big data and cognitive computing

MDPI

Review

Atmospheric Turbulence Mitigation in the Deep Learning Era: A Critical Review from CNNs and GANs to Transformers, Diffusion, Mamba, and Physics-Informed Models

Nurul Jannah [1], Teddy Surya Gunawan [2,*], Mira Kartiwi [3], Nadirah Abdul Rahim [2] and Ali Sophian [1]

[1] Department of Mechatronics Engineering, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), Jalan Gombak, Kuala Lumpur 53100, Malaysia; jannah.nurul@live.iium.edu.my (N.J.); ali_sophian@iium.edu.my (A.S.)
[2] Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), Jalan Gombak, Kuala Lumpur 53100, Malaysia; nadirahabdulrahim@iium.edu.my
[3] Department of Information Systems, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia (IIUM), Jalan Gombak, Kuala Lumpur 53100, Malaysia; mira@iium.edu.my
* Correspondence: tsgunawan@iium.edu.my

Abstract

Anyone who has watched a distant scene shimmer above hot pavement has seen atmospheric turbulence destroy image detail. In long-range imaging, turbulence produces spatially varying blur, geometric warping, scintillation, and temporal instability. Recovering the underlying scene is therefore an ill-posed inverse problem, and learned priors must compensate for distortions that simplified optical models capture only partially. We use *turbulence mitigation* as the umbrella term for all countermeasures and *turbulence restoration* for its computational core, the estimation of a clean image from degraded observations. From a cognitive-computing perspective, mitigation is not merely image enhancement. It is an uncertainty-constrained visual inference problem in which an intelligent system must reconstruct, interpret, and act on observations relayed through a stochastic physical channel. This critical review examines how the deep learning era has reshaped turbulence mitigation, with physics as the foundation for understanding degradation and designing inductive biases. It traces the architectural progression from convolutional and adversarial networks to Transformers, denoising diffusion models, Mamba and other state-space architectures, and physics-informed frameworks. For each family, we ask a common question: How does it treat the aleatoric uncertainty intrinsic to a random optical channel and the epistemic uncertainty introduced by scarce and simulator-dominated training data? The review also analyzes datasets, simulation strategies, loss functions, and evaluation metrics, and it separates the small body of shared-protocol benchmark evidence from the far larger body of self-reported results that cannot be compared across studies. Persistent obstacles include the synthetic-to-real domain gap, the scarcity of paired real turbulence data, the mismatch between fidelity metrics and downstream task performance, and the computational cost that limits operational deployment. We close with an AI-centered agenda in which uncertainty quantification stands alongside domain adaptation as a first-order priority.

Keywords: atmospheric turbulence; image restoration; deep learning; uncertainty quantification; Transformer; diffusion model; Mamba; physics-informed learning; computer vision; cognitive computing

Check for updates

Academic Editor: Balazs Harangi

Received: 17 July 2026
Revised: 13 August 2026
Accepted: 18 August 2026
Published: 21 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The same atmospheric physics that makes stars twinkle degrades every long-range imaging system on Earth. Surveillance, astronomy, aerospace engineering, autonomous navigation, free-space optical communication, remote sensing, defense, and environmental monitoring all depend on imaging over long optical paths [1,2]. Along those paths, random fluctuations in temperature, pressure, humidity, and air density perturb the refractive index of the air itself [3]. The resulting wavefront errors produce spatially varying blur, geometric distortion, scintillation, pixel displacement, and temporal instability [4–6]. These degradations are doubly damaging. They reduce human visual interpretability, and they undermine the machine-vision tasks that increasingly consume such imagery, including detection, recognition, tracking, optical character recognition (OCR), and scene understanding [7].

Terminology. Two terms recur throughout this review and are used in a strict hierarchical relationship. *Atmospheric turbulence mitigation* is the umbrella concept: the full set of countermeasures that reduce the impact of turbulence on an imaging system. It spans acquisition-side measures (adaptive optics, aperture and wavelength selection, and frame-rate and exposure design), selection and fusion strategies (lucky imaging and multi-frame registration), computational recovery, and task-side measures such as the training of a downstream recognizer directly on degraded imagery. *Turbulence restoration* is the computational core of mitigation: the estimation of a clean image or video from turbulence-degraded observations. Every restoration method is therefore a mitigation method but not conversely. Where a statement applies specifically to image estimation, we say restoration; where it applies to the broader system-level problem, we say mitigation. Related phrases in the literature, such as "turbulence removal" and "turbulence correction", are treated as synonyms of restoration.

Restoration under turbulence differs from conventional denoising or deblurring. The degradation is stochastic, anisoplanatic, temporally dynamic, and often only partially observed. The inverse problem is consequently ill-posed and scene-dependent [8,9]. Classical mitigation approaches, including adaptive optics, lucky imaging, multi-frame registration, blind deconvolution, and wavefront correction, remain important foundations. However, they typically require specialized hardware, high frame rates, expensive computation, or stationarity assumptions that are difficult to maintain in field conditions [10–12].

Deep learning has changed the center of gravity of the field. Over the past several years, the architectural progression has run from convolutional neural networks (CNNs) and generative adversarial networks (GANs) to Transformers, diffusion models, Mamba and other state-space models (SSMs), and physics-informed networks. Each learns image and motion priors directly from data, which makes it possible to model complex spatiotemporal distortions without fully specifying the atmospheric state [13–18]. Data-driven restoration also exposes a central weakness. Models trained on simulated turbulence often fail to generalize to real optical paths, sensors, ranges, and weather conditions [15,19]. This synthetic-to-real domain gap makes atmospheric turbulence mitigation a useful test case for AI research on inverse problems, domain adaptation, physically informed learning, generative restoration, and task-aware evaluation.

The uncertainty lens. The problem also sits naturally within cognitive computing and machine perception, and this review adopts uncertainty as its organizing analytical instrument rather than as a framing slogan. Because the scene is observed through a stochastic physical channel and the clean image is only partially identifiable, an intelligent imaging system cannot simply enhance pixels. It must infer scene content under uncertainty, quantify how much of its reconstruction is evidence and how much is prior, and deliver outputs that downstream recognition and decision processes can trust. Following

the standard decomposition in Bayesian deep learning [20], we distinguish two sources throughout the architectural analysis:

- **Aleatoric uncertainty** is irreducible ambiguity in the observation itself. In turbulence, it is intrinsic and dominant: a random, spatially varying point-spread function (PSF) and random tilt field destroy information, so many distinct clean scenes are consistent with a single degraded frame. No amount of additional training data removes it. It can only be represented, for example, by predicting a distribution rather than a point estimate.
- **Epistemic uncertainty** is reducible ignorance of the model. In turbulence, it is driven by scarce paired real data and by training corpora that are overwhelmingly simulator-generated, so the model is poorly constrained on severities, sensors, ranges, and atmospheric regimes it has never seen. Additional or better-matched data reduces it, and a network that stays confident on out-of-distribution input is the characteristic symptom of leaving it unmeasured.

The two sources call for different remedies and fail in different ways downstream, yet they are routinely conflated. Sections 5–10 therefore close each architectural family with an explicit assessment of how it handles or ignores each source; Table 1 consolidates that assessment, and Section 16 treats uncertainty quantification as a first-order research priority ranked alongside the synthetic-to-real domain gap rather than beneath it.

Table 1. Treatment of aleatoric and epistemic uncertainty by architectural family, as assessed across the studies reviewed in Sections 5–10.

Family	Aleatoric (Irreducible Channel Ambiguity)	Epistemic (Model Ignorance, Data Scarcity)	Characteristic Failure Mode	Most Tractable Improvement
CNN/residual (Section 5)	Implicit; averaged into the conditional mean. Explicit per-pixel variance head in [13], used as training weight only	MC dropout in AT-Net [21], used internally as guidance; not reported or calibrated	Oversmoothing, i.e., unrepresented ambiguity made visible	Expose the existing variance head as a calibrated output; report MC-dropout maps under distribution shift
GAN/adversarial (Section 6)	Implicit; the objective refuses the conditional mean, but the generator resolves ambiguity silently and returns one sample	None. Manifold constraint reduces exposure [6,22] without measuring it	Confident hallucination; under shift, synthesis of training-domain structure rather than blur	Deep ensemble over generators [23]; per-sample confidence labeling of recovered detail
Transformer (Section 7)	Implicit; multi-frame attention *reduces* ambiguity by aggregating channel realizations but does not represent the remainder	None; highest fitting capacity to simulator statistics of any family reviewed	Residual temporal jitter; confident failure on out-of-distribution sequences	Report ensemble or dropout disagreement; evaluate on held-out real sequences, not simulator holdouts
Diffusion (Section 8)	**Explicit by construction:** $p_\theta(\mathbf{x}_0 \mid \mathbf{y})$ is a posterior; per-pixel variance available from K samples at no design cost	None; sample agreement reflects the learned prior and can be tight and wrong under shift	Hallucinated detail; single-sample evaluation conceals the spread the model encodes	Report $\mathrm{Var}[\hat{\mathbf{x}}]$ over K samples and calibrate it [24]; the cheapest high-value gain in the field
Mamba/SSM (Section 9)	None; deterministic sequence model with reconstruction objectives. In TSTRNet, the probabilistic burden sits entirely in the diffusion stage [9]	None; behavior of the input-dependent selection gate under shift is unexamined	Same averaging as CNN, with an additional unquantified traversal-order bias	Pair the efficient SSM stage with an explicit uncertainty head rather than deferring to a costly diffusion refiner
Physics-informed (Section 10)	None intrinsic; the forward model constrains the admissible set but does not resolve within it, hence the stochastic refiner in [25]	**Reduced** (its principal mechanism) via forward-model consistency and C_n^2/r_0 conditioning; not measured and relocated into model assumptions	Confident enforcement of a wrong constraint under non-Kolmogorov or slant-path conditions (Section 3.3)	Propagate the uncertainty of estimated r_0/C_n^2 into the output to yield physically grounded intervals

Hill et al. [26] provide a valuable recent review of deep neural networks for atmospheric turbulence removal, including datasets, evaluation metrics, and comparative experiments involving CNN, Swin–Transformer, and Mamba variants. This review is positioned differently. It follows a structured, protocol-driven literature selection process (Section 2) and offers an architecture-centered critical synthesis for readers in artificial intelligence, computer vision, and cognitive computing. It emphasizes seven dimensions that remain underdeveloped in prior review coverage: physics-informed and physically constrained learning, diffusion-based restoration, Mamba and selective state-space models,

synthetic-to-real generalization, the aleatoric/epistemic treatment of uncertainty, downstream task coupling, and operational deployment constraints. Table 2 sets out the contrast dimension by dimension.

Table 2. Positioning of this review relative to Hill et al. [26].

Dimension	Hill et al. [26]	This Review
Primary aim	Deep-network survey with comparative experiments	Protocol-driven critical synthesis for AI restoration and deployment
Architectural scope	CNN, Swin–Transformer, and Mamba-centered coverage	CNN, GAN, Rransformer, diffusion, Mamba/SSM, and physics-informed paradigms
Physics role	Background and selected model design context	Foundation for inverse-problem framing, inductive bias, and domain transfer
Domain gap	Discussed as an evaluation concern	Treated as a central barrier to deployable AI systems
Uncertainty	Not a structuring dimension	Aleatoric/epistemic assessment applied to every architectural family (Table 1)
Benchmark evidence	Comparative experiments run by the authors	Shared-protocol evidence separated from non-comparable self-reported results (Tables 3 and 4)
Downstream tasks	Limited coverage	Dedicated analysis of restore-then-task and task-aware pipelines
Deployment	Secondary emphasis	Explicit discussion of latency, robustness, and edge constraints

Table 3. Benchmark anchor: shared-protocol results only. Panels **A** and **B** report directly comparable results within but not across their respective evaluation protocols.

Method/Family	PSNR	SSIM	CW-SSIM	LPIPS	Frames	Params	Runtime	Report
Panel A–TMT synthetic static-scene benchmark (all values as evaluated in [16]; 540 × 960, RTX 2080 Ti)								
CNN, single-frame (TurbNet) [15]	22.76	0.6923	N.R.	0.4012	1	N.R.	N.R.	Re-eval. [16]
GAN, multi-frame (TSRWGAN) [27]	25.29	0.7784	N.R.	0.2243	15	46.28 M	1.15 s/f	Re-eval. [16]
Transformer, multi-frame (TMT) [16]	**27.73**	**0.8341**	N.E. [§]	**0.1815**	12	N.R.	1.52 s/f	Orig.
Panel B–ATSyn-dynamic benchmark (all values as evaluated in [28]; overall average over weak/medium/strong; 960 × 540, RTX 2080 Ti)								
GAN, multi-frame (TSRWGAN) [27]	26.45	N.R.	N.R.	N.R.	15	46.28 M	N.R.	Re-eval. [28]
Transformer, multi-frame (TMT) [16]	28.27	N.R.	N.R.	N.R.	12	N.R.	0.80 FPS	Re-eval. [28]
Recurrent (DATUM) [28]	**29.43**	N.R.	**0.8787**	N.R.	recurrent	**5.754, M**	**9.17, FPS**	Orig.

Interpretation: Every row within a panel was evaluated on the same benchmark with the same metrics by the same evaluating study, so *rows within a panel are directly comparable*. Panels A and B use different protocols and must not be compared with each other. The Report column distinguishes values from the original method paper (Orig.) from later benchmark re-evaluations (Re-eval., with the re-evaluating source). SSIM and CW-SSIM occupy separate columns because the two variants are not numerically interchangeable. This table contains the complete set of directly comparable quantitative evidence identified in this review; all other results appear in Table 4. [§] N.E. = reported in the cited source but not extracted for this table; N.R. = not reported by the source (never estimated here). **Panel A sources:** PSNR/SSIM/LPIPS from Table V of [16]; TSRWGAN runtime and parameter count from Table VII of [16], measured in the same setup as the TMT runtime. **Panel B sources:** PSNR and CW-SSIM from [28]; DATUM throughput and parameter count measured in the same study on the same hardware as the TMT figure, which is the only controlled efficiency comparison in this review. **How to read this table:** Rows within a panel differ simultaneously in temporal input, alignment, architecture, loss design, training strategy, and capacity. The deltas therefore compare *complete systems*, not architectural components; see Section 15.4 for what may and may not be concluded.

The objectives of this review are to: (i) explain the physical degradation mechanisms that motivate AI model design; (ii) organize learning-based turbulence mitigation by architectural paradigm, tracing the progression from CNNs and GANs to Transformer, diffusion, Mamba/state-space, and physics-informed models; (iii) analyze datasets, simulation protocols, losses, and metrics as sources of generalization risk; (iv) compare restoration families in terms of supervision, uncertainty treatment, computational cost, robustness, and task relevance; and (v) identify open research directions for deployable AI systems. The main contributions are:

1. An AI-centered taxonomy of atmospheric turbulence mitigation methods covering CNN, GAN, Transformer, diffusion, Mamba/state-space, and physics-informed architectures;
2. A cross-cutting uncertainty analysis that asks of every architectural family how it treats aleatoric and epistemic uncertainty, as consolidated in Table 1, and that shows that the field currently represents the first only implicitly and measures the second almost not at all;
3. A synthesis of how physical turbulence models can support learned priors, data simulation, loss design, and domain adaptation, including the observation that essentially all learned methods are trained on Kolmogorov or von Kármán statistics and are therefore untested under non-Kolmogorov and slant-path conditions (Section 3.3);
4. A critical analysis of datasets, evaluation metrics, and the synthetic-to-real gap that limits current benchmark conclusions, including a deliberate separation of shared-protocol evidence (Table 3) from non-comparable self-reported results (Table 4);
5. A discussion of task-aware restoration, including how turbulence mitigation interacts with detection, recognition, tracking, OCR, and operational latency constraints;
6. An evidence-based decision framework (Section 15.5) linking turbulence conditions, frame availability, supervision constraints, computational budgets, uncertainty requirements, and downstream objectives to defensible restoration paradigms and the safeguards each requires.

Table 4. Self-reported results under heterogeneous protocols. △ CROSS-ROW COMPARISONS ARE NOT VALID. For the field's directly comparable evidence, see Table 3.

Method/Family	PSNR	SSIM	LPIPS	Frames	Runtime	Evaluation Protocol (Unique to This Row)	Severity Regime/Caution
CNN face (TDRN) [13]	25.42 †	0.815 †	N.R.	1	N.R.	CelebA + synthetic turbulence [13]	Face-only domain; strong appearance priors
GAN (AE-GAN) [29]	28.43 #	0.845 #	N.R.	1	1.21 s/img #	Custom simulated astronomical set (NASA Solar System images; $D/r_0 = 1$–30; 512×512)	Astronomical domain; not comparable with terrestrial imaging
Diffusion (AT-DDPM) [30]	22.41	0.6262	0.3223	1	N.R. ‖	CelebA + turbulence, $D/r_0 = 1.6$	Face-only; single fixed severity
SSM+Diff. (TSTRNet) [9]	27.71 *	0.819 *	N.R. **	1	4.12 s ***	Custom synthetic 5 scenes (Face/City/Forest/Mountain/Sea); Places365 + Kaggle, $D/r_0 \sim \mathcal{U}(1,3)$, 256×256	Self-defined benchmark introduced in the same paper
Mamba/SSM (MAMAT) [18]	30.99 §	0.6514 §	N.R.	N.R.	N.R.	Custom video corpus, mild turbulence	**Mild regime**: baseline ≈30.4 dB vs. ≈22–24 dB for TMT synthetic. The high PSNR reflects an easier problem, *not* a stronger method
Physics-inf. (PiRN-SR) [25]	25.61 ††	0.7204 ††	0.3978 ††	1	N.R.	PiRN-Syn synthetic test set	Self-defined synthetic protocol

† Values as reported in the preprint version (arXiv:2007.08404) of [13]; LPIPS and runtime not reported. Provenance is stated explicitly because the peer-reviewed version may report different figures. # AE-GAN: PSNR/SSIM averaged across four aberration types (defocus, spherical, coma, and astigmatism) at $D/r_0 = 10$ and 1.5λ amplitudes (Table 4 of [29]); runtime on GTX 1080 Ti for 512×512 images. ‖ Runtime not reported in [30]; inference uses $T = 60$ DDPM sampling steps. * TSTRNet $\mathcal{N}^R$ on the Face synthetic scenario; perceptual quality assessed via MUSIQ↑ = 41.86 and NIQE↓ = 7.21. ** LPIPS not reported in [9]; MUSIQ and NIQE used throughout. *** $\mathcal{N}^R$ (patch-based DDPM, 64×64 patches) on RTX 3090; the $\mathcal{N}^G$ Mamba U-Net stage alone runs at 0.047 s/img with 23.13 G FLOPs. § LPIPS and runtime not reported in [18]. †† PiRN-SR on PiRN-Syn: PSNR/SSIM from Table 2 and LPIPS from Table 5 of [25]. **Reporting gaps visible in this table are themselves a finding:** runtime is unreported in four of six rows, LPIPS in four of six, parameter counts in all six, and uncertainty or calibration in all six. No row reports evaluation under a shared protocol with published baselines.

The remainder of this paper is organized as follows. Section 2 describes the review methodology, including the research questions, search strategy, and study selection process. Section 3 summarizes the physical foundations of atmospheric turbulence. Section 4 reviews classical mitigation methods. Sections 5–10 analyze the major deep-learning paradigms, and Section 11 consolidates their uncertainty treatment into a single cross-architecture assessment. Section 12 discusses datasets and simulation resources, Section 13 reviews losses and evaluation metrics, and Section 14 discusses downstream task integration and

deployment evidence. Section 15 synthesizes recent methodological trends, Section 16 identifies open challenges, and Section 17 concludes the paper.

2. Review Methodology

2.1. Review Design

This paper is a structured critical review of the literature on atmospheric turbulence removal and mitigation in imaging systems. Unlike a systematic meta-analysis, which quantitatively synthesizes homogeneous experimental outcomes, a structured critical review is appropriate here because the primary studies span heterogeneous methodological paradigms, ranging from classical wave-optics-based compensation to generative deep-learning restoration, and report performance metrics that are not directly commensurable across all subdomains. The primary studies use different datasets, turbulence simulators, degradation strengths, frame counts, metrics, and runtime settings, which prevents a statistically meaningful pooled comparison across the entire field. The review was nonetheless conducted with a structured, reproducible protocol (predefined search strings, explicit inclusion and exclusion criteria, and documented screening stages) to minimize selection bias and to support transparency. In addition, each core study was appraised with a technical evidence-quality rubric (Section 2.9); this is a reporting and external-validity appraisal appropriate to engineering and computer-vision studies, not a clinical-style risk-of-bias assessment.

2.2. Research Questions

Five research questions (RQs) were formulated a priori to scope the review and to structure the subsequent evidence mapping (Section 2.7):

RQ1. What are the major physical mechanisms and optical effects associated with atmospheric turbulence in imaging systems?

RQ2. What classical and deep learning-based approaches are commonly employed for atmospheric turbulence mitigation and image restoration?

RQ3. What types of datasets and turbulence simulation methods are used for training and evaluation of turbulence restoration models?

RQ4. What recent innovations have shaped atmospheric turbulence removal methodologies in recent years?

RQ5. How do different turbulence mitigation frameworks compare in terms of restoration quality, computational efficiency, robustness, and real-world applicability?

2.3. Search Strategy

The search covered IEEE Xplore and Scopus, selected for their coverage of peer-reviewed work in image processing, computer vision, computational imaging, and optical sensing. Both were searched on 11 June 2026, with an update search on 15 July 2026. Boolean queries paired seven turbulence phrases ("atmospheric turbulence mitigation", "atmospheric turbulence removal", "atmospheric turbulence correction", "turbulence image restoration", "turbulence-degraded image restoration", "image restoration through turbulence", and "through-turbulence imaging") with learning- and architecture-related terms spanning deep learning, CNN, GAN, Transformer, diffusion/DDPM/DDIM, Mamba and state-space models, and physics-informed/-guided/-based methods. Scopus was queried over `TITLE-ABS-KEY` and IEEE Xplore over *All Metadata*; both were limited to 2019–2026 and to journal articles and conference papers. The complete strings as executed appear in Appendix A.

The 2019–2026 window reflects the period in which deep learning and physics-guided restoration began to shape this field. Records carrying a 2026 publication year, including online-first articles, were eligible if indexed on or before the final search date. Pre-2019 foundational works were not retrieved by the queries; they entered selectively through forward and backward citation tracing where they supplied essential theory, benchmark methods, or widely adopted turbulence models, and they are used as background rather than being counted among the core studies. The English-language restriction was applied during screening rather than as a database filter.

2.4. Study Selection

Articles were evaluated according to predefined inclusion and exclusion criteria, applied first during title-and-abstract screening, then during full-text assessment.

2.4.1. Inclusion Criteria

Studies were included if they satisfied all of the following criteria:

- Published between 2019 and 2026, or identified as a foundational pre-2019 reference through citation tracing;
- Addressed atmospheric turbulence in imaging, computational imaging, or optical sensing systems;
- Proposed, evaluated, or surveyed methods related to turbulence mitigation, turbulence simulation, or turbulence-degraded image restoration;
- Reported quantitative experimental findings or provided substantial conceptual, theoretical, or methodological contributions;
- Published as peer-reviewed journal articles, conference papers, benchmark studies, or authoritative review papers or admitted as preprints under the explicit policy of Section 2.4.3;
- Available in full-text English format.

2.4.2. Exclusion Criteria

Studies were excluded if any of the following conditions applied:

- Focused exclusively on non-atmospheric blur or unrelated degradation mechanisms;
- Addressed underwater, biomedical, microscopy, or other imaging contexts without relevance to atmospheric propagation;
- Represented duplicate publications or substantially overlapping versions of the same study;
- Consisted only of editorials, posters, abstracts, or brief communications without sufficient technical depth;
- Lacked adequate methodological details or experimental evidence for meaningful evaluation;
- Fell outside the defined scope of atmospheric turbulence imaging and restoration.

2.4.3. Preprint Policy

Peer-reviewed publications were prioritized throughout. A preprint was eligible for inclusion only when (i) it was directly relevant to a recent or under-represented architectural category and (ii) no peer-reviewed version was available on the final search date. When a work initially disseminated as a preprint subsequently appeared in a peer-reviewed venue, the published version was used. This applies, for example, to the Turbulence Mitigation Transformer [16], DATUM [28], Mamba [31], and the RLR-AT long-range dataset [32]. Accordingly, the final reference list contains no preprint-only sources; where an arXiv identifier is retained alongside a citation, it serves only as an additional locator for a work cited in its published form.

In a small number of cases, quantitative values reported in the comparison tables originate from an earlier preprint version of a study for which a peer-reviewed publication is now available. Such cases are identified explicitly. For example, the PSNR and SSIM values reported for TDRN in Table 4 were taken from an earlier version of the work subsequently published in *IEEE Transactions on Biometrics, Behavior, and Identity Science* [13]. Evidence derived specifically from an earlier manuscript version is distinguished from evidence reported in the final peer-reviewed publication.

2.5. Screening Procedure

Primary screening at both the title-and-abstract and full-text stages was performed by one author (N.J.), who applied the inclusion and exclusion criteria of Section 2.4 to every record. To provide an independent check on the primary screening, a second author (T.S.G.) independently re-screened a random 26% subset of the records entering the screening process ($n = 55$ of 213) against the same criteria, blinded to the first author's decisions. Disagreements between the two screeners were resolved through discussion until consensus was reached, with the agreed upon decision recorded in the screening log. Full-text screening was not independently duplicated.

No inter-rater agreement statistic is reported. Cohen's κ was not calculated during screening, and computing one retrospectively over a partial subset would imply a precision that the screening record does not support. The duplicate screening should therefore be read as a targeted verification of the primary screening rather than as a full dual-reviewer protocol with quantified reliability, which remains the standard recommended by PRISMA 2020. Section 2.10 states the residual deviation.

