SeeNN: Leveraging Multimodal Deep Learning for In-Flight Long-Range Atmospheric Visibility Estimation in Aviation Safety

Taha Bouhsine [*], Giuseppina Carannante.[†], Nidhal C. Bouaynaya.[‡]
Electrical and Computer Engineering Department, Henery M.Rowan College of Engineering, Rowan University, Glassboro, New Jersey, 08028

Soufiane Idbraim [§]
IRF-SIC Laboratory, Computer Science Department, Faculty of Sciences Agadir, Ibn Zohr University, Agadir, Morocco

Phuong Tran, Grant Morfit, Maggie Mayfield, Charles Cliff Johnson
William J. Hughes Technical Center, Federal Aviation Administration, Atlantic City, NJ, USA

Accurate, real-time estimation of atmospheric visibility is a critical yet challenging task in aviation safety. While deep learning has shown promise, unimodal approaches relying solely on RGB imagery often fail to capture the complexity of atmospheric conditions, leading to limitations in reliability and accuracy. This paper introduces SeeNN, a novel multimodal deep learning framework designed for robust, long-range, in-flight visibility estimation. SeeNN integrates information from five diverse modalities: RGB imagery, depth maps, normal surface maps, edge maps, and entropy maps. To facilitate the development and evaluation of such models, we also present SeeSet V1, a new, comprehensive, and publicly available benchmark dataset featuring a wide range of altitudes, land covers, and visibility conditions. Our extensive experiments demonstrate the superiority of the multimodal approach. The SeeNN framework achieves a classification accuracy of over 97%, a significant improvement upon the 87.92% accuracy of a baseline unimodal RGB model. This work underscores the substantial potential of multimodal fusion to enhance the reliability of automated visibility estimation systems, representing a key advancement toward improving safety and operational efficiency in aviation and other domains where visibility is a critical factor.

I. Introduction

Atmospheric visibility is a critical determinant of aviation safety, directly influencing a pilot's capacity for navigation and critical decision-making [1–6]. The tragic 2020 accident involving Kobe Bryant, which the National Transportation Safety Board (NTSB) attributed to the pilot's decision to continue flight under Visual Flight Rules (VFR) into Instrument Meteorological Conditions (IMC), starkly underscores the severe consequences of impaired visibility. This incident highlights the urgent and unmet need for accurate, real-time, in-flight visibility estimation technologies.

The development of automated visibility estimation systems for aviation is fraught with challenges. Pilots frequently rely on their familiarity with local landmarks and terrain, a dependency that complicates the creation of automated systems requiring broad geographical adaptability [7]. Moreover, the dynamic nature of atmospheric conditions, including fluctuating cloud cover and abrupt weather changes, necessitates solutions that are both robust and versatile.

While deep learning presents promising avenues for addressing complex problems, unimodal approaches, particularly those reliant on RGB imagery, have demonstrated significant limitations in the context of visibility estimation. These models are often susceptible to overfitting, exhibit poor generalization to new environments, and are affected by inherent biases in the training data. RGB data alone is frequently insufficient for capturing the nuanced characteristics of the atmosphere or for mitigating confounding factors such as glare, low-light conditions, or rapid meteorological shifts [8–16].

To overcome these obstacles, multimodal deep learning has emerged as a demonstrably superior paradigm. By integrating data from diverse sources, these techniques augment the capabilities of the model and address the intrinsic

[*] Graduate Research Fellow

[†] Postdoctoral Fellow

[‡] Associate Dean for Research & Graduate Studies and Professor of Electrical & Computer Engineering

[§] Computer Science Professor and Head of IRF-SIC Laboratory shortcomings of single-modality systems [17–20]. Each modality contributes unique information, fostering a more holistic and veridical perception of the environment, which in turn leads to more precise and reliable predictions. The value of multimodal deep learning in visibility estimation is increasingly recognized, as evidenced by a growing body of literature [9, 21–29]. As detailed in table 1, the fusion of multiple data streams enhances the robustness, safety, and reliability of deep learning systems, rendering them suitable for mission-critical real-world applications [30–32].

Table 1 Modalities Employed in On-Ground Atmospheric Visibility Estimation Literature

Modality	**[25]**	**[33]**	**[24]**	**[26]**	**[27]**	**[28]**	**[29]**
Depth Map			X				
Transmission Map	X		X		X		
Disparity Map	X						
Entropy						X	X
Edge Detection					X		
Contrast Computation					X		
Koschmieder Law					X		X
FFT		X					
Spectral Filter		X					
Dark Channel Prior				X	X		X

Despite the progress in ground-based visibility estimation, a significant gap persists in addressing in-flight scenarios. This deficiency is primarily attributable to the scarcity of comprehensive in-flight visibility datasets, which are essential for training and validating deep learning models in realistic aviation contexts [9]. Existing datasets are often constrained to short-range, ground-level visibility and lack the necessary diversity in scenery and land cover. This limitation severely impedes the development of universally applicable and robust in-flight visibility estimation models.

This paper introduces a multimodal framework for training visibility estimation systems, with the goal of improving the accuracy, trustworthiness, and robustness of deep learning models for atmospheric visibility assessment. We demonstrate that by integrating diverse data modalities, the limitations of unimodal RGB approaches can be substantially mitigated, thereby advancing the development of versatile and reliable deep learning applications in environmentally dynamic fields. Furthermore, we address the critical dataset gap by introducing a comprehensive, publicly available dataset that captures visibility degradation across a wide range of land covers and altitudes.

The primary contributions of this work are twofold:

- A meticulously curated dataset, SeeSet V1, for the benchmarking of visibility estimation, dehazing, and visibility restoration algorithms [34]. This dataset, which is publicly available at `https://github.com/skywolfmo/seeNN-paper`, was generated using the X-Plane 11 flight simulator. It includes a wide array of images captured under diverse visibility conditions and at various altitudes, from ground level to 2,000 feet Above Ground Level (AGL). Its comprehensiveness provides a robust foundation for the training and evaluation of advanced in-flight visibility estimation and restoration techniques.
- A multimodal fusion framework for atmospheric visibility estimation. This framework is employed to train and validate deep learning models, with results demonstrating the superior accuracy of the multimodal approach when compared to single-modality RGB models.

