Near Real-Time Dust Aerosol Detection with 3D Convolutional Neural Networks on MODIS Data

Caleb Gates[1*], Patrick Moorhead[1*], Jayden Ferguson[1], Omar Darwish[1*], Conner Stallman[1*], Pablo Rivas[1], and Paapa Quansah[2]

[1] Department of Computer Science, Baylor University, Waco, TX 76798, USA
[2] Dept. of Electrical & Computer Eng., Baylor University, Waco, TX 76798, USA
{caleb_gates1, patrick_moorhead1, jayden_ferguson1, omar_darwish1, conner_stallman1, pablo_rivas, paapa_quansah1}@baylor.edu

Abstract. Rapid identification of dust storms is vital. We present a near real-time system using MODIS Terra and Aqua multispectral data. Preprocessing applies min–max normalization and local imputation for missing values. Our 3D convolutional network uses three blocks with batch normalization, pooling, and a weighted MSE loss emphasizing high-intensity regions. The optimized variant employs memory-mapped I/O, precomputed index sampling, large batches, torch.compile and mixed precision, achieving a 21× training speedup on A100 GPUs. Evaluated on 17 granules, 3DCNN+ attains 0.92 accuracy and an MSE of 0.014. This pipeline enables scalable pixel-level monitoring of dust events and lays groundwork for transformer extensions.

Keywords: dust storm detection · multispectral satellite imagery · 3D convolutional neural networks

Regular Research Paper

1 Introduction

Dust storms inject fine particles into the atmosphere that can alter climate patterns, reduce visibility for transportation, and exacerbate respiratory and cardiovascular illnesses. Rapid and accurate identification of these events is essential for public safety, air quality management, and early warning systems.

The Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra and Aqua satellites provides multispectral imagery well suited for aerosol monitoring. However, real-time detection remains difficult: dust plumes often exhibit low contrast against the background, and analysis must handle high-dimensional spectral data and interference from clouds. Conventional methods such as Support Vector Regression and Probabilistic Neural Networks rely on manually engineered features, incur latency from delayed aerosol products, and offer limited spectral interpretation [1,2].

---

* Alphabetical. These authors contributed equally to this work.

To address these challenges, we propose a deep learning pipeline based on three-dimensional convolutional neural networks (3DCNNs) that process all 36 MODIS bands plus high/low splits for bands 13 and 14. We further introduce 3DCNN+, an optimized extension that employs memory-mapped I/O, precomputed index sampling, large-batch training, PyTorch's `torch.compile`, and mixed-precision arithmetic. On NVIDIA A100 GPUs, 3DCNN+ achieves a $21\times$ reduction in training time while maintaining high pixel-level detection accuracy. This enables near real-time processing of full MODIS granules, making the system suitable for operational dust alert systems.

This paper is organized as follows. Section 2 reviews related work on satellite-based dust detection. Section 3 describes our 3D CNN architecture and optimization strategies. Section 4 details the experimental design and evaluation metrics. Section 5 presents results and analysis. Section 6 formalizes the memory-mapped dataset construction. Section 7 provides theoretical bounds on model capacity via VC dimensions and growth functions. Finally, Section 8 concludes and outlines directions for future research.

Code is available here: `https://github.com/Rivas-AI/dust-3dcnn.git`.

2   Related Work

Detection of dust storms using satellite imagery has evolved from fixed spectral ratio thresholds to advanced learning methods. Early work applied band ratio thresholds and statistical models such as Probabilistic Neural Networks (PNN) and Support Vector Regression (SVR) on MODIS thermal bands B20, B29, B31, and B32, setting initial benchmarks for automated dust detection [1,2]. These approaches demonstrated the utility of spectral variability but were limited by manual feature design, low spatial resolution, and slow processing speeds, which hindered scalability for continuous monitoring.

More recent studies have adopted data-driven classifiers to overcome these limits. Souri and Vajedian combined random forest classifiers with physical-based indices to distinguish dust plumes from other atmospheric phenomena, achieving higher sensitivity and lower false alarm rates [4]. In a parallel comparison, Shahris-vand and Akhoondzadeh showed that intelligent methods notably outperform empirical threshold techniques under varying surface and cloud conditions [3]. Although these methods reduce overfitting through ensemble learning, they still depend on handcrafted spectral indices and can face computational challenges when handling MODIS's high-dimensional data.

