

# A Unified Framework for Meteorologically Physical Parameter-Driven Weather Generation

Jiantao Yu, Diyue Zhang, Ting Luo, Xiangyu Wu, Jing Hu, Xia Yuan\*, Xi Wu, Feihu Huang  
Chengdu University of Information Technology  
School of Computer Science, Sichuan Province, China  
Email: xyuan623@163.com

**Abstract**—Converting continuous meteorological data into controllable visual scenes remains challenging due to scarce supervision and unstable physical control. We propose a collaborative solution bridging simulation and generation. We introduce WeatherImage, a dataset of 20,000 pixel-aligned pairs providing the first continuous physical supervision across five weather conditions. We develop WeatherDiff to address two key bottlenecks: a Meteorological Semantic Adapter overcomes semantic insensitivity to fine-grained parameters, while a Structure-First Dynamic Optimization Strategy resolves the texture-structure conflict. Experiments demonstrate significant improvements in FID and CLIP, critically, the first monotonic response to continuous meteorological inputs. This enables quantitatively controllable, photorealistic weather synthesis for autonomous driving and disaster management.

**Index Terms**—Weather Image Generation, Paired Dataset, Structural Consistency, Generative Models, Meteorological Quantitative Semantics

## I. INTRODUCTION

Visual representation of meteorological scenes is a critical bridge connecting atmospheric physics to machine perception, essential for autonomous driving robustness and disaster management [1], [2]. However, real weather is defined by continuous physical parameters rather than discrete labels. These parameters include precipitation intensity, visibility, snow depth, and snowfall. This makes controllable visual scene generation fundamentally challenging.

Current approaches suffer from distinct limitations. Traditional methods rely on abstract representations like contour maps [3], which pose interpretability barriers for lay users despite their scientific rigor. Alternatively, graphics-based rendering offers controllability but struggles with unrealistic fidelity due to domain gaps and excessive computational costs, rendering it impractical for large-scale applications.

Methods dedicated to meteorological visualization typically target specific phenomena, with rain approaches focusing on streak synthesis [4], [5] and haze models on scattering simulation [6], [7]. However, lacking paired data and continuous physical supervision, these approaches rely on discrete style transfers, failing to ensure monotonic responses to numerical parameters or preserve structural

integrity. While recent diffusion models [8] offer superior photorealism, achieving controllable generation from continuous physical data remains challenging due to three persistent constraints. At the data level, general models rely on discrete text labels such as LAION [9] which lack physical continuity, whereas real-world datasets miss the paired annotations required for precise intensity guidance [10], [11]. In terms of modeling, standard text encoders fail to capture the non-linear coupling between multi-dimensional parameters and visual effects, leading to insensitivity toward fine-grained physical changes. During optimization, conflicts arise between local high-frequency weather textures and global structural stability, where gradient competition often causes background degradation and detail drift.

To address the aforementioned challenges, we synergize the physical precision of graphics simulation with the visual realism of generative models to achieve quantitatively controllable meteorological synthesis. By bridging the gap between continuous physical parameters and photorealistic appearance via WeatherImage and WeatherDiff, our contributions are summarized as follows:

- 1) WeatherImage Dataset: The first large-scale, pixel-aligned dataset providing continuous physical supervision across five weather conditions, establishing a benchmark for intensity-controllable scene generation.
- 2) Parameterized Condition Modeling: A novel mechanism within the WeatherDiff framework that maps continuous numerical variables to stable and predictable visual responses.
- 3) Dynamic Structure-Texture Optimization: A strategy that balances high-frequency weather textures with low-frequency geometric structures, effectively suppressing structural degradation under extreme conditions.

Code and dataset are available at <https://github.com/Yjt12138/WeatherDiff-WeatherImage>.