Three features of the review design bound the consequences of the remaining single-reviewer component. First, the inclusion and exclusion criteria (Section 2.4) and the executed search strings (Appendix A) are reported in full, so any reader can re-execute the search and compare the resulting set with the one used here. Second, this is a critical narrative synthesis rather than a meta-analysis, so no pooled effect estimate depends on the exact composition of the included set, and the substantive conclusions rest on the architecture-level evidence assessed in Sections 5–10 rather than on record counts. Third, the highest-consequence records namely, the studies contributing quantitative evidence to Tables 3 and 4 were re-examined against their primary sources during data extraction, and every extracted value is traced to a specific table or section of the cited paper in the accompanying footnotes.

2.6. PRISMA Selection Process

The study selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 reporting guidelines [33] and is summarized in Figure 1. Database search identified 213 records from Scopus and IEEE Xplore. After the removal of duplicate records and clearly irrelevant records during title-and-abstract screening, 114 reports were retained for full-text assessment. The archived screening log records a combined total of 99 records removed before full-text assessment; however, the original record did not separately distinguish duplicate removals from title-and-abstract exclusions. These two counts are therefore reported together in the PRISMA flow diagram as records removed before full-text assessment. For the same reason, the log does not support a breakdown of yields by database or of exclusions by reason, and neither breakdown is estimated here. Section 2.10 acknowledges this reporting limitation.

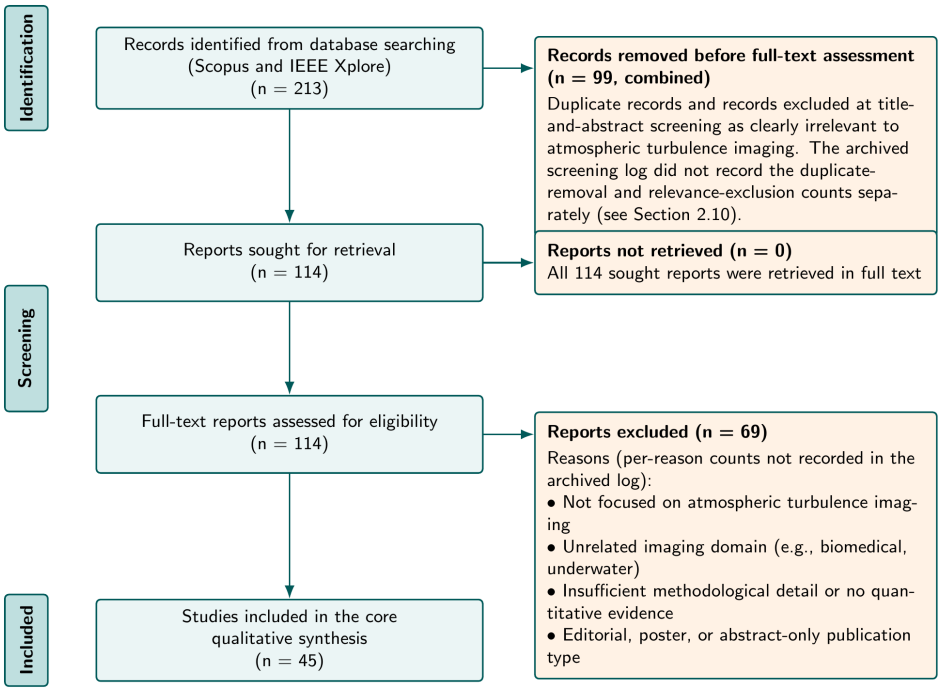


Figure 1. PRISMA 2020 flow diagram of the literature selection process, from the 213 records identified in Scopus and IEEE Xplore to the 45 studies included in the core qualitative synthesis.

All 114 reports sought for retrieval were available and assessed for eligibility. Following full-text review, 69 reports were excluded because they did not meet the predefined inclusion criteria. The main reasons for exclusion were lack of direct focus on atmospheric turbulence imaging; focus on unrelated imaging domains such as biomedical or underwater imaging; insufficient methodological detail or quantitative evidence; and publication type not suitable for full review synthesis, such as editorial, poster, or abstract-only records. Ultimately, 45 studies were included in the core qualitative synthesis. Thirty of these are analyzed individually in the body of the review: Tables 5 and 6 enumerate the method studies with their datasets, metrics, contributions, and limitations; Tables 3 and 4 consolidate their quantitative evidence; and Sections 5–14 discuss them in detail. The complete inclusion list is maintained in the authors' screening records.

Table 5. Comparison of atmospheric turbulence removal methods in imaging systems (2020–2022).

Author, Year [Ref.]	Dataset Used	Method	Metric Used	Novelty	Limitation
Physics-Based and Classical Atmospheric Turbulence Removal					
Mao et al. 2020 [34]	Synthetic, NATO RTO, long-range (4 km)	Space-time non-local averaging + lucky fusion	PSNR	Stable reference-frame construction for optical flow under dynamic turbulence	Weak integration with physical turbulence modeling
Parmar et al. 2021 [12]	Door and car sequences (OTIS)	Blind PSF estimation via central moments	PSNR, SSIM	Blind turbulence restoration using normalized central moments	Real-time deployment was not demonstrated
Chan et al. 2022 [35]	Zernike-based simulations	Tilt-then-blur turbulence modeling	Physical consistency analysis	Clarifies physically correct turbulence formation sequence	Focused on theoretical modeling rather than inversion
Guo et al. 2022 [36]	Aero-optical dataset (300 k + frames)	Mean filtering + conditional GAN	PSNR, SSIM, tracking efficiency	Lightweight real-time aero-optical restoration framework	Occasional ghosting artifacts remain
Nair et al. 2022 [37]	Synthetic turbulence datasets	Comparative turbulence simulation study	PSNR, SSIM	Systematic comparison of turbulence simulation approaches	Limited emphasis on restoration architecture design
Hua et al. 2022 [2]	Multiframe video sequences	Geometric distortion and blur representation	No-reference blur, sharpness	Lucky-region fusion using robust error estimation	Requires multiple frames and depends on local region quality

Table 5. *Cont.*

Author, Year [Ref.]	Dataset Used	Method	Metric Used	Novelty	Limitation
Deep Learning-Based Atmospheric Turbulence Removal					
Lau et al. 2021 [38]	AFLW, 300W	ATKey semi-supervised restoration	NME	Semantic landmark priors for turbulence restoration	Limited generalization to non-face scenes
Feng et al. 2022 [14]	ATSyn, Hubble, OTIS	TurbuGAN	PSNR, SSIM	Adversarial multiframe blind deconvolution	Performance decreases under severe turbulence
Mao et al. 2022 [15]	OTIS, turbulence benchmark datasets	Physics-inspired Transformer	PSNR, SSIM	Benchmark study integrating turbulence physics with Transformers	Limited real-world generalization
Rai et al. 2022 [39]	ImageNet, FGCV, PASCAL3D+	DTD-GAN+	PSNR, SSIM	Improves restoration for downstream visual tasks	Less effective for pure deblurring scenarios
Yasarla et al. 2022 [13]	Single-face turbulence datasets	Uncertainty-guided CNN restoration	PSNR, SSIM	Semantic uncertainty-guided restoration	Highly dependent on face priors

Table 6. Comparison of atmospheric turbulence removal methods in imaging systems (2023–2026).

Author, Year [Ref.]	Dataset Used	Method	Metric Used	Novelty	Limitation
Jaiswal et al. 2023 [25]	Synthetic and real turbulence data	Physics-driven stochastic refinement	PSNR, SSIM, LPIPS	Physics-guided stochastic turbulence refinement	Computationally intensive training
Liu et al. 2023 [6]	Simulated and real Australian data	EAF-WGAN	PSNR, SSIM, BRISQUE, NIQE	Alignment-fusion GAN for stable restoration	Generalization gap between synthetic and real data
Mei et al. 2023 [22]	FFHQ, LR-FID	LTT-GAN	LPIPS, FID	GAN inversion for turbulence restoration	Weak reconstruction of fine details
Wang et al. (a) 2024 [8]	P2S simulator, HeatChamber	Monte Carlo-based restoration (FTM-Net)	PSNR, SSIM, inference time	Resource-efficient turbulence restoration for constrained hardware	Dependent on Monte Carlo sampling frequency
Chu et al. 2024 [19]	SHTM dataset	Semi-supervised hybrid model	PSNR, SSIM, GMSD, FSIM	Semi-supervised restoration under limited labels	Real-world evaluation lacks ground truth
Lopez-Tapia et al. 2024 [40]	REDS, heat chamber videos	Variational turbulence correction (VarAT)	PSNR, SSIM, LPIPS	Temporal consistency via variational modeling	Lower perceptual quality than GAN approaches
Wang et al. (b) 2024 [41]	REDS, BRIAR	Teacher–student domain adaptation	PSNR, SSIM	Bridges synthetic-to-real turbulence gap	Training complexity is high
Xia et al. 2024 [17]	Paired RGB and narrow-band images	NB-GTR	PSNR, SSIM	Guided turbulence removal using dual imaging	Requires specialized acquisition hardware
Zhang et al. (TMT) 2024 [16]	TMT dataset	Turbulence mitigation Transformer	PSNR, SSIM, CW-SSIM	Transformer-based multiframe restoration	Visual jitter remains in outputs
Zhang et al. (DATUM) 2024 [28]	ATSyn-dynamic, ATSyn-static	Recurrent turbulence restoration	PSNR, SSIM	First recurrent turbulence restoration framework	Requires larger frame sequences for best performance
Khowaja et al. 2025 [11]	Synthetic and real scenes	Deep neurofuzzy restoration	PSNR, SSIM, ADLCS	Neurofuzzy integration for adaptive restoration	Increased architectural complexity
Wu et al. 2025 [9]	5-scene synthetic (Places365, Kaggle), CLEAR, OTIS	TSTRNet: Mamba U-Net + edge-cond. DDPM	PSNR, SSIM, MUSIQ, NIQE	Two-stage SSM–diffusion hybrid; edge-aware stochastic refinement	DDPM stage is slowest among compared methods (4.12 s/image); no LPIPS reported
Zhang et al. 2025 [42]	Synthetic, physical, and real data	Adaptive deep feature perception	PSNR, SSIM, LPIPS, FID	Adaptive attention guided by turbulence intensity	Still computationally demanding
Hancer et al. 2026 [7]	Field-collected RGB/MWIR sequences	CPU multiframe fusion	BRISQUE, NIQE, PIQE	Real-time turbulence mitigation without GPU	Reduced reliability in low-texture scenes

Additional foundational references were identified through backward and forward citation tracing. These references were used to support background discussion on turbulence theory, classical optical models, generic deep-learning architectures, and dataset documentation. They were not counted among the 45 studies included in the core qualitative synthesis.

2.7. Evidence Mapping

Because the evidence types below and the comparison tables throughout this review are stated in terms of standard image-quality metrics, we define the abbreviations once here and use them thereafter without expansion. Full-reference metrics, which require a clean ground-truth image, comprise the peak signal-to-noise ratio (PSNR); the structural similarity index measure (SSIM) and its complex-wavelet variant (CW-SSIM); and the learned perceptual image-patch similarity (LPIPS), a deep-feature perceptual distance. No-reference metrics, which do not, comprise the blind/referenceless image spatial quality evaluator (BRISQUE), the naturalness image quality evaluator (NIQE), and the multi-scale image quality transformer (MUSIQ). The Fréchet inception distance (FID) is a distribution-level measure computed over sets of images rather than image pairs. Formal definitions of

PSNR and SSIM and a caution on the non-interchangeability of SSIM and CW-SSIM appear in Section 13. A consolidated abbreviation list is provided at the end of the manuscript.

Table 7 maps each research question (Section 2.2) to the corresponding review section(s), the types of evidence consulted, and the primary literature categories that address it. This mapping anchors the review structure to its stated objectives and allows readers to locate relevant evidence directly. Figure 2 summarizes the organizing taxonomy of the field along three orthogonal dimensions. The first is the architectural family of the restoration backbone, covering CNN and residual designs, GANs, recurrent models, Transformers, diffusion models, and Mamba and other selective state-space models (Sections 5–9). The second is the mechanism, if any, by which turbulence physics enters the learning pipeline: physics-based simulation of training data; conditioning on physical parameters such as C_n^2, r_0, and θ_0; forward-model or observation-consistency objectives; wave-optics and PDE constraints; physically informed losses; and hybrid model-based and data-driven reconstruction (Section 10). The third is the dimension along which restoration quality is judged: pixel fidelity, perceptual fidelity, temporal consistency, downstream task performance, uncertainty and calibration, and efficiency under deployment constraints (Sections 13 and 14). Treating these as independent axes matters for correct interpretation. Any architectural family can adopt any physics-integration mechanism, and state-space models, in particular, are an architectural choice rather than an inherently physics-informed one.

Table 7. Evidence mapping: research questions, review coverage, and primary evidence types.

RQ	Question (Summary)	Section(s)	Primary Evidence Types	Literature Categories
RQ1	Physical mechanisms and optical effects of atmospheric turbulence	Section 3	Theoretical analyses; wave-optics models; experimental characterization; Kolmogorov/von Kármán turbulence statistics	Journal articles; optics textbooks; conference papers
RQ2	Classical and deep learning-based mitigation approaches	Sections 4–10	Algorithm proposals; ablation studies; comparative benchmarks; survey papers	IEEE/CVPR/ICCV/ECCV; optics journals; reviews
RQ3	Datasets and turbulence simulation methods	Section 12	Benchmark dataset papers; simulation framework descriptions; evaluation protocols	Dataset papers; challenge reports; software papers
RQ4	Recent innovations shaping the field (post 2019)	Sections 15 and 16	State-of-the-art method papers; Transformer- and diffusion-based models; real-world deployment studies	Top-venue papers; preprints (2019–2026)
RQ5	Comparative analysis across quality, efficiency, and applicability	Sections 13 and 15	Quantitative benchmark results (PSNR, SSIM, and LPIPS); runtime comparisons; real-world trials	Benchmark papers; multi-method comparison studies

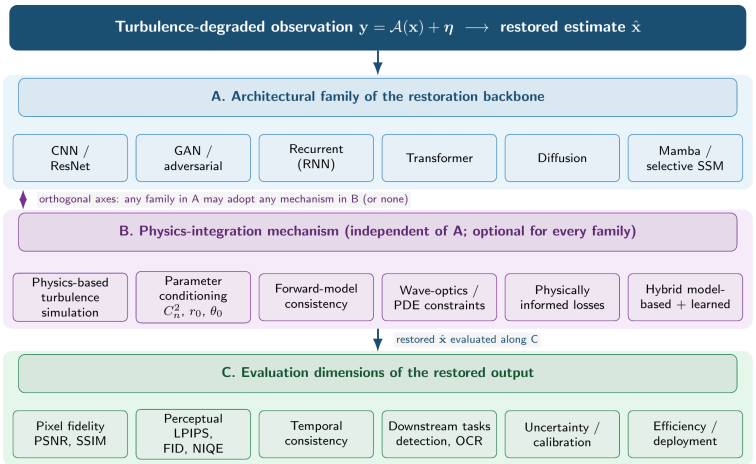


Figure 2. Three-dimensional taxonomy of learning-based atmospheric turbulence mitigation: (**A**) architectural family of the restoration backbone, (**B**) physics-integration mechanism, and (**C**) evaluation dimension.

2.8. Data Extraction and Synthesis

For each core study, the following information was extracted into a structured evidence table: publication metadata (authors, year, venue, and publication status, with preprints flagged per Section 2.4.3); method family and specific architecture; input configuration (single-frame or multi-frame, and the number of input frames); training and evaluation data type (real, synthetic, or mixed) and the named datasets; the turbulence simulator or physical degradation model used; supervision type; loss functions; evaluation metrics; reported turbulence-severity variables (e.g., D/r_0 or C_n^2); model parameters, FLOPs or MACs, runtime, and hardware, where reported by the primary source; public code and dataset availability; the main finding; the main limitation; the research question(s) the study informs; and an evidence-quality rating assigned with the rubric of Section 2.9. The extracted information is synthesized directly in the body of the review: Tables 5 and 6 present the study-level characteristics, contributions, and limitations; Tables 3 and 4 consolidate the quantitative and efficiency evidence, deliberately separated by whether a shared protocol licenses cross-row comparison; and the architecture-family assessments in Sections 5–10 carry the corresponding evidence-strength judgments. Throughout, values that the primary source does not report are marked N.R. rather than inferred. Where multiple studies reported the same metric on the same benchmark, results were tabulated for direct comparison (Table 3); all remaining results were tabulated separately and explicitly marked as non-comparable (Table 4).

2.9. Evidence-Quality Appraisal

Formal risk-of-bias instruments developed for clinical research do not transfer meaningfully to engineering and computer-vision restoration studies, so no clinical-style risk-of-bias assessment is claimed. Instead, each core study was appraised against a technical reporting and external-validity rubric with sixteen criteria: (1) evaluation on at least one publicly available dataset; (2) public code release; (3) precise train/validation/test split specification; (4) evaluation across multiple turbulence severities; (5) evaluation on real (not only synthetic) turbulence; (6) paired real ground truth; (7) more than one scene domain; (8) evaluation under a shared protocol with published baselines; (9) controlled ablation of the proposed components; (10) model parameters and FLOPs/MACs reported; (11) hardware, input resolution, runtime measurement, and timing boundary clearly defined; (12) statistical variation reported over multiple runs; (13) temporal-consistency evaluation; (14) downstream-task evaluation; (15) uncertainty or calibration evaluation; and (16) explicit discussion of limitations. Ratings are categorical and defined by explicit conditions: *High* evidence strength requires a shared-protocol comparison with published baselines, evaluation on both synthetic and real turbulence, and complete efficiency reporting (parameters, runtime, and hardware); *Moderate* denotes a peer-reviewed study with quantitative evaluation that misses one or more of the High conditions; and *Preliminary* denotes studies with limited external validation, such as evaluation restricted to a single custom dataset, absence of shared-protocol baselines, or incomplete experimental reporting. Per-study ratings are reflected in the evidence-strength qualifiers that accompany the architecture-level assessments in Sections 5–10. Across the 30 core studies analyzed individually in the body of this review, three rate High (the shared-protocol studies, i.e., TMT [16], DATUM [28], and TSTRNet [9]), twenty-two rate Moderate, and five rate Preliminary; the distribution confirms that shared-protocol, efficiency-complete reporting remains the exception in this literature.

2.10. Review Limitations

The following limitations should be considered when interpreting this review:

1. Publication bias: The search was restricted to two databases and to peer-reviewed sources, supplemented by explicitly labeled preprints under the policy of Section 2.4.3. Unpublished negative results and proprietary industrial techniques are not captured.
2. Recency cutoff: The search closes at 2026. Given the pace of development in foundation models and diffusion-based restoration, methods published after the final search date are not included.
3. Incomplete PRISMA disaggregation: The archived screening log recorded duplicate removal and title-and-abstract exclusion as a single combined count and did not record per-database yields or per-reason full-text exclusion counts (Section 2.6). The selection process is therefore reproducible at the aggregate level only.
4. Partial duplicate screening: Primary screening at both stages was performed by one author, with independent re-screening by a second author covering a random 26% subset of the identified records ($n = 55$ of 213) at the title-and-abstract stage only; full-text screening was not duplicated, and no inter-rater agreement statistic was calculated (Section 2.5). This falls short of the full dual-reviewer protocol that PRISMA 2020 recommends and remains the most consequential methodological limitation of this review. It threatens completeness, in that a relevant study may have been missed among the records screened by a single reviewer, more than it threatens the architecture-level judgments, which rest on primary-source evidence that is individually cited and traceable. The included set should therefore be read as one defensible sample of the literature rather than as an exhaustive census.
5. Heterogeneous evaluation protocols: Primary studies use a wide variety of datasets, turbulence simulation settings, and metric thresholds, making cross-study comparisons inherently imprecise.
6. Search-vocabulary coverage: The executed queries paired seven turbulence phrases with learning- and architecture-related terms (Appendix A) and did not include adjacent terminology such as “heat haze”, “atmospheric seeing”, “aero-optical distortion”, “refractive distortion”, “anisoplanatic imaging”, or “turbulence-degraded video”. Relevant studies indexed only under such terms or classical studies without learning-related indexing may have been missed unless recovered through citation tracing. The executed queries were not retrospectively expanded, so the reported search remains the one actually performed.
7. English-language restriction: Only English-language full texts were retained; relevant work published exclusively in other languages may have been missed.
8. Scope exclusions: Turbulence mitigation in non-passive optical domains (e.g., LiDAR and free-space optical communication) is addressed only briefly.
9. Critical narrative synthesis: No formal meta-analysis was performed, and the evidence-quality appraisal of Section 2.9 is a technical reporting and external-validity appraisal rather than a clinical-style risk-of-bias assessment; conclusions represent a qualitative synthesis subject to the authors’ interpretive judgments.

3. Physics of Atmospheric Turbulence

This section provides the minimum physics an AI reader needs for the rest of the review: where turbulence comes from and the three quantities the structure constant (C_n^2), the Fried coherence length (r_0), and the isoplanatic angle (θ_0) that reappear throughout dataset design, turbulence simulation, and model evaluation.

3.1. Fundamentals of Atmospheric Turbulence

Atmospheric turbulence arises from random velocity fluctuations within a viscous fluid medium. Air moves in one of two dynamical states: laminar flow, in which adjacent air layers slide past one another without lateral mixing, and turbulent flow, in which dynamic mixing generates a spectrum of random subflows termed turbulent eddies. The onset of turbulence is quantified by the Reynolds number, i.e., $Re = Vl/\nu$, where V denotes the bulk flow velocity, l is the characteristic length scale of the flow, and ν is the kinematic viscosity of the fluid [43]. Near the Earth's surface, Reynolds numbers on the order of 10^5 are routinely observed, placing the lower atmosphere well beyond the critical threshold for turbulent transition. This sustained turbulence is continuously fed by energy inputs from wind shear and thermal convection, both of which generate large-scale flow instabilities that subsequently fragment into progressively smaller structures.

Kolmogorov's theoretical framework formalizes this fragmentation through the energy-cascade hypothesis: large unstable air masses at the outer scale of turbulence (L_0) break apart under inertial forces into smaller eddies spanning a continuous hierarchy of scales, i.e., the inertial range, until the eddy size falls below the inner scale (l_0), at which point viscous forces dominate and the remaining kinetic energy is dissipated as heat [43]. A central postulate of this theory is that small-scale fluctuations within the inertial range are statistically homogeneous, isotropic, and structurally independent of the large-scale flow configuration, a simplification that has proven effective for the modeling of real atmospheric behavior. For imaging systems, the optical consequence of this cascade is that temperature and humidity fluctuations accompanying turbulent mixing produce spatially and temporally varying perturbations in the refractive index of air. A propagating wavefront accumulates phase errors traversing this randomly structured medium, yielding the spatially variant blur, geometric distortion, and intensity scintillation that constitute the primary degradation modes addressed throughout this survey.

3.2. Kolmogorov Refractive-Index Structure Function

Kolmogorov turbulence theory remains the fundamental framework for the modeling of atmospheric turbulence, describing the transfer of energy from large turbulent eddies toward progressively smaller eddies until viscous dissipation occurs at molecular scales. The refractive-index structure parameter (C_n^2) is commonly used to characterize turbulence strength, where larger values indicate stronger atmospheric turbulence.

The refractive-index structure function over a spatial separation ($\boldsymbol{r}$) is given by

$$\begin{aligned} D_n(\mathbf{r}) &= \mathbb{E}\Big[\big(n(\mathbf{x}+\mathbf{r}) - n(\mathbf{x})\big)^2\Big] = C_n^2 |\mathbf{r}|^{2/3}, \\ & l_0 \ll |\mathbf{r}| \ll L_0, \end{aligned} \tag{1}$$

where C_n^2 [$\mathrm{m}^{-2/3}$] denotes the refractive-index structure parameter characterizing turbulence strength, l_0 is the inner scale, and L_0 is the outer scale of turbulence.

To account for finite inner and outer scales, atmospheric refractive-index fluctuations are commonly modeled using the von Kármán power spectral density (PSD):

$$\Phi_n(\kappa) = 0.033\, C_n^2 \left(\kappa^2 + \kappa_0^2\right)^{-11/6} \exp\left(-\frac{\kappa^2}{\kappa_m^2}\right), \tag{2}$$

where $\kappa = |\boldsymbol{\kappa}|$ is the spatial frequency, $\kappa_0 = 2\pi/L_0$ denotes the outer-scale cutoff, and $\kappa_m = 5.92/l_0$ represents the inner-scale cutoff [43].

3.3. Non-Kolmogorov Generalization

The classical Kolmogorov turbulence model assumes a fixed power-law exponent of $\alpha = 11/3$ for the refractive-index fluctuation spectrum, which works reasonably well under idealized horizontal propagation conditions. Real atmospheric channels, particularly those involving non-horizontal paths or varying altitudes, often deviate from this assumption in measurable ways [44]. To accommodate this, a generalized non-Kolmogorov spectrum is adopted, expressed as

$$\phi_n(\sigma, \alpha, z) = \Lambda(\alpha)\, C_n^2(z, \alpha)\, \sigma^{-\alpha}, \quad \sigma > 0, \quad 3 < \alpha < 4, \tag{3}$$

where σ is the spatial wavenumber, α is the power-law exponent controlling the turbulence regime, $\Lambda(\alpha) = \Gamma(\alpha - 1)\cos(\alpha\pi/2)/(4\pi^2)$ is a consistency function that preserves dimensional correctness across the generalized spectrum, and $C_n^2(z, \alpha)$ is the generalized refractive-index structure parameter dependent on both propagation distance and zenith angle [44]. The standard Kolmogorov case is naturally recovered when $\alpha = 11/3$.

Beyond the spectral model, the generalized Fried-like coherence parameter ($r_0(\alpha, z)$) plays a central role in quantifying phase distortion under non-Kolmogorov conditions, serving as the non-Kolmogorov analog of the classical Fried parameter. As shown in [44], when $\alpha < 11/3$, the generalized structure parameter remains below the Kolmogorov baseline and decreases monotonically along the propagation path, indicating weaker coherence loss relative to the standard model. Conversely, for $\alpha > 11/3$, the structure parameter develops a pronounced peak along the path and exceeds the Kolmogorov value, reflecting a fundamentally different spatial energy distribution within the turbulence spectrum. This non-linear sensitivity to α underscores the importance of using physically accurate turbulence models when evaluating system performance in real free-space optical channels.