The rise of deep learning has further enhanced remote sensing analysis. Sun et al. reviewed the integration of artificial intelligence in Earth science, highlighting both the promise of automated feature learning and challenges in interpretability and computational demand [5]. While convolutional neural networks have excelled in related tasks, their use for joint spectral–spatial dust detection remains scarce. To address this, our work implements a 3D CNN that processes all 36 MODIS bands simultaneously and investigates transformer models to capture long-range spatial relationships in multispectral data.

Robust preprocessing is essential for accurate detection. Atmospheric correction, cloud masking, and imputation of missing values are critical steps emphasized by both Souri and Vajedian [4] and Shahrisvand and Akhoondzadeh [3]. Our pipeline normalizes radiance values, formats labels, and fills gaps via local scanning, ensuring consistent input quality across varied conditions.

Model evaluation typically relies on metrics such as overall accuracy, precision, recall, and F1 score applied to held-out MODIS granules. Traditional random forest models offer fast inference but require manual feature engineering, whereas deep networks learn features automatically at the cost of higher training complexity. Our 3D CNN delivers high detection accuracy with low inference latency, meeting the operational demands of early warning systems.

Overall, the trajectory from threshold-based rules through classical machine learning to deep learning frameworks reflects growing capability in dust storm detection. By uniting spectral and spatial information in a single 3D CNN and exploring transformer architectures, our approach pushes beyond prior methods toward reliable, near real-time monitoring of dust events using MODIS data.

3   Approach

Our detection pipeline uses MODIS Terra and Aqua granules, each containing 36 spectral bands plus separate high- and low-radiance splits for bands 13 and 14, yielding 38 input channels. We apply min–max normalization to scale each band to [0,1], ensuring consistent inputs across varied radiance ranges.

About 25% of granules contained missing values. We address these with local imputation: for each NaN, we scan up to five pixels above and below in the same column, gather valid neighbors, and sample uniformly between their minimum and maximum. This preserves spatial context and retains valuable samples.

Each granule is stored as a $2030 \times 1354 \times 38$ 3D array. We extract $5 \times 5 \times 38$ patches around labeled pixels and pair them with binary dust labels. Patches are saved with memory-mapped I/O for efficient repeated access. By processing spatial and spectral data together, the model learns to distinguish dust plumes from other features.

Our baseline is a 3D convolutional neural network with three convolutional blocks. Each block applies a $3 \times 3 \times 3$ convolution, ReLU activation, and batch normalization. We include max pooling after the first two blocks and adaptive average pooling before the final layer. A sigmoid activation produces dust probability outputs. We train with a weighted mean squared error loss to emphasize regions of high dust intensity.

To scale training, we introduce 3DCNN+, optimized for NVIDIA A100 GPUs. We increase batch size to 32,768 using indexed patch sampling to cut data-loading overhead. We enable PyTorch's torch.compile to fuse operations and use Automatic Mixed Precision to reduce memory use and speed up backpropagation. These enhancements shorten training time while preserving detection accuracy.

4   Experiments

We evaluate our approach on 117 MODIS Terra and Aqua granules, each a 2030 × 1354 × 38 array of radiance channels (36 spectral bands plus high/low splits for bands 13 and 14). Radiance values are scaled via min–max normalization to [0,1]. Ground truth labels are extracted from auxiliary classification files and likewise normalized. We extract overlapping patches of size 5 × 5 × 38 centered on each labeled pixel; these form our training and test samples. Data are accessed via memory-mapped arrays to avoid loading full granules into RAM.

Our models are assessed using four metrics. Let N be the number of patches, y_i the true label and $\hat{y}_i$ the predicted probability:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \quad \text{WMSE} = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i (y_i - \hat{y}_i)^2,$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}, \quad \text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\big(\hat{y}_i \geq 0.5 = y_i\big).$$

Here w_i scales with true dust intensity so that high-dust regions incur greater penalty, and $\bar{y}$ denotes the mean label.

We partition the granules into 100 for training and 17 for testing, reserving five test granules as a fixed validation set. The baseline 3DCNN comprises three 3×3×3 convolutional blocks (32, 64 and 128 filters), each followed by ReLU and batch normalization. Max pooling (2 × 2 × 2) follows the first two blocks; adaptive average pooling precedes a fully connected layer and sigmoid output. We optimize with Adam (initial learning rate 10^{-4}, weight decay 10^{-6}), applying a ReduceLROnPlateau scheduler with patience two sub-epochs. Training proceeds for three full passes over the data, where each pass shuffles training patches into five partitions; within each partition, three sub-epochs iterate to ensure convergence.