II. Methodology

This section details the methodology employed in this study. We introduce a novel framework that leverages multiple modalities for the development of atmospheric visibility estimation solutions. A key component of this work is the construction of a new dataset, SeeSet V1, which encompasses both ground-level and elevated altitude conditions, addressing a critical gap in existing resources.

A. SeeSet V1 Dataset

To address the limitations of existing datasets and to encompass a broader range of real-world operational scenarios, we have developed a novel aerial imagery dataset designated SeeSet V1. This dataset has been meticulously curated to include dynamic views from multiple locations, capturing scenery from both ground-based and aerial perspectives.

This section provides a comprehensive description of the data collection and labeling procedures (section II.A.1). In section II.A.2, we detail the techniques utilized to generate the supplementary image modalities.

1. Dataset Collection Process

The generation of our synthetic dataset was accomplished using an FAA-approved flight simulator. The use of this advanced simulator enabled the systematic and controlled acquisition of images, showcasing a diverse range of viewpoints and visibility degradation levels. The data collection process, as depicted in Figure 1, commenced at ground level. Visibility was incrementally increased in discrete steps, up to a maximum of 100 miles. Upon reaching this limit, the viewpoint's altitude was elevated, and the visibility was reset to zero. This iterative procedure was continued up to a maximum altitude of 2,000 feet Above Ground Level (AGL).

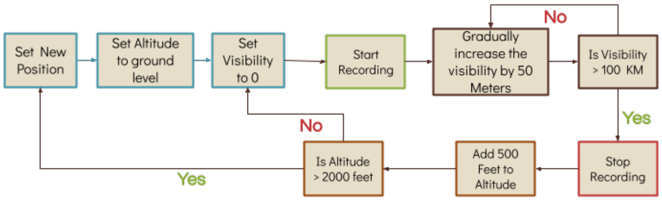


Fig. 1 Automatic Dataset Collection Process using X-Plane 11

The collected images are automatically labeled into five discrete bins, each tailored to specific FAA requirements. This categorization is based on visibility conditions and regulations relevant to both ground-based and aerial environments. The designated bins serve as the basis for the five labels utilized in training our DL models. We report the classes (bins) specifications and the corresponding counts in table 2.

Table 2 Visibility Categories and Images Count

Category	Visibility in miles	Visibility in meters	Count
4	≥ 5 miles	≥ 8046.72m	67002
3	3 to 5 miles	4828.03m to 8046.72m	19584
2	1 to 3 miles	1609.34m to 4828.03m	19648
1	0.5 to 1 mile	804.672m to 1609.34m	4928
0	≤ 0.5 mile	≤ 804.672m	4938
Total			116100

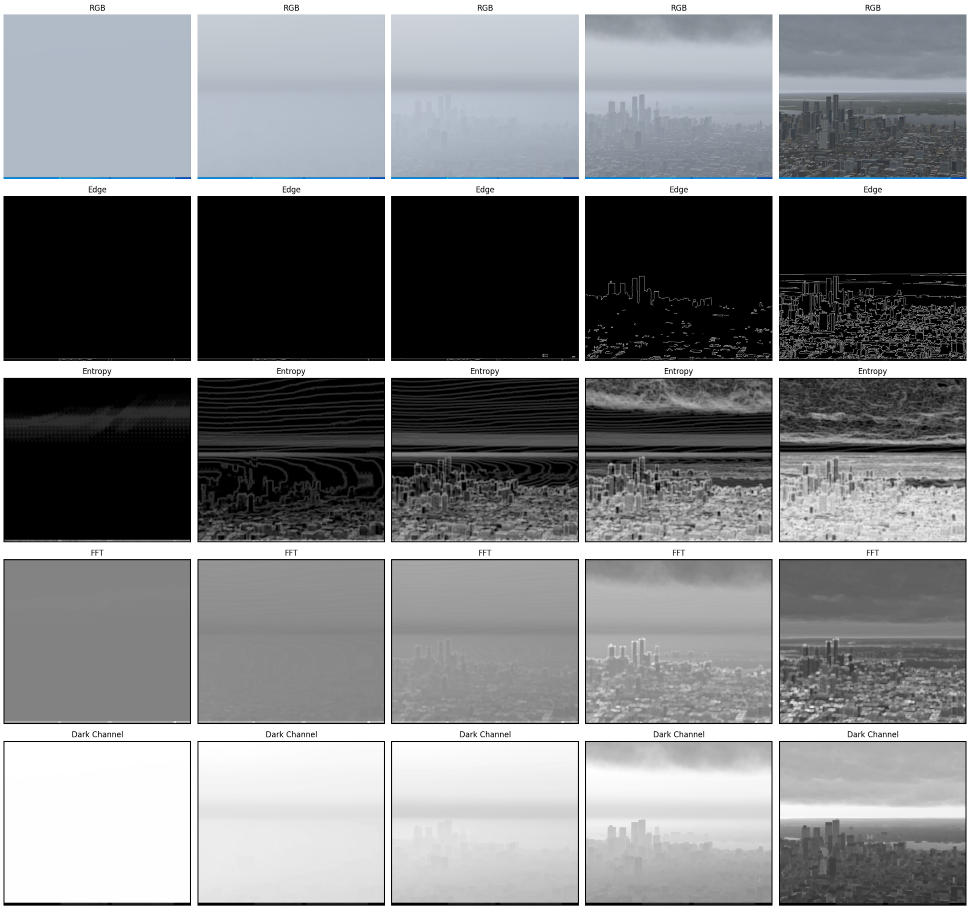
2. Modalities

Monocular Depth Estimation:

Monocular depth maps were extracted using the Omnidata toolkit [35, 36]. This toolkit provides a scalable and comprehensive method for depth estimation, which is essential for understanding the spatial arrangement of a scene. The resulting depth maps furnish a pixel-wise measurement of distance from the viewpoint, thereby facilitating an accurate representation of the three-dimensional scene structure.

It is important to note a specific limitation of the depth estimation models employed. The training methodology for these models involves masking the sky and exclusively considering the ground for depth estimation. This may present challenges for certain images within our dataset that were captured at varying altitudes.