## II. RELATED WORK

### A. Meteorological Visualization and Simulation

Meteorological visualization bridges atmospheric physics and machine perception. Early physical models

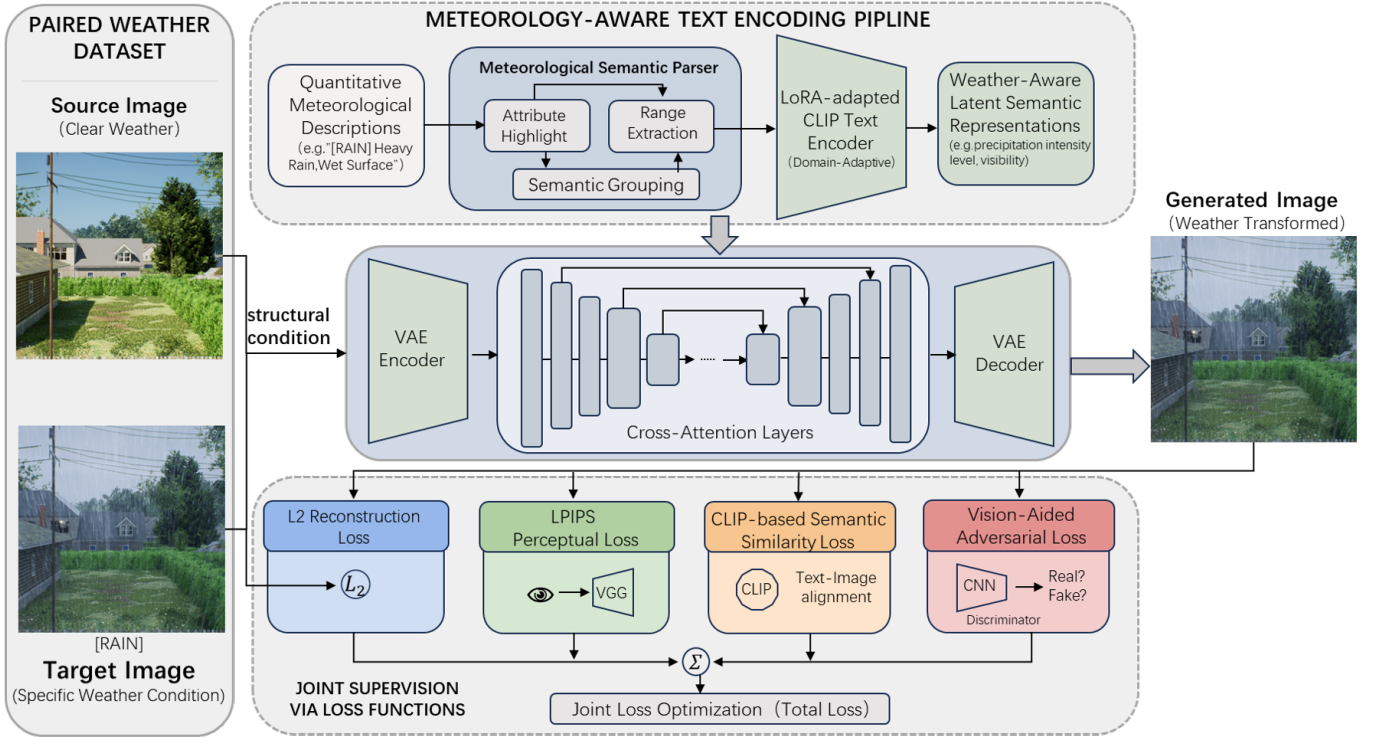


Fig. 1: The WeatherDiff architecture. Our framework achieves orthogonal decoupling between scene structure and meteorological textures. A LoRA-adapted CLIP encoder captures quantitative physical semantics from text, while an SD-Turbo backbone serves as a geometric anchor to preserve the source image’s topology. The model is optimized via Joint Supervision of reconstruction, perceptual, semantic, and adversarial losses.

[12] and modern graphics engines [1], [13] enable parameterized simulation of weather effects. While ensuring physical consistency, these methods suffer from high computational costs and significant sim-to-real gaps, making them difficult to substitute for real-world semantic mapping directly.

### B. Image-to-Image Translation for Weather

GAN-based approaches established the paradigm for weather translation. While Pix2Pix [14] and CycleGAN [15] addressed paired and unpaired translation, domain-specific models like WeatherGAN [4] and ClimateGAN [16] focused on category-level transfer. However, GANs remain limited to discrete style transfer, failing to achieve precise control over continuous physical quantities and suffering from inherent training instability and mode collapse.