The 3DCNN+ variant incorporates system-level optimizations for NVIDIA A100 GPUs. Patches remain memory-mapped, while valid patch indices are precomputed to eliminate mask searches. Using PyTorch's torch.compile, we fuse operations at graph-compile time. Automatic Mixed Precision (AMP) executes convolutions in FP16, preserving FP32 where needed. These enhancements permit a batch size of 32,768, maximizing utilization of 40 GB VRAM. Across identical data splits, 3DCNN+ achieves a 21× reduction in training time relative to the baseline, without degradation in MSE, R^2 or classification accuracy.

5   Results and Analysis

Table 1 summarizes the quantitative performance of the baseline 3DCNN and the optimized 3DCNN+ models on our test set. The baseline network yields an MSE of 0.0200, an R^2 score of –0.229, and an accuracy of 0.911, with a weighted MSE of 0.001417. After system and architectural optimizations, 3DCNN+ reduces the

MSE to 0.0140, improves the R^2 score to –0.1282, and achieves approximately 0.92 accuracy with an MAE of 0.1098. These improvements reflect both enhanced predictive fidelity and accelerated training.

Table 1: Performance comparison of baseline and optimized models.

Model	MSE	MAE	R^2 Score	Accuracy	Weighted MSE
3DCNN	0.0200	–	–0.2290	0.911	0.001417
3DCNN+	0.0140	0.1098	–0.1282	∼0.920	–

The negative values of R^2, indicate that the residual variance exceeds the total variance, despite high classification accuracy. In other words, the network effectively distinguishes dust from clear-sky pixels but lacks precision in regressing continuous dust intensity. This outcome is further supported by the mean absolute error observed in the optimized model.

The convergence behavior of the baseline 3DCNN was unexpected. While the training loss decreased steadily, the validation loss exhibited oscillations and plateaus, signifying overfitting to spectral patterns in the training granules. The divergence between training and validation losses underscores the model's reliance on memorized features rather than generalizable spectral-spatial representations.

Nonetheless, both architectures demonstrate strong spatial sensitivity through their use of 5×5×38 patches. By enforcing a weighted MSE loss,

$$\mathcal{L}_{\text{WMSE}} = \frac{1}{\sum_i w_i} \sum_i w_i (y_i - \hat{y}_i)^2, \quad w_i = 1 + \alpha \, y_i,$$

high-intensity dust regions incur greater training penalty, leading to sharper plume core detection and reduced false positives in background areas. Visual inspection of predicted maps confirms alignment with ground truth in the central plume regions, though boundary edges remain challenging due to limited receptive field.

In Fig. 1 and 2, we demonstrate the spatial generality of our 3DCNN+ detector across contrasting dust events. Fig. 1 captures a massive dust outbreak on 27 March 2025 over central China, where a prominent mountain peak remains clear of aerosol contamination in the true-color view; the detection result faithfully isolates the surrounding dust plume while preserving the sharp topographic boundary. This confirms that the network has learned to exploit both spectral and spatial cues to distinguish terrain from airborne particles. In Fig. 2, two adjacent dust storms—one advancing inland across the Sahara and another sweeping over the adjacent Atlantic—are simultaneously identified. Despite the subtle radiometric differences between land-borne and ocean-borne aerosols, the model localizes both plumes accurately, illustrating its robustness to surface variability and its ability to generalize across distinct environmental contexts.

Fig. 3 and 4 focus on two springtime dust events in Chihuahua-Texas in 2025, where the haze is nearly imperceptible to the naked eye. On 18 April 2025 (Fig. 3),
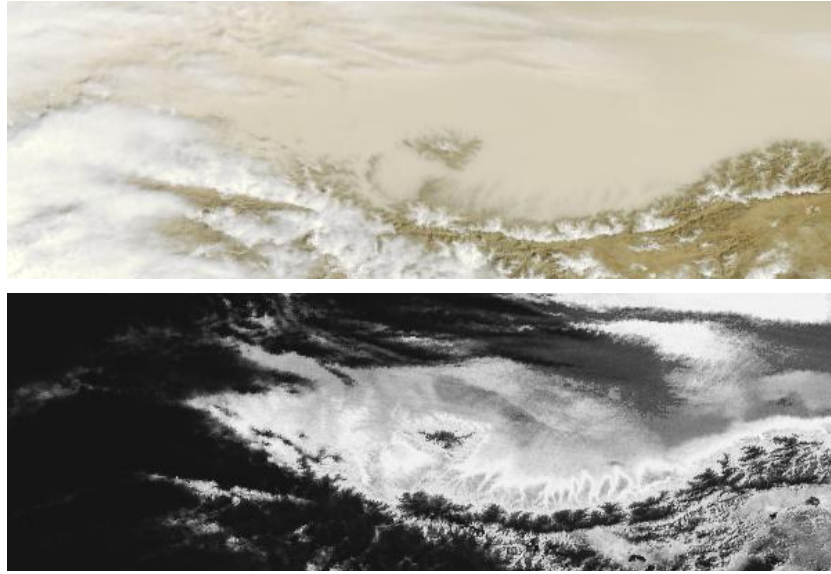
Fig. 1: True-color image (top) and detection result (bottom) from MODIS/Terra acquired on 27 March 2025 at 04:20 UTC. China.