Normal Surface Estimation:



(a) < 0.5 mile **(b) (0.5, 1] miles** **(c) (1, 3] miles** **(d) (3, 5] miles** **(e) > 5 miles**

Fig. 2 The impact of visibility on the multiple modalities for the 6N7 Sealane 01 View. Each row shows one modality: RGB, edge map, entropy map, FFT magnitude, and dark channel prior. Each column refers to a visibility bin.

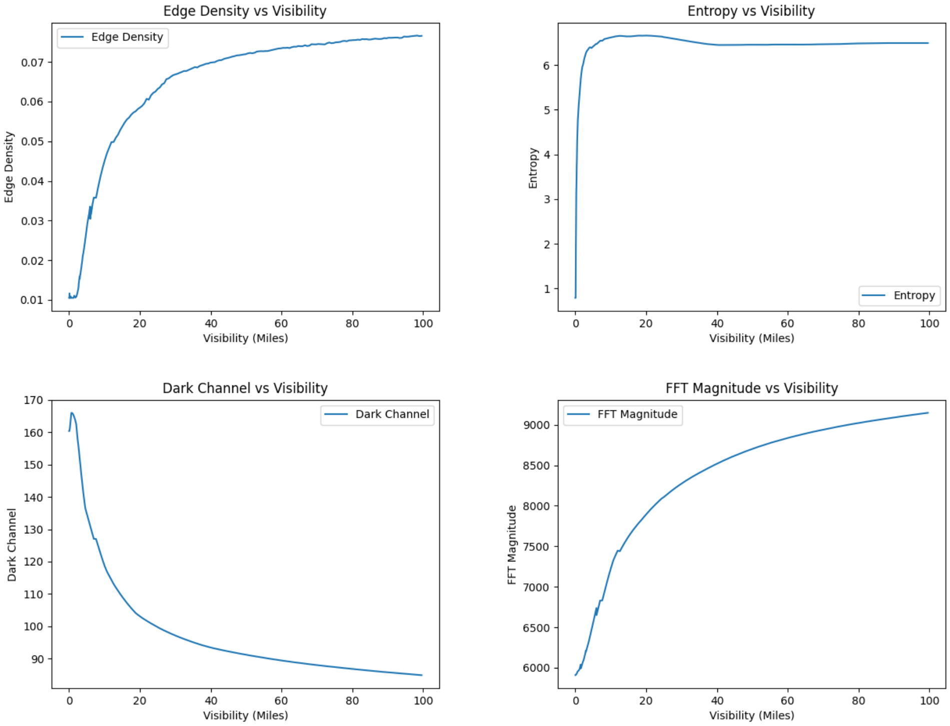


Fig. 3 Impact of Visibility Degradation on Edge Density (a), Entropy map (b), Dark Channel Prior (c), and FFT Magnitude (d) vs Visibility in Miles

In addition to depth maps, the Omnidata toolkit was also utilized for normal surface estimation [35]. This modality provides information regarding the orientation of surfaces within the image, which is crucial for discerning the geometric properties of the scene. In contrast to the depth estimation model, the normal surface estimator considers both sky and ground details.

Entropy Map:

An image entropy map is incorporated as a modality to enhance the model's sensitivity to variations in visibility, particularly under low-visibility conditions. The entropy map quantifies the amount of information, or uncertainty, present in different regions of an image.

Edge Detection:

Edge detection serves as another key modality, particularly well-suited for long-range visibility scenarios where the delineation of objects and scene boundaries is critical. By highlighting the contours and edges within an image, this modality aids in defining shapes and structures, thereby providing a clearer distinction between different objects and features in the scene.

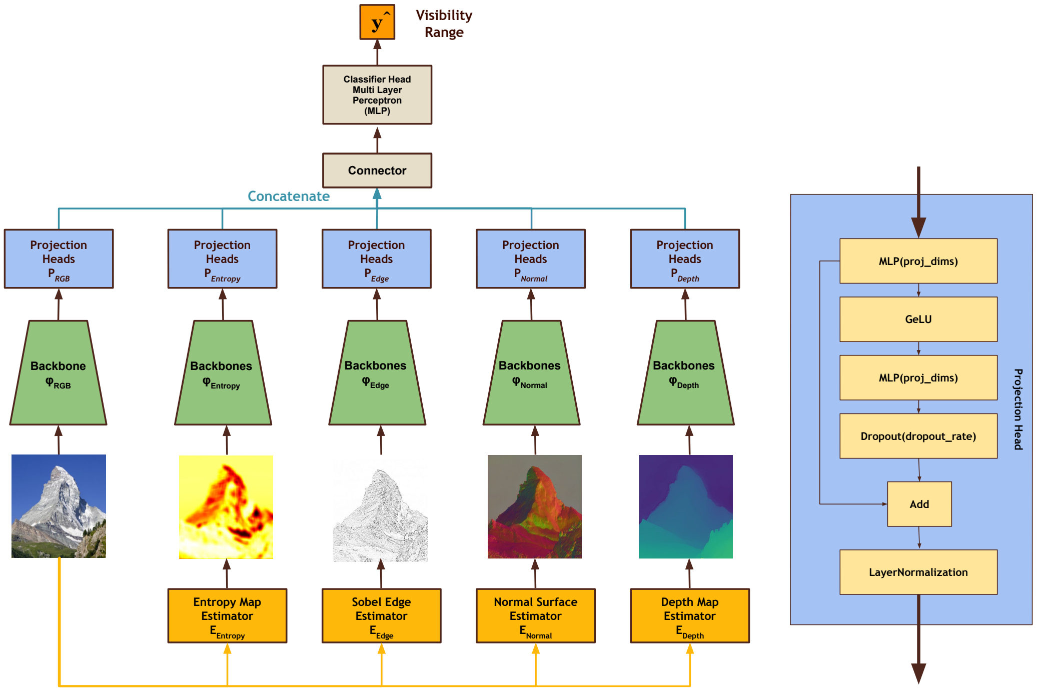
In Figures 2 and 3, we illustrate the impact of visibility degradation on various modalities for the same scene. Each row displays a single modality, while each column corresponds to a specific visibility bin.