### C. Diffusion Models for Controllable Generation

Diffusion models like OmniGen2 [17] offer superior synthesis but suffer from Semantic Insensitivity to fine-grained physical instructions. Furthermore, high-frequency texture injection often triggers Gradient Competition, causing structural drift. While methods like InstructPix2Pix [18] and ControlNet [19] improve controllability, they struggle to balance precise physical texture rendering with structural preservation—a gap our WeatherDiff framework addresses via parameterized adaptation and dynamic optimization.

## III. METHOD

### A. WeatherImage: Strictly Paired Dataset with Physical Annotations

To enable controllable generation driven by continuous meteorological data, we constructed WeatherImage, a dataset containing 20,000 strictly paired samples (Fig 2). Each sample is organized as a triplet  $(I_{\text{clr}}, \mathbf{q}, I_{\text{gt}})$ , where both images are rendered via UE under fixed intrinsics and extrinsics.  $I_{\text{clr}}$  is the sunny reference image, selected as the geometric anchor due to its optimal structural clarity free from meteorological occlusion;  $\mathbf{q}$  is the continuous meteorological vector; and  $I_{\text{gt}}$  is the target ground truth. This format enables the model to learn fine-grained control laws beyond discrete labels (Table I).

**Physical Parameter Space.** To bridge the sim-to-real gap, we mapped real-world physical quantities to engine parameters. (Table II) Crucially, we treat Visibility ( $V$ , Level 0-5) and Cloud Coverage ( $C$ ) as foundational dependent variables, strictly coupled with specific weather intensities to ensure physical consistency. We defined five sub-tasks with specific coupling logic:

**Rain and Snow:** Focus on high-frequency particle perturbations. Rain couples instantaneous intensity ( $PI$ , 0-30mm/h) with cumulative precipitation ( $P$ , 0-100mm), while Snow links snowfall ( $SF$ , 0-5cm/h) with snow depth ( $SD$ , 0-20cm). In both tasks,  $C$  is locked