Implication for Learned Restoration

The generalized spectrum is not introduced here for completeness. It exposes a limitation shared by essentially the entire learning-based literature reviewed in Sections 5–10 and one that the field has not confronted.

Across the reviewed corpus, we did not identify a single deep-learning restoration method trained or evaluated on data generated under a non-Kolmogorov spectrum with $\alpha \neq 11/3$. Every simulator that supplies training data to these models, including the phase-to-space (P2S) pipeline released with TMT [16], the Zernike-based tilt-then-blur formulation [35], the simulation frameworks compared by Nair et al. [37], and the phase-screen methods they build on, assumes classical Kolmogorov statistics or the finite-scale von Kármán refinement of Equation (2). The consequence is a *modeling monoculture*: the learned prior over degradations is conditioned on a fixed power-law exponent, and α is not a variable the models have ever been asked to generalize over.

Three practical risks follow. First, real channels deviate measurably from $\alpha = 11/3$, particularly on non-horizontal and slant paths, at varying altitudes, and in the upper atmosphere [44]; these are precisely the geometries of air-to-ground, ground-to-air, satellite, and astronomical imaging. Second, the deviation is not a small perturbation. As noted above, $\alpha > 11/3$ produces a pronounced peak in the generalized structure parameter along the path and exceeds the Kolmogorov value, whereas $\alpha < 11/3$ falls below it and decays monotonically. The spatial distribution of turbulence energy and, hence, the statistics of the PSF that a restoration network must invert, therefore differ in kind rather than in degree. Third and most consequentially for this review's argument, this is a source of *epistemic* uncertainty that current evaluation practice cannot even detect: a model trained exclusively under Kolmogorov statistics has no mechanism for recognizing that a slant-

path observation lies outside its training distribution, and no reviewed study reports the out-of-distribution behavior that would reveal it.

We therefore treat the Kolmogorov/von Kármán monoculture as a distinct and under-recognized component of the synthetic-to-real domain gap, and it is carried forward as such in Challenge 1 of Section 16. The resolving evidence is inexpensive to obtain relative to its value: simulators already parameterize the spectral exponent, so training and test corpora that sweep α across the $3 < \alpha < 4$ range of Equation (3), together with reported degradation curves as α departs from $11/3$, would establish whether current architectures degrade gracefully or fail outright under non-standard atmospheres.

3.4. Fried Coherence Length

The Fried parameter, denoted as r_0, represents the atmospheric coherence length and is widely used to characterize the strength of optical turbulence. Physically, r_0 quantifies the effective aperture diameter over which the optical wavefront remains approximately coherent. A smaller value of r_0 indicates stronger turbulence and more severe wavefront distortions, whereas larger values correspond to weaker turbulence and improved imaging conditions. Under atmospheric turbulence, the achievable imaging resolution is often limited by the coherence length rather than the physical aperture size of the optical system.

For a plane wave under the flat-Earth approximation, the Fried coherence length is expressed as

$$r_0 = \left[0.423\, k^2 \sec(\zeta) \int_0^L C_n^2(z)\, \mathrm{d}z\right]^{-3/5}, \tag{4}$$

where $k = 2\pi/\lambda$ denotes the optical wavenumber, λ is the optical wavelength, ζ is the zenith angle, $C_n^2(z)$ is the refractive-index structure parameter as a function of altitude (z), and L represents the propagation path length. The secant term ($\sec(\zeta)$) accounts for the increased atmospheric path length for off-zenith propagation [45].

3.5. Isoplanatic Angle

The isoplanatic angle (θ_0) defines the angular region over which the atmosphere can be treated as a uniform optical system, producing an approximately constant point-spread function (PSF) across the field of view. Formally, it corresponds to the angular separation between two propagation paths that induces a mean-square wavefront error of 1.0 rad^2. For a given refractive-index structure constant profile ($C_n^2(z)$) under a flat-Earth assumption, the isoplanatic angle is given by [45]

$$\theta_0 = \left[2.91 k^2 \sec^{8/3}\zeta \int_{\mathrm{Path}} C_n^2(z)\, z^{5/3}\, dz\right]^{-3/5} \tag{5}$$

where $k = 2\pi/\lambda$ is the optical wavenumber and ζ is the zenith angle. The $z^{5/3}$ weighting in the integral indicates that turbulence at higher altitudes contributes disproportionately to anisoplanatism because higher-altitude layers subtend larger angular displacements across off-axis propagation paths. The $\sec^{8/3}\zeta$ factor reflects the increased effective path length at larger zenith angles, which substantially reduces θ_0 as the observation moves away from zenith.

The practical significance of θ_0 for turbulence restoration is direct: when the angular extent of the imaged scene exceeds the isoplanatic angle, the turbulence-induced PSF becomes spatially varying across the field of view. This anisoplanatic behavior is a primary reason why atmospheric turbulence constitutes a fundamentally harder inverse problem than space-invariant blur. Since $\theta_0 \propto \lambda^{6/5}$, longer wavelengths exhibit larger isoplanatic patches, making correction more tractable in the infrared than in the visible regime. At

a zenith angle of 60°, θ_0 is smaller than at zenith because of the $\sec^{8/3}\zeta$ dependence in Equation (5), further constraining the spatial region over which any single correction remains valid. These characteristics directly inform the design of turbulence simulation frameworks, spatially varying deep learning architectures, and the evaluation of restoration methods under realistic operational conditions.

4. Classical Turbulence Mitigation: Formulations

Before deep learning, two ideas dominated turbulence mitigation: wait for the moments when the atmosphere is briefly calm and keep only those frames (lucky imaging) or model the blur mathematically and invert it (blind deconvolution). Both formulations still shape modern learned methods, so we state them precisely here. A third classical strand, adaptive optics, corrects the wavefront in hardware before detection rather than the image after it and is therefore a mitigation strategy that is not a restoration strategy in the sense defined in Section 1. It falls outside the computational scope of this review, and readers seeking the current state of learning-augmented wavefront sensing and correction are referred to the dedicated review by Liu et al. [46]. The distinction matters for system design: adaptive optics and computational restoration address the same physics at different points in the imaging chain, and they are complementary rather than competing.

4.1. Lucky Imaging

Lucky imaging, also known as *frame selection*, is a technique employed for high-resolution imaging in astronomy. It involves recording a time series of short-exposure images and subsequently selecting the sharpest frames from the sequence. These selected sharp images are then aligned and combined to generate a final, higher-quality image. The effectiveness of this method arises from the fact that atmospheric turbulence causes rapid fluctuations in image quality. By capturing numerous short exposures, there is a probability of obtaining moments in which atmospheric distortions are temporarily reduced, thereby producing clearer observations. This technique is particularly effective for achieving diffraction-limited stellar images, even when using large telescopes.

The probability of obtaining a sharp image, defined as an image with wavefront variance below 1 rad^2, depends on the telescope aperture (D) and the Fried parameter (r_0), which characterizes atmospheric seeing conditions. The probability is expressed as [47]

$$P = 5.6 \exp\left[-0.1557\left(\frac{D}{r_0}\right)^2\right], \tag{6}$$

where lower values of D/r_0 correspond to a higher likelihood of acquiring high-quality frames. This expression should be interpreted within the aperture-to-seeing regime for which the Fried approximation is valid and should not be extrapolated to small D/r_0 values, where the probability expression can become nonphysical. For example, the probability of obtaining a sharp image is approximately 1 in 9 when $D/r_0 = 5$ but decreases rapidly as D/r_0 increases. Consequently, frame selection techniques are generally most effective when D/r_0 ranges between 6 and 7, as this interval provides a reasonable likelihood of capturing multiple high-quality images within a sequence of thousands of frames. Furthermore, advances in imaging technologies, including electron-multiplying charge-coupled devices (EM-CCDs) and high-speed processors, have enabled rapid image acquisition, readout, and storage of large datasets, thereby increasing the practicality and broader adoption of lucky imaging techniques [48].

4.2. Blind Deconvolution: MAP Objective

Blind deconvolution for atmospheric turbulence presents a fundamental challenge because neither the degradation kernel nor the underlying clean image is known *a priori*. Following the formulation in [49], the observed image spectrum, denoted by $y(\omega)$, is modeled as the product of the Fried kernel ($M_{\boldsymbol{a}}(\omega)$) and the latent image spectrum ($x(\omega)$), corrupted by additive white Gaussian noise with a standard deviation of $\sigma > 0$. Since both the latent image and blur kernel must be recovered from a single degraded observation, the problem is inherently ill-posed. To address this issue, the joint estimation of x and the kernel parameter vector ($\boldsymbol{a}$) is formulated as a regularized maximum *a posteriori* (MAP) optimization problem:

$$\underset{x,\,\boldsymbol{a},\,\lambda}{\arg\min}\ \frac{1}{2\sigma^2}\|y - M_{\boldsymbol{a}} \cdot x\|_2^2 + R(x) + \lambda\|M_{\boldsymbol{a}}\|_2^2 + \frac{\alpha}{2\lambda} \tag{7}$$

Each term in (7) serves a distinct role within the estimation framework. The first term enforces consistency between the reconstructed blurred observation and the measured image, arising directly from the Gaussian noise assumption. The regularization term ($R(x)$), modeled using a zero-mean Gaussian mixture model (GMM) prior over natural image patches, constrains the recovered image toward perceptually realistic structures. The third and fourth terms jointly regulate the total energy of the estimated kernel. Specifically, the penalty term ($\lambda\|M_{\boldsymbol{a}}\|_2^2$) prevents the kernel energy from becoming unbounded, while the auxiliary regularization term ($\alpha/(2\lambda)$) avoids degenerate solutions. Importantly, λ is not fixed but optimized jointly with x and $\boldsymbol{a}$, enabling adaptive behavior across varying turbulence conditions without manual parameter tuning [49].

5. CNN-Based Atmospheric Turbulence Restoration

CNNs were the first deep architectures applied to turbulence restoration, and they set the template that every later family refines: learn the mapping from degraded to clean images directly from data. Three building blocks recur across CNN-based turbulence networks, as summarized in Figure 3: residual learning, dilated convolution, and channel attention. In the arrangement shown there, a 3×3 convolution with batch normalization and ReLU feeds a dilated convolution, which enlarges the receptive field at no additional parameter cost; a squeeze-and-excitation module then recalibrates the channels through global average pooling and two fully connected layers, and the recalibrated features are added back to the identity shortcut. The three subsections below define each block.

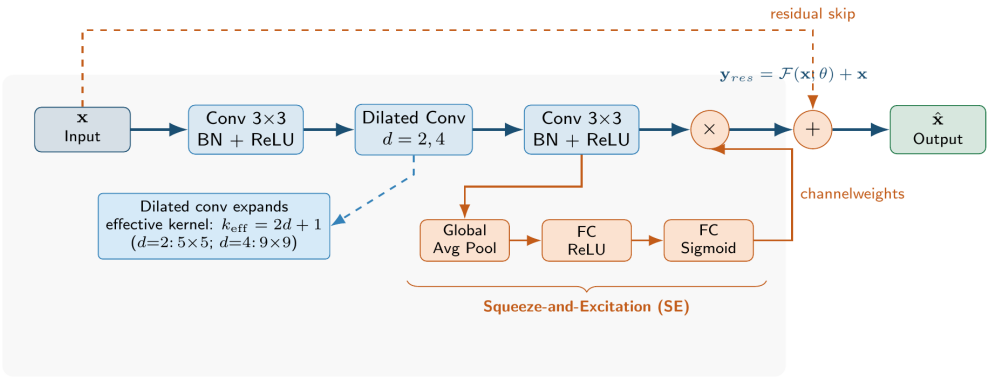


Figure 3. Building blocks of a CNN turbulence restoration block: dilated convolution, squeeze-and-excitation (SE) channel attention, and an identity skip connection.

5.1. Residual Learning Formulation

Residual learning [50] asks stacked layers to fit the residual ($\mathcal{F}(\mathbf{x}) = \mathcal{H}(\mathbf{x}) - \mathbf{x}$) of a target mapping rather than the mapping $\mathcal{H}(\mathbf{x})$ itself, with a parameter-free identity shortcut recovering the output of each block as

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}, \tag{8}$$

where $\{W_i\}$ represents the weights of the stacked layers. When an identity mapping is locally optimal, the residual path only needs to be driven toward zero, which is far easier for gradient-based optimization than fitting a near-identity transformation; this is what allows restoration networks to be trained at substantial depth without degradation [50]. The formulation is particularly natural for turbulence restoration, where the clean image and the degraded observation share most of their content: the network learns to predict the turbulence-induced perturbation rather than re-synthesize the entire image.

5.2. Dilated Convolution for Expanded Receptive Field

Dilated convolution is a generalization of the standard discrete convolution operator, where a dilation factor, denoted as l, is introduced to control the spacing between the filter elements. This modification allows the convolution operator to apply the same filter at different ranges, effectively expanding the receptive field without losing resolution or coverage [51]. The familiar discrete convolution is simply a 1-dilated convolution. The mathematical definition of a dilated convolution, denoted as $*_l$, for a discrete function ($F : \mathbb{Z}^2 \to \mathbb{R}$) and a discrete filter ($k : \Omega_r \to \mathbb{R}$) is given by

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)k(t). \tag{9}$$

For a kernel of width k and dilation factor d, the effective kernel size is $k_{\text{eff}} = k + (k-1)(d-1)$; a 3×3 kernel therefore covers an effective $(2d+1) \times (2d+1)$ window, e.g., 5×5 at $d = 2$ and 9×9 at $d = 4$ (Figure 3). Stacking dilated layers grows the receptive field rapidly without pooling, which is why dilation appears throughout turbulence CNNs, whose blur and tilt correlations extend well beyond a plain 3×3 neighborhood.

5.3. Channel Attention (Squeeze and Excitation)

The Squeeze-and-Excitation (SE) block [52] recalibrates feature channels using global context, letting a turbulence network emphasize channels that carry scene structure and suppress channels dominated by turbulence-induced responses. Global average pooling first squeezes each channel of a feature map (U) with spatial dimensions of $H \times W$ into a scalar descriptor, i.e.,

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), \tag{10}$$

and a bottlenecked gating function then maps the descriptor ($z \in \mathbb{R}^C$) to per-channel weights, i.e.,

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)), \tag{11}$$

where $\delta(\cdot)$ is the ReLU activation, $\sigma(\cdot)$ is the sigmoid, and $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ form a bottleneck with the reduction ratio (r). The weights (s) rescale the original channels multiplicatively, adding global-context sensitivity at negligible computational cost.

5.4. CNN Methods for Turbulence Restoration

The building blocks described above underpin dedicated CNN architectures that address atmospheric turbulence in application-specific ways.

Yasarla and Patel [13] address single-face image restoration, observing that turbulence severity is spatially non-uniform within any given frame: some regions suffer strong blur and geometric warping while others remain relatively intact. Their convolutional network estimates a per-pixel uncertainty map alongside the restored output, directing training supervision toward pixels where the ground truth is most reliable and reducing the influence of severely distorted regions where the inverse problem is most ambiguous. This uncertainty-guided strategy exploits the structured appearance priors of face imagery, such as the geometric regularity of facial landmarks, to improve spatial consistency in the restored output. Evaluated on a CelebA-based synthetic turbulence dataset, the approach achieved a PSNR/SSIM of 25.42 dB/0.815 [13]. The earlier AT-Net formulation from the same authors is, to the best of our knowledge, the only method in the reviewed corpus to target *epistemic* uncertainty explicitly: it uses Monte Carlo (MC) dropout to identify regions where the network itself is unreliable and feeds those maps back as guidance [21]. That distinction matters for the argument developed below, and it is worth stating precisely because the two works are often cited interchangeably. MC dropout approximates model uncertainty by sampling several stochastic forward passes and measuring their disagreement [53], whereas a directly regressed per-pixel variance head captures observation noise. The first is epistemic and shrinks with more or better-matched data; the second is aleatoric and does not.

Uncertainty treatment. The CNN family is the only one in this review with a genuine if narrow uncertainty tradition, and it illustrates both what is achievable and how far current practice falls short. Aleatoric uncertainty is addressed implicitly and poorly: a network trained with an $\mathcal{L}_1$ or $\mathcal{L}_2$ objective converges toward the conditional mean of the plausible clean images, so the intrinsic one-to-many ambiguity of the random channel is not represented but silently averaged away. The characteristic oversmoothing of CNN outputs is precisely this averaging made visible, which is why it should be read as a symptom of unmodeled aleatoric uncertainty rather than as a mere aesthetic defect. Only the heteroscedastic variance head of Yasarla and Patel [13] represents it explicitly and only as a spatial weighting for training rather than as a calibrated output for downstream consumers. Epistemic uncertainty is addressed in the earlier AT-Net [21] through MC dropout, but its use there is internal: the maps guide restoration, and neither study reports whether the resulting confidence is *calibrated*, that is, whether a stated confidence of 90% corresponds to a 90% empirical hit rate. Neither is evaluated under the distribution shift, unseen severities, unseen sensors, or unseen scene domains in which epistemic uncertainty is supposed to become large and useful. The result is that the family with the best uncertainty machinery still ships a bare point estimate to any downstream detector or recognizer.

Mao et al. [15] approach single-frame turbulence mitigation from two directions simultaneously. They introduce the Heat Chamber and Turbulence Text datasets with standardized evaluation protocols, creating shared test beds for subsequent methodological comparison. As a reference baseline within the same work, TurbNet, a convolutional encoder–decoder with residual connections, is trained end to end and achieves a PSNR/SSIM/LPIPS of 22.76 dB/0.6923/0.4012 on the TMT synthetic evaluation suite, as re-evaluated in [16]. The calibration value of TurbNet is its role as the anchor for subsequent comparisons: on the same benchmark, the multi-frame TMT Transformer configuration scores 27.73 dB. Because the two systems differ simultaneously in temporal input, alignment, architecture, loss design, training strategy, and model capacity, this roughly 5 dB

gap measures the difference between two complete configurations; it cannot be attributed uniquely to self-attention or to temporal fusion as individual components.

Evidence and limitations. The CNN family carries the longest evaluation record of any learning-based paradigm in this domain, and its efficiency advantages (small parameter counts, hardware-friendly operations, and low inference latency) remain unmatched, which keeps CNNs relevant wherever deployment constraints dominate. Its structural limitations are equally consistent across the reviewed studies. Convolution is local, so the modeling of the long-range blur and tilt correlations of anisoplanatic turbulence depends on receptive-field engineering (dilation, downsampling, and deep stacking) that trades off spatial resolution or capacity; pixel-wise regression objectives bias outputs toward the conditional mean, producing the oversmoothing documented across benchmarks; and face-specific uncertainty-guided methods raise pixel-fidelity scores by exploiting strong domain priors that do not transfer to arbitrary outdoor scenes. The strength of evidence for these conclusions is moderate to high: the TurbNet baseline has been re-evaluated under shared protocols by independent later studies [16,28], while the TDRN face-restoration results are supported by preprint-reported values whose published counterparts remain to be cross-checked (Table 4). These limitations motivate the attention-based and temporal architectures described in the following sections.

6. GAN-Based Turbulence Restoration

Generative adversarial networks (GANs) train a generator (G) and a discriminator D against each other: G produces outputs meant to be indistinguishable from real data, and D learns to tell them apart [54]. Figure 4 shows the configuration most often used for turbulence restoration. A U-Net generator (G_θ) with encoder, bottleneck, decoder, and skip connections maps the degraded observation ($\mathbf{y}$) to a restored image ($\hat{\mathbf{x}}$), while a PatchGAN discriminator (D_ϕ) distinguishes patches of $\hat{\mathbf{x}}$ from patches of the clean reference ($\mathbf{x}$). Training minimizes a weighted sum ($\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{adv}} + \lambda_3 \mathcal{L}_{\text{perc}}$) that balances pixel reconstruction, adversarial realism, and perceptual fidelity. The competition is formalized as a minimax problem over a value function ($V(D, G)$), i.e.,

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[\log D(\boldsymbol{x})\right] + \mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}\left[\log\left(1 - D(G(\boldsymbol{z}))\right)\right] \tag{12}$$

where G maps a noise vector ($\boldsymbol{z} \sim p_z$) to the data space and $D(\boldsymbol{x})$ estimates the probability that a sample comes from the real data distribution (p_{data}) [54]. The framework matters for turbulence restoration because the degradation is stochastic and spatially varying, so the clean image is only partially determined by the degraded observation: an adversarial objective rewards outputs that lie on the natural-image manifold instead of the oversmoothed conditional mean that pixel regression produces. Several GAN architectures have been applied to this problem, each addressing a distinct facet of the restoration challenge.

Jin et al. [27] provide the earliest widely adopted supervised GAN baseline for complex (non-face) scenes, a multi-frame adversarial restoration network published in *Nature Machine Intelligence* and commonly abbreviated as TSRWGAN in later benchmark studies. Its significance for this review is largely calibrational: because subsequent works re-evaluated it under shared protocols, it anchors the GAN family in cross-architecture comparisons. On the TMT synthetic static-scene benchmark, it reaches a PSNR/SSIM/LPIPS of 25.29 dB/0.7784/0.2243 using 15 frames, with 46.28 M parameters and 1.15 s per frame at a 540×960 resolution [16], and on ATSyn-dynamic, it averages 26.45 dB [28]. These shared-protocol results show that the evaluated multi-frame TSRWGAN configuration outperforms

the single-frame TurbNet baseline while trailing the Transformer and recurrent configurations tested on the same data; because the compared systems differ simultaneously in temporal input, alignment, architecture, loss design, and capacity, the deltas characterize whole configurations rather than the isolated contribution of adversarial training or any other single component.

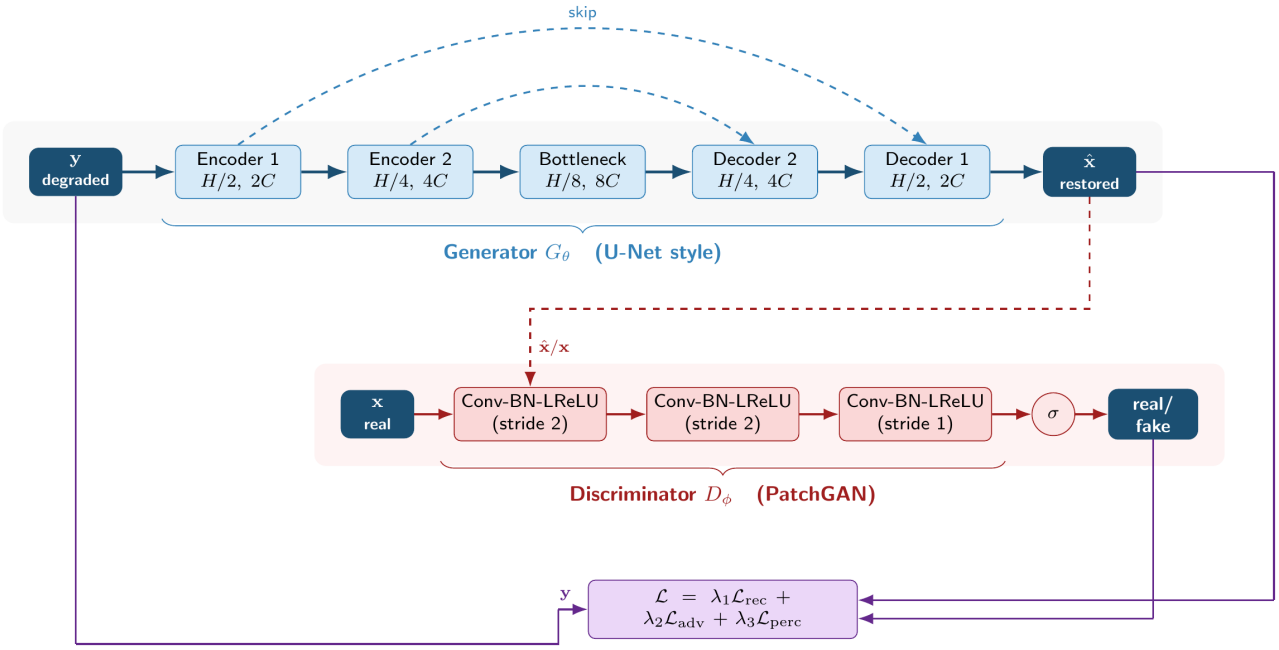


Figure 4. GAN-based turbulence restoration with a U-Net generator; a PatchGAN discriminator; and a combined reconstruction, adversarial, and perceptual objective.

Feng et al. [14] proposed TurbuGAN, an unsupervised multi-frame adversarial approach. Rather than requiring paired clean references, TurbuGAN infers a shared sharp latent image from a collection of degraded frames through a combination of adversarial feedback and a cross-frame consistency objective. Because it is self-supervised, TurbuGAN can be applied to real outdoor sequences where paired ground truth is unavailable, a practical advantage for field settings. The limitation of this design is that the absence of pixel-level supervision makes quantitative comparison difficult: reported per-image PSNR estimates of approximately 27 dB on four face images cannot be averaged over a standard benchmark and should not be compared directly against supervised methods [14].

Mei and Patel [22] approach turbulence restoration through GAN inversion. Their LTT-GAN method projects the degraded input into the latent space of a face-specific pretrained generator (StyleGAN2), then optimizes the latent code to produce a restoration that simultaneously lies on the natural image manifold and is consistent with the observed degradation. Because the generator has already learned the distribution of sharp face images, the resulting output inherits high-frequency fidelity without turbulence-specific training pairs. LTT-GAN is evaluated using LPIPS and FID rather than PSNR, reflecting the view that perceptual fidelity on real turbulence sequences is more meaningful than pixel-level agreement against synthetic references [22].