B. Fusing Modalities

In the literature, numerous methods have been proposed for the fusion of different modalities within multi-stream networks [37–39]. These methods range from the simple concatenation of input streams in the input space to more complex fusion strategies at various levels of the model architecture.

Early fusion [40] involves concatenating or otherwise preprocessing all input streams in the input space. The combined data is then fed into a single feature extractor. While this method is the simplest to implement, it is often limited, as the feature extractor may learn to disregard some modalities, with the feature representation being dominated by a single modality.

Intermediate fusion [40], a widely adopted approach, involves feeding the different modalities into separate encoder


ŷ
Visibility Range
Classifier Head Multi Layer Perceptron (MLP)
Connector
Concatenate
Projection Heads P_{RGB}
Projection Heads $P_{Entropy}$
Projection Heads P_{Edge}
Projection Heads P_{Normal}
Projection Heads P_{Depth}
Backbone φ_{RGB}
Backbones $\varphi_{Entropy}$
Backbones φ_{Edge}
Backbones φ_{Normal}
Backbones φ_{Depth}
Entropy Map Estimator $E_{Entropy}$
Sobel Edge Estimator E_{Edge}
Normal Surface Estimator E_{Normal}
Depth Map Estimator E_{Depth}
MLP(proj_dims)
GeLU
MLP(proj_dims)
Dropout(dropout_rate)
Add
LayerNormalization
Projection Head

(a) **(b)**

Fig. 4 (a) SeeNN Framework: The framework first extracts features (entropy map, surface normals map, edge map, depth map) from the input image. Separate encoders $\phi_m(\cdot)$ ($\phi_m(\cdot)$ denotes modality encoders) process these features followed by a projection head (b), followed by fusion of these features through a Connector and prediction via a classifier ŷ. (b) Projection Head: The input vector is transformed by an MLP (Multi-Layer Perceptron) with a non-linear activation function (GeLU) and dropout for regularization.

layers before fusing the extracted embeddings. In this paradigm, the model learns to extract salient features from each modality before they are combined, thereby preventing any single modality from dominating the feature space. A significant advantage of this architecture is its compatibility with recent advancements in representation learning, such as contrastive learning or unsupervised representation learning, where fusion occurs between the encoder and decoder layers or at the initial stages of processing.

Late fusion [40] represents another fusion strategy, wherein each modality is passed through its own complete network until the decision layer (e.g., a classifier). Fusion is then performed at the decision level, either through a voting mechanism between the different models or by averaging their respective outputs.

1. Multimodal Fusion Methods

Various techniques for multimodal fusion have been proposed in the literature, including Tensor Fusion [41], Low-Rank Fusion [42], and attention mechanisms [43]. Although each method possesses its own set of advantages and disadvantages, self-attention has emerged as a foundational component for many recent large-scale models. While it typically requires a larger volume of training data, its computational cost is significantly lower compared to methods such as tensor fusion.

2. The SeeNN Multimodal Fusion Framework

The proposed SeeNN framework, illustrated in Figure 4, integrates multimodal deep learning techniques to process images concurrently with multiple derived modalities.

Initially, each input RGB image I undergoes a series of transformations via modality estimators to generate a depth map $E_d(I)$, a normal surface map $E_n(I)$, an edge detection map $E_e(I)$, and an entropy map $E_s(I)$. Each of these modalities captures distinct characteristics of the input, providing a diverse set of perspectives on the image's content.

Let m denote a specific modality (i.e., generated depth map $depth$, normal surface $normal$, entropy map $entropy$, edge map $edge$, and RGB image rgb). We employ different backbone models $\Phi_m(\cdot)$ for each modality input X_m. In this work, we utilize DenseNet121 [44] as the architecture for all Φ_m. The resulting embedding from each encoder is fed to a projection head P_m, which consists of a Multi-Layer Perceptron (MLP) with a non-linear activation function (GeLU) and dropout for regularization. This is followed by a layer normalization step, which is crucial for aligning the feature representations and mitigating the risk of dominance by any single modality. This process yields a feature vector F_m.

This procedure is applied to the RGB image X_{rgb}, depth map X_{depth}, normal surface map X_{normal}, entropy map $X_{entropy}$, and edge map X_{edge} to obtain the feature vectors F_{rgb}, F_{depth}, F_{normal}, $F_{entropy}$, and F_{edge}, respectively.

Following the projection heads, the SeeNN framework concatenates these embeddings into a single, comprehensive feature vector F. This concatenation is represented as $F = [F_{rgb}, F_{depth}; F_{normal}; F_{entropy}; F_{edge}]$. This composite vector is then fed to a connector module, C, which is responsible for fusing these modalities.

Finally, an MLP classifier head is applied to the fused feature vector to obtain the final prediction, $\hat{y}$.

For the connector module, we explored two primary methods. The first method involves passing the flattened feature vector F directly to the MLP, representing a simple yet effective fusion of the different features. The second method utilizes an attention block to perform self-attention on F, followed by flattening the output and feeding it to the MLP head.

3. Experimental Setup

For this study, we utilized our custom-collected dataset, SeeSet V1 (II.A), which comprises 320 distinct views collected across 20 locations with varying land covers, each with visibility ranging from 0 to 100 miles. The dataset was partitioned into training and validation subsets using a holdout approach. Specifically, all views from a predefined set of locations were reserved for the validation set, ensuring that the model does not overfit to specific sceneries and instead learns to estimate visibility based on image degradation [8]. This resulted in a training set of $100,350$ instances and a validation set of $15,750$ instances. All images in the dataset were preprocessed to an input resolution of 224×224 pixels.

We employed the Omnidata models to preprocess the RGB images and extract the estimated Depth Map and Normal Surface [35]. This approach, based on the DPT-Hybrid architecture [36], is analogous to methods used in the literature to generate pseudo-labels from RGB data for pre-training multimodal models [45, 46].

For the other modalities, namely the edge map and the entropy map, the RGB images were processed through handcrafted estimators, as depicted in Figure 4.

All models were trained for 100 epochs using the Adam optimizer with a learning rate of 0.001. A batch size of 32 was used for all training procedures.