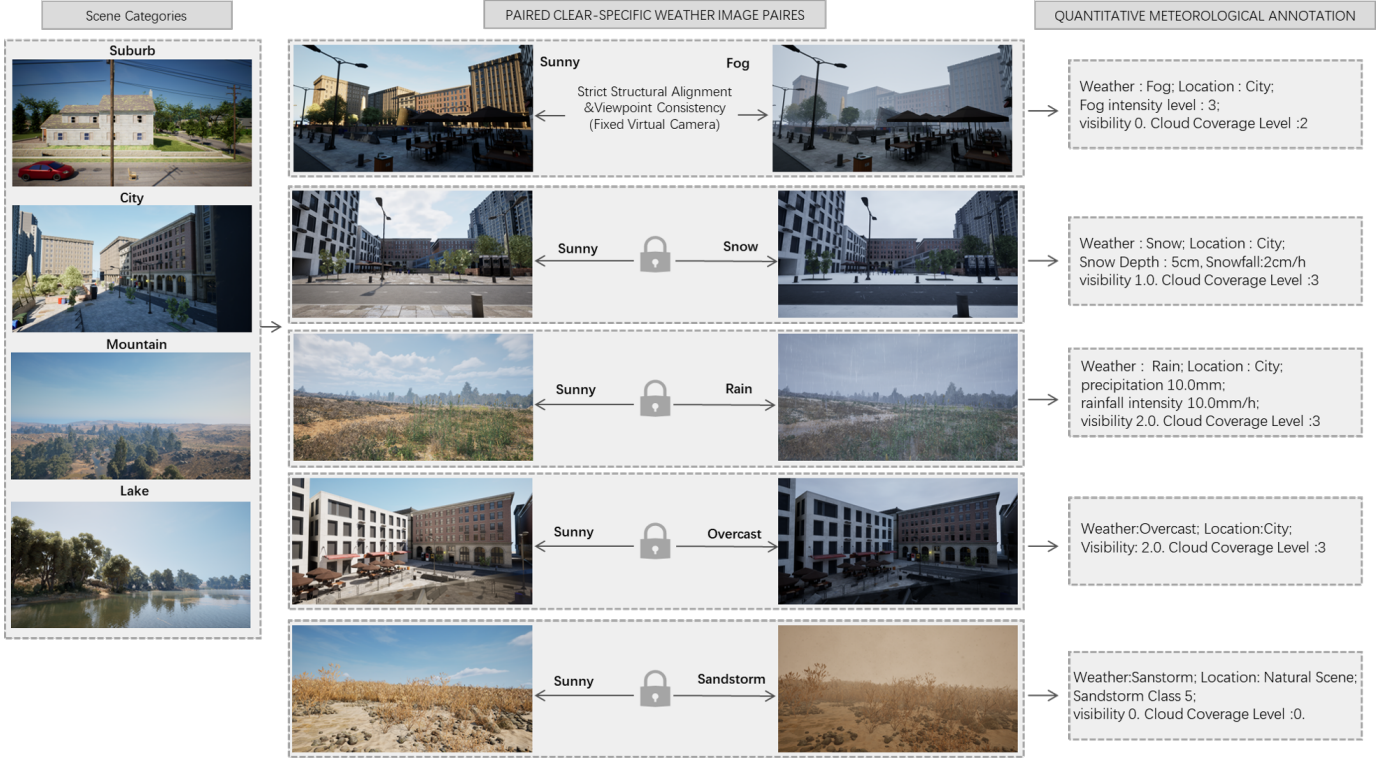


Fig. 2: WeatherImage dataset overview. Our dataset provides: (Left) Diverse scene coverage, including urban and natural landscapes. (Middle) Strict pixel-level alignment between clear-day reference and weather images via fixed virtual cameras. (Right) Fine-grained quantitative annotations supporting continuous intensity supervision.

to maximum, and  $V$  is negatively correlated with precipitation rates.

**Fog and Sandstorm:** Focus on volumetric medium scattering. Fog links Fog Intensity ( $FI$ , Level 0-5) inversely to Visibility to generate depth-dependent decay. Sandstorm simulates extreme dry medium interference ( $SS$ , Level 0-4), excluding precipitation parameters while forcing  $C$  to zero.

**Overcast:** Focuses on illumination changes. It sets Cloud Coverage to full saturation to remove direct light for diffuse skylight, while maintaining normal Visibility.

WeatherImage shows similar low-order photometric statistics to real-world imagery, such as brightness and contrast, providing preliminary support for its visual realism.

### B. WeatherDiff: Unified Generation Architecture via Structure-Semantic Decoupling

To reconcile meteorological semantic consistency with structural integrity, we observe that weather effects manifest primarily as mutable high-frequency textures like rain streaks, while the background constitutes stable low-frequency structures like road topology. Based on this, we propose WeatherDiff (Fig 1). We propose a 'structure-first, semantics-driven' principle for weather generation. This principle decouples the task into Scene Structure Locking and Meteorological Texture Rendering. Consequently, it achieves unified modeling via decoupled pathways.

1) *Meteorological Semantic Domain Adaptation Module:* Standard encoders, trained on natural images, struggle to capture quantitative physical attributes. We employ Low-Rank Adaptation (LoRA) [23] to construct a lightweight projection layer without disrupting pre-trained knowledge. Given a structured meteorological sequence  $T_{quant}$ , we inject LoRA operators into the frozen CLIP backbone  $\Theta_{pre}$ . The feature mapping is optimized via low-rank matrices  $\mathbf{A} \in \mathbb{R}^{r \times d}$  and  $\mathbf{B} \in \mathbb{R}^{d \times r}$  ( $r \ll d$ ):

$$z_t = \mathcal{E}_{txt}(T_{quant}; \Theta_{pre} + \mathbf{BA}) \quad (1)$$

where  $\mathcal{E}_{txt}$  denotes the text encoder and  $z_t$  is the adapted embedding. This establishes a meteorology-specific subspace, explicitly encoding scalar parameters into discriminative semantics  $z_t$  aligned with fine-grained physical laws.

2) *Structure-Preserving Backbone:* To minimize structural distortion, we employ SD-Turbo [24] as the core generative backbone, explicitly coupled with VAE skip connections [25], [26]. These connections constitute a dedicated transmission channel for low-frequency structural information. Distinct from direct generation from noise, the backbone synthesizes high-frequency meteorological textures directly on top of the existing low-frequency structural foundation. This effectively locks the underlying geometric topology, allowing the generator to be relieved of scene reconstruction tasks.