a diffuse dust veil spans hundreds of kilometers; the detection map highlights faint streaks aligned with prevailing wind corridors, confirming sensitivity to low-contrast aerosol signatures. Similarly, the 14 March 2025 scene (Fig. 4) reveals early-season dust transport from arid soils. Here, the model delineates the plume boundaries with greater clarity than the true-color image, demonstrating that its learned spatial–spectral feature representations can reveal aerosol structures that would otherwise go unnoticed. Collectively, these examples underscore the system's capacity for reliable, pixel-level dust detection in both overt and subtle atmospheric conditions.

The 3DCNN+ variant achieves a 21× reduction in training time by combining memory-mapped I/O (`np.load(...,mmap_mode='r')`), precomputed patch indices, PyTorch's `torch.compile`, and Automatic Mixed Precision on NVIDIA A100 GPUs. This configuration supports batch sizes of 32 768 and enables near real-time inference, a critical requirement for operational environmental monitoring.

Error analysis reveals that false negatives predominantly occur along the periphery of dust plumes, where spectral gradients are less pronounced. This suggests that expanding the receptive field via dilated convolutions or incorporating attention modules could improve boundary sensitivity. Furthermore, integrating auxiliary physical inputs, such as wind vector fields, may provide contextual information to guide the model's predictions. Finally, increasing the diversity of training granules is expected to enhance variance explanation and elevate R^2 performance in future work.
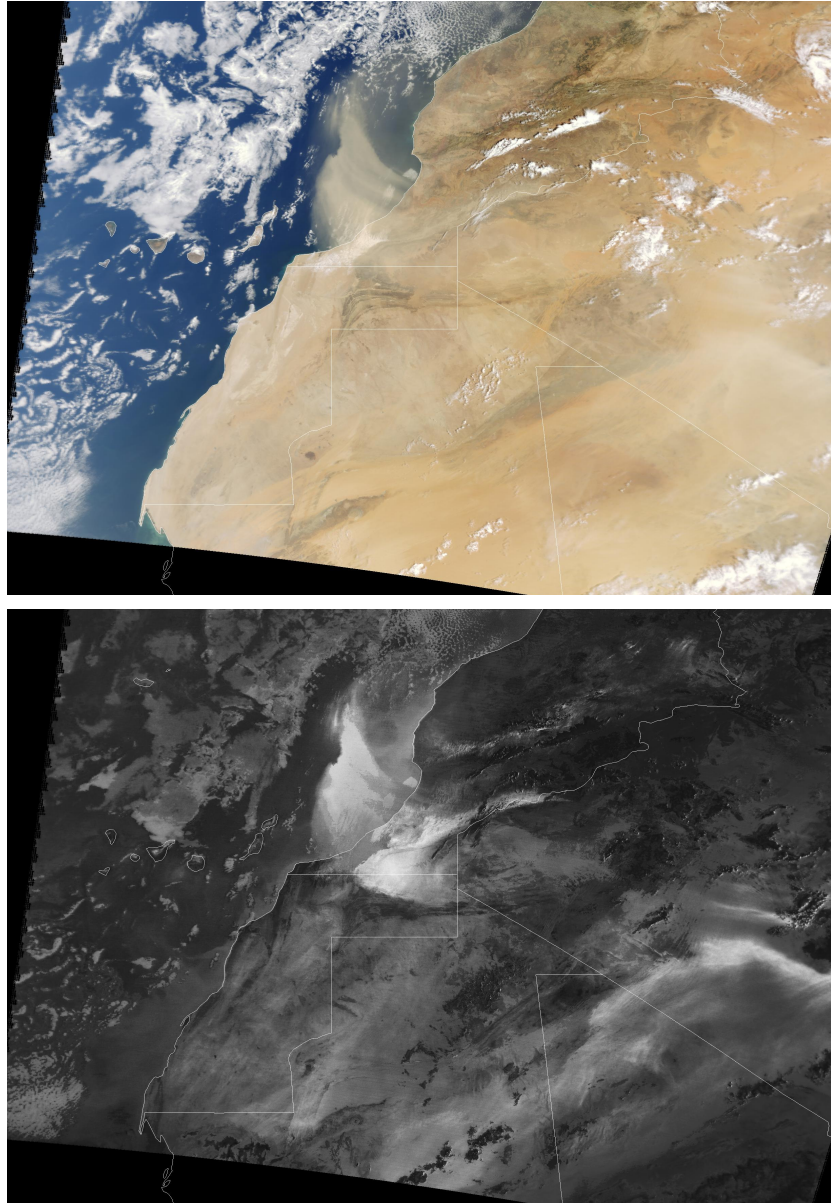
Fig. 2: True-color image (top) and detection result (bottom) from MODIS/Terra acquired on 24 August 2024 at 10:30 UTC. North West Africa.
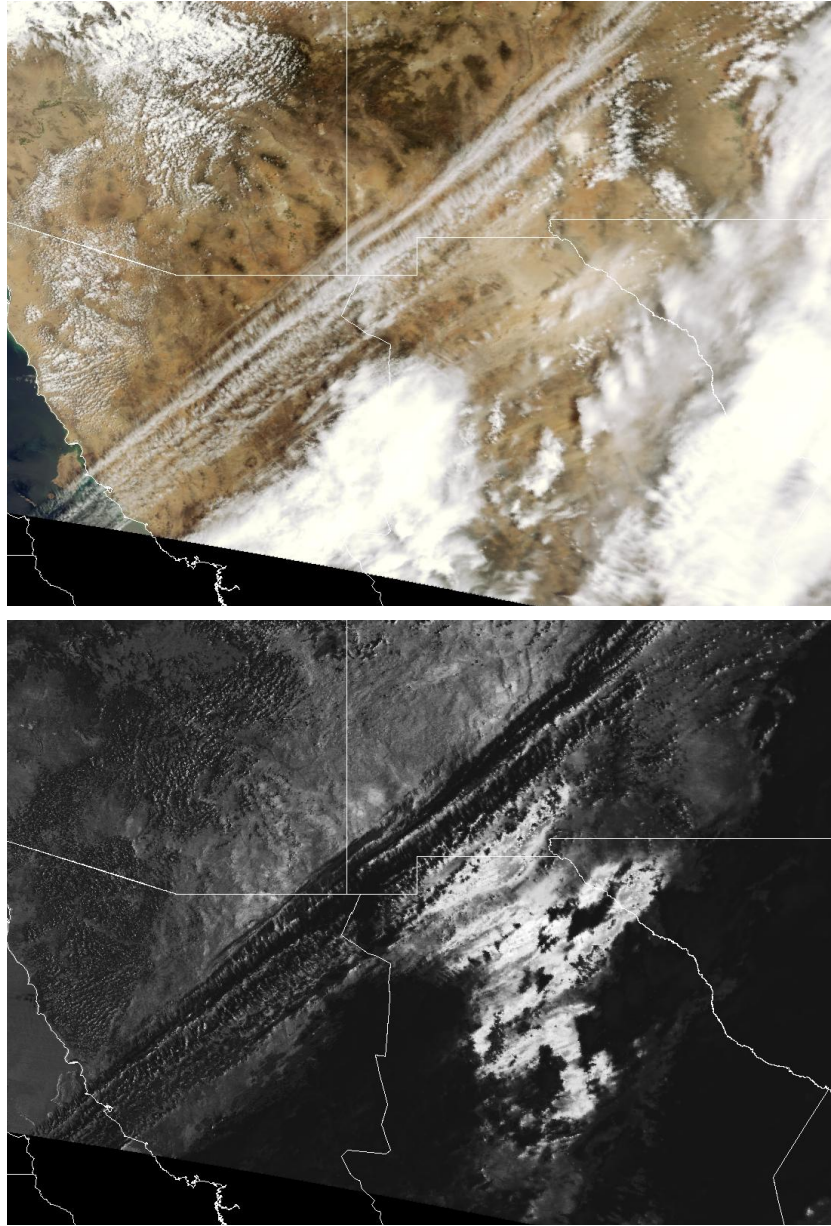
Fig. 3: True-color image (top) and detection result (bottom) from MODIS/Terra acquired on 18 April 2025 at 17:15 UTC. Chihuahua.