III. Results and Discussion

This section presents the experimental results of our study. We begin by describing the performance of a unimodal RGB-based model, followed by a detailed analysis of the multimodal SeeNN framework. The discussion encompasses a comparative analysis of different fusion strategies, an examination of model performance through confusion matrices, and a consideration of computational costs and dataset limitations.

A. Unimodal Model Performance

A baseline model was established using only the RGB modality. This unimodal model achieved an overall accuracy of 87.92% on the validation set. This performance, while reasonable, highlights the inherent limitations of relying on a single data source, particularly when tested on previously unseen views. To ensure the robustness of our evaluation and prevent data leakage, we employed a strict holdout validation strategy, as described in the Experimental Setup, where entire geographical locations were withheld from the training set. This rigorous approach prevents the model from overfitting to specific sceneries and provides a more realistic assessment of its generalization capabilities.

B. Multimodal Model Performance

In contrast to the unimodal baseline, the multimodal models developed within the SeeNN framework demonstrated a significant improvement in performance. As shown in Table 3 and Figure 5, the fusion of multiple modalities resulted in a substantial increase in prediction accuracy, with gains of up to 10 percentage points.

Table 3 Ablation study comparing the performance of different modality combinations and fusion connectors. The highest accuracy for each connector type is highlighted in bold.

Connector	RGB	Entropy	Edge	Depth	Normal Surface	# Trainable Params.	Val. Acc. (%)
Unimodal	✓					7M	87.92
Concatenate	✓	✓				14M	96.40
	✓		✓			14M	96.53
	✓			✓		14M	**97.57**
	✓			✓	✓	21M	97.14
	✓	✓	✓	✓	✓	38M	96.30
Self-Attention	✓		✓			14M	96.86
	✓			✓		14M	96.31
	✓			✓	✓	21M	97.47
	✓	✓	✓	✓	✓	38M	**97.63**

For instance, a model combining RGB and Depth modalities, using a simple concatenation connector, achieved an accuracy of 97.57%. The highest performing model, which fused all five modalities (RGB, Entropy, Edge, Depth, and Normal Surface) using a self-attention connector, reached an impressive validation accuracy of 97.63%. These results strongly support the hypothesis that integrating diverse data sources enables the model to form a more comprehensive and robust understanding of the atmospheric conditions, leading to more accurate visibility estimations.

C. Analysis of Misclassifications

The confusion matrices presented in Figure 5 provide further insight into the performance of the top-performing multimodal models. While the overall accuracy is high, the models exhibit some difficulty in distinguishing between adjacent visibility categories. Specifically, there is a tendency to misclassify instances of "Class 3" (3 to 5 miles visibility) as "Class 4" (>= 5 miles visibility). This suggests that the visual cues differentiating these two classes are subtle and challenging for the models to discern.

Interestingly, the combination of RGB and Depth modalities yielded improved performance for this specific class, indicating that depth information is particularly valuable for resolving ambiguity in this visibility range. Future work could explore the integration of additional modalities or the development of more sophisticated fusion mechanisms to address this specific challenge.

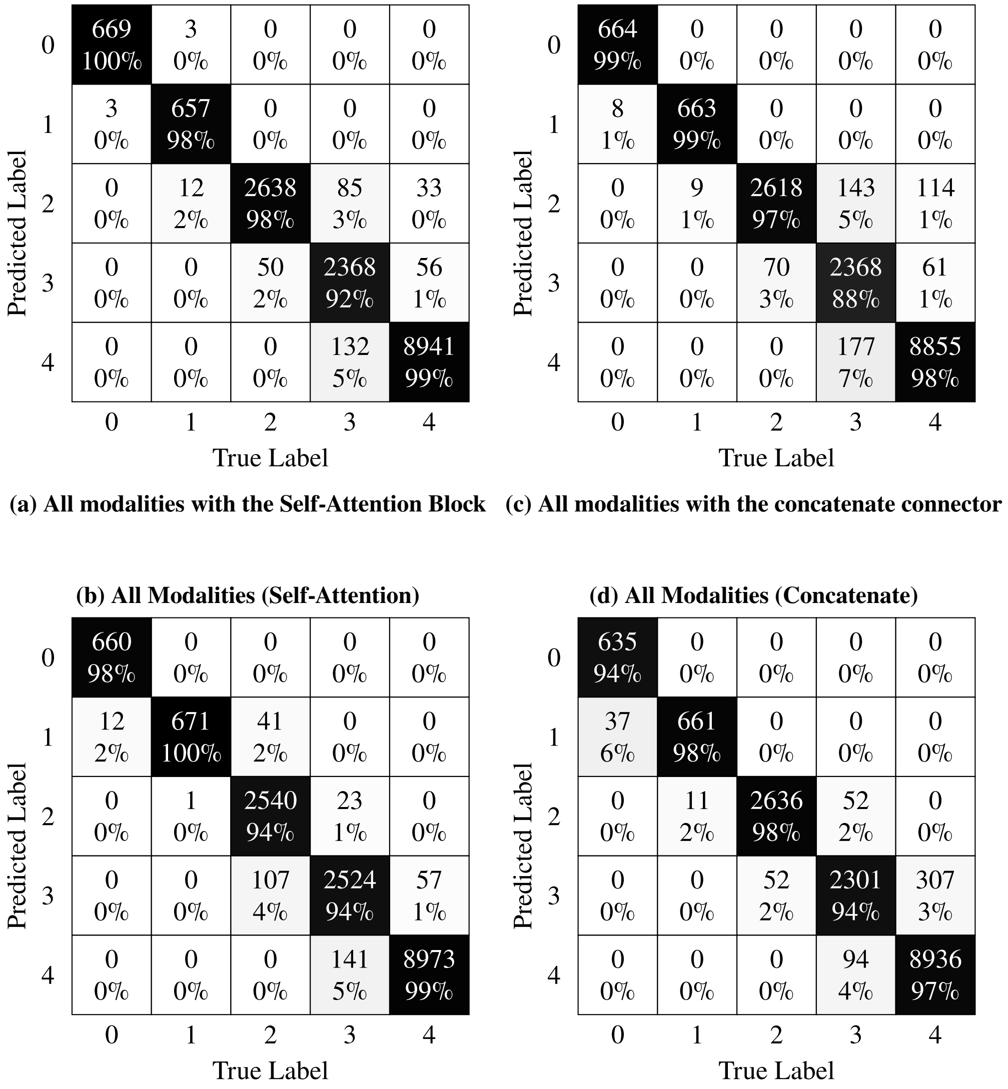
D. Discussion

1. Computational Cost and Deployment Considerations

A critical consideration in the practical application of these models is the trade-off between performance and computational cost. While the model that fused all available modalities achieved the highest accuracy, it also has the highest computational overhead, requiring the execution of multiple modality estimators and backbone networks. When deploying such models, particularly on resource-constrained hardware such as embedded devices, it is essential to consider these limitations. The results suggest that a carefully selected subset of modalities (e.g., RGB and Depth) can provide a favorable balance between accuracy and efficiency.