TABLE I: Comparison of Meteorological Image Datasets

| Dataset                    | Weather Types |     |          |      |      |       |       | Granularity                | Paired | Source           |
|----------------------------|---------------|-----|----------|------|------|-------|-------|----------------------------|--------|------------------|
|                            | Rain          | Fog | Overcast | Snow | Sand | Flood | Sunny |                            |        |                  |
| BDD100k [10]               | ✓             | ✓   | ✓        | ✓    |      |       | ✓     | Discrete Labels            | ×      | Real             |
| Multitask Weather [20]     | ✓             |     | ✓        |      |      |       | ✓     | Discrete Labels            | ×      | Real             |
| SPA+ [21]                  | ✓             |     |          |      |      |       | ✓     | Discrete Labels            | ▲      | Real             |
| RealSnow [11]              |               |     |          | ✓    |      |       | ✓     | Discrete Labels            | ▲      | Real             |
| Rain1400 [22]              | ✓             |     |          |      |      |       | ✓     | Discrete Labels            | ✓      | Synthetic        |
| Mila-Simulated [16]        |               |     |          |      |      | ✓     | ✓     | Discrete Labels            | ✓      | simulated        |
| <b>WeatherImage (Ours)</b> | ✓             | ✓   | ✓        | ✓    | ✓    |       | ✓     | <b>Quantitative labels</b> | ✓      | <b>simulated</b> |

<sup>a</sup>✓: Paired; ×: Unpaired; ▲: Partially paired/Synthetic. Our dataset supports continuous physical parameter control.

TABLE II: Parameter Distribution in WeatherImage

| Weather   | V | C | P | PI | SD | SF | FI | SS | Num   |
|-----------|---|---|---|----|----|----|----|----|-------|
| sunny     | ✓ | ✓ | ✓ | -  | ✓  | -  | -  | -  | 20000 |
| fog       | ✓ | ✓ | ✓ | -  | ✓  | -  | ✓  | -  | 4000  |
| rain      | ✓ | ✓ | ✓ | ✓  | -  | -  | -  | -  | 4000  |
| snow      | ✓ | ✓ | - | -  | ✓  | ✓  | -  | -  | 4000  |
| overcast  | ✓ | ✓ | ✓ | -  | ✓  | -  | -  | -  | 4000  |
| sandstorm | ✓ | ✓ | - | -  | -  | -  | -  | ✓  | 4000  |

✓ indicates the parameter is active.

**Abbreviations:** V=Visibility; C=Cloud Coverage; P=Precipitation;

PI=Precipitation Intensity; SD=Snow Depth; SF=Snowfall;

FI=Fog Intensity; SS=Sandstorm Severity.

### 3) Cross-Modal Injection of Meteorological Semantics:

With the geometric structure anchored by the channel, the SD-Turbo [24] generator is fully dedicated to Meteorological Texture Rendering. We inject the semantic vectors  $\mathbf{z}_t$  into the generator’s intermediate features  $\mathbf{F}_i$  via Cross-Attention:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left( \frac{\mathbf{Q}(\mathbf{F}_i) \mathbf{K}(\mathbf{z}_t)^\top}{\sqrt{d_k}} \right) \mathbf{V}(\mathbf{z}_t) \quad (2)$$

where  $d_k$  represents the scaling dimension. Guided by  $\mathbf{z}_t$ , the generator focuses on synthesizing high-frequency textures such as rain density and fog granularity. By superimposing dynamic weather effects onto the fixed geometric topology, the method achieves precise rendering while maintaining structural integrity.

4) *Dynamic Balance Optimization:* We design a Structure-First Dynamic Loss strategy to reconcile the competition between structure and texture. This strategy synergizes three key objectives: Structural Loss ( $\mathcal{L}_{\text{STRUCT}}$ ) to constrain pixel-level geometric consistency, Perceptual Loss ( $\mathcal{L}_{\text{LPIPS}}$ ) to ensure feature-level visual realism, and Semantic Alignment Loss ( $\mathcal{L}_{\text{CLIP}}$ ) to strictly align the synthesized effects with meteorological inputs. The total objective integrates these components with time-dependent weights  $\lambda(t)$ :

$$\mathcal{L}_T = \lambda_s(t) \cdot \mathcal{L}_{\text{STRUCT}} + \lambda_p(t) \cdot \mathcal{L}_{\text{LPIPS}} + \lambda_c(t) \cdot \mathcal{L}_{\text{CLIP}} \quad (3)$$

We employ a dynamic scheduling strategy for  $\lambda(t)$ . In the early stages, we prioritize  $\lambda_s$  to firmly anchor the geometric structure, while progressively increasing  $\lambda_c$  to guide the precise injection of meteorological textures as training proceeds.