Liu et al. [6] target the synthetic-to-real domain gap through EAF-WGAN (Enhanced Alignment Fusion Wasserstein GAN). The Wasserstein distance formulation stabilizes adversarial training and enables feature distribution matching between synthetic training data and real turbulence sequences, while an alignment-fusion module addresses temporal geometric inconsistency across frames. Evaluated using PSNR, SSIM, BRISQUE, and NIQE

on Australian outdoor turbulence data, EAF-WGAN demonstrates that domain-adaptive GAN training can narrow the performance gap between simulator-trained models and real optical environments [6].

For downstream visual task improvement, Rai and Jawahar [39] propose DTD-GAN+, an adversarial turbulence removal framework that optimizes restoration jointly with detection and recognition objectives. The adversarial objective guides the generator toward outputs that preserve task-salient structure rather than minimizing pixel error alone. While gains in task-oriented metrics are reported, the authors note that these do not necessarily transfer to blind deblurring quality, reflecting the fidelity–task metric mismatch discussed in Section 14.

In the context of astronomical imaging under combined atmospheric turbulence and optical aberrations, Peng et al. [29] extend the GAN framework with a Channel Attention Block (CAB) for channel-wise recalibration and a Semantic Attention Module (SAM) for cross-scale semantic guidance within a feature pyramid backbone. This design suppresses turbulence-dominated feature responses while preserving the fine structures of astronomical targets that plain GANs tend to smooth over.

Uncertainty treatment. The adversarial family illustrates a model class that can respond to ambiguity in the restoration problem without explicitly quantifying that ambiguity. On aleatoric uncertainty, the adversarial objective encourages outputs that are perceptually plausible rather than forcing the generator toward the conditional mean of all possible clean images. However, the turbulence GANs reviewed here generally return a single restoration and do not report a predictive distribution or calibrated uncertainty measure. Consequently, the output does not indicate which recovered details are strongly supported by the observation and which depend more heavily on the learned image prior. This creates a recognized risk of plausible but unfaithful high-frequency detail, particularly in recognition-critical applications (Section 14), and motivates explicit uncertainty reporting rather than treating perceptual realism as evidence of fidelity.

On epistemic uncertainty, adversarial methods carry an additional and less-recognized exposure. The learned manifold is the manifold of the *training* distribution, so under distribution shift, a GAN does not degrade toward blur, as a regression network does, but toward confident synthesis of training-domain structure that the observation does not support. This makes epistemic error harder to detect by inspection than in the CNN case, not easier. It is instructive that the two reviewed works most concerned with the domain gap attack it through the feature distribution rather than through uncertainty: EAF-WGAN aligns synthetic and real feature distributions using a Wasserstein formulation [6], and LTT-GAN constrains the solution to the latent space of a pretrained face generator [22]. Both reduce the epistemic exposure, and neither measures it. No reviewed GAN study reports output variance across seeds, ensemble disagreement, a discriminator-derived confidence score, or any calibration assessment. The practical recommendation that follows is concrete: adversarial restorers intended for evidential or recognition-critical use should be paired with an explicit uncertainty channel, whether from a deep ensemble [23], MC dropout in the generator [53], or a distribution-free interval procedure [24], so that recovered detail can be labeled with the confidence it actually warrants.

Evidence and limitations. The adversarial family delivers the perceptual sharpness that regression objectives suppress, but four limitations recur across the reviewed studies. The first is hallucination: adversarial gradients push outputs toward the learned image manifold, so recovered high-frequency detail may be plausible rather than faithful, which limits reliability for evidential, forensic, or safety-critical imagery (Section 14). Second is training instability and sensitivity to distribution shift, which Wasserstein formulations mitigate but do not eliminate [6]. Third is a trade-off between perceptual quality and pixel

fidelity: LTT-GAN is evaluated on LPIPS/FID precisely because its pixel-fidelity scores do not capture its perceptual gains [22]. Fourth is temporal inconsistency across frames unless an explicit alignment or fusion module is added [6]. Evidence strength is moderate overall: TSRWGAN has been re-evaluated under two shared protocols by independent studies [16,28], whereas the unsupervised and inversion-based variants report results on small or method-specific test sets that preclude quantitative cross-comparison, and the astronomical AE-GAN results come from a custom simulated domain not comparable with terrestrial benchmarks.

7. Transformer-Based Turbulence Restoration

Convolution is local, but turbulence is not: blur and tilt correlations span distances far beyond any practical convolutional receptive field. Transformers resolve this mismatch through self-attention, which lets every image patch consult every other patch, and through temporal attention, which lets every frame consult its neighbors. The Turbulence Mitigation Transformer (TMT) is the reference architecture of this family [16] and is outlined in Figure 5. Input frames ($\{\mathbf{y}_t\}_{t=1}^{N}$) are first geometrically aligned by a tilt alignment module ($\hat{\mathbf{y}}_t = \mathcal{W}(\mathbf{y}_t, \delta_t)$), then patch-embedded. Stacked transformer blocks apply spatial multi-head self-attention and temporal cross-frame attention with LayerNorm residuals, and a blur decoder followed by temporal aggregation, i.e., $\hat{\mathbf{x}} = \sum_t w_t \mathbf{F}_t$, produces the restored output. Unlike conventional video restoration Transformers that treat restoration as a single-stage process, TMT explicitly incorporates turbulence physics by decoupling image restoration into separate de-tilting and de-blurring stages. This physics-guided design is further strengthened through a multi-scale training strategy that supervises restoration at different image resolutions. By progressively minimizing reconstruction errors across multiple scales, the model achieves scale-consistent tilt removal, since turbulence-induced tilts become more structured and less severe at lower resolutions. The original study reports improved cross-domain performance, from synthetic training data to real turbulence scenarios, for the complete TMT configuration; the isolated contribution of the multi-scale supervision component should be interpreted according to the ablation evidence reported there [16]. The tilt-removal loss is defined as

$$\mathcal{L}_{\text{tilt}} = \sum_{l=1}^{L} \gamma_l \mathcal{L}_{\text{char}}\left(\tilde{J}_e^l(x,t), J_e^l(x,t)\right), \tag{13}$$

where $J_e^l(x,t)$ denotes the ground-truth, tilt-free blur-only image at scale l, $\tilde{J}_e^l(x,t)$ represents the corresponding estimate produced by the tilt-removal module, γ_l is the weighting coefficient for scale l, and $\mathcal{L}_{\text{char}}(\cdot)$ denotes the Charbonnier loss function. The tilt-free blur-only images are generated from the original image ($J(x,t)$) according to

$$J_e^l(x,t) = \mathcal{B}_e^l\big(J(x,t)\big), \qquad l = 1, \ldots, L, \tag{14}$$

where $\mathcal{B}_e^l$ denotes the blur operator at scale l.

To further improve computational efficiency, TMT introduces a Temporal-Channel Joint Attention (TCJA) module that efficiently captures temporal information while significantly reducing memory consumption. This mechanism enables the network to process longer frame sequences, improve inference speed, and enhance generalization performance. In addition, TMT incorporates an advanced turbulence simulator based on an improved phase-to-space transformation framework to generate realistic training data. The simulator employs a Fourier-based dense-field sampling strategy, incorporates temporal correlations,

and supports a wider range of turbulence kernel sizes. Temporal correlation is modeled using an auto-regressive process for the white-noise vector (w_t):

$$w_t = aw_{t-1} + \sqrt{1-a^2}\, z, \tag{15}$$

where z is an independent and identically distributed Gaussian random variable and a is the regression coefficient, typically set to 0.9 to represent strong temporal correlation. Consequently, the Zernike coefficients evolve according to

$$\bar{a}_t = a\bar{a}_{t-1} + \sqrt{1-a^2}\, z, \tag{16}$$

which ensures that the simulated turbulence exhibits realistic temporal dependencies. Through the integration of physics-guided restoration, multi-scale supervision, efficient temporal attention, and realistic turbulence simulation, TMT achieves robust and effective image restoration under challenging atmospheric turbulence conditions.

The family has continued to develop along the axis of geometric handling. DeTurb pairs deformable 3D convolutions, which estimate flow and remove random geometric distortion across frames, with a U-Net of 3D Swin–Transformer blocks that performs the subsequent sharpening [55]. The design is instructive for the argument of Section 14.7 because it reinforces a pattern visible throughout the multi-frame literature: what these systems share is not a backbone but an architectural commitment to separating geometric correction from photometric restoration, a decomposition that follows directly from the tilt-then-blur structure of the physics [35] rather than from any property of attention. TMT realizes it through explicit de-tilting before de-blurring; DeTurb realizes it through deformable alignment before Swin-based refinement. That two independently designed systems converge on the same factorization is stronger evidence for the value of the decomposition than either system's benchmark score is for the value of its backbone.

Uncertainty treatment. The reviewed Transformer methods remain predominantly deterministic and do not introduce a dedicated mechanism for reporting either aleatoric or epistemic uncertainty, so strong restoration scores should not be interpreted as calibrated reliability. TMT is trained with a Charbonnier reconstruction objective and returns a point estimate, so aleatoric ambiguity is again averaged rather than represented, and the residual visual jitter reported by the original authors [16] is a temporal manifestation of that averaging: successive frames resolve the same underlying ambiguity differently, and the model has no mechanism to express that the variation is intrinsic rather than erroneous. What the family does contribute is an implicit *reduction* in aleatoric uncertainty rather than a representation of it, and this is the more accurate way to read its benchmark gains. Multi-frame attention with prior tilt alignment aggregates independent realizations of the random channel, so genuinely new information about the scene is accumulated across frames and the posterior narrows. That is a real advantage over single-frame restoration, and it is a different achievement from quantifying what ambiguity remains. On epistemic uncertainty, dependence on large simulator-generated training corpora leaves these models exposed to distribution shift, and none of the reviewed Transformer studies reports ensemble disagreement or another calibrated confidence signal designed to flag inputs outside the training distribution.

Evidence and limitations. In reporting terms, the Transformer evidence is among the strongest in the field: TMT was released with its benchmark, simulator, and efficiency figures (1.52 s per 540×960 frame on an RTX 2080 Ti), enabling the shared-protocol comparisons used throughout this review [16]. The family's limitations are also well documented. Self-attention has quadratic cost in token count, which drives high memory consumption at operational resolutions and motivated TMT's own temporal-channel attention approx-

imation; performance depends on large-scale synthetic training corpora, inheriting the simulator's assumptions; restoration quality is sensitive to the quality of the preceding tilt alignment and to the number of input frames; and residual visual jitter in outputs is reported by the original authors [16]. On the ATSyn-dynamic protocol, the recurrent DATUM configuration surpasses the evaluated TMT configuration in both quality and throughput [28], showing that multi-frame attention is not uniquely capable of exploiting temporal structure.

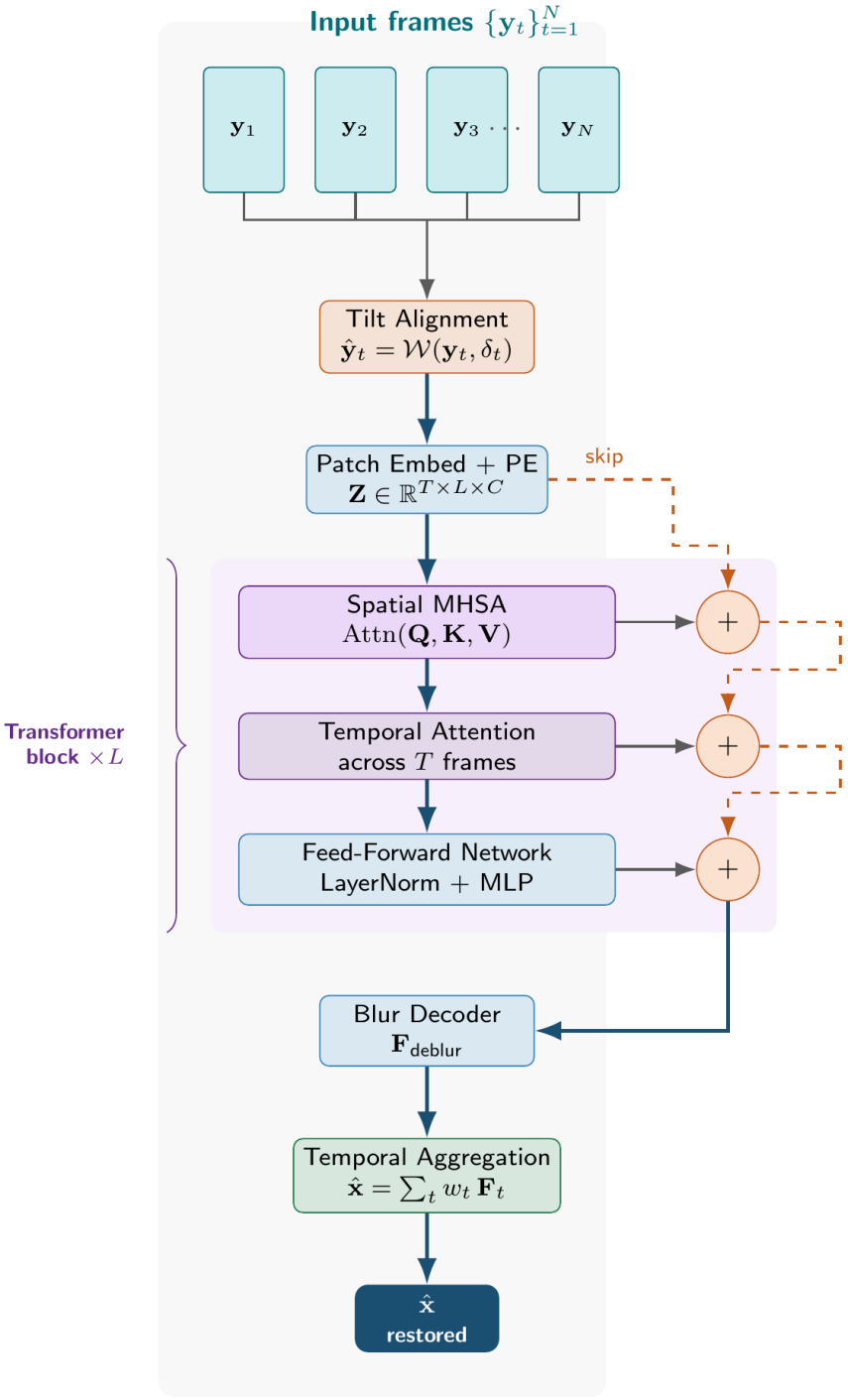


Figure 5. Transformer-based turbulence restoration in the TMT style, from tilt alignment through spatial and temporal attention to blur decoding and temporal aggregation.

8. Diffusion Probabilistic Models

Diffusion models bring an explicitly probabilistic answer to the one-to-many nature of turbulence restoration: rather than predicting a single best image, they learn to sample from the distribution of clean images consistent with the degraded observation. The conditional formulation used for restoration is illustrated in Figure 6, where a fixed forward process ($q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$) adds noise over T steps and a U-Net denoiser conditioned on the degraded observation ($\mathbf{y}$) learns the reverse transition ($p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$). Denoising Diffusion Probabilistic Models (DDPMs) [56] corrupt a clean image ($\mathbf{x}_0$) over T timesteps with Gaussian noise under a fixed variance schedule ($\{\beta_t\}_{t=1}^{T}$); the corrupted sample at any step follows in closed form, i.e.,

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{17}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ and a denoising network ($\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$) trained with a simple noise-prediction MSE objective reverses the process step by step [56,57]. For restoration, the network is additionally conditioned on the degraded observation, steering the sampled trajectory toward images consistent with the input [30]. Two turbulence-specific adaptations matter in practice. AT-DDPM initializes the reverse chain from a noise-perturbed turbulent image rather than pure Gaussian noise, which reduces inference time and anchors coarse facial structure already present in the input; its inference still requires $T = 60$ sampling steps [30]. The deterministic DDIM formulation [57] removes per-step stochastic injection and reaches comparable sample quality within 20–100 function evaluations, an essential efficiency lever for any deployment-oriented diffusion restorer.

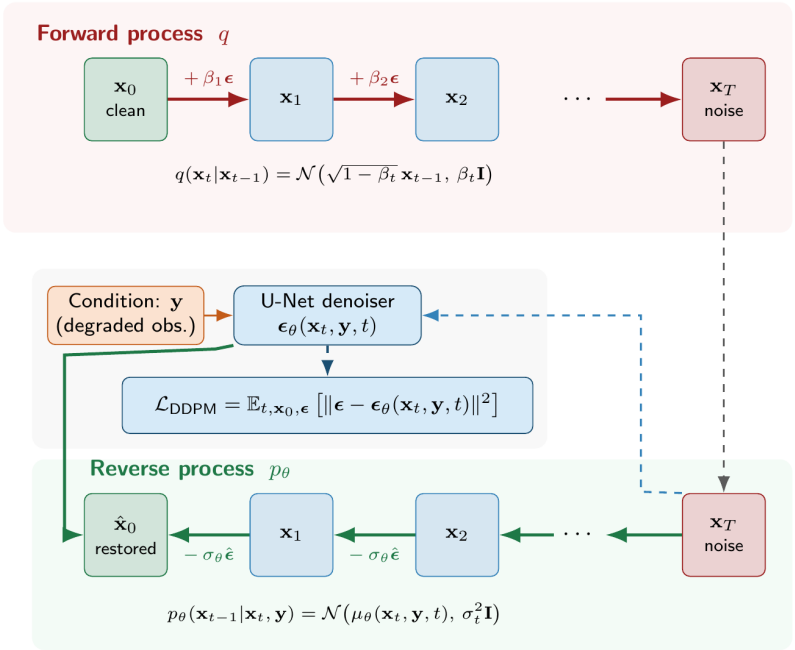


Figure 6. Conditional DDPM for turbulence restoration: a fixed forward noising process and a learned reverse process conditioned on the degraded observation.

Uncertainty treatment. Diffusion models provide a natural mechanism for representing multiple plausible restorations, but this capability is only sparsely used for uncertainty analysis in the turbulence literature reviewed here.

The principled part is real. A conditional diffusion model defines a distribution ($p_\theta(\mathbf{x}_0 \mid \mathbf{y})$) over clean images, given the degraded observation, so aleatoric uncertainty is not merely acknowledged but is the object the model is trained to represent. Drawing K

samples with independent initial noise yields an empirical posterior, from which a per-pixel variance map ($\mathrm{Var}[\hat{\mathbf{x}}]$) follows immediately and at no cost beyond the extra sampling passes. Pixels where the samples disagree are exactly the pixels where the turbulence has destroyed the information, so the variance map is a direct, spatially resolved readout of what the atmosphere took away. This is the natural home of the uncertainty-constrained visual inference framing adopted in this review.

The practice does not follow. No reviewed turbulence diffusion study reports such a map. AT-DDPM evaluates single samples against PSNR, SSIM, and LPIPS [30], and TSTRNet uses its DDPM stage as a deterministic-in-effect refiner scored on the same reference metrics [9]. Three obstacles explain the omission, and each suggests its own remedy. The first is cost: sampling K times multiplies an already expensive inference procedure, which is acute at the 4.12 s per image measured for the TSTRNet refinement stage, and it is the main reason deterministic samplers such as DDIM [57] are preferred, a choice that discards the posterior spread the model was trained to represent. The second is evaluative: reference metrics reward a single well-centered sample, so a model that honestly reports high variance is penalized relative to one that confidently commits, and the incentive runs against uncertainty reporting. The third is that sample spread is not automatically calibrated, since the empirical variance of a diffusion posterior is a model-dependent quantity that requires validation against actual error before it can be trusted, which is what distribution-free procedures for image restoration are designed to supply [24].

On epistemic uncertainty, the family offers no advantage. Sample diversity reflects the learned prior, and a diffusion model trained on simulator output will produce confidently tight posteriors on an out-of-distribution real sequence because the samples agree with each other about a prior that is, itself, wrong. Posterior spread and model correctness are distinct quantities, and conflating them would be the most dangerous available misreading of a diffusion variance map. Separating them requires an orthogonal mechanism such as an ensemble over independently trained models [23], and no reviewed study attempts it.

Evidence and limitations. Diffusion evidence in turbulence restoration is best characterized as promising but preliminary to moderate: AT-DDPM is peer-reviewed but evaluated on face-centric synthetic turbulence [30], and the strongest diffusion results in this review arise inside a hybrid system (Section 9) rather than from a standalone diffusion model. Four structural concerns apply to the family. Iterative sampling is slow relative to feed-forward alternatives (the diffusion refinement stage of TSTRNet is the most expensive component measured in this review, at 4.12 s per image on an RTX 3090 [9]), which makes suitability for time-critical systems uncertain. Output variability depends on the sampling configuration: stochastic DDPM-style sampling generally yields a different restoration on each run, whereas deterministic samplers such as DDIM with fixed initial noise are repeatable, with diversity governed by the sampler, the initial noise, and the noise schedule. Where stochastic sampling is used, single-sample evaluation understates output variability, and no reviewed study reports calibrated posterior or uncertainty evaluation over its samples. The sampled detail is drawn from a learned prior and can be hallucinated, inheriting the recognition hazards discussed in Section 14. Finally, the one-to-many formulation means restoration non-uniqueness is a feature of the model class, which complicates reference-based metric comparison against deterministic baselines.

9. Mamba State-Space Architecture

Self-attention solves the long-range problem but pays a quadratic price in sequence length, a cost that quickly becomes prohibitive for multi-frame restoration at operational resolutions. State-space models promise the same global context at linear cost, through the selective-scan mechanism sketched in Figure 7. Structured state-space sequence (S4)

models treat sequence modeling as a continuous-time dynamical system, mapping a one-dimensional input ($x(t) \in \mathbb{R}$) to an output ($y(t) \in \mathbb{R}$) through a latent state ($h(t) \in \mathbb{R}^d$), governed by $h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t)$ and $y(t) = \mathbf{C}h(t)$, where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are learnable parameter matrices [31]. For practical computation on discrete sequences, the continuous parameters are converted to their discrete counterparts ($(\bar{\mathbf{A}}, \bar{\mathbf{B}})$) via the zero-order hold (ZOH) discretization rule:

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}\big(\exp(\Delta\mathbf{A}) - \mathbf{I}\big) \cdot \Delta\mathbf{B}, \tag{18}$$

where $\Delta \in \mathbb{R}$ is a learnable timescale parameter that controls how strongly the model focuses on the current input versus persisting the prior state [31]. What separates the Mamba framework from earlier SSMs is its selection mechanism: rather than holding $\mathbf{B}$, $\mathbf{C}$, and Δ fixed across all timesteps, Mamba parameterizes them as functions of the input, enabling the model to selectively retain or discard information based on content [31]. This removes the linear time-invariance (LTI) constraint that had historically limited SSMs to time-invariant recurrences or global convolutions. The original authors report substantially higher inference throughput than size-matched Transformers on language-modeling benchmarks [31], but that advantage depends on model size, sequence length, batching regime, and a hardware-aware scan implementation, and it has not been measured for turbulence restoration; the efficiency of selective SSMs in this domain should therefore be treated as a promising, implementation-dependent potential rather than an established multiple.

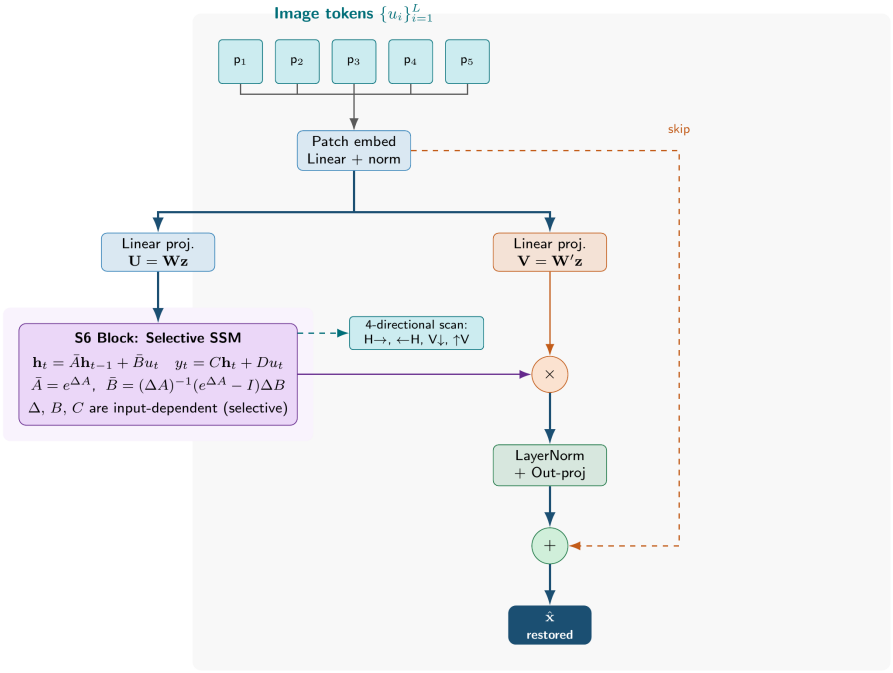


Figure 7. Mamba selective state-space block for image restoration, with input-dependent gating and a four-directional, two-dimensional scan.

The selective SSM is directly leveraged in MAMAT through the Enhancement and Detail Preservation (EDP) module, where a 3D Mamba block is deployed to recover contrast, sharpness, and fine texture lost to atmospheric turbulence distortion [18]. Within the module, residual Mamba (Res-Mamba) blocks sandwich the SSM operation between convolutional layers with a skip connection, expressed as $F_o = \text{Conv}_1(\text{SSM}(\text{Conv}_1(F_i))) + \text{Conv}_1(F_i)$, where F_i and F_o are the 3D input and output feature maps, respectively [18]. The selective SSM layer is trained using four direction-augmented inputs, each flipped along a distinct spatial axis, to capture

long-range spatiotemporal dependencies from multiple orientations without incurring the quadratic memory cost that would accompany a comparable Transformer-based design [18].