2. Dataset Limitations and Future Directions

While the SeeSet V1 dataset addresses a significant gap in the availability of public, multi-view datasets for atmospheric visibility research, it has certain limitations. The diversity of landscapes and land covers could be expanded to enhance the model's generalizability. Future work should focus on enriching the dataset using the latest generation of flight simulators (e.g., Microsoft Flight Simulator, X-Plane 12), which offer near-photorealistic rendering and more sophisticated atmospheric models.



	0	1	2	3	4
0	669 100%	3 0%	0 0%	0 0%	0 0%
1	3 0%	657 98%	0 0%	0 0%	0 0%
2	0 0%	12 2%	2638 98%	85 3%	33 0%
3	0 0%	0 0%	50 2%	2368 92%	56 1%
4	0 0%	0 0%	0 0%	132 5%	8941 99%

(a) All modalities with the Self-Attention Block

	0	1	2	3	4
0	664 99%	0 0%	0 0%	0 0%	0 0%
1	8 1%	663 99%	0 0%	0 0%	0 0%
2	0 0%	9 1%	2618 97%	143 5%	114 1%
3	0 0%	0 0%	70 3%	2368 88%	61 1%
4	0 0%	0 0%	0 0%	177 7%	8855 98%

(c) All modalities with the concatenate connector

(b) All Modalities (Self-Attention)

(d) All Modalities (Concatenate)

	0	1	2	3	4
0	660 98%	0 0%	0 0%	0 0%	0 0%
1	12 2%	671 100%	41 2%	0 0%	0 0%
2	0 0%	1 0%	2540 94%	23 1%	0 0%
3	0 0%	0 0%	107 4%	2524 94%	57 1%
4	0 0%	0 0%	0 0%	141 5%	8973 99%

(e) RGB and Depth Map Model with the Concatenate Connector

	0	1	2	3	4
0	635 94%	0 0%	0 0%	0 0%	0 0%
1	37 6%	661 98%	0 0%	0 0%	0 0%
2	0 0%	11 2%	2636 98%	52 2%	0 0%
3	0 0%	0 0%	52 2%	2301 94%	307 3%
4	0 0%	0 0%	0 0%	94 4%	8936 97%

(g) RGB and Depth Map Model with the Self-Attention Block

(f) RGB + Depth (Concatenate)

(h) RGB + Depth (Self-Attention)

Fig. 5 Confusion matrices for the top-performing multimodal models. The models show high accuracy but struggle with adjacent classes, particularly misclassifying Class 3 as Class 4.

Further research should also explore advanced pre-training techniques to improve the quality of the learned feature representations. Many state-of-the-art multimodal systems leverage self-supervised or unsupervised pre-training on large-scale datasets, which has been shown to improve downstream task performance.

Finally, a deeper investigation into the impact of visibility degradation on the feature representations extracted by different network architectures is warranted. A thorough understanding of this relationship is crucial for improving the trustworthiness and reliability of these models in safety-critical, real-world applications.

IV. Conclusion

In this paper, we have presented a novel multimodal deep learning framework, SeeNN, for the estimation of atmospheric visibility in challenging in-flight scenarios. Our work makes two primary contributions to the field.

First, we introduced the SeeNN framework, which effectively fuses information from RGB imagery, entropy maps, edge maps, depth maps, and normal surface maps. Our extensive experimental results demonstrate that this multimodal approach significantly outperforms traditional unimodal models that rely solely on RGB data. The superior performance of SeeNN underscores the value of integrating diverse data modalities to overcome the inherent ambiguities and limitations of single-source systems, thereby achieving more accurate and reliable visibility estimation.

Second, we have developed and released a comprehensive, open-source benchmark dataset for atmospheric visibility estimation. This dataset, a key contribution of our work, is distinguished by its diversity, encompassing a wide range of altitudes, land cover types, and visibility conditions. It provides a much-needed resource for the research community, enabling the robust training, validation, and comparative evaluation of visibility estimation algorithms.

Our empirical results show that the proposed multimodal framework offers substantial improvements in estimation accuracy over unimodal RGB methods. The release of our benchmark dataset addresses a critical gap in the field, providing a standardized platform for future research and development. We anticipate that this resource will catalyze further innovation in the domain, spurring the development of increasingly sophisticated multimodal deep learning techniques for atmospheric visibility estimation.

Future research could explore several promising avenues, including the integration of additional sensor modalities, the investigation of more advanced fusion architectures, and the application of our framework to related problems in atmospheric science. Furthermore, the potential for leveraging transfer learning and domain adaptation techniques in this context remains a compelling area for future investigation.

In conclusion, this work contributes to the growing body of research at the intersection of deep learning and atmospheric science, offering both methodological advancements and a valuable resource to the research community. As the field continues to evolve, we believe that multimodal approaches, such as the one presented in this paper, will play an increasingly pivotal role in addressing complex environmental perception tasks, with far-reaching implications for aviation safety and other domains.

Acknowledgments

This research was supported by the Federal Aviation Administration (FAA). The authors wish to express their gratitude to the William J. Hughes Technical Center for providing the resources and support necessary for this work. The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official position, policy, or decision of the FAA, unless so designated by other documentation.

References

[1] Malm, W., *Visibility: The Seeing of Near and Distant Landscape Features*, 2016.