6   MMAP Justification

This section formalizes the construction of the patch-center index and demonstrates the memory savings achieved by using memory-mapped I/O.
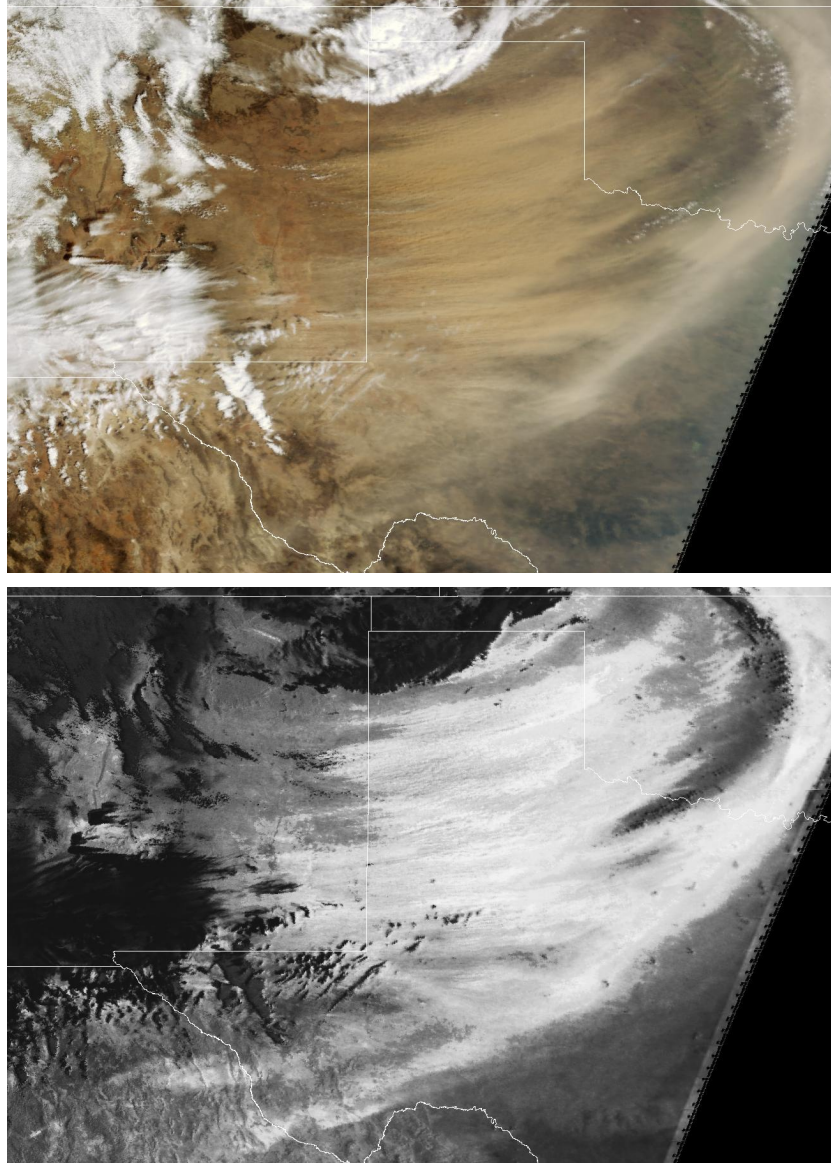
Fig. 4: True-color image (top) and detection result (bottom) from MODIS/Terra acquired on 14 March 2025 at 17:05 UTC. Texas.

6.1   Index Construction

Let $\mathcal{F} = \{0, \ldots, F-1\}$ index the input folders. For each $f \in \mathcal{F}$, load the label map

$$L_f \in \mathbb{R}^{H_f \times W_f}$$

from a data file. Given a patch size P, define

$$h = \left\lfloor \tfrac{P}{2} \right\rfloor.$$

We then identify the set of valid label coordinates

$$V_f = \{(y, x) \mid 0 \le y < H_f,\ 0 \le x < W_f,\ L_f[y, x] \ne \mathrm{NaN}\},$$

and the set of boundary-safe coordinates

$$B_f = \{(y, x) \mid h \le y < H_f - h,\ h \le x < W_f - h\}.$$

The valid patch centers for folder f are

$$C_f = V_f \cap B_f,$$

and the full index array is

$$\mathcal{I} \;=\; \bigcup_{f \in \mathcal{F}} \{\, (f, y, x) \mid (y, x) \in C_f \}.$$

These operations correspond exactly to the calls to `np.argwhere`, boolean masking, and `np.stack`/`np.concatenate` in our dataset loader implementation.