9.1. Hybrid SSM–Diffusion: TSTRNet

A more recent approach combines Mamba and conditional diffusion into a unified two-stage framework. Wu et al. [9] propose TSTRNet, in which a Mamba-based U-Net ($\mathcal{N}^G$) performs coarse global restoration, followed by a conditional denoising–diffusion probabilistic model ($\mathcal{N}^R$) that refines high-frequency detail and corrects residual geometric distortions through stochastic sampling.

In the first stage, a Vision State-Space Module (VSSM) is embedded into each encoder–decoder layer of the U-Net. The VSSM processes features through two parallel branches: one applies depth-wise convolution, SiLU activation, and a 2D-selective scan (2D-SSM) to capture long-range dependencies across four spatial directions (top left to bottom right and three additional traversals); the other branch applies a linear projection with SiLU activation. The outputs are fused via element-wise product and projected back to the original channel count. A Channel and Fourier Processing Block (CFPB) follows each VSSM layer, processing features in two parallel paths: spatial channel attention with a bottleneck network and channel attention module, as well as a frequency-domain path that decomposes features into amplitude and phase via FFT and applies separate convolutional branches to capture low- and high-frequency turbulence characteristics. This frequency-domain path directly exploits the physical relationship between PSFs and pupil functions expressed as $|\mathcal{F}[P_{\mathcal{D}}(r)e^{ik\phi(r)}]|^2$. The first-stage training objective is

$$\mathcal{L}_{\mathcal{N}^G} = \mathcal{L}_{pix} + \lambda_1 \mathcal{L}_{per} + \lambda_2 \mathcal{L}_{fft}, \quad \mathcal{L}_{fft} = \frac{1}{hw} \sum_{u,v} \|Re(\Delta\mathcal{F}) + Im(\Delta\mathcal{F})\|_1 \tag{19}$$

where $\mathcal{L}_{pix} = \|S_1 - I_{gt}\|_1$, $\mathcal{L}_{per}$ is a VGG-19 perceptual loss, $\mathcal{L}_{fft}$ penalizes patch-wise FFT residuals in amplitude and phase, $\lambda_1 = 0.01$, and $\lambda_2 = 0.1$.

The second stage uses a WideResNet-based conditional DDPM ($\mathcal{N}^R$) to refine the coarse output (S_1). Edge-aware conditioning is introduced by computing a Sobel gradient map ($S_2 = \sqrt{G_x^2 + G_y^2}$) from S_1, which is concatenated with S_1 to form a four-channel condition (S) that directs the denoising trajectory toward geometrically sharp outputs. A Frequency Filtering of Skip (FFS) module suppresses low-frequency redundancy in the U-Net skip connections of $\mathcal{N}^R$ using a learnable factor ($\mathbf{v} = 0.95$), preventing low-frequency noise from interfering with stochastic high-frequency recovery. The reverse process is then conditioned as follows:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, S) = \mathcal{N}(\mathbf{x}_{t-1};\, \mu_\theta(\mathbf{x}_t, S, t),\, \Sigma_\theta(\mathbf{x}_t, S, t)) \tag{20}$$

Evaluated on five synthetic scenes (Face, City, Forest, Mountain, and Sea; $D/r_0 \sim \mathcal{U}(1,3)$, 256×256) and five real turbulence benchmarks (CLEAR Hill House, CLEAR Mirage, Monument, Heat Chamber, and OTIS Fixed Patterns), TSTRNet achieves a PSNR/SSIM of 27.71 dB/0.819 on the synthetic Face scenario and 25.17 dB/0.821 on the real Heat Chamber real, outperforming the standalone CNN, GAN, Transformer, and SSM baselines evaluated in the same study. The $\mathcal{N}^R$ stage is the computationally heaviest of all compared methods, at 4.12 s per image (RTX 3090, patch-based DDPM with 64×64 patches), while $\mathcal{N}^G$ alone runs at 0.047 s per image with only 23.13 G FLOPs, offering a deployment-flexible design in which the diffusion refinement stage is applied selectively when perceptual quality or geometric precision is paramount.

9.2. Critical Assessment

In this review, Mamba is treated as one of six major architectural families, but its empirical footing is not comparable to that of the other five. Tables 5 and 6 contain one dedicated SSM study (MAMAT [18]) and one hybrid in which the SSM is half of a two-stage system (TSTRNet [9]), against a dozen or more entries for the CNN, GAN, and Transformer families. We include Mamba because its architectural argument is genuinely distinct and because the field is actively investing in it, not because the evidence yet supports a favorable verdict. This subsection states the verdict we believe the current evidence does support, organized around efficiency, which we take to be the decisive dimension for this family.

9.2.1. Efficiency: Theory Versus Measurement

Everything that distinguishes Mamba from a Transformer in this domain reduces to a claim about cost. Both deliver global context; only the SSM claims to do so in $O(L)$ rather than $O(L^2)$ time in the number of tokens (L). This makes efficiency the dimension on which the family must be judged, and the reviewed evidence on that dimension is thin to the point of being anecdotal.

Asymptotic complexity is not runtime, and the difference between the two is where the practically relevant behavior lies. The linear-time advantage is realized only through a hardware-aware selective-scan kernel that keeps the recurrent state in fast on-chip memory, and its benefit is contingent on sequence length, batch size, channel width, and the specific accelerator. The crossover point below which an optimized quadratic-attention implementation is simply faster in wall-clock terms is genuinely favorable to Transformers at the token counts typical of patch-based image restoration, and it has not been located for any turbulence workload. The throughput multiples reported by the original authors were established on the basis of autoregressive language modeling [31], a regime with sequence lengths orders of magnitude longer than an image patch and with a decoding pattern that plays to recurrence in a way image restoration does not. Transferring those multiples to turbulence restoration is an extrapolation, not a measurement.

What the turbulence literature actually provides is one number: the TSTRNet first stage runs at 0.047 s per image with 23.13 G FLOPs on an RTX 3090 [9]. That figure is encouraging, but it is a single measurement of a single configuration on a single high-end desktop GPU, and it describes a coarse restoration stage whose output the same paper deems insufficient without a 4.12 s diffusion refiner. It is not a controlled efficiency comparison. It does not isolate the SSM contribution from the depth, width, and channel-attention choices made around it, and it says nothing about the platforms where linear complexity should matter most. Selective scan is a memory-bandwidth-bound sequential recurrence with limited parallelism across the sequence axis, so it maps to embedded accelerators, mobile neural processing units, and integer-quantized inference far less predictably than the dense convolutions and matrix multiplications that CNN and Transformer deployment toolchains have spent a decade optimizing. A linear-complexity architecture that no edge runtime executes efficiently delivers none of its promised advantage.

The comparison the field most needs is also the one most conspicuously absent. DATUM, a recurrent design, sustains 9.17 frames per second at 960×540 with 5.754 M parameters while exceeding the evaluated TMT configuration's quality on the same protocol [28]. Recurrence is, itself, $O(L)$. Linear-time global context is therefore not exclusive to selective SSMs, and DATUM demonstrates that a conventional recurrent architecture already achieves it, at competitive quality, with published throughput on a shared benchmark. No reviewed study compares a Mamba turbulence restorer against DATUM under matched conditions. Until one does, the efficiency case for SSMs in this domain rests on an asymptotic argument that an existing, simpler, better-evaluated alternative also satisfies.

9.2.2. One-Dimensional Scan, Two-Dimensional Image

A second reservation is structural rather than empirical, and we regard it as the deeper of the two. The selective scan is a causal recurrence over an ordered sequence, whereas an image is an unordered two-dimensional field in which spatial dependency is isotropic and bidirectional. Reconciling them requires flattening of the image into a sequence, and both MAMAT and TSTRNet do so with four fixed traversals [9,18].

This compromise has costs that the reviewed studies do not measure. The traversal set is a fixed inductive bias chosen a priori, not learned, so the model's ability to propagate information between two pixels depends on their positions in the scan orders rather than on their spatial or physical relationship. Two pixels that are near neighbors in the image can be separated by an entire row in every traversal. This sits awkwardly with the physics: the isoplanatic angle (θ_0) defines an approximately isotropic patch within which the PSF is roughly constant (Section 3), and there is no reason for that patch to align with any axis-parallel raster order. Self-attention has no such difficulty, since it is permutation-equivariant over tokens and imposes no traversal at all. Multi-directional scanning is also not free: four traversals multiply the scan cost fourfold, which erodes the very efficiency advantage that motivated the architecture, and this trade-off is neither ablated nor reported in the turbulence studies. Whether four fixed orders adequately approximate isotropic two-dimensional dependency is an open modeling question, and it is currently answered by assertion rather than by ablation.

9.2.3. Uncertainty Treatment

The state-space family offers nothing distinctive on either uncertainty axis. Selective SSMs are deterministic sequence models trained with reconstruction objectives, so they average over aleatoric ambiguity exactly as CNNs do; MAMAT reports no uncertainty output [18], and in TSTRNet, the entire probabilistic burden is carried by the diffusion stage, not by the Mamba stage [9]. That division is, itself, informative and supports the deployment reading offered below: the SSM contributes efficient global context, and the ambiguity of the inverse problem is handled elsewhere. On epistemic uncertainty, the family is exposed like every other, with the added consideration that the input-dependent selection mechanism, which gates what the recurrence retains, has an unknown failure mode under distribution shift and is not examined in any reviewed study.

9.2.4. Verdict

We therefore adopt an explicit wait-and-see position. The genuine advantages of Mamba for turbulence restoration are *not yet validated*. The architectural argument is coherent, but the empirical case rests on two studies, neither of which isolates the SSM contribution through controlled ablation on a shared public benchmark, and the efficiency claim that constitutes the family's entire rationale has never been measured against a strong recurrent baseline or on a platform where efficiency is binding. Where we do see present value is narrow and specific: as an efficient coarse-restoration stage that supplies global context cheaply within a hybrid pipeline, which is exactly the role $\mathcal{N}^G$ plays in TSTRNet [9]. A coarse stage of under 0.05 s feeding a selectively applied refinement stage is a defensible deployment pattern. We do not consider the evidence sufficient to recommend selective SSMs over Transformer or recurrent alternatives for turbulence restoration in general, and practitioners should demand three specific results before adopting one: a controlled ablation replacing the SSM with an attention or recurrent block of matched capacity on a public benchmark, wall-clock and memory measurements on the actual target hardware rather than a desktop GPU, and a matched comparison against DATUM-class recurrent

baselines. Absent those results, the family's status in this review is promising but unproven, and we would rather say so than grant it parity with paradigms that have earned it.

Evidence and limitations. Summarizing the above, turbulence-specific evidence for the state-space family is emerging and comparatively limited, resting on one conference study (MAMAT, evaluated on a custom mild-turbulence video corpus without shared-protocol baselines [18]) and one journal study in which the Mamba component was half of a hybrid design (TSTRNet [9]). Component-level conclusions about the SSM contribution await controlled ablation on public benchmarks. Note also that MAMAT's 30.99 dB reported in Table 4 is not evidence of superiority over the Transformer or GAN entries: its baseline corpus is substantially milder (baseline PSNR ≈30.4 dB against ≈22–24 dB for the TMT synthetic suite), so the number reflects an easier problem rather than a stronger method, which is exactly the confusion that separating Table 3 from Table 4 is intended to prevent.

10. Physics-Informed Deep Learning Frameworks

Encoding physical laws into neural network training provides a principled way to compensate for limited labeled data and to improve generalization to turbulence regimes not seen during training. Figure 8 summarizes how the physical model enters the pipeline. A parameter estimator infers the Fried coherence length ($\hat{r}_0$) and the isoplanatic angle ($\hat{\theta}_0$) from the degraded input ($\mathbf{y}$); these values parameterize a turbulence model ($H(r_0, \theta_0, \ell_0)$) and the associated degradation operator ($\mathbf{y} \approx H * \mathbf{x} + \boldsymbol{\eta}$), which, in turn, supply the physical prior ($\phi_{\text{phys}}(\hat{\mathbf{x}}, H)$). The encoder–decoder network ($D_\theta(E_\theta(\mathbf{y}))$) is then trained with a hybrid objective, i.e., $\mathcal{L} = \|\mathbf{y} - H * \hat{\mathbf{x}}\|^2 + \lambda_1 \phi_{\text{phys}}(\hat{\mathbf{x}}) + \lambda_2 \mathcal{L}_{\text{percep}}$, that combines data fidelity, physical plausibility, and perceptual quality. Raissi et al. formalized this principle in the physics-informed neural network (PINN) framework [58]: a network is trained to simultaneously fit observed measurements and satisfy a governing PDE at a set of collocation points, with the PDE residual acting as a physics-grounded regularizer. The composite PINN loss takes the form of

$$\mathcal{L} = \underbrace{\frac{1}{N_u}\sum_{i=1}^{N_u}\left|u\left(t_u^i, x_u^i\right) - u^i\right|^2}_{\mathcal{L}_u} + \underbrace{\frac{1}{N_f}\sum_{i=1}^{N_f}\left|f\left(t_f^i, x_f^i\right)\right|^2}_{\mathcal{L}_f} \tag{21}$$

where $\mathcal{L}_u$ enforces data consistency at N_u measurement points and $\mathcal{L}_f$ penalizes the PDE residual at N_f collocation points [58].

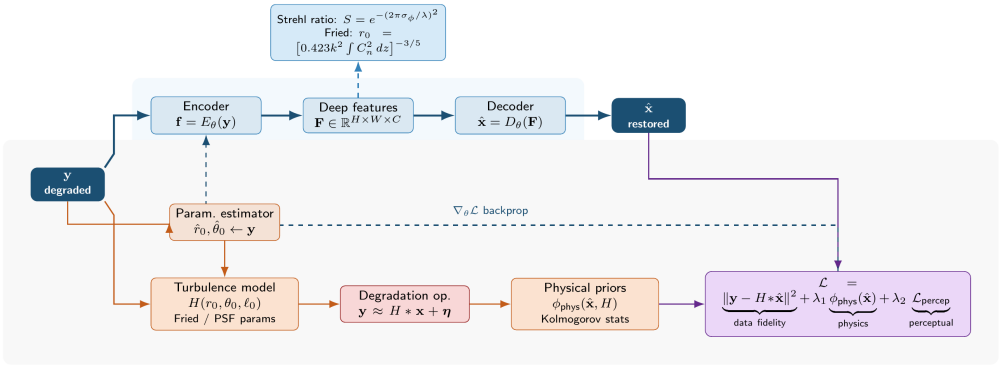


Figure 8. Conceptual schematic of the mechanisms by which turbulence physics enters a learned restoration pipeline. The figure synthesizes design patterns from the cited literature and does not depict a single implemented system.

In turbulence image restoration, this logic transfers to three concrete design patterns: (i) enforcement of consistency between the restored image and the known forward tur-

bulence degradation model, (ii) the use of physics-derived parameters such as r_0 and θ_0 as inductive biases injected into the network at inference time, and (iii) generation of physics-parameterized synthetic training data to reduce the domain gap. Each pattern uses physical knowledge in a different stage of the learning pipeline.

Jaiswal et al. [25] implemented the first pattern through a physics-driven stochastic refinement framework for turbulence restoration. Their approach constructs the forward degradation model using the Tilt-Then-Blur (TTB) pipeline [35], which simulates geometrically displaced, spatially varying blur consistent with atmospheric optics. A restoration network is trained with a loss that penalizes inconsistency between the degraded network output when re-degraded through the TTB simulator and the observed turbulent input. This forward-model consistency objective acts as the turbulence analog of the PINN collocation loss: it does not require paired clean references for every training sample only a differentiable forward model. A stochastic refinement stage then samples plausible high-frequency details from a learned conditional distribution, improving perceptual quality beyond what deterministic regression achieves. The framework is evaluated against PSNR, SSIM, and LPIPS on both synthetic and real turbulence sequences (see Table 4), demonstrating that physics-consistency constraints improve generalization to real optical paths compared to purely data-driven baselines [25].

Yin et al. [59] implemented the second pattern for satellite video turbulence. They constructed a physics-informed turbulence signature derived from the refractive-index structure constant (C_n^2), estimated in real time directly from image gradients and localized temporal intensity variance within the video stream, eliminating dependence on external atmospheric sensors. This signature is injected as an inductive bias into three dedicated modules for feature alignment, reference-feature updating, and inter-frame fusion, guiding the network toward physically consistent restorations. Evaluated across five satellite platforms and 18 test scenes, the framework achieves average PSNR and SSIM improvements of 0.27 dB and 0.0015 over the strongest competing baseline on synthetic data, alongside perceptual quality improvements exceeding 4% in BRISQUE and NIQE on real turbulence sequences [59].

Together, both works indicate that when a reliable physical model of the degradation process is available, integrating it into the loss function or the architectural inductive bias can improve accuracy and generalization relative to the purely data-driven baselines evaluated in the same studies, particularly under turbulence conditions outside the training distribution. The size of the reported gains is modest in absolute terms (e.g., $+0.27$ dB average PSNR for TMTS over its strongest baseline [59]), and neither study isolates the physics component through an ablation that holds all other design factors fixed.

Uncertainty treatment. Physics-informed methods primarily constrain model uncertainty through prior knowledge of the degradation process; this should be distinguished from directly quantifying the irreducible ambiguity of the observation. Their central mechanism is a reduction in *epistemic* uncertainty: a forward-model consistency term or a C_n^2-derived conditioning signal injects knowledge the training data does not contain, which is why these methods report their gains under conditions outside the training distribution [25,59]. Constraining the solution to be consistent with the degradation physics also bounds hallucination, which is the same reduction viewed from the output side. On *aleatoric* uncertainty, a physical consistency constraint does not, by itself, provide a probabilistic representation of the remaining ambiguity. A forward model constrains which reconstructions are admissible; it does not resolve which admissible reconstruction is correct because the turbulence PSF is genuinely information-destroying, and the residual ambiguity survives any consistency constraint. This is precisely why the stochastic refine-

ment stage of PiRN-SR [25] is needed at all: the physics narrows the feasible set, and a generative component must still sample within it.

Two cautions follow, and they qualify the family's apparent advantage. First, the epistemic reduction is only as sound as the assumed model, so a physics constraint built on a stationary C_n^2 profile, an idealized aperture, or Kolmogorov statistics does not eliminate epistemic uncertainty but relocates it into the model assumptions, where it is harder to detect because the output now looks physically principled. The Kolmogorov monoculture of Section 3.3 is the sharpest instance: a slant-path observation violates the assumed spectrum, and a physics-constrained network will enforce the wrong constraint confidently. Second, physical parameters such as r_0 or C_n^2 are, themselves, estimated, often from the degraded imagery, and no reviewed study propagates the uncertainty of those estimates into the restoration output. That propagation is a well-posed and tractable piece of future work, and it would make this family the natural place to deliver physically grounded confidence intervals rather than point estimates.

Evidence and limitations. Physics-informed evidence is moderate: the reviewed studies are peer-reviewed and span two application domains, but each evaluates its physics mechanism inside a distinct architecture, dataset, and protocol, so the generality of the benefit is not yet established. The family also carries risks specific to its premise. Its value depends directly on the fidelity of the assumed forward model; the simplified propagation models used in practice (Zernike-based tilt-then-blur pipelines, stationary C_n^2 profiles, and idealized apertures) deviate from field conditions, so a physics constraint can import simulator bias into the network rather than remove it. Physical parameters such as C_n^2 or r_0 must, themselves, be estimated, often from the degraded imagery, and errors in these estimates propagate through the conditioning or consistency terms. These caveats do not negate the approach; they define its scope: physics integration is most defensible where the degradation physics is well characterized and least defensible as a claimed substitute for real training data.

11. How the Field Treats Uncertainty

Table 1 consolidates the family-level assessments of Sections 5–10 into a single view. Two entries in that table carry a specific meaning. "Implicit" denotes a design that responds to the uncertainty source without representing or reporting it, and "None" denotes that no mechanism was identified in the reviewed studies. No reviewed study in any family reports a *calibrated* uncertainty estimate, that is, one validated against realized error. Reading down the columns rather than across the rows exposes three patterns that no individual paper reveals, and together, they constitute the central diagnostic finding of this review.

First, **aleatoric uncertainty is pervasive but almost never represented.** Every family concedes, in its design, that the degraded observation underdetermines the clean image. Adversarial objectives exist because the conditional mean is unsatisfactory; diffusion models are formulated as posterior samplers, and multi-frame aggregation exists to narrow the posterior, yet with the single exception of the heteroscedastic variance head of Yasarla and Patel [13], no reviewed system *outputs* that ambiguity. The field has built an entire architectural progression around the one-to-many structure of the problem while continuing to deliver one-to-one outputs.

Second, **epistemic uncertainty is essentially unmeasured**, which is remarkable, given that the synthetic-to-real domain gap is the field's most cited problem and is precisely an epistemic phenomenon. The domain gap is a statement that models are unreliable outside their training distribution; epistemic uncertainty is the quantity that measures exactly that unreliability, yet the reviewed corpus attacks the gap entirely through mitigation (adversarial feature alignment, teacher–student transfer, semi-supervised training on unlabeled real

data, and physics conditioning [6,19,25,41]) and never through measurement. AT-Net's MC dropout [21] is the sole exception in the corpus, and it uses the estimate internally as restoration guidance rather than reporting it as a reliability signal. The practical consequence is that no deployed turbulence restorer in this literature can currently answer the question an operator most needs answered: *Is this scene one my model understands?*

Third, **calibration is absent everywhere.** Even where an uncertainty-like quantity is produced, whether a variance head, an MC-dropout map, or the spread of diffusion samples, no reviewed study validates it against realized error. An uncalibrated uncertainty map is not merely uninformative but actively hazardous because it invites downstream systems to trust a number that has never been checked. This is the gap that distribution-free procedures developed for image restoration are designed to close [24], and importing them is among the most tractable contributions available in this area.

These three patterns motivate the elevation of uncertainty quantification to a first-order priority in Challenge 2 of Section 16, ranked alongside rather than beneath the synthetic-to-real domain gap.

12. Recent Dataset Sources

A persistent challenge in atmospheric turbulence restoration research is the scarcity of datasets that are simultaneously large-scale, physically representative, and accompanied by clean ground truth. Real turbulence is inherently uncontrolled, making paired acquisition difficult, while fully synthetic datasets carry the risk of distributional mismatch with field conditions. The ten resources described below collectively span laboratory-controlled captures, genuine outdoor sequences, physics-based synthetic generation, and domain-specific collections from solar and satellite imaging. Figure 9 organizes them by acquisition type, and Table 8 consolidates their key attributes. The real-acquisition sets (OTIS, CLEAR, RLR-AT, and SUTO-Solar) capture genuine distortions but typically lack pixel-aligned ground truth. The simulated sets (Heat Chamber, TMT synthetic, and the Turbulence Text Dataset) supply paired training data through physics-based turbulence simulation. The specialized-domain sets (SILO, BRIAR, and CDnet-2014) target satellite imagery, long-range biometrics, and change detection. Five attributes govern the choice among them: ground-truth availability, pairing, turbulence-severity range, licensing, and still-image versus video format. Uncertainty in licensing terms or repository persistence is noted explicitly, as these practical factors directly affect reproducibility.

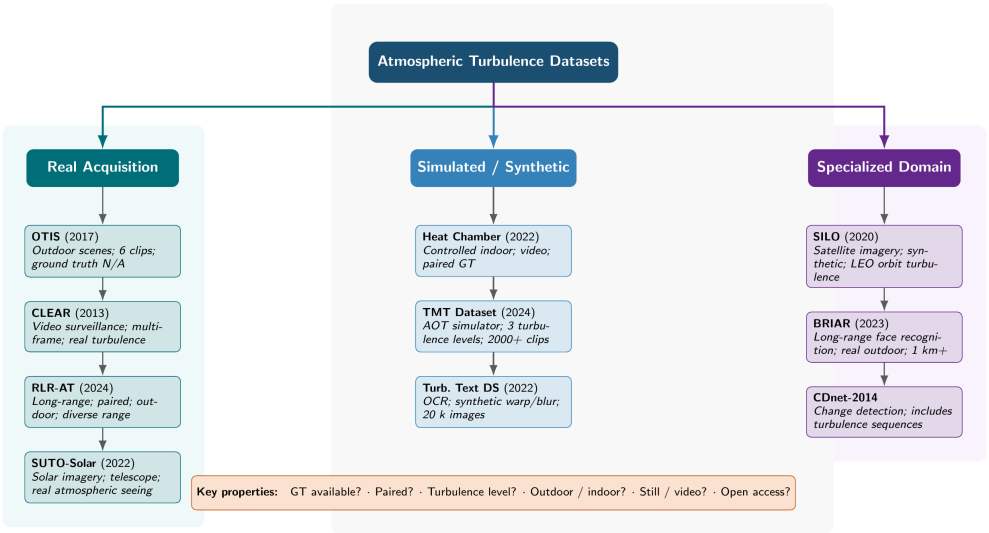


Figure 9. Schematic taxonomy of atmospheric turbulence datasets organized by acquisition type.

Table 8. Standardized summary of atmospheric turbulence datasets used in recent literature.