[2] Kulesa, G., "WEATHER AND AVIATION: HOW DOES WEATHER AFFECT THE SAFETY AND OPERATIONS OF AIRPORTS AND AVIATION, AND HOW DOES FAA WORK TO MANAGE WEATHER-RELATED EFFECTS?" 2003. URL `https://api.semanticscholar.org/CorpusID:108023423`.

[3] Fultz, A. J., and Ashley, W. S., "Fatal weather-related general aviation accidents in the United States," *Physical Geography*, Vol. 37, No. 5, 2016, pp. 291–312.

[4] Long, T., "Analysis of weather-related accident and incident data associated with Section 14 CFR Part 91 Operations," *The Collegiate Aviation Review International*, Vol. 40, No. 1, 2022.

[5] Fujita, T. T., and Caracena, F., "An analysis of three weather-related aircraft accidents," *Bulletin of the American Meteorological Society*, Vol. 58, No. 11, 1977, pp. 1164–1181.

[6] Ramee, C., Speirs, A., Payan, A. P., and Mavris, D., "Analysis of weather-related helicopter accidents and incidents in the United States," *AIAA Aviation 2021 Forum*, 2021, p. 2954.

[7] Ahlstrom, U., Racine, N., and Hallman, K., "Assessments of Flight and Weather Conditions during General Aviation Operations," Tech. Rep. DOT/FAA/TC-19/33, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, 2019. Available from the Federal Aviation Administration William J. Hughes Technical Center: `https://actlibrary.tc.faa.gov`.

[8] Bouhsine, T., Idbraim, S., Bouaynaya, N. C., Alfergani, H., Ouadil, K. A., and Johnson, C. C., "Atmospheric Visibility Image-Based System for Instrument Meteorological Conditions Estimation: A Deep Learning Approach," *Proc. 2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Rabat, Morocco, 2022, pp. 1–6. [Online]. Available: https://doi.org/10.1109/WINCOM55661.2022.9966454.

[9] Ait Ouadil, K., Idbraim, S., Bouhsine, T., et al., "Atmospheric visibility estimation: a review of deep learning approach," *Multimedia Tools and Applications*, 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-16855-z.

[10] Li, S., Fu, H., and Lo, W., "Meteorological Visibility Evaluation on Webcam Weather Image Using Deep Learning Features," 2017. https://doi.org/10.7763/IJCTE.2017.V9.1186.

[11] Chaabani, H., Werghi, N., Kamoun, F., Taha, B., Outay, F., and Yasar, A.-U.-H., "Estimating meteorological visibility range under foggy weather conditions: A deep learning approach," *Procedia Computer Science*, Vol. 141, 2018, pp. 478–483. https://doi.org/10.1016/j.procs.2018.10.139, URL https://www.sciencedirect.com/science/article/pii/S1877050918317885.

[12] Palvanov, A., and Im Cho, Y., "DHCNN for Visibility Estimation in Foggy Weather Conditions," *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 2018, pp. 240–243. https://doi.org/10.1109/SCIS-ISIS.2018.00050.

[13] Choi, Y., Choe, H.-G., Choi, J. Y., Kim, K. T., Kim, J.-B., and Kim, N.-I., "Automatic Sea Fog Detection and Estimation of Visibility Distance on CCTV," *Journal of Coastal Research*, , No. 85 (10085), 2018, pp. 881–885. https://doi.org/10.2112/SI85-177.1, URL https://doi.org/10.2112/SI85-177.1.

[14] You, Y., Lu, C., Wang, W., and Tang, C.-K., "Relative CNN-RNN: Learning Relative Atmospheric Visibility From Images," *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, Vol. 28, No. 1, 2019, pp. 45–55. https://doi.org/10.1109/TIP.2018.2857219.

[15] Li, Q., Tang, S., Peng, X., and Ma, Q., "A Method of Visibility Detection Based on the Transfer Learning," *Journal of Atmospheric and Oceanic Technology*, Vol. 36, 2019. https://doi.org/10.1175/JTECH-D-19-0025.1.

[16] Outay, F., Taha, B., Chaabani, H., Kamoun, F., Werghi, N., and Yasar, A. U.-H., "Estimating ambient visibility in the presence of fog: a deep convolutional neural network approach," *Personal and Ubiquitous Computing*, Vol. 25, No. 1, 2021, pp. 51–62. https://doi.org/10.1007/s00779-019-01334-w, URL https://doi.org/10.1007/s00779-019-01334-w.

[17] Liu, K., Li, Y., Xu, N., and Natarajan, P., "Learn to Combine Modalities in Multimodal Deep Learning," , 2018. [Online]. Available: https://arxiv.org/abs/1805.11730.

[18] Castanedo, F., Ursino, D., and Takama, Y., "A Review of Data Fusion Techniques," *The Scientific World Journal*, Vol. 2013, 2013, p. Article ID 704504. [Online]. Available: https://doi.org/10.1155/2013/704504.

[19] Molino-Minero-Re, E., Aguileta, A. A., Brena, R. F., and Garcia-Ceja, E., "Improved Accuracy in Predicting the Best Sensor Fusion Architecture for Multiple Domains," *Sensors*, Vol. 21, No. 7007, 2021. [Online]. Available: https://doi.org/10.3390/s21217007.

[20] Blasch, E., Pham, T., Chong, C. Y., Koch, W., Leung, H., Braines, D., and Abdelzaher, T., "Machine Learning/Artificial Intelligence for Sensor Data Fusion-Opportunities and Challenges," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 36, No. 7, 2021, pp. 80–93. [Online]. Available: https://doi.org/10.1109/MAES.2020.3049030.

[21] Palvanov, A., and Cho, Y. I., "VisNet: Deep Convolutional Neural Networks for Forecasting Atmospheric Visibility," *Sensors*, Vol. 19, No. 6, 2019, p. 1343. https://doi.org/10.3390/s19061343, URL https://www.mdpi.com/1424-8220/19/6/1343.

[22] Department of Computer Science, Chu Hai College of Higher Education, 80 Castle Peak Road, Castle Peak Bay, Tuen Mun, N.T. Hong Kong, Lo, W. L., Zhu, M., and Fu, H., "Meteorology Visibility Estimation by Using Multi-Support Vector Regression Method," *Journal of Advances in Information Technology*, 2020, pp. 40–47. https://doi.org/10.12720/jait.11.2.40-47, URL http://www.jait.us/index.php?m=content&c=index&a=show&catid=198&id=1091.