“A photo of city taken at morning, with weather information as follows: fog, with a cloud coverage level of 0, Fog intensity level of 3, visibility 0.”

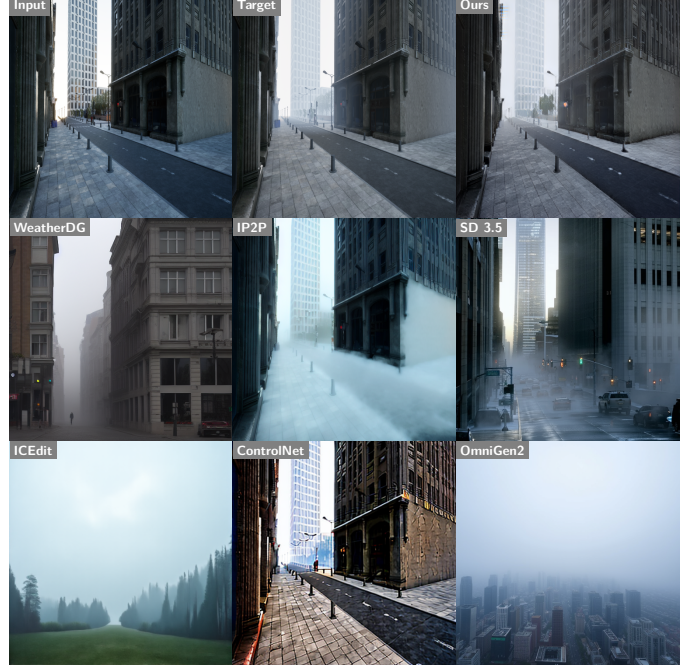


Fig. 3: **Qualitative comparison for Fog generation.** Responding to the Fog Intensity input, our method renders realistic volumetric fog with depth-dependent decay. In contrast, baselines suffer from structural collapse or fail to produce atmospheric depth.

## IV. EXPERIMENTS

### A. Experimental Settings

1) *Experimental Datasets:* All experiments are conducted on the WeatherImage dataset, featuring 20,000 strictly paired samples with continuous physical annotations across five conditions: rain, fog, snow, overcast, and sandstorm. Rain is controlled by precipitation intensity; fog by volumetric scattering; snow by airborne and surface accumulation; overcast by diffuse illumination; and sandstorm by medium interference. These tasks collectively span a wide range of weather-induced appearance variations.

2) *Implementation Details:* Our model is implemented on a single NVIDIA RTX 4090 GPU and trained using the AdamW optimizer. The learning rate decays from  $1 \times 10^{-6}$  to  $1 \times 10^{-8}$  with a batch size of 2 at  $512 \times 512$  resolution.



TABLE III: Quantitative Comparison of Different Methods Under Various Weather Conditions

| Method          | Fog             |                  | Rain            |                  | Snow            |                  | Overcast        |                  | Sandstorm       |                  |
|-----------------|-----------------|------------------|-----------------|------------------|-----------------|------------------|-----------------|------------------|-----------------|------------------|
|                 | CLIP $\uparrow$ | FID $\downarrow$ | CLIP $\uparrow$ | FID $\downarrow$ | CLIP $\uparrow$ | FID $\downarrow$ | CLIP $\uparrow$ | FID $\downarrow$ | CLIP $\uparrow$ | FID $\downarrow$ |
| SD 2.1 [8]      | 30.04           | 120.24           | 30.90           | 148.50           | 31.02           | 137.18           | 29.52           | 155.58           | 31.13           | 163.26           |
| IP2P [18]       | 29.99           | 118.38           | 23.41           | 70.90            | 25.85           | 93.82            | 27.19           | 82.03            | 29.84           | 217.47           |
| OmniGen2 [17]   | 31.44           | 112.12           | 30.66           | 130.64           | 28.86           | 125.82           | 29.90           | 129.50           | 30.10           | 150.82           |
| WeatherDG [27]  | 28.08           | 153.23           | 27.35           | 159.01           | 30.28           | 164.84           | 28.24           | 167.81           | 29.10           | 185.02           |
| SD 3.5 [28]     | 28.62           | 126.71           | 28.40           | 134.49           | 27.39           | 153.89           | 26.38           | 143.63           | 25.56           | 165.35           |
| ICEdit [29]     | 23.97           | 279.90           | 21.83           | 287.37           | 19.74           | 278.34           | 31.54           | 324.54           | 23.09           | 350.35           |
| ControlNet [19] | 22.58           | 63.53            | 26.38           | 102.42           | 26.07           | 107.07           | 23.88           | 108.40           | 28.42           | 113.67           |
| <b>Ours</b>     | <b>34.95</b>    | <b>42.63</b>     | <b>34.08</b>    | <b>57.92</b>     | <b>36.77</b>    | <b>65.06</b>     | <b>36.77</b>    | <b>60.66</b>     | <b>39.74</b>    | <b>77.02</b>     |

