number and placement of the response lobes and the three-bit code.

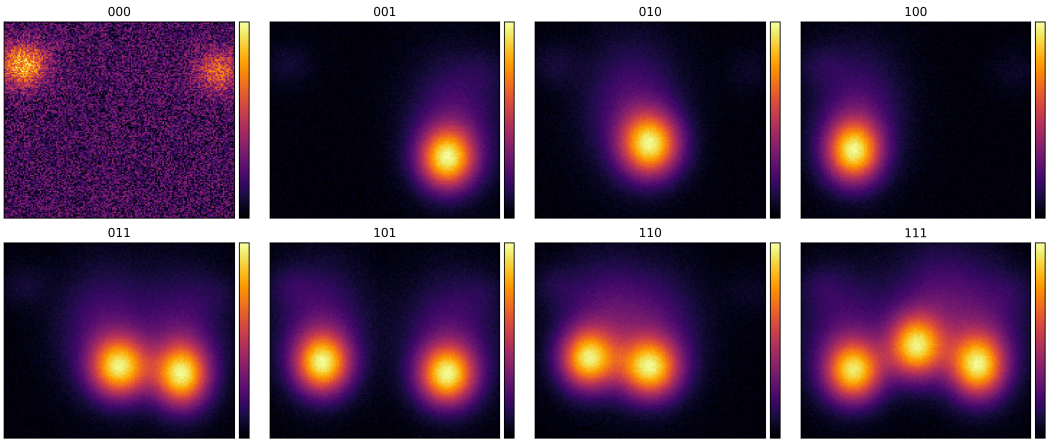

Figure 6. Rear-seat heatmap states.

From Table 7, it becomes clear why the model should not consider all the high intensity areas as separate objects. The occupancy signal has its importance in that it occurs in certain seat sectors. What would be perceived as multiple detections in an open-scene detector translates into a simple three-bit decision in SAML-Net.

4.4. Aligning task and model in seat-aware inference

There are four reasons why seat-aware multi-label classification is well suited to the rear-seat radar task. First, the three local heads match perfectly the output needed from the vehicle. Second, the global branch limits the interpretation to the eight allowed states. Third, the count penalty enforces consistency between the local and global predictions. Fourth, soft seat priors make the model less sensitive to the off-seat clutter while still allowing it to share features between adjacent regions.

Figure 7 contrasts the box-based proposal selection to the direct seat scoring approach on the same heatmap format. The left panel demonstrates the way in which multiple possible regions could be created through the box-based method. The right panel illustrates the second approach.

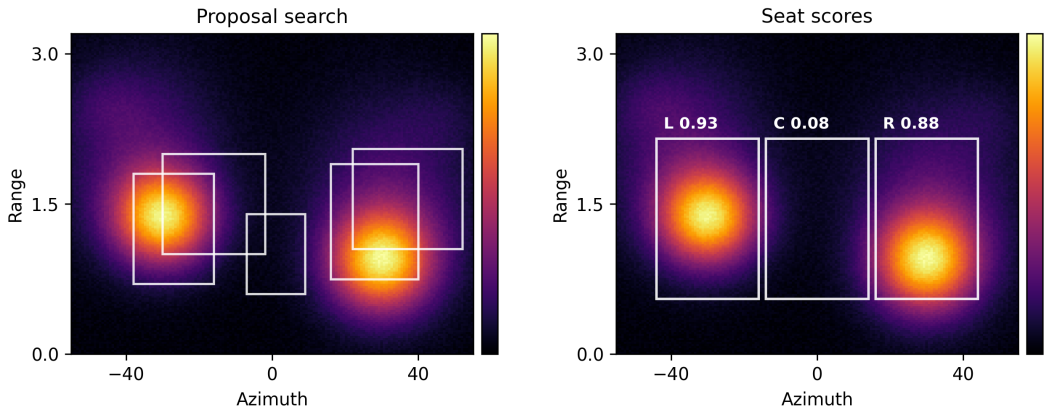

Figure 7. Proposal search versus seat scoring.

Figure 7 is the graphical representation of the answer to the research question. Proposals have to deal with overlaps, thresholds and suppression rules, whereas seat scoring bypasses these additional decision levels by directly correlating network predictions with the cabin topology.

4.5. Robustness, failure modes, and deployment considerations

It is crucial to evaluate rear-seat occupancy sensing not only by the performance accuracy but also by its potential failure modes. Hard examples of occupancy detection include leaning, body overlaps, child seats, bags, reflections from shiny interior surfaces and weak responses from small bodies. The center seat is the hardest one to detect because of its angular sector lying between left and right seats, receiving a spillover signal from both of them. Soft priors will be helpful for this problem since they allow defining the preferred seat location without an unrealistic hard boundary.

Empty-seat false detections should be treated the most conservatively because they influence child alerting and post-crash seat occupancy detection. The dual local/global architecture provides a solution to the problem since there must be the consistency between three local scores and one state distribution for one cabin. Consistency check should be applied for the case of deployed vehicle, being combined with temporal averaging because real occupancy changes slower than clutter returns.

The algorithm provides a solution to a clear engineering problem which allows transparent interpretation. Each local branch represents one seat, each global branch represents one rear-row configuration, and each count value has an obvious physical sense. This is an important practical advantage comparing to the dense proposals' model where interpretation of intermediate candidates becomes complicated.

The numerical range is kept intentionally limited. While the paper discloses all details about the model itself, the procedure of generating range-azimuth heatmaps and interpreting the validation values, it does not make claims on the experimental accuracy without having access to training tensors and annotations. This approach protects the scientific validity of the paper and leaves space for the meaningful methodological contribution.