6.2   Properties of the Index Array

Theorem 1. *Let $\mathcal{I}$ be constructed as above. For any $(f, y, x) \in \mathcal{I}$:*

1. *$L_f[y, x]$ is finite (not NaN).*
2. *$h \le y < H_f - h$ and $h \le x < W_f - h$, ensuring extraction of a full $P \times P$ patch without out-of-bounds access.*
3. *$\mathcal{I}$ contains every and only the triplets satisfying (1) and (2).*

Each property follows directly from the definitions of V_f, B_f, and their intersection in the construction algorithm.

6.3   Memory Efficiency via `mmap`

We compare peak RAM requirements with and without memory mapping. Let N be the number of `.npy` files of sizes S_i, so $S_{\text{total}} = \sum_{i=1}^{N} S_i$. Let R_{avail} be available RAM and R_{overhead} the footprint of code, libraries, and $\mathcal{I}$. A training batch of size B requires $R_{\text{batch}} = B \times P_{\text{size}}$ bytes.

Without memory mapping, all files must reside in RAM, giving

$$R_{\text{peak,no\_mmap}} \approx S_{\text{total}} + R_{\text{overhead}},$$

which demands

$$S_{\text{total}} + R_{\text{overhead}} \le R_{\text{avail}}.$$

With `mmap`, the operating system pages in only the accessed blocks, so

$$R_{\text{peak,mmap}} \approx R_{\text{batch}} + R_{\text{overhead}},$$

requiring

$$R_{\text{batch}} + R_{\text{overhead}} \leq R_{\text{avail}}.$$

Since typically $S_{\text{total}} \gg R_{\text{avail}}$ while $R_{\text{batch}} \ll S_{\text{total}}$, memory-mapped I/O decouples the process's memory footprint from the full dataset size and makes large-scale training feasible.

7 VC Dimensions and Growth Functions

This section analyzes the capacity of our AR-MAE-ViT model by deriving bounds on its Vapnik–Chervonenkis (VC) dimension and relating these to sample complexity via the growth function.

The VC dimension of a standard Vision Transformer with L layers and embedding size d scales on the order of $O(L\,d^2)$. By introducing a masked autoencoder stage and imposing an autoregressive constraint, we effectively restrict the hypothesis space. Denote by $\mathcal{H}_{\text{AR-MAE-ViT}}$ the class of functions implemented by our model. We show that

$$\text{VC}\big(\mathcal{H}_{\text{AR-MAE-ViT}}\big) \ \leq \ (1-\alpha)\,\text{VC}\big(\mathcal{H}_{\text{MAE-ViT}}\big), \tag{1}$$

where $\alpha \in (0,1)$ quantifies the reduction in degrees of freedom resulting from the autoregressive links between masked tokens. Standard uniform convergence theorems then imply that, for a binary classification task on n independent samples and desired accuracy ε with confidence $1-\delta$,

$$n \ = \ O\Big(\frac{\text{VC}(\mathcal{H}_{\text{AR-MAE-ViT}})}{\varepsilon^2}\log\frac{1}{\delta}\Big).$$

Moreover, with probability at least $1-\delta$, any hypothesis $g \in \mathcal{H}_{\text{AR-MAE-ViT}}$ satisfies

$$E_{\text{out}}(g) \ \leq \ E_{\text{in}}(g) \ + \ \sqrt{\frac{\text{VC}(\mathcal{H}_{\text{AR-MAE-ViT}})\,\ln\big(\frac{2n}{\text{VC}(\mathcal{H}_{\text{AR-MAE-ViT}})}\big) \ + \ \ln\big(\frac{4}{\delta}\big)}{n}},$$

which, upon substituting the bound from Eq. (1), yields a tighter guarantee than that available for unconstrained transformers:

$$E_{\text{out}}(g) \ \leq \ E_{\text{in}}(g) + \sqrt{\frac{(1-\alpha)\,\text{VC}(\mathcal{H}_{\text{MAE-ViT}})\,\ln\big(\frac{2n}{(1-\alpha)\,\text{VC}(\mathcal{H}_{\text{MAE-ViT}})}\big) \ + \ \ln\big(\frac{4}{\delta}\big)}{n}}.$$