[23] Li, J., Lo, W. L., Fu, H., and Chung, H. S. H., "A Transfer Learning Method for Meteorological Visibility Estimation Based on Feature Fusion Method," *Applied Sciences*, Vol. 11, No. 3, 2021. https://doi.org/10.3390/app11030997, URL https://www.mdpi.com/2076-3417/11/3/997.

[24] Zhang, F., Yu, T., Li, Z., Wang, K., Chen, Y., Huang, Y., and Kuang, Q., "Deep Quantified Visibility Estimation for Traffic Image," *Atmosphere*, Vol. 14, No. 1, 2023, p. 61. https://doi.org/10.3390/atmos14010061.

[25] You, J., Jia, S., Pei, X., and Yao, D., "DMRVisNet: Deep Multihead Regression Network for Pixel-Wise Visibility Estimation Under Foggy Weather," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 11, 2022, pp. 22354–22366. https://doi.org/10.1109/TITS.2022.3180229.

[26] Chen, X.-H., and Li, Z., "Dark Channel Based Visibility Measuring from Daytime Scene Videos," *Journal of Internet Technology*, Vol. 23, No. 3, 2022, pp. 593–599.

[27] Wauben, W., and Roth, M., "Exploration of Fog Detection and Visibility Estimation from Camera Images," *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation, CIMO TECO*, 2016, pp. 1–14.

[28] Cheng, X., Liu, G., Hedman, A., Wang, K., and Li, H., "Expressway Visibility Estimation Based on Image Entropy and Piecewise Stationary Time Series Analysis," , 2018. [Online]. Available: arXiv:1804.04601.

[29] Zhou, H., Dai, M., Shi, D., Meng, Y., Peng, B., and Chen, T., "Research on visibility detection model optimization based on dark channel prior and image entropy and visibility development trend prediction," *IOP Conference Series: Earth and Environmental Science*, Vol. 826, 2021, p. 012031. https://doi.org/10.1088/1755-1315/826/1/012031.

[30] Chen, X.-W., and Lin, X., "Big Data Deep Learning: Challenges and Perspectives," *IEEE Access*, Vol. 2, 2014, pp. 514–525. https://doi.org/10.1109/ACCESS.2014.2325029.

[31] Liang, P. P., "Foundations of Multisensory Artificial Intelligence," , 2024. URL https://arxiv.org/abs/2404.18976.

[32] Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H., and Huang, L., "What Makes Multi-modal Learning Better than Single (Provably)," , 2021. URL https://arxiv.org/abs/2106.04538.

[33] Palvanov, A., and Cho, Y. I., "VisNet: Deep Convolutional Neural Networks for Forecasting Atmospheric Visibility," *Sensors*, Vol. 19, No. 6, 2019, p. 1343. https://doi.org/10.3390/s19061343.

[34] Gui, J., Cong, X., Cao, Y., Ren, W., Zhang, J., Zhang, J., Cao, J., and Tao, D., "A comprehensive survey and taxonomy on single image dehazing based on deep learning," *ACM Computing Surveys*, Vol. 55, No. 13s, 2023, pp. 1–37.

[35] Eftekhar, A., Sax, A., Bachmann, R., Malik, J., and Zamir, A., "Omnidata: A Scalable Pipeline for Making Multi-Task Mid-Level Vision Datasets from 3D Scans," , 2021. [Online]. Available: https://arxiv.org/abs/2110.04994.

[36] Ranftl, R., Bochkovskiy, A., and Koltun, V., "Vision Transformers for Dense Prediction," , 2021.

[37] Akkus, C., Chu, L., Djakovic, V., Jauch-Walser, S., Koch, P., Loss, G., Marquardt, C., Moldovan, M., Sauter, N., Schneider, M., Schulte, R., Urbanczyk, K., Goschenhofer, J., Heumann, C., Hvingelby, R., Schalk, D., and Aßenmacher, M., "Multimodal Deep Learning," , 2023.

[38] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., "Learning Transferable Visual Models From Natural Language Supervision," , 2021.

[39] Jia, C., Yang, Y., Xia, Y., Chen, Y.-T., Parekh, Z., Pham, H., Le, Q. V., Sung, Y., Li, Z., and Duerig, T., "Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision," , 2021.

[40] Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M. P., "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines," *npj Digital Medicine*, Vol. 3, No. 1, 2020, p. 136. https://doi.org/10.1038/s41746-020-00341-z, URL https://doi.org/10.1038/s41746-020-00341-z.

[41] Zadeh, A., Chen, M., Poria, S., Cambria, E., and Morency, L.-P., "Tensor Fusion Network for Multimodal Sentiment Analysis," , 2017. URL https://arxiv.org/abs/1707.07250.

[42] Liu, Z., Shen, Y., Lakshminarasimhan, V. B., Liang, P. P., Zadeh, A., and Morency, L.-P., “Efficient Low-rank Multimodal Fusion with Modality-Specific Factors,” , 2018. URL https://arxiv.org/abs/1806.00064.

[43] Nagrani, A., Yang, S., Arnab, A., Jansen, A., Schmid, C., and Sun, C., “Attention Bottlenecks for Multimodal Fusion,” *Advances in Neural Information Processing Systems*, Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Curran Associates, Inc., 2021, pp. 14200–14213. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/76ba9f564ebbc35b1014ac498fafadd0-Paper.pdf.

[44] Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q., “Densely Connected Convolutional Networks,” , Jan. 2018. https://doi.org/10.48550/arXiv.1608.06993, URL http://arxiv.org/abs/1608.06993.

[45] Bachmann, R., Mizrahi, D., Atanov, A., and Zamir, A., “MultiMAE: Multi-modal Multi-task Masked Autoencoders,” , 2022.

[46] Wang, X., Chen, G., Qian, G., Gao, P., Wei, X.-Y., Wang, Y., Tian, Y., and Gao, W., “Large-scale Multi-Modal Pre-trained Models: A Comprehensive Survey,” , 2024.