Dataset	Type	Content	Paired GT	Reported Size	Severity Spec.	Access	License/Locator	Recommended Use/Known Limitation
Heat Chamber [15]	Real (lab)	Static images	Yes (heater-off ref.)	N.R.	$D/r_0 \approx 3$	Open (direct)	No license declared; no data-use agreement. Locator: project repo. cited in [15]	Paired evaluation at moderate to strong severity; lab heating may differ from field propagation
Turbulence Text [15]	Real (outdoor, 300 m)	Static images	Yes (known text)	N.R.	$D/r_0 \approx 2.5$–4	Open (direct)	No license declared; same terms as Heat Chamber; UG2+ (CVPR 2022) corpus	Task-driven OCR protocol plus pixel metrics; severity varies with ambient conditions
TMT synthetic [16]	Synthetic (P2S)	Video (12-frame clips)	Yes	N.R.; corpus regenerable	Config. C_n^2	Open (direct)	No license declared. Locator: https://xg416.github.io/TMT/ (accessed on 16 July 2026)	Multi-frame training and shared-protocol comparison; simulator assumptions transfer into models
OTIS [60]	Real (outdoor)	Video, ≈300 frames *per seq.*	No	N.R. aggregate; ≈300 fr./seq.	Not quantified	Open (direct)	No license declared; distributed with [60]	Qualitative or no-reference consistency check on genuine turbulence
CLEAR [61]	Real (outdoor)	Video	No	N.R.	Not quantified	Open (TBC)	No license declared; confirm terms with maintainers	Qualitative comparison; some frames $<400 \times 400$ px, excluding fixed-crop protocols
CDnet 2014 [62]	Real (outdoor)	Video	Masks only	11 categories, 4–6 seq. each	Not quantified	Open (direct)	Free for academic research; http://changedetection.net	Temporal coherence and robustness on non-turbulent inputs; PSNR not applicable
SUTO-Solar [63]	Real (solar)	Solar imagery	No	N.R.	Varying, not param.	Open (direct)	**CC BY 4.0**; Zenodo `10.5281/zenodo.7064324`; verified 16 Jul 2026	Domain-specific solar and astronomical evaluation only
RLR-AT [32]	Real (long range)	Video	No	1500 sequences	Range 1–13 km; no D/r_0	Unverified	No license declared. Locator: project page in [32]; confirm access	Operationally relevant stand-off ranges; no-reference evaluation only
SILO [64]	Synthetic (LEO)	Static images	Yes	90,000 images	Controlled sim. params	Open (TBC)	AFRL release terms; consult originating documentation [64]	Satellite-domain evaluation with exact ground truth
BRIAR [28]	Real (extended range)	Video	Partial (biometric labels)	N.R.	Extended range, not param.	Restricted	Application via the IARPA BRIAR program; eligibility constraints apply	Biometric task evaluation; independent replication infeasible for most groups

Reading the table: Size and severity are stated only as given by the primary source, and N.R. marks a value that the source does not report; nothing is estimated here. The severity column states whichever variable the source specifies, whether D/r_0, capture range, or none. Access follows a fixed vocabulary: *Open (direct)* is openly downloadable without registration, *Open (TBC)* is publicly hosted with license or usage conditions still to be confirmed with the maintainers, *Unverified* means availability could not be verified at the final search date, and *Restricted* means formal application and eligibility constraints apply. The License/Locator column reports only terms stated explicitly by the maintainer or the originating publication, since the absence of a stated license is not equivalent to permissive terms. **Reproducibility note:** Only SUTO-Solar carries an explicit machine-readable license (CC BY 4.0). The remaining nine resources are distributed without a declared license, which is, itself, a finding: the field's dominant benchmarks rest on informal availability rather than on terms that guarantee continued redistribution. Users should obtain written permission from maintainers before redistributing derived data, and journals should not treat "publicly downloadable" as equivalent to "openly licensed".

12.1. Heat Chamber Dataset

Introduced by Mao et al. in their ECCV 2022 study on single-frame turbulence mitigation [15], the Heat Chamber dataset achieves paired ground truth through a carefully constructed laboratory arrangement rather than simulation. A target image is displayed on a screen 20 m from the camera, and three heat chambers are positioned at equal intervals along the optical path to induce spatially varying, anisoplanatic distortion that resembles real atmospheric degradation. Clean reference frames are then captured by switching off the heating elements, preserving the identical scene geometry and enabling pixel-accurate correspondence. The dataset is characterized by $D/r_0 \approx 3$, placing it squarely in the moderate-to-strong turbulence regime that is practically relevant for surveillance and long-range imaging tasks. Published evaluations employ crops of 256×256 and 400×400 pixels to match the input requirements of competing architectures. Paired RGB image folders are freely available through the maintainers' public repository without a data-use agreement [15].

12.2. Turbulence Text Dataset

Released alongside the Heat Chamber data in the same ECCV 2022 work [15], the Turbulence Text Dataset pursues a complementary design philosophy by moving outdoors. Printed text targets are photographed at a camera-to-subject distance of 300 m under genuine atmospheric conditions, with D/r_0 varying between approximately 2.5 and 4, depending on the ambient temperature at the time of collection. This variability, while a practical limitation for controlled comparison, accurately reflects the range of turbulence strengths encountered in operational settings. The textual target content enables two complementary evaluation protocols that many restoration benchmarks lack: conventional pixel-fidelity metrics (PSNR and SSIM) against the known reference pattern, and a task-driven protocol that measures the OCR detection rate and longest common subsequence against the output of a Tesseract engine. That second pathway provides a functionally grounded quality measure independent of pixel-level statistics, which is useful for assessing methods intended to support downstream scene understanding. The dataset served as the primary evaluation corpus for the 5th UG2+ Prize Challenge at CVPR 2022 and is hosted in the same Purdue repository under identical open research terms.

12.3. TMT Dataset and Simulator

Zhang et al. accompany their Turbulence Mitigation Transformer (TMT) [16] with both a pre-generated training corpus and a publicly released simulation framework, making this one of the more complete dataset packages available to the community. Synthetic sequences are produced by a Fourier-domain sampler that incorporates inter-frame temporal correlation, a property essential for the training of multi-frame video restoration models whose performance degrades when trained on temporally independent samples. Each clip contains twelve frames paired with a clean reference, and the simulator source code is made available so that researchers can generate additional data matched to specific C_n^2 profiles or scene types. This flexibility is practically significant: fixed pre-generated datasets impose a fixed turbulence distribution at training time, whereas an accessible simulator allows for domain-specific adaptation without external data collection. The pre-generated corpus and simulation code are maintained in the authors' public project repository [16].

12.4. OTIS Dataset

Predating the deep-learning era of turbulence restoration, the Open Turbulent Image Set (OTIS) remains a widely used real-world evaluation resource [60]. It contains both static and dynamic sequences acquired through atmospheric turbulence. The sequence lengths are not uniform: many of the static pattern sequences contain 300 frames, while some contain fewer, and the fixed-background Door sequence contains 300 frames. Importantly, OTIS is not uniformly without reference information. The published dataset provides ground-truth images for most static pattern sequences, whereas the Door sequence has no clean ground truth; for the dynamic sequences, frame-wise bounding boxes support tracking evaluation rather than pixel-aligned restoration assessment [60]. Accordingly, the appropriate metric depends on the selected subset and task. OTIS appears consistently across recent literature, including in the TMT [16] and Turb-Seg-Res [65] models, where it provides a recurring real-world evaluation reference. Its principal value is that it exposes restoration methods to genuine outdoor turbulence, while its heterogeneous subsets and annotations require care when comparing results reported under different protocols.

12.5. CLEAR Dataset

The CLEAR dataset provides real outdoor turbulence sequences captured under natural atmospheric conditions, of which the hillhouse sequence has become a near-universal visual reference for qualitative comparison in the restoration literature [61]. Clean ground truth is unavailable, and published results therefore appear predominantly in qualitative form or report no-reference image quality scores. A spatial constraint worth recording explicitly when using this dataset is that certain sequences contain frames smaller than 400×400 pixels, which excludes them from standardized comparisons against architectures such as AT-Net (which requires 256×256 input crops) or TurbNet (400×400 input). Inconsistent handling of this cropping boundary across papers can affect reported rankings; authors adopting this dataset should document the exact frame selection and crop strategy applied.

12.6. CDnet 2014 Change Detection Benchmark

Originally compiled for the Change Detection Workshop at CVPR 2014 [62], the CDnet 2014 benchmark covers eleven video categories with four to six sequences each. Four of those categories contain real outdoor footage affected by visible heat-haze and refractive distortion, making them directly applicable to turbulence restoration evaluation. The remaining non-turbulent categories serve a complementary purpose: they allow for assessment of whether a restoration model introduces unwanted smoothing or color artifacts on inputs that do not require correction, an aspect of robustness that purely turbulence-focused datasets cannot probe. Ground truth in CDnet takes the form of pixel-level binary segmentation masks for change detection rather than clean image references, so PSNR-based restoration evaluation is not applicable to the turbulence subset. This dataset is most appropriate for assessing temporal coherence and dynamic-scene handling and is well suited to experiments that require camera or object motion to coexist with turbulent degradation. The collection is available without charge for academic research through the benchmark's public portal [62].

12.7. SUTO Solar Turbulence Dataset

Hosted on Zenodo under identifier 7064323, the SUTO dataset consists of solar surface imagery acquired under varying atmospheric turbulence intensities [63]. The photometric statistics of solar observation differ substantially from terrestrial scenes: dynamic range, spectral profile, and sensor noise characteristics all deviate from those of standard optical imaging hardware. Consequently, SUTO is unsuitable as primary training data for general turbulence mitigation models but serves as a meaningful evaluation resource for domain-specific demonstrations in solar or astronomical imaging contexts. The Zenodo record is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits redistribution and reuse with attribution (license verified against the Zenodo record on 16 July 2026).

12.8. RLR-AT Long-Range Atmospheric Turbulence Dataset

The Real Long-Range Atmospheric Turbulence (RLR-AT) dataset reportedly contains 1500 real turbulence sequences with camera-to-target distances spanning 1 km to 13 km across a variety of outdoor scenes. This capture range makes it one of the most operationally relevant real-data resources available, directly addressing the stand-off distances encountered in ground-based surveillance, aerospace imaging, and reconnaissance applications. No clean ground truth accompanies the sequences, so evaluation must rely on no-reference metrics or task-specific proxies. Licensing terms and data-access arrangements should be confirmed with the project maintainers prior to use [32], as long-range outdoor captures

occasionally carry site-specific or institutional restrictions that are not always reflected in the public project page.

12.9. SILO Satellite Imagery Dataset

The Scored Images of LEO Objects (SILO) dataset provides 90,000 fully synthetic images of low-Earth-orbit satellite objects observed through simulated atmospheric turbulence with controlled parameters and known scene geometry [64]. Because all images are computer-generated, pixel-accurate paired ground truth is available for every sample, enabling PSNR-based evaluation without the ambiguity inherent in real outdoor captures. SILO is particularly relevant for satellite remote sensing, where turbulence characteristics differ fundamentally from the ground-level propagation paths captured in other datasets. Researchers should consult the originating Air Force Research Laboratory documentation for the appropriate citation key and any conditions governing academic publication.

12.10. BRIAR Dataset

Produced under the U.S. IARPA Biometric Recognition and Identification at Altitude and Range (BRIAR) program, this dataset contains face and whole-body biometric imagery of consenting subjects captured at extended distances under real atmospheric turbulence. It constitutes the primary validation corpus for the DATUM (Deep Atmospheric Turbulence Mitigation) network [28], and its inclusion here reflects its prominence in several key references in the turbulence mitigation literature. Access is not publicly available: researchers must apply through IARPA channels, and access has historically been restricted to U.S.-based or formally partnered institutions. This dataset is included for completeness and because understanding its scope is necessary for correct interpretation of DATUM evaluation results, but it is not a resource that most international research groups can access independently. Methods that report results on BRIAR should be noted as such during cross-paper comparison, since the restricted access makes independent replication infeasible.

13. Loss Functions and Evaluation Metrics

13.1. Summary of Loss Functions

Table 9 consolidates the principal loss functions employed across atmospheric turbulence restoration networks, spanning pixel-level, perceptual, structural, adversarial, and physics-aware objectives. At the lowest level of supervision, $\mathcal{L}_1$ and $\mathcal{L}_2$ penalties enforce direct pixel fidelity, though $\mathcal{L}_2$ is prone to over-smoothing under the spatially non-uniform degradations characteristic of turbulence, making the more outlier-robust $\mathcal{L}_1$ norm a common default. Perceptual losses computed over intermediate VGG feature maps and the structure-sensitive SSIM term complement pixel-wise terms by penalizing deviations in mid-level texture and structure, properties that are perceptually salient but poorly captured by scalar norms. Adversarial losses drive the generator toward the manifold of photorealistic images, recovering high-frequency details that purely regression-based objectives tend to suppress, while total variation and frequency-domain terms provide complementary regularization against spatial noise artifacts and spectral distortion, respectively.

Physics-informed terms explicitly couple the network output to the forward imaging model, providing domain-consistent supervision, even when clean ground truth is unavailable. Finally, diffusion-based networks replace the conventional reconstruction objective with a noise-prediction loss derived from the denoising diffusion probabilistic model (DDPM) framework, enabling iterative stochastic refinement of fine structural detail.

Table 9. Loss functions used in atmospheric turbulence restoration networks.

Loss Term	Formula	Role/Notes
$\mathcal{L}_1$ pixel	$\|\hat{I} - I\|_1$	Pixel fidelity; sharper than MSE; robust to outliers
$\mathcal{L}_2$ MSE	$\|\hat{I} - I\|_2^2$	Smooth gradients; over-smoothing risk for turbulence
Perceptual (VGG)	$\frac{1}{C_j H_j W_j} \left\|\phi_j(\hat{I}) - \phi_j(I)\right\|_2^2$	Feature-space matching; preserves textures and edges
SSIM	$1 - \mathrm{SSIM}(\hat{I}, I)$	Structure-preserving; combines luminance, contrast, and structure
Adversarial	$-\log D(G(T))$	Forces photorealistic high-frequency detail during GAN training
Total Variation	$\sum_{i,j} \sqrt{(\nabla_x \hat{I})^2 + (\nabla_y \hat{I})^2}$	Spatial smoothness; suppresses noise artifacts
Frequency	$\|\mathcal{F}(\hat{I}) - \mathcal{F}(I)\|_2^2$	Frequency fidelity; prevents spectral distortion
Physics	$\|T - h * \hat{I}\|_2^2$	Consistency with the forward model
Diffusion (DDPM)	$\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}}\, I_0 + \sqrt{1-\bar{\alpha}}\, \epsilon, T, t\right)\right\|_2^2$	Noise-prediction objective for diffusion training
Domain-adaptation objectives (synthetic-to-real transfer)		
MMD (domain)	$d^2_{\mathrm{MMD}}(\mathcal{S}, \mathcal{R})$	Aligns synthetic and real feature distributions by kernel moment matching; requires no additional trainable network
Domain classification (adversarial)	$\mathcal{L}_d = -\sum_i \left[d_i \log \hat{d}_i + (1 - d_i) \log(1 - \hat{d}_i)\right]$	A domain discriminator predicts whether a feature came from synthetic ($d = 0$) or real ($d = 1$) data; the basis of EAF-WGAN-style adaptation [6]
Gradient reversal (GRL)	$\min_{\theta_f, \theta_y} \left[\mathcal{L}_y - \lambda \mathcal{L}_d\right]$; $\mathcal{R}(\mathbf{x}) = \mathbf{x}$ on the forward pass and $\partial \mathcal{R}/\partial \mathbf{x} = -\lambda \mathbf{I}$ on the backward pass	Makes one feature extractor simultaneously restoration-useful and domain-*indiscriminate* by negating the domain-loss gradient [66]; a single-stage alternative to alternating adversarial optimization, with λ typically ramped up during training
Wasserstein (domain critic)	$\mathcal{L}_W = \mathbb{E}_{\mathcal{S}}[D(f)] - \mathbb{E}_{\mathcal{R}}[D(f)]$	Earth-mover form of the domain critic; stabilizes adaptation where the Jensen–Shannon form saturates, as adopted by EAF-WGAN [6]

A distinct group of objectives deserves separate comment because it addresses the synthetic-to-real domain gap directly at the level of the training signal and because its mechanisms are easily conflated. Maximum mean discrepancy (MMD) penalizes the distance between synthetic and real feature distributions using a kernel moment statistic and requires no additional trainable component. Domain-adversarial objectives instead introduce a domain classifier that tries to predict whether a feature originated from synthetic or real data and train the feature extractor to defeat it. The gradient reversal layer (GRL) is the mechanism that makes this practical in a single training stage [66]: it is an identity map on the forward pass and multiplies the gradient by $-\lambda$ on the backward pass, so one gradient descent step simultaneously minimizes restoration loss and *maximizes* domain-classification loss with respect to the shared features. The result is a representation that is useful for restoration while carrying no information about which domain it came from, which is the property that transfers a simulator-trained restorer to real optical paths. Wasserstein formulations of the same idea replace the classifier's cross-entropy with an Earth-mover critic, which is the route EAF-WGAN takes to stabilize adaptation between simulated and real Australian outdoor turbulence data [6]. Readers examining domain-adaptive turbulence methods should note the practical distinction: MMD aligns distributions by matching statistics, whereas GRL and Wasserstein critics align them by adversarial games and therefore inherit adversarial training's sensitivity to the balance between the two players.

In practice, state-of-the-art systems combine several of these terms, typically as a weighted sum of pixel, perceptual, and adversarial losses augmented by one or more domain-specific regularizer. The optimal combination depends on the available paired data, the severity of the turbulence regime, and whether the architecture follows a deterministic regression or generative sampling paradigm.

13.2. Evaluation Metrics

Quantitative evaluation of atmospheric turbulence mitigation methods requires metrics that reflect both pixel-level reconstruction accuracy and perceptual image quality. A persistent challenge in this domain is the limited availability of paired ground-truth imagery, as real-world turbulence sequences rarely have clean reference images. When ground-truth images are available, either through controlled acquisition or synthetic turbulence generation, full-reference metrics can be applied. The two most widely adopted reference-based metrics in the turbulence mitigation literature are the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

PSNR measures pixel-level reconstruction fidelity by comparing the restored image (f_{rest}) against the ground-truth image (f_{gt}) in terms of squared error. It is defined as [67]

$$\mathrm{PSNR}(f_{gt}, f_{rest}) = 10 \log_{10} \left(\frac{MAX_I^2}{\frac{1}{N} \| f_{gt} - f_{rest} \|_2^2} \right), \tag{22}$$

where MAX_I is the maximum possible pixel intensity (255 for 8-bit encoded images) and N is the total number of pixels. Higher PSNR values indicate greater fidelity to the ground truth.

SSIM provides a perceptually motivated complement to PSNR by jointly evaluating luminance, contrast, and structural consistency between the restored and reference images. It is defined as [67]

$$\mathrm{SSIM}(f_{gt}, f_{rest}) = \frac{(2\mu_{gt}\mu_{rest} + c_1)(2\sigma_{gt,rest} + c_2)}{(\mu_{gt}^2 + \mu_{rest}^2 + c_1)(\sigma_{gt}^2 + \sigma_{rest}^2 + c_2)} \tag{23}$$

where μ_{gt} and μ_{rest} denote the mean intensities of f_{gt} and f_{rest} respectively; σ_{gt}^2, σ_{rest}^2, and $\sigma_{gt,rest}$ are their respective variances and cross-covariance; and c_1 and c_2 are stabilizing constants derived from the dynamic image range. SSIM returns values in the range of $[0, 1]$, where a value of 1 denotes perfect structural agreement with the reference [67].

A caution applies to structural metrics in this literature: several turbulence benchmarks, including the ATSyn protocol used by DATUM [28], report the complex-wavelet variant, i.e., CW-SSIM, rather than the conventional SSIM. CW-SSIM is computed in a complex-wavelet domain and is deliberately less sensitive to small geometric distortions, so its values are not numerically interchangeable with those of conventional SSIM and must not be pooled in a single column or averaged across studies. Throughout this review, the comparison tables identify the exact structural metric reported for each entry.

14. Downstream Task Integration and Operational Deployment

In the great majority of operational settings, turbulence restoration is not an end in itself but a preprocessing stage whose output is consumed by a downstream computer-vision task such as object detection, recognition, or tracking. This distinction matters because the metrics that dominate the restoration literature—PSNR; SSIM; and, increasingly, LPIPS—measure reconstruction fidelity rather than task utility, and the two are not guaranteed to move together. A method that improves pixel-level fidelity may leave downstream accuracy unchanged or, in the case of strongly generative restorers, may even degrade it by introducing plausible but incorrect structure. Viewed from a cognitive-computing standpoint, this is the point where turbulence mitigation stops being an image-enhancement problem and becomes a visual-inference problem: the system observes a scene through a stochastic physical channel, the clean image is only partially identifiable from the observation, and what ultimately matters is whether the reconstruction supports reliable

interpretation and decisions downstream, with restoration quality, task accuracy, uncertainty, and latency evaluated jointly rather than in isolation. This section organizes the surveyed evidence around the question of how restoration integrates with downstream pipelines, then identifies the systematic challenges that remain open.

14.1. Integration Paradigms

Two broad design patterns recur across the literature. The first, which we term *restore-then-task*, treats restoration and the downstream model as decoupled stages: a turbulence-mitigation network is trained independently against clean references, and its output is fed to an off-the-shelf detector, recognizer, or tracker. The appeal of this pattern is modularity; the restorer can be swapped or upgraded without retraining the downstream model, but its weakness is that the restoration objective is agnostic to the downstream task, so fidelity gains need not translate into task gains.

The second pattern, which we term *task-aware restoration*, couples the two stages, either by guiding restoration with a task-derived signal or by optimizing restoration and the downstream objective jointly. Landmark and semantically guided restoration, in which a localization or recognition cue steers the network toward task-salient regions, exemplifies this pattern [38]. Task-aware restoration can yield better downstream accuracy at equal or lower pixel fidelity, at the cost of reduced modularity and a dependence on task-specific supervision that may not transfer across applications.

14.2. Task-Level Evidence Across the Literature

Several of the works surveyed above evaluate restoration directly through a downstream task rather than or in addition to pixel metrics, and together, they sketch the current state of task-coupled evaluation.

Object detection. The MAMAT framework explicitly reports object detection performance as a measure of restoration quality, positioning detection accuracy on restored imagery as a primary evaluation axis rather than a secondary check [18].

Biometric recognition. The DATUM network is validated on the BRIAR corpus, which targets face and whole-body recognition of subjects imaged at extended range under real turbulence [28]. Because BRIAR access is restricted (Section 12.10), independent replication of these recognition results is infeasible, which, itself, is a constraint on how task-level claims in this subdomain can be verified.

Scene text and OCR. The Turbulence Text Dataset supplies a task-driven protocol that is independent of pixel statistics: OCR detection rate and longest common subsequence against the output of a Tesseract engine [15]. This protocol is significant precisely because it provides a functionally grounded quality measure that can diverge from PSNR/SSIM, and it constitutes the clearest existing evidence in the surveyed set that pixel fidelity and task utility are not interchangeable.

General downstream visual tasks. Adversarial restoration aimed at improving performance on downstream visual tasks has also been reported, with the trade-off that gains on task-oriented restoration do not necessarily extend to pure deblurring quality [39].

14.3. Real-World Deployment Evidence

Evidence from real-world deployment is thin, and the thinness is a property of the literature rather than of the search. We examined the reviewed corpus specifically for such evidence and found no study that reports a sustained operational deployment with field-measured reliability over time. What exists is a small set of studies whose evaluation conditions approach operational realism along one axis each, and their limitations are as informative as their results.

CPU-only field deployment. The closest approach to an operational report is the CPU-deployable multi-frame fusion pipeline evaluated on field-collected RGB and mid-wave infrared sequences [7]. Its significance is that it inverts the field's usual priorities: it accepts reduced restoration capability in exchange for real-time operation without a GPU, which is the trade-off a fielded surveillance system typically must make. It also reports the failure mode such a system encounters namely, reduced reliability in low-texture scenes. Notably, it is evaluated entirely with no-reference metrics (BRISQUE, NIQE, and PIQE) because field capture admits no clean reference. This is the characteristic evidential situation of genuine deployment and the reason deployment evidence and quantitative rigor are in tension throughout this literature.

Long-range stand-off capture. The RLR-AT collection spans camera-to-target distances of 1 to 13 km [32], which covers the stand-off geometry of ground-based surveillance and reconnaissance. It supplies operationally realistic inputs but no clean ground truth, so it supports qualitative and no-reference assessment only.

Biometric recognition at range. The BRIAR corpus supports face and whole-body recognition of consenting subjects at extended range under real turbulence, and it is the validation corpus for DATUM [28]. It is the strongest task-level evidence in the corpus and simultaneously the least verifiable because restricted access (Section 12.10) makes independent replication infeasible for most groups.

Domain-specific operational imaging. Satellite video turbulence mitigation evaluated across five platforms and 18 scenes [59] and aero-optical restoration with tracking-efficiency reporting [36] both evaluate in application-specific settings closer to fielded use than generic benchmarks, though neither reports longitudinal operational performance.

Three gaps follow, and we state them as a deployment-evidence agenda. First, no reviewed study reports performance drift over a sustained deployment period, across changing seasons, weather, and diurnal thermal cycles, even though C_n^2 varies by orders of magnitude over exactly those cycles; benchmark performance in one atmospheric state predicts little about a system that must run continuously. Second, no study reports failure statistics from operational use, that is, how often the restorer produced output that misled rather than assisted an operator or a downstream model, which is the quantity an acquiring organization most needs. Third, and connecting to Section 11, no deployment-adjacent study reports the calibrated confidence that would let an operational system act differently when restoration is unreliable. These are not architectural problems, and they will not be solved by a better backbone. They require field programs with instrumented, longitudinal reporting, and their absence is a more serious constraint on operational adoption than any remaining gap in restoration quality.

14.4. Open Challenges in Task-Coupled Restoration

14.4.1. Metric Mismatch

The most consistent gap is the weak coupling between reconstruction metrics and task performance. Pixel and perceptual fidelity metrics reward outputs that resemble a clean reference, but downstream detectors and recognizers are sensitive to different cues, and a restorer optimized for SSIM may not maximize detection mAP or the recognition rate. The OCR protocol of the Turbulence Text Dataset [15] is the most direct demonstration of this divergence and argues for task-driven evaluation as a standard complement to pixel metrics.

14.4.2. Hallucinated Detail and Recognition Hazard

Generative restoration methods, particularly GAN- and diffusion-based approaches,can produce high-frequency detail through learned image priors. This creates a risk that

visually plausible structure may not be fully supported by the observation. For recognition tasks, especially face recognition, such hallucinated detail is an active hazard: it can raise perceptual scores while corrupting identity-bearing features. This tension between perceptual realism and task-faithful reconstruction is underexplored in current evaluation practice.

14.4.3. Restore-or-Not and Degraded-Domain Training

A largely open question is whether explicit restoration is the right intervention at all, as opposed to training the downstream model directly on turbulence-degraded data so that it learns turbulence-invariant features. The surveyed literature is dominated by the restore-then-task assumption, and a systematic comparison against degraded domain training is absent.