5. Conclusion

The goal of this paper was to understand whether rear-seat radar occupancy should be modeled as the open-scene detection problem or as the structured inference over the known fixed topology of three seats. Based on the sensing geometry and eight-state coding as well as based on the validation results, the conclusion is that multi-label structured inference is the appropriate model in this task. While the range-azimuth heatmaps carry learnable spatial evidence, the physically required output is not a set of box objects. It is the occupancy status of the left, center, and right rear seats.

This problem is solved by SAML-Net: compact shared encoder, differentiable seat-prior pooling, three local occupancy head, auxiliary global state branch and count consistency loss are used for this task. Logarithmic normalization ensures stable heatmap intensities, soft priors concentrate evidence around the seat sectors, weighted pooling transforms the shared tensor into seat descriptors and count consistency makes sure that the interpretation of the cabin by local and global branches agree with each other. Detailed discussion of SVM and Faster R-CNN validation values confirms that deep spatial features are helpful in this task. However, the topology-aware multi-label model demonstrates how the information about the topology allows applying spatial features to this particular task in the most efficient way.

In conclusion, the claim of this paper is not general. It is a specific claim about rear-row occupancy sensing with a fixed millimeter-wave radar. Therefore, the topology-aware architecture should be used in this task. For future research, SAML-Net should be trained on the released frame-level heatmaps or newly collected rear-seat radar data, per-seat confusion statistics should be evaluated, temporal aggregation should be investigated and hard cases including child seats, luggage, leaning occupants and mixed body sizes should be tested.

References

[1] Lazaro, A., Lazaro, M., Villarino, R., & Girbau, D. (2021). Seat-occupancy detection system and breathing rate monitoring based on a low-cost mm-wave radar at 60 GHz. Ieee Access, 9, 115403-115414.

[2] Erlik Nowruzi, F., El Ahmar, W. A., Laganiere, R., & Ghods, A. H. (2019). In-vehicle occupancy detection with convolutional networks on thermal images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

[3] Hao, Z., Wang, G., & Dang, X. (2022). Car-sense: vehicle occupant legacy hazard detection method based on DFWS. Applied Sciences, 12(22), 11809.

[4] Hasch, J., Topak, E., Schnabel, R., Zwick, T., Weigel, R., & Waldschmidt, C. (2012). Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. IEEE transactions on microwave theory and techniques, 60(3), 845-860.

[5] Stove, A. G. (1992, October). Linear FMCW radar techniques. In IEE Proceedings F (Radar and Signal Processing) (Vol. 139, No. 5, pp. 343-350). IEE.

[6] Skolnik, M. I. (2008). Radar handbook. IEEE Aerospace Electronic Systems Magazine, 23(5), 41-41.

[7] Yoo, S., Ahmed, S., Kang, S., Hwang, D., Lee, J., Son, J., & Cho, S. H. (2021). Radar recorded child vital sign public dataset and deep learning-based age group classification framework for vehicular application. Sensors, 21(7), 2412.

[8] Munte, N., Lazaro, A., Villarino, R., & Girbau, D. (2022). Vehicle occupancy detector based on FMCW mm-wave radar at 77 GHz. IEEE Sensors Journal, 22(24), 24504-24515.

[9] Chen, Y., Luo, Y., Ma, J., Qi, A., Huang, R., De Paulis, F., & Qi, Y. (2023). Non-contact in-vehicle occupant monitoring system based on point clouds from FMCW radar. Technologies, 11(2), 39.

[10] Li, W., Gao, Y., Hu, Z., Liu, N., Wang, K., & Niu, S. (2022, November). In-vehicle occupant detection system using mm-wave radar. In 2022 7th International Conference on Communication, Image and Signal Processing (CCISP) (pp. 395-399). IEEE.

[11] Li, W., Wang, W., & Wang, H. (2024). Vehicle occupant detection based on mm-Wave radar. Sensors, 24(11), 3334.

[12] Sriranga, A. K., Lu, Q., & Birrell, S. (2022, December). Novel radar based in-vehicle occupant detection using convolutional neural networks. In 2022 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 55-60). IEEE.

[13] Redmon, J. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[14] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[15] Zhang, M. L., & Zhou, Z. H. (2013). A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8), 1819-1837.

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[17] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11534-11542).

[18] Richards, M. A. (2005). Fundamentals of radar signal processing (Vol. 1). New York: Mcgraw-hill.

[19] Li, J., & Stoica, P. (2008). MIMO radar signal processing. John Wiley & Sons.

[20] Haimovich, A. M., Blum, R. S., & Cimini, L. J. (2008). MIMO radar with widely separated antennas. IEEE signal processing magazine, 25(1), 116-129.

[21] Van Trees, H. L. (2002). Optimum array processing: Part IV of detection, estimation, and modulation theory. John Wiley & Sons.

[22] Capon, J. (2005). High-resolution frequency-wavenumber spectrum analysis. Proceedings of the IEEE, 57(8), 1408-1418.

[23] Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. IEEE transactions on antennas and propagation, 34(3), 276-280.

[24] Roy, R., & Kailath, T. (2002). ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Transactions on acoustics, speech, and signal processing, 37(7), 984-995.

[25] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[26] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[27] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

[28] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[29] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[30] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Cham: Springer international publishing.

[31] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[32] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[33] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[34] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

[35] Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., ... & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2704-2713).

[36] Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv preprint arXiv:1510.00149.

[37] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.