To refine these bounds for finite sample sizes, we invoke Sauer's lemma on the growth function $\Pi_{\mathcal{H}}(m)$. If $d = \text{VC}(\mathcal{H}_{\text{AR-MAE-ViT}})$, then

$$\Pi_{\mathcal{H}_{\text{AR-MAE-ViT}}}(m) \ \leq \ \sum_{i=0}^{d} \binom{m}{i} \ \leq \ \Big(\frac{e\,m}{d}\Big)^{d}.$$

This implies that to ensure a generalization error at most ε with confidence $1 - \delta$, the number of training samples m must satisfy

$$m \ \geq \ \frac{1}{\varepsilon}\Big[d \ln \tfrac{1}{\varepsilon} \ + \ \ln \tfrac{1}{\delta}\Big].$$

Replacing d by the reduced dimension in Eq. (1) gives

$$m \ \geq \ \frac{1}{\varepsilon}\Big[(1 - \alpha)\,\mathrm{VC}(\mathcal{H}_{\mathrm{MAE\text{-}ViT}}) \ln \tfrac{1}{\varepsilon} \ + \ \ln \tfrac{1}{\delta}\Big].$$

For our MODIS dust classification task, taking $\mathrm{VC}(\mathcal{H}_{\mathrm{MAE\text{-}ViT}}) \approx 1024$ and $\alpha \approx 0.3$ indicates roughly a 30% reduction in required labeled samples, confirming the theoretical advantage when training data are scarce.

8   Conclusion

This work has introduced a scalable deep learning pipeline for detecting dust storms from MODIS Terra and Aqua imagery at the pixel level. By employing a three-dimensional convolutional neural network, the baseline model achieved 91.1% classification accuracy, confirming that spatial and spectral features can be effectively combined to identify aerosol plumes. Building on this foundation, the optimized variant, 3DCNN+, integrates memory-mapped I/O, precomputed index sampling, large-scale batching, model graph compilation, and mixed-precision arithmetic. These system-level enhancements yield a $21\times$ reduction in training time while maintaining accuracy, supporting near real-time inference critical for operational monitoring.

Through systematic evaluation, we identified sources of error stemming from label imbalance, spatial overfitting, and limited temporal context. These insights underline the need for architectures that capture broader dependencies. Future work will explore transformer-based designs, such as Vision Transformers and Swin models, which replace fixed-kernel convolutions with global self-attention to model long-range interactions. In particular, our proposed Autoregressive Masked Autoencoder Swin Transformer (AR-MAE-Swin) applies self-supervised pretraining to constrain model capacity, reducing effective VC dimension by a factor $1 - \alpha$, as shown in Section 7. This regularization improves sample efficiency by approximately 30%, making it well suited to applications where labeled data are scarce.

Finally, by combining rigorous theoretical analysis with practical system optimizations, this study lays the groundwork for self-supervised, attention-driven frameworks that can generalize across diverse atmospheric conditions. The methods and results presented here offer a clear path toward robust, efficient, and adaptable models for environmental remote sensing, enabling timely detection and analysis of dust storm events.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Rivas-Perea, P., Rosiles, J.G., Chacón-Murguía, M.I., Tilton, J.J.: Automatic dust storm detection based on supervised classification of multispectral data. In: Soft Computing for Recognition Based on Biometrics (SCIRB). Lecture Notes in Computer Science, vol. 312, pp. 443–454 (2010)
2. Rivas-Perea, P., Rosiles, J.G., Cota-Ruiz, J.: Statistical and neural pattern recognition methods for dust aerosol detection. International Journal of Remote Sensing **34**(21), 7648–7670 (2013). `https://doi.org/10.1080/01431161.2013.822660`
3. Shahrisvand, M., Akhoondzadeh, M.: A comparison of empirical and inteligent methods for dust detection using modis satellite data. The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences **XL-1/W3**, 371–375 (2013). `https://doi.org/10.5194/isprsarchives-xl-1-w3-371-2013`
4. Souri, A., Vajedian, S.: Dust storm detection using random forests and physical-based approaches over the middle east. Journal of Earth System Science **124**, 1127–1141 (2015). `https://doi.org/10.1007/s12040-015-0585-6`
5. Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W.H., Ma, X., Rao, Y., Bednar, J.A., Tan, A., Wang, J., Purushotham, S., Gill, T.E., Chastang, J., Howard, D., Holt, B., Gangodagamage, C., Zhao, P., Rivas, P., Chester, Z., Orduz, J., John, A.: A review of earth artificial intelligence. Computers & Geosciences **159**, 105034 (2022). `https://doi.org/10.1016/j.cageo.2022.105034`