14.4.4. Latency and Real-Time Constraints

Inserting restoration as a preprocessing stage compounds end-to-end latency, which is acute for the real-time surveillance, reconnaissance, and navigation scenarios that motivate the field (Section 1). High-capacity Transformers and diffusion models are particularly affected; the CPU-deployable, GPU-free pathway reported in [7] illustrates the opposite end of this trade-off, achieving real-time operation at the cost of reduced reliability in low-texture scenes.

14.4.5. Robustness on Non-Turbulent Inputs

An operational pipeline cannot assume every frame requires correction. A restorer that introduces smoothing or color artifacts on clean inputs harms downstream accuracy on exactly the frames that needed no intervention. The non-turbulent categories of the CDnet 2014 benchmark provide a means of probing this failure mode [62], yet such robustness checks are rarely reported.

14.5. Summary

Taken together, the surveyed evidence shows that downstream integration is recognized as important but remains evaluated unevenly: a handful of works report task-level result detection [18], recognition [28], and OCR [15], while the majority rely on pixel and perceptual metrics alone. The principal open problems are the mismatch between fidelity and task metrics, the recognition hazard posed by hallucinated detail, the unresolved restore-or-not question, latency under real-time constraints, and robustness on inputs that do not require correction. Closing these gaps will require task-driven evaluation protocols to become a standard component of turbulence restoration benchmarking rather than an occasional addition.

14.6. Two Classes of Quantitative Evidence

A single cross-architecture results table is the conventional centerpiece of a review such as this one, and in this field, it is actively misleading. The reason is worth stating before the numbers are presented. Almost every turbulence restoration paper evaluates on a protocol of its own choosing, and the resulting scores are therefore measurements of different quantities that happen to share a unit. A 30.99 dB result on a mild custom corpus and a 27.73 dB result on a strong shared benchmark do not stand in the relation that their numerical order suggests. Placing them in adjacent rows of one table invites exactly the comparison that the underlying evidence forbids, and annotating every cell with a caveat does not repair this because readers compare rows before they read footnotes.

We therefore separate the evidence into two tables that are read in different ways and support different kinds of claims.

Table 3 **is the benchmark anchor.** It contains only results obtained under the two protocols that multiple independent studies have actually shared: the TMT synthetic static-scene benchmark and the ATSyn-dynamic benchmark. Within each panel, every method was evaluated on the same data with the same metrics by the same evaluating study, so rows may be compared directly. This is the entirety of the field's directly comparable quantitative evidence, and it is small.

Table 4 **collects the self-reported results.** Each row is a valid characterization of one method on the evaluation its own authors selected, and no comparison across rows is licensed. The datasets, turbulence severities, scene domains, image sizes, metrics, and hardware differ between every pair of rows, so a numerical difference between two rows carries no information about relative methodological quality. The table therefore documents what each study reports and what it omits, rather than ranking methods. Cells marked N.R. indicate metrics the cited source does not report; none is estimated here.

14.7. Reading the Shared-Protocol Evidence

Table 3 supports one robust conclusion and, we argue, only one.

What it establishes. Under both shared protocols, multi-frame configurations outperform single-frame configurations, and the margin is substantial. On the TMT synthetic static-scene benchmark, PSNR rises from 22.76 dB for the single-frame TurbNet baseline to 25.29 dB for a 15-frame adversarial configuration and 27.73 dB for a 12-frame Transformer configuration. Panel B reproduces the ordering of the two multi-frame systems on independent data and adds a recurrent configuration above both. Taken together, this is genuine, independently replicated evidence, and it is the strongest quantitative statement this literature currently supports.

What it does not establish. The comparison is between *entire systems*, not between architectural components, and this distinction is not a formality. Any two rows in Panel A differ simultaneously in the number of input frames, the presence and quality of a tilt-alignment stage, the backbone architecture, the loss composition, the training corpus and schedule, and the model capacity. At least six factors vary at once across a single comparison. Consequently, the 5 dB gap between TurbNet and TMT cannot be attributed to self-attention, and the ordering of TSRWGAN below TMT cannot be attributed to adversarial training being inferior to attention. The intrinsic superiority of any individual architectural component simply cannot be disentangled from these data.

This is a stronger and more defensible claim than the familiar one that "Transformers outperform GANs, which outperform CNNs", and it is also the claim the evidence actually licenses. Panel B makes the point concrete in a way that should give component-level narratives pause: DATUM, a *recurrent* architecture of the kind the field had largely moved past, exceeds the Transformer configuration by roughly 1.2 dB at about eleven times the throughput. Had architecture family been the operative variable, that result would be difficult to explain. It is straightforward to explain once the comparison is understood as one between complete, separately engineered systems.

Resolving component-level questions requires controlled ablation: fixing the frame count, alignment, loss, capacity, and training data and varying only the backbone. No study in the reviewed corpus performs this experiment across architectural families, and we regard it as the single most valuable missing experiment in the field. Until it is done, claims of the form of "architecture X is better suited to turbulence than architecture Y" should be read as claims about particular published systems, not about architectures.

15. Discussion

15.1. From Deterministic to Stochastic, Physics-Aware Restoration

The architectural progression documented across the preceding sections reflects an evolving understanding of turbulence restoration as a one-to-many inverse problem rather than a deterministic mapping. Early CNN-based approaches treat restoration as a pixel-level regression task: given a degraded input, minimize the expected reconstruction error over a training distribution. This formulation is computationally efficient and produces stable results when turbulence is mild, but it systematically suppresses high-frequency detail by averaging over the distribution of plausible restorations, biasing outputs toward the conditional mean of the training data. GAN-based methods partially address this limitation by replacing the regression objective with an adversarial signal that rewards perceptual realism, recovering texture and edge sharpness that regression objectives smooth away. The cost is training instability and sensitivity to distribution shift, particularly on outdoor scenes outside the training domain.

Transformer architectures advance the approach by modeling long-range spatial and temporal dependencies through self-attention, capturing the non-stationary blur and tilt correlations that span the isoplanatic patch and cannot be reached by practical dilation rates in convolutional networks. Multi-frame Transformers such as TMT [16] reduce tilt-induced pixel displacement across frames before applying attention, exploiting the turbulence physics explicitly as a preprocessing step. Diffusion-based systems go further still: by sampling from the posterior distribution over clean images, given the degraded observation, each stochastic sampling run can produce a distinct plausible restoration (deterministic samplers trade this diversity for repeatability), formally acknowledging the fundamental ambiguity of the underdetermined inverse problem. Physics-informed frameworks constrain this posterior with the known turbulence forward model, reducing the risk of hallucinated structure and improving generalization to turbulence strengths not seen during training. The hybrid SSM–diffusion design of TSTRNet [9] illustrates where this trajectory currently stands: a Mamba encoder provides the global context and linear-complexity efficiency needed for coarse restoration, while a DDPM refiner handles the stochastic fine-detail recovery that deterministic networks cannot reliably produce. This trajectory should be read as an emerging direction rather than an accomplished displacement: deterministic CNN, recurrent, and Transformer frameworks remain the most extensively evaluated approaches in the reviewed corpus, and the stochastic and physics-aware systems have yet to accumulate a comparable evidence base.

The practical cost of this progression is visible in Tables 3 and 4: lightweight state-space stages execute in tens of milliseconds per image (0.047 s for TSTRNet's Mamba U-Net alone), multi-frame GAN and Transformer systems require approximately 1–2 s per frame (1.15 s for TSRWGAN and 1.52 s for TMT under the same measurement setup), and diffusion-based refinement stages require 4 or more seconds on high-end GPUs. The recurrent DATUM shows that this cost curve is not fixed, sustaining 9.17 frames per second at 960×540 with 5.754 M parameters while exceeding the evaluated TMT configuration's restoration quality on the same benchmark [28]. Two qualifications bound this picture: except within Panel B of Table 3, the reported runtimes were measured with different hardware, resolutions, and timing boundaries, so they characterize configurations rather than establishing a controlled ranking, and because the compared systems differ simultaneously in architecture, input frames, losses, and capacity, none of these deltas isolates the contribution of a single design component. Within those limits, architecture choice is a deployment decision as much as a quality decision, a trade-off made explicit in the selection framework of Section 15.5.

15.2. Evolution of the Field, 2020–2026

The preceding sections also show that atmospheric turbulence mitigation has evolved from classical physics-based correction toward data-driven and hybrid physics-aware learning systems. Tables 5 and 6 span the 2020–2026 period of atmospheric turbulence restoration research and show a clear shift from handcrafted physics-based methods toward increasingly data-driven, diverse, and multi-objective deep learning frameworks. During the 2020–2022 period, classical approaches such as space-time non-local averaging [34] and blind PSF estimation via central moments [12] established principled restoration baselines grounded in optical turbulence theory. However, these methods struggled with real-time implementation and with the highly non-stationary spatial characteristics of natural turbulence. At the same time, deep learning methods began to emerge, including semi-supervised landmark-guided CNNs [38] and adversarial multi-frame deconvolution networks [14]. These approaches achieved better perceptual quality on benchmark datasets but often depended on face priors or scene-specific training data, limiting their transferability to more general outdoor environments.

From 2023 onward, the literature shows a wider range of architectural developments. Transformer-based systems, particularly TMT [16], used global self-attention across temporal frame sequences to address spatially correlated blur and tilt. Recurrent architectures such as DATUM [28] introduced a dedicated recurrent framework for turbulence restoration, highlighting the importance of temporal information across frames as an underutilized source of restoration signal. At the same time, generative modeling methods progressed from conventional GANs to physics-guided stochastic refinement [25] and hybrid state-space and conditional diffusion architectures [9]. TSTRNet [9] exemplifies this trend by composing a Mamba-based U-Net ($\mathcal{N}^G$, exploiting the linear-complexity 2D selective scan for coarse global restoration) with a patch-based conditional DDPM ($\mathcal{N}^R$) guided by Sobel edge maps to correct residual geometric distortions through stochastic refinement. This two-stage design achieves a PSNR/SSIM of 27.71 dB/0.819 on synthetic face turbulence and 25.17 dB/0.821 on the real Heat Chamber dataset, outperforming standalone CNN, GAN, Transformer, and diffusion baselines on both reference-based and perceptual metrics (MUSIQ and NIQE). However, the $\mathcal{N}^R$ stage runs at 4.12 s per image on an RTX 3090, reinforcing the quality–latency tension observed across the reviewed literature. In addition, the increasing use of perceptual and no-reference metrics such as LPIPS, FID, BRISQUE, NIQE, and MUSIQ alongside PSNR and SSIM reflects a growing recognition that pixel-level accuracy alone is insufficient to assess image quality for downstream tasks including surveillance, recognition, and tracking.

15.3. Cross-Cutting Limitations

Three limitations cut across every architectural family in Tables 5 and 6 and therefore qualify all of the comparative statements above. First is the synthetic-to-real generalization gap: methods trained on simulator-generated degradations experience performance drops on real field data, and domain adaptation strategies [6,19,41] narrow but do not close this gap. Second is evaluation fragmentation: the scarcity of paired real ground truth pushes many studies toward no-reference metrics and custom test sets, which is why the directly comparable evidence of Table 3 is so much smaller than the self-reported body in Table 4. Third is the quality–cost tension: the configurations with the strongest reported quality are generally the most expensive at inference, while the only CPU-deployable pipeline in the corpus trades reliability in low-texture scenes for its real-time operation [7]. These three constraints are diagnostic findings of the synthesis; the corresponding research agenda, including what evidence would resolve each constraint, is developed in Section 16 rather than repeated here.

15.4. Answers to the Research Questions

The evidence assembled across this review supports direct answers to the five research questions posed in Section 2.2.

RQ1 (physical mechanisms and optical effects). Atmospheric turbulence originates in random refractive-index fluctuations whose spatial statistics are described by Kolmogorov theory and its non-Kolmogorov generalizations (Section 3). At the image plane, these fluctuations manifest as spatially varying anisoplanatic blur, geometric tilt and warping, scintillation, and temporal instability. The governing physical quantities, the structure constant (C_n^2), the Fried coherence length (r_0), and the isoplanatic angle (θ_0) are not merely background theory: D/r_0 sets the effective degradation severity that datasets must span, and θ_0 bounds the patch size over which a single point-spread function is valid, which directly motivates the patch-based and attention-based processing found in modern architectures.

RQ2 (classical and deep learning-based approaches). Classical mitigation, including adaptive optics, lucky imaging, multi-frame registration and fusion, and blind deconvolution (Section 4), remains foundational but is constrained by specialized hardware, multi-frame requirements, or stationarity assumptions. Learning-based restoration now spans six architectural paradigms (Sections 5–10): CNNs and residual regression, adversarial (GAN) restoration, Transformer-based multi-frame attention, diffusion-based posterior sampling, Mamba/state-space models with linear-complexity global context, and physics-informed hybrids. No single paradigm dominates; each occupies a different point on the fidelity–perceptual quality–latency–generalization trade-off surface.

RQ3 (datasets and simulation methods). The ten dataset resources surveyed in Section 12 divide into laboratory-controlled paired capture (Heat Chamber), outdoor sequences with partial or no ground truth (Turbulence Text, OTIS, CLEAR, and RLR-AT), physics-based synthetic generation (the TMT/P2S simulation pipeline and Zernike- and phase-screen-based simulators), and domain-specific collections (SUTO, SILO, BRIAR, and CDnet). The critical structural weakness is the near absence of large, paired, real datasets, which forces training onto simulators whose simplifying assumptions are a root cause of the synthetic-to-real gap discussed in Section 15.3.

RQ4 (recent innovations). Tables 5 and 6 trace the 2020–2026 trajectory: from classical space-time averaging and blind PSF estimation, through landmark- and uncertainty-guided CNNs and adversarial multi-frame deconvolution to temporal Transformers (TMT), dedicated recurrent frameworks (DATUM), physics-guided stochastic refinement, hybrid state-space and diffusion designs (TSTRNet), and CPU-only real-time fusion pipelines. Two quieter innovations accompany the architectural ones: the deliberate reuse of turbulence physics as an inductive bias rather than as a full forward model and the shift toward perceptual and no-reference metrics (LPIPS, FID, BRISQUE, NIQE, and MUSIQ) as the field acknowledges that pixel fidelity alone does not predict downstream utility.

RQ5 (comparative performance). Table 3 consolidates the directly comparable evidence, which two shared protocols anchor, and Table 4 records the remaining self-reported results that no cross-row comparison may use (Section 14.6). On the TMT synthetic static-scene benchmark, PSNR rises from 22.76 dB for the single-frame TurbNet baseline to 25.29 dB for the evaluated 15-frame TSRWGAN configuration and 27.73 dB for the 12-frame TMT configuration. These shared-protocol results establish that the multi-frame configurations outperform the single-frame baseline, but because the systems differ simultaneously in temporal input, alignment, architecture, loss design, training strategy, and model capacity, the observed gains cannot be attributed uniquely to temporal aggregation, self-attention, or any other single component; isolating those contributions would require controlled ablations that the current literature does not provide. On ATSyn-dynamic, the

recurrent DATUM configuration exceeds the evaluated TMT configuration by about 1.2 dB (29.43 versus 28.27 dB PSNR) at approximately 11× the throughput (9.17 versus 0.80 frames per second at 960 × 540, both measured in the same study on the same hardware [28]), showing that restoration quality and efficiency are not inherently opposed. Reported runtimes across both tables span roughly two orders of magnitude, from 0.047 s per image for a lightweight state-space stage to 4.12 s for diffusion refinement, although these figures come from different hardware and timing setups and are indicative only. Beyond such shared-protocol pairs, the honest answer to RQ5 is that a global ranking is not currently possible: most methods are evaluated on incompatible datasets, turbulence strengths, and metrics. This evaluation fragmentation is, itself, a principal finding of the review, and it motivates the benchmark standardization called for in Section 16.

15.5. Method Selection Under Operational Constraints

The synthesis above supports guidance that is conditional rather than absolute: no family is universally best, and the evidence reviewed here identifies which families are *defensible* under a given operational constraint, what safeguards their known failure modes require, and what evaluation evidence a practitioner should demand before deployment. Table 10 condenses this into a decision-oriented reference. Two caveats govern its use. The family recommendations reflect the configurations actually evaluated in the reviewed studies, not intrinsic architectural superiority (Section 15.4), and the safeguards column is integral to each recommendation: a generative restorer without hallucination controls, for example, is not an acceptable substitute for a deterministic one in evidential use.

Table 10. Decision-oriented mapping from operational constraints to methodological families, based on the evidence reviewed in Sections 5–14.

Operational Scenario	Suitable Families (Evidence Herein)	Required Safeguards	Principal Limitation	Recommended Evaluation
Single-frame input, low latency	Lightweight CNN/residual regression (Section 5)	Restrict use to severity range covered in training	Oversmoothing; weak recovery under severe turbulence	PSNR/SSIM plus latency on the target hardware
Multi-frame video, strong geometric distortion	Recurrent (DATUM-style); multi-frame Transformer with tilt alignment (Sections 7 and 9)	Explicit alignment/registration stage; sufficient frame budget	Latency grows with frame count; sensitivity to alignment quality	Full-reference metrics plus temporal-consistency assessment
No paired ground truth	Self-/unsupervised adversarial (TurbuGAN-style); forward-model consistency objectives (Sections 6 and 10)	Physics-consistency constraints to bound hallucination	Quantitative validation inherently limited	No-reference metrics plus task proxies; avoid direct comparison with supervised scores
Severe synthetic-to-real mismatch	Domain adaptation (teacher–student); semi-supervised hybrids; physics conditioning [6,19,41]	Validate on real sequences, not simulator held-out sets	Residual domain gap persists in all reviewed studies	Real-sequence evaluation (OTIS/CLEAR/RLR-AT) with no-reference metrics
High perceptual-quality requirement	GAN; diffusion refinement stages (Sections 6–8)	Hallucination controls (edge/physics conditioning); label output as generative	Perceptual-fidelity trade-off; stochastic output variability	LPIPS/FID over multiple samples; human review for critical content
Evidential or safety-critical restoration	Deterministic regression (CNN/recurrent/Transformer) with uncertainty estimation; avoid unconstrained generative detail	Confidence maps; provenance audit of recovered detail	Hallucination risk unacceptable; calibration rarely reported in current literature	Calibration and task-reliability evaluation, not perceptual scores alone
Downstream recognition or tracking	Task-aware restoration; restore-then-task with task-level validation (Section 14)	Verify robustness on non-turbulent inputs	Fidelity gains need not transfer to task gains	Task metrics (mAP, recognition rate, and OCR) alongside fidelity metrics
Edge or embedded deployment	Lightweight CNN; SSM coarse stages; CPU multiframe fusion [7,9]	Quantization/distillation; measurement on the target device	Reduced reliability in low-texture scenes (CPU pipeline); limited severe-turbulence evidence	On-device FPS, memory, and quality at the target severity

Table 10. *Cont.*

Operational Scenario	Suitable Families (Evidence Herein)	Required Safeguards	Principal Limitation	Recommended Evaluation
Uncertainty-sensitive applications	Diffusion posterior sampling; uncertainty-guided CNN [13]	Multiple posterior samples; calibrated confidence reporting	Calibration methodology immature in this field	Distributional evaluation over repeated samples; calibration curves
High-resolution offline restoration	Multi-frame Transformer; hybrid SSM + diffusion (Sections 7 and 9)	Patch-based processing with seam handling	Memory and runtime at native resolution	Full-reference and perceptual metrics at native resolution

16. Challenges and Future Directions

Despite remarkable progress, atmospheric turbulence restoration remains an open research problem. The answers to RQ3 and RQ5 (Section 15.4) point to the underlying pattern: architectural innovation remains important, but progress is increasingly constrained by real-data scarcity, simulator mismatch, fragmented evaluation, uncertainty calibration, and deployment cost. The seven challenges below elaborate on that diagnosis, and each closes by stating what evidence or experiment would resolve it. They complement the task-coupled restoration challenges identified in Section 14, which concern how restored imagery is consumed, whereas the items here concern how restoration models are trained, compared, and deployed; the comparative findings themselves are not restated here (see Section 15). The ordering is deliberate. The synthetic-to-real domain gap and uncertainty quantification are presented first and as co-equal priorities because they are the same deficiency approached from two directions and because the second has been persistently treated as a longer-horizon concern in this literature, while the evidence of Section 11 suggests it is both urgent and unusually tractable.

1. **Synthetic-to-Real Domain Gap and Limited Propagation-Model Coverage.** The majority of deep-learning turbulence restoration models are trained on synthetic data generated by simulators such as the P2S framework [16] or phase-screen methods. These simulators implement specific Kolmogorov or von Kármán assumptions and commonly use simplified C_n^2 profiles, aperture geometries, and noise models that may not reflect the scintillation, partial coherence, and sensor-specific noise present in real field systems. As a result, performance on simulator holdouts does not by itself establish performance on real outdoor sequences, and several studies introduce real-data adaptation or separate qualitative evaluation for this reason. Domain adaptation strategies such as teacher–student transfer [41], adversarial alignment via domain classification and gradient reversal [6,66], and semi-supervised learning on unlabeled real data [19] have reported improvements under their respective protocols, but the available evidence does not establish general closure of the gap.
   One component of this gap has received limited controlled evaluation and deserves separate emphasis (Section 3.3). Most simulator-based learned methods in the reviewed corpus use classical Kolmogorov statistics or the finite-scale von Kármán refinement, commonly with the spectral exponent fixed at $\alpha = 11/3$; evidence under alternative exponents is limited. Real channels can deviate from this value, particularly on slant paths, at varying altitudes, and in the upper atmosphere [44]. The consequence is a narrow modeling convention whose generalization properties remain uncertain, and the reviewed benchmarks provide little controlled evidence across values of α. Future progress therefore requires larger real paired datasets, more realistic simulation of sensor and atmospheric co-variation, self-supervised consistency objectives applicable without clean references, and controlled experiments that sweep the spectral exponent across the $3 < \alpha < 4$ range while holding the remaining simulator and training settings fixed.

2. **Uncertainty Quantification and Calibrated Restoration.** We place uncertainty quantification immediately after the domain gap and deliberately ahead of benchmarking and efficiency because the two leading challenges are the same problem viewed from opposite ends. The domain gap is the statement that models are unreliable outside their training distribution; uncertainty quantification is the machinery that would let a system *know* when it is in that situation. A field that invests heavily in the first while neglecting the second is building restorers that fail silently, and Section 11 documents that this is the present state of affairs: aleatoric ambiguity is almost never represented in an output, epistemic uncertainty is essentially never measured, and no reviewed study calibrates any uncertainty estimate it produces.
The two sources require different instruments, and conflating them is the most common conceptual error in this area. *Aleatoric* uncertainty is irreducible: the random, information-destroying PSF means many clean scenes are consistent with one degraded frame. Representing it requires a model that outputs a distribution rather than a point, whether through a heteroscedastic variance head trained with a Gaussian likelihood [20] or through posterior sampling, which diffusion models already provide for free. Drawing K samples from a conditional diffusion restorer and reporting the per-pixel variance ($\mathrm{Var}[\hat{\mathbf{x}}]$) costs only additional sampling passes and yields a spatially resolved map of exactly where the atmosphere destroyed information. That no reviewed turbulence diffusion study reports such a map is, in our assessment, the largest unexploited opportunity in the field. *Epistemic* uncertainty is reducible and requires disagreement among plausible models: MC dropout, which approximates a Bayesian posterior over weights by sampling stochastic forward passes [53] and which AT-Net has already shown to be feasible for turbulence [21], or deep ensembles of independently initialized models, which are simple, parallelizable, and typically better calibrated than single-model approximations [23]. Only ensembles and related model-diversity methods detect the case that matters most operationally namely, a diffusion or adversarial restorer whose samples agree tightly with each other while the underlying prior is wrong for the observed scene.
Neither estimate is usable until it is calibrated, and calibration is where the field should start rather than finish. A variance map is a claim about error rates, and it is worth nothing until validated against realized error. Distribution-free procedures developed for image restoration provide per-pixel intervals with finite-sample coverage guarantees under exchangeability [24], and importing them into turbulence restoration is tractable today: it requires a held-out calibration set rather than a new architecture. The resolving evidence is correspondingly concrete. Benchmarks should score calibration alongside fidelity, for instance, through the empirical coverage of stated 90% intervals, reliability diagrams of predicted against realized error, and the correlation between predicted uncertainty and downstream task failure.
That last quantity is the point of the entire exercise, and it is what makes uncertainty an operational rather than an academic concern. The value of a confidence map is realized only when a downstream consumer acts on it: a face-recognition pipeline that suppresses matches computed over high-variance identity regions rather than treating hallucinated detail as evidence, a tracker that widens its association gate where restoration is unreliable, a change-detection system that abstains rather than raising a false alarm on invented structure, or an operator display that visually marks which recovered detail is evidence and which is prior. Each is straightforward to build once a calibrated per-pixel uncertainty accompanies the restored image and impossible without one. This is where the cognitive-computing framing of this review becomes operational rather than rhetorical: a system acting on observations received

through a stochastic channel must know how much of its reconstruction is evidence and how much is prior, and that requirement connects turbulence mitigation directly to trustworthy AI and machine perception. We recommend that per-pixel calibrated confidence maps become a standard reported output of turbulence restoration systems, in the same way that runtime and parameter counts are now expected.