The best results are highlighted in bold.  $\uparrow$ : higher is better;  $\downarrow$ : lower is better.

"A photo of city taken at morning, with weather information as follows: rain, with a cloud coverage level of 3, precipitation of 10(mm), precipitation intensity of 15.0(mm/h), visibility 1."



Fig. 4: **Qualitative comparison for Rain generation.** Responding to the precipitation intensity of 15(mm/h), our method synthesizes high-frequency details including rain streaks and wet surface reflections. Conversely, baselines either without generating rain streaks or destroy the scene geometry.

The dataset is split into training, validation, and test sets with an 8:1:1 ratio. For a fair comparison, both ControlNet and our model are trained on WeatherImage until convergence, while other baselines utilize their official pre-trained weights. All evaluations are performed on the independent test set with unseen scenes.

3) *Evaluation Metrics:* To assess generation fidelity and controllability, we employ Fréchet Inception Distance (FID) [30] to measure the distributional similarity between generated and real meteorological images, serving as a proxy for visual realism and structural fidelity. Additionally, we use CLIPScore [31] to quantify the semantic alignment between the output images and meteorological data,

"A photo of city taken at day, with weather information as follows: sandstorm, Class level 4 sandstorm, visibility level of 0, cloud coverage level of 0."

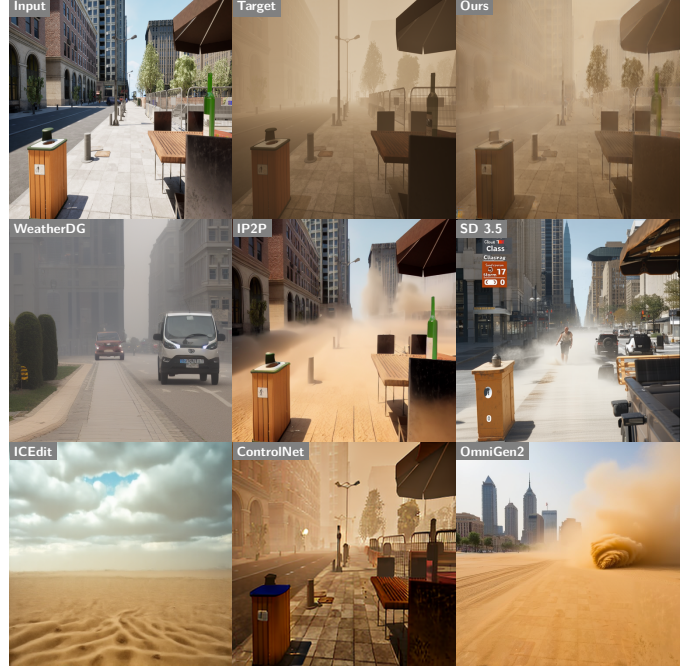


Fig. 5: **Qualitative comparison for Sandstorm generation.** Responding to the extreme sandstorm instruction, our method accurately renders suspended particles and tonal shifts while strictly preserving urban topology. This effectively avoids the hallucination of non-existent dunes observed in baselines.

reflecting the model's precision in intensity-controllable generation.

### B. Quantitative Analysis

To comprehensively verify effectiveness, we benchmarked six representative models. Stable Diffusion 3