3. **Lack of Standardized Benchmarks.** A persistent obstacle to meaningful progress comparisons is the absence of a single shared benchmark with fixed turbulence strength, scene content, and evaluation metrics. As Table 4 illustrates, current methods are evaluated on incompatible protocols: TDRN and AT-DDPM use CelebA-based face turbulence, TMT uses its own synthetic suite, MAMAT uses a custom mild-turbulence video corpus, and TSTRNet introduces a new five-scene dataset. Direct cross-method comparison is therefore impossible without re-evaluation, and reported rankings are highly protocol-dependent. The field needs agreed-upon benchmark datasets covering multiple turbulence levels (D/r_0 ranging from mild to severe), multiple scene categories (face, outdoor scene, and satellite), and both single-frame and multi-frame evaluation modes, alongside mandatory reporting of runtime, parameter count, and inference hardware.
4. **Computational Efficiency and Edge Deployment.** High-performing restoration methods, particularly large Transformer models and diffusion-based systems, incur inference costs that preclude deployment on the edge platforms most relevant to surveillance, reconnaissance, and autonomous navigation. The reported runtimes span 0.047 s per image for a lightweight Mamba U-Net stage to over four seconds per image for diffusion refinement (TSTRNet $\mathcal{N}^R$), a range of roughly two orders of magnitude, although the heterogeneous hardware, resolutions, and timing conventions behind these figures prevent a controlled efficiency ranking. Latency requirements are application-dependent: offline scientific restoration tolerates minutes per frame, whereas live human monitoring, tracking, and navigation are commonly discussed against illustrative targets on the order of 15–30 frames per second, which only lightweight CNN, state-space, and classical methods currently approach; the recurrent DATUM's 9.17 frames per second at 960×540 [28] is the closest a full multi-frame restoration network comes in Table 3. Future research should investigate knowledge distillation from large turbulence-restoration models into compact student networks, quantization-aware training for integer-arithmetic inference, and architectural choices (such as Mamba/SSM's linear complexity) that reduce the scaling cost of attention without sacrificing global context.
5. **Multimodal and Multi-Spectral Turbulence Mitigation.** Atmospheric turbulence affects different spectral bands differently: the longer wavelengths of thermal infrared (MWIR/LWIR) exhibit larger isoplanatic patches ($\theta_0 \propto \lambda^{6/5}$) and lower scintillation, while visible imagery captures finer spatial detail. Combining optical and infrared channels can, in principle, provide complementary information that neither modality alone can supply. Narrow-band guided turbulence removal [17] demonstrates that dual-band imaging can improve restoration under turbulence, but broader multi-spectral and sensor-fusion approaches remain underexplored. Future work should investigate joint restoration across spectral bands, exploiting cross-spectral coherence as a physical prior that reduces the ill-posedness of the single-frame inverse problem.
6. **Foundation Models and Zero-Shot Generalization.** Large-scale pretrained vision foundation models such as diffusion-based image priors and vision–language models represent an underutilized resource for turbulence restoration. A foundation model trained on billions of natural images encodes extremely rich priors over image structure, and several recent works in blind image restoration have demonstrated that

this prior can be leveraged for zero-shot or few-shot correction of degradations not seen during pretraining. For turbulence restoration, a foundation-model approach could enable generalization to turbulence regimes, scene types, and sensor configurations without retraining, which is a critical requirement for operational systems that must function under diverse and unpredictable atmospheric conditions. The main open questions are how to condition a general foundation model on turbulence-specific physics without full retraining and how to ensure that restoration outputs are physically admissible rather than photorealistic but physically incorrect.

7. **Explainability and Physics Interpretability.** Turbulence restoration is applied in safety-critical and high-stakes contexts (surveillance, reconnaissance, and scientific observation) where the failure modes of the restoration model must be understood and auditable. Current deep-learning systems offer limited interpretability: it is not generally possible to determine from the output whether a restored feature (an edge, a texture element, or a face contour) was recovered from the degraded input or hallucinated by the generative prior. Physics-informed architectures provide partial transparency by tying network outputs to interpretable physical quantities such as r_0 or the isoplanatic angle, but this transparency is limited to the physical parameterization rather than the full restoration pathway. Turbulence-aware attention mechanisms that highlight which image regions most influenced the restoration decision, combined with visualization of the inferred turbulence map, would substantially improve operator trust and enable targeted quality control in operational pipelines.

These challenges are not independent, and their coupling suggests an ordering for the research agenda. Benchmark standardization (Challenge 3) is the enabling step: without shared protocols spanning multiple D/r_0 levels, multiple spectral exponents, and multiple scene categories, progress on the synthetic-to-real gap (Challenge 1) cannot even be measured consistently, calibration claims (Challenge 2) cannot be scored, and cross-architecture efficiency claims (Challenge 4) remain anecdotal, as the separation between Tables 3 and 4 makes plain. The domain gap and uncertainty quantification are the twin near-term priorities, and we emphasize, again, that they are two views of a single deficiency: the first asks how to make models work outside their training distribution, and the second asks how a model can recognize that it is outside it. Neither is complete without the other, and of the two, uncertainty quantification is the more tractable today because the required machinery (posterior sampling, ensembles, MC dropout, and distribution-free calibration) already exists in the broader machine-learning literature and awaits transfer rather than invention. Multimodal sensing, foundation-model priors, and physics interpretability (Challenges 5–7) form the longer-horizon agenda. Uncertainty quantification and interpretability, together, are prerequisites for adoption in the safety-critical applications that motivate much of this field: a restoration system whose confidence cannot be calibrated and whose recovered detail cannot be distinguished from hallucination will not be trusted in surveillance, forensic, or scientific pipelines, regardless of its benchmark scores. Physics-informed learning cuts across the entire agenda, serving simultaneously as a regularizer against hallucination, a bridge across the domain gap, and a source of interpretable intermediate quantities.

17. Conclusions

Atmospheric turbulence remains one of the most difficult challenges in long-range optical imaging, and the underlying inverse problem is ill-posed, stochastic, and scene-dependent. This review followed a structured selection protocol (Section 2) to synthesize 45 core studies, tracing the field's architectural progression from CNNs and GANs to Transformers, diffusion models, Mamba/state-space models, and physics-informed frameworks and answering each of its five research questions explicitly in Section 15.4. Its central

contribution is a configuration-level, evidence-calibrated map of what current methods have and have not demonstrated.

Four findings carry the most weight. First, the field is moving beyond deterministic pixel regression toward stochastic, uncertainty-aware, and physics-informed restoration. Deterministic CNN, recurrent, and Transformer frameworks nonetheless remain the most extensively evaluated approaches, and no paradigm is universally dominant. Second and more sharply than is usually acknowledged, the comparative evidence base is far smaller than the literature's volume suggests. Only two shared protocols support direct comparison at all (Table 3), and what they establish is that multi-frame configurations outperform single-frame ones *as complete systems*. Because the compared systems differ simultaneously in frame count, alignment, architecture, loss design, training, and capacity, the intrinsic superiority of any individual architectural component cannot be disentangled from these data, and the controlled ablation that would disentangle it has not been performed by any study we reviewed.

Third, the field has organized itself around the one-to-many structure of turbulence restoration while continuing to deliver one-to-one outputs. Aleatoric ambiguity shapes the designs but is almost never represented; epistemic uncertainty, which is precisely what the much-discussed synthetic-to-real domain gap consists of, is essentially never measured; and no reviewed study calibrates any uncertainty estimate it produces (Section 11). We regard this as the field's most consequential and most tractable gap because the necessary machinery already exists and awaits transfer: posterior sampling, deep ensembles, MC dropout, and distribution-free calibration. Fourth, architecture choice has become a deployment decision as much as a quality decision: reported inference costs span roughly two orders of magnitude under non-uniform measurement conditions, and efficient recurrent designs show that quality and throughput are not inherently opposed. On the newest of the six families, we deliberately withhold a favorable verdict: the case for selective state-space models in turbulence restoration rests on an asymptotic efficiency argument that has never been measured against a strong recurrent baseline or on hardware where efficiency binds (Section 9.2).

The requirements for reliable deployment are correspondingly few and concrete: paired real turbulence data with which generalization can actually be measured; shared multi-severity benchmarks, spanning spectral exponents as well as D/r_0 levels, with which methods can actually be compared; and calibrated uncertainty with which restored detail can actually be trusted. Progress on these three fronts would convert turbulence mitigation from a collection of benchmark results into a deployable capability and would consolidate this uncertainty-constrained visual-inference problem as a proving ground for cognitive computing and trustworthy machine perception.

Author Contributions: Conceptualization, N.J., T.S.G. and M.K.; methodology, N.J., T.S.G. and N.A.R.; validation, T.S.G., N.A.R. and A.S.; formal analysis, N.J., N.A.R. and A.S.; investigation, N.J. and N.A.R.; resources, T.S.G. and A.S.; data curation, N.J. and N.A.R.; writing—original draft preparation, N.J.; writing—review and editing, T.S.G., M.K., N.A.R. and A.S.; visualization, N.J. and A.S.; supervision, T.S.G., M.K., N.A.R. and A.S.; project administration, T.S.G. and N.A.R.; funding acquisition, T.S.G. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Asian Office of Aerospace Research and Development (AOARD) through research grant FA2386-25-1-4061, project reference SPI25-295-0295.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This work is fully supported by and also conceptually aligned with ongoing research under AOARD projects FA2386-25-1-4061 and SPI25-295-0295, "Adaptive Physics-Guided Transformer with Explainable Attention for Atmospheric Turbulence Mitigation". Nurul Jannah gratefully acknowledges the Kulliyyah of Engineering, IIUM, Engineering Merit Scholarship 2025 (KOEIEMS25). The authors used generative artificial intelligence tools only to improve grammar, readability, and sentence structure during manuscript preparation. All AI-assisted revisions were critically reviewed, edited, and verified by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AO	Adaptive Optics
CNN	Convolutional Neural Network
CW-SSIM	Complex-Wavelet Structural Similarity Index Measure
DDIM	Denoising Diffusion Implicit Model
DDPM	Denoising Diffusion Probabilistic Model
FID	Fréchet Inception Distance
FLOPs	Floating-Point Operations
GAN	Generative Adversarial Network
GRL	Gradient Reversal Layer
LPIPS	Learned Perceptual Image-Patch Similarity
MAP	Maximum *A Posteriori*
MC	Monte Carlo
MMD	Maximum Mean Discrepancy
MWIR	Mid-Wave Infrared
OCR	Optical Character Recognition
P2S	Phase-to-Space (turbulence simulation)
PINN	Physics-Informed Neural Network
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD	Power Spectral Density
PSF	Point-Spread Function
PSNR	Peak Signal-to-Noise Ratio
RQ	Research Question
SE	Squeeze and Excitation
SSIM	Structural Similarity Index Measure
SSM	State-Space Model
TTB	Tilt Then Blur
UQ	Uncertainty Quantification
VSSM	Vision State-Space Module

Appendix A. Full Database Search Strings

This appendix reproduces the database search strings supporting Section 2, exactly as executed on 11 June 2026 (final update search: 15 July 2026).

Appendix A.1. Scopus

The Scopus search was conducted using the *TITLE-ABS-KEY* field. The publication year was limited to 2019–2026, and the search was restricted to journal articles and conference papers. The final Scopus search string was

> TITLE-ABS-KEY ((“atmospheric turbulence mitigation” OR “atmospheric turbulence removal” OR “atmospheric turbulence correction” OR “turbulence image restoration” OR “turbulence-degraded image restoration” OR “image restoration through turbulence” OR “through-turbulence imaging”) AND (“deep learning” OR “neural network” OR CNN OR “convolutional neural network” OR GAN OR “generative adversarial network” OR transformer OR “vision transformer” OR diffusion OR “diffusion model” OR DDPM OR DDIM OR Mamba OR “state space model” OR “selective state space” OR “physics-informed” OR “physics-guided” OR “physics-based”)) AND PUBYEAR > 2018 AND PUBYEAR < 2027 AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “cp”))

Appendix A.2. IEEE Xplore

The IEEE Xplore search was conducted using the *All Metadata* field. The publication year was limited to 2019–2026, and the search was restricted to journal articles and conference papers. The final IEEE Xplore search string was

> ((“atmospheric turbulence mitigation” OR “atmospheric turbulence removal” OR “atmospheric turbulence correction” OR “turbulence image restoration” OR “turbulence-degraded image restoration” OR “image restoration through turbulence” OR “through-turbulence imaging”) AND (“deep learning” OR “neural network” OR CNN OR “convolutional neural network” OR GAN OR “generative adversarial network” OR transformer OR “vision transformer” OR diffusion OR “diffusion model” OR DDPM OR DDIM OR Mamba OR “state space model” OR “selective state space” OR “physics-informed” OR “physics-guided” OR “physics-based”))

References

1. Abdukirim, A.; Ren, Y.; Tao, Z.; Liu, S.; Li, Y.; Deng, H.; Rao, R. Effects of Atmospheric Coherent Time on Inverse Synthetic Aperture Ladar Imaging through Atmospheric Turbulence. *Remote Sens.* **2023**, *15*, 2883. [CrossRef]
2. Hua, X.; Pan, C.; Shi, Y.; Liu, J.; Hong, H. Removing Atmospheric Turbulence Effects Via Geometric Distortion and Blur Representation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4100113. [CrossRef]
3. Medlej, M.J.; Giordano, C.; Ziad, A.; Rafalimanana, A.; Aristidi, E. Optimizing Ground-layer Turbulence Prediction Using an Instrumented Drone. *Publ. Astron. Soc. Pac.* **2025**, *137*, 104508. [CrossRef]
4. Tonpe, S.; Reddy, J.S.; Bhar, C.; Pratap, A.; Nayak, J. Development of a plenoptic imaging system for turbulent atmospheric applications. *J. Opt.* **2025**, *54*, 240–253. [CrossRef]
5. Yang, C.; Shan, K.; Chen, J.; Hou, J.; Chen, S. CNN-Based Phase Matching for the OAM Mode Selection in Turbulence Heterodyne Coherent Mitigation Links. *IEEE Photonics J.* **2020**, *12*, 7905913. [CrossRef]
6. Liu, X.; Li, G.; Zhao, Z.; Cao, Q.; Zhang, Z.; Yan, S.; Xie, J.; Tang, M. EAF-WGAN: Enhanced Alignment Fusion-Wasserstein Generative Adversarial Network for Turbulent Image Restoration. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 5605–5616. [CrossRef]
7. Hançer, Z.; Şenyayla, O.; Bakırcıoğlu, V. Real-Time Atmospheric Turbulence Mitigation Using CPU-Based Multiframe Fusion on Embedded Electro-Optical Systems. *IEEE Access* **2026**, *14*, 42663–42673. [CrossRef]
8. Wang, C.; Sun, G.; Wang, C.; Gao, Z.; Wang, H. Monte Carlo-Based Restoration of Images Degraded by Atmospheric Turbulence. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 6610–6620. [CrossRef]
9. Wu, Y.; Cheng, K.; Chai, T.; Lyu, G.; Zhao, S.; Jia, W. Boosting Restoration of Turbulence-Degraded Images With State Space Conditional Diffusion. *IEEE Trans. Signal Process.* **2025**, *73*, 2631–2645. [CrossRef]
10. Ji, F.; Wang, J.; Cui, S.; Li, J.; Tang, X.; Xu, F. Physics-Guided Optical Simulation and PSF Analysis for Remote Sensing Images Deblurring. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5633015. [CrossRef]
11. Khowaja, S.A.; Ali, U.; Dev, K.; Lee, I.H. Image Registration and Deep NeuroFuzzy Networks for Mitigating Atmospheric Turbulence Effects in Consumer-Based Optical Imaging. *IEEE Trans. Consum. Electron.* **2025**, *71*, 7086–7094. [CrossRef]

12. Parmar, B.; Israni, D.; Shah, A. An Efficient Blind Approach for Restoration of Atmospheric Turbulence Degraded Images. In Proceedings of the 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysore, India, 10–11 December 2021; pp. 690–695. [CrossRef]
13. Yasarla, R.; Patel, V.M. CNN-Based Restoration of a Single Face Image Degraded by Atmospheric Turbulence. *IEEE Trans. Biom. Behav. Identity Sci.* **2022**, *4*, 222–233. [CrossRef]
14. Feng, B.Y.; Xie, M.; Metzler, C.A. Imaging Through Turbulence with GANs. In Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 31 October–2 November 2022. [CrossRef]
15. Mao, Z.; Jaiswal, A.; Wang, Z.; Chan, S.H. Single Frame Atmospheric Turbulence Mitigation: A Benchmark Study and a New Physics-Inspired Transformer Model. In *Proceedings of the Computer Vision–ECCV 2022*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13679, pp. 430–446. . [CrossRef]
16. Zhang, X.; Mao, Z.; Chimitt, N.; Chan, S.H. Imaging Through the Atmosphere Using Turbulence Mitigation Transformer. *IEEE Trans. Comput. Imaging* **2024**, *10*, 115–128. [CrossRef]
17. Xia, Y.; Zhou, C.; Zhu, C.; Teng, M.; Xu, C.; Shi, B. NB-GTR: Narrow-Band Guided Turbulence Removal. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 24934–24943. [CrossRef]
18. Hill, P.; Liu, Z.; Anantrasirichai, N. MAMAT: 3D Mamba-Based Atmospheric Turbulence Removal and its Object Detection Capability. In Proceedings of the 2025 IEEE International Conference on Advanced Visual and Signal-Based Systems (AVSS), Tainan, Taiwan, 11–13 August 2025; pp. 1–6. [CrossRef]
19. Chu, W.; Cheng, Z.; He, L. Semi-Supervised Atmospheric Turbulence Mitigation Based on Hybrid Models. *IEEE Access* **2024**, *12*, 174527–174538. [CrossRef]
20. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5574–5584.
21. Yasarla, R.; Patel, V.M. Learning to Restore Images Degraded by Atmospheric Turbulence Using Uncertainty. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AL, USA, 19–22 September 2021; pp. 1694–1698. [CrossRef]
22. Mei, K.; Patel, V.M. LTT-GAN: Looking Through Turbulence by Inverting GANs. *IEEE J. Sel. Top. Signal Process.* **2023**, *17*, 587–598. [CrossRef]
23. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In Proceedings of the Advances in Neural Information Processing Systems 30 (NeurIPS), Long Beach, CA, USA, 4–6 December 2017, pp. 6402–6413.
24. Belhasin, O.; Romano, Y.; Freedman, D.; Rivlin, E.; Elad, M. Principal Uncertainty Quantification With Spatial Correlation for Image Restoration Problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 3321–3333. [CrossRef] [PubMed]
25. Jaiswal, A.; Zhang, X.; Chan, S.H.; Wang, Z. Physics-Driven Turbulence Image Restoration with Stochastic Refinement. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 12170–12181. [CrossRef]
26. Hill, P.; Anantrasirichai, N.; Achim, A.; Bull, D.R. Deep learning techniques for atmospheric turbulence removal: A review. *Artif. Intell. Rev.* **2025**, *58*, 101. [CrossRef]
27. Jin, D.; Chen, Y.; Lu, Y.; Chen, J.; Wang, P.; Liu, Z.; Guo, S.; Bai, X. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. *Nat. Mach. Intell.* **2021**, *3*, 876–884. [CrossRef]
28. Zhang, X.; Chimitt, N.; Chi, Y.; Mao, Z.; Chan, S.H. Spatio-Temporal Turbulence Mitigation: A Translational Perspective. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 2889–2899. [CrossRef]
29. Peng, C.; Li, J.; Bao, J.; Luo, L. Attention-Enhanced GAN for Astronomical Image Restoration Under Atmospheric Turbulence and Optical Aberrations. *Sensors* **2026**, *26*, 2135. [CrossRef] [PubMed]
30. Nair, N.G.; Mei, K.; Patel, V.M. AT-DDPM: Restoring Faces Degraded by Atmospheric Turbulence Using Denoising Diffusion Probabilistic Models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–7 January 2023; pp. 3434–3443.
31. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the First Conference on Language Modeling (COLM), Philadelphia, PA, USA, 7 October 2024.
32. Xu, S.; Sun, R.; Chang, Y.; Cao, S.; Xiao, X.; Yan, L. Long-range Turbulence Mitigation: A Large-scale Dataset and A Coarse-to-fine Framework. In *Proceedings of the Computer Vision–ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024. [CrossRef]
33. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *PLoS Med.* **2021**, *18*, e1003583. [CrossRef] [PubMed]

34. Mao, Z.; Chimitt, N.; Chan, S.H. Image Reconstruction of Static and Dynamic Scenes Through Anisoplanatic Turbulence. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1415–1428. [CrossRef]
35. Chan, S.H. Tilt-Then-Blur or Blur-Then-Tilt? Clarifying the Atmospheric Turbulence Model. *IEEE Signal Process. Lett.* **2022**, *29*, 1833–1837. [CrossRef]
36. Guo, Y.; Wu, X.; Qing, C.; Su, C.; Yang, Q.; Wang, Z. Blind Restoration of Images Distorted by Atmospheric Turbulence Based on Deep Transfer Learning. *Photonics* **2022**, *9*, 582. [CrossRef]
37. Nair, N.G.; Mei, K.; Patel, V.M. A Comparison of Different Atmospheric Turbulence Simulation Methods for Image Restoration. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022. [CrossRef]
38. Lau, C.P.; Kumar, A.; Chellappa, R. Semi-Supervised Landmark-Guided Restoration of Atmospheric Turbulent Images. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 204–215. [CrossRef]
39. Rai, S.N.; Jawahar, C. Removing Atmospheric Turbulence via Deep Adversarial Learning. *IEEE Trans. Image Process.* **2022**, *31*, 2633–2646. [CrossRef] [PubMed]
40. López-Tapia, S.; Wang, X.; Katsaggelos, A.K. Video Variational Deep Atmospheric Turbulence Correction. *IEEE Access* **2024**, *12*, 3455289. [CrossRef]
41. Wang, X.; López-Tapia, S.; Katsaggelos, A.K. Real-World Atmospheric Turbulence Correction via Domain Adaptation. In Proceedings of the 2024 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 27–30 October 2024; pp. 1466–1472. [CrossRef]
42. Zhang, Y.; Yao, H.; Cao, Z.; Wang, W.; Zhang, T.; Hao, Q. Research on Turbulent Image Correction Methods Based on Deep Feature Perception. In Proceedings of the 2025 17th International Conference on Advanced Infocomm Technology (ICAIT), Liaocheng, China, 24–27 October 2025; pp. 224–227. [CrossRef]
43. Andrews, L.C. *Field Guide to Atmospheric Optics*, 2nd ed.; SPIE: Cergy, France, 2019. [CrossRef]
44. Yang, Y.; Yang, D.; Wang, J.; Zhu, Y.; Hu, Z.; Zhang, Y. Non-Kolmogorov atmospheric turbulence and orbital angular momentum of entangled states for optical communication. *Results Phys.* **2019**, *15*, 102676. [CrossRef]
45. Tyson, R.K.; Frazier, B.W. *Field Guide to Adaptive Optics*, 2nd ed.; SPIE: Cergy, France, 2012. [CrossRef]
46. Liu, Q.; Di, Y.; Zhang, M.; Ren, Z.; Di, J.; Zhao, J. Research Progress on Atmospheric Turbulence Perception and Correction Based on Adaptive Optics and Deep Learning. *Adv. Photonics Res.* **2025**, *6*, 2400204. [CrossRef]
47. Fried, D. Optical heterodyne detection of an atmospherically distorted signal wave front. *Proc. IEEE* **1967**, *55*, 57–77. [CrossRef]
48. Smith, A.; Bailey, J.; Hough, J.H.; Lee, S. An investigation of lucky imaging techniques. *Mon. Not. R. Astron. Soc.* **2009**, *398*, 2069–2073. [CrossRef]
49. Deledalle, C.; Gilles, J. Blind atmospheric turbulence deconvolution. *IET Image Process.* **2020**, *14*, 3422–3432. [CrossRef]
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
51. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, PR, USA, 2–4 May 2016.
52. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef] [PubMed]
53. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; Proceedings of Machine Learning Research; Volume 48, pp. 1050–1059.
54. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
55. Zou, Z.; Anantrasirichai, N. DeTurb: Atmospheric Turbulence Mitigation with Deformable 3D Convolutions and 3D Swin Transformers. In *Proceedings of the Computer Vision–ACCV 2024*; Lecture Notes in Computer Science; Springer: Singapore, 2024; Volume 15475, pp. 904–921. [CrossRef]
56. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 6840–6851.
57. Song, J.; Meng, C.; Ermon, S. Denoising Diffusion Implicit Models. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021.
58. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
59. Yin, J.; Sun, T.; Zhang, X.; Zhang, G.; Wan, X.; He, J. TMTS: A Physics-Based Turbulence Mitigation Network Guided by Turbulence Signatures for Satellite Video. *Remote Sens.* **2025**, *17*, 2422. [CrossRef]
60. Gilles, J.; Ferrante, N.B. Open Turbulent Image Set (OTIS). *Pattern Recognit. Lett.* **2017**, *86*, 38–41. [CrossRef]

61. Anantrasirichai, N.; Achim, A.; Kingsbury, N.G.; Bull, D.R. Atmospheric Turbulence Mitigation Using Complex Wavelet-Based Fusion. *IEEE Trans. Image Process.* **2013**, *22*, 2398–2408. [CrossRef] [PubMed]
62. Wang, Y.; Jodoin, P.M.; Porikli, F.; Konrad, J.; Benezeth, Y.; Ishwar, P. CDnet 2014: An Expanded Change Detection Benchmark Dataset. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 393–400. [CrossRef]
63. Popowicz, A.; Orlov, V. *SUTO-Solar Through-Turbulence Open Image Dataset*; Zenodo: Geneva, Switzerland, 2022. [CrossRef]
64. Werth, M.; Lucas, J.; Kyono, T.; McQuaid, I.; Fletcher, J. SILO: A Machine Learning Dataset of Synthetic Ground-Based Observations of LEO Satellites. In Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MN, USA, 7–14 March 2020; pp. 1–8. [CrossRef]
65. Saha, R.K.; Qin, D.; Li, N.; Ye, J.; Jayasuriya, S. Turb-Seg-Res: A Segment-then-Restore Pipeline for Dynamic Videos with Atmospheric Turbulence. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 25286–25296. [CrossRef]
66. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
67. Jain, T.; LuBien, M.; Gilles, J. Evaluation of Neural Network Algorithms for Atmospheric Turbulence Mitigation. In *Proceedings of the Signal Processing, Sensor/Information Fusion, and Target Recognition XXXI*; SPIE: Cergy, France, 2022; p. 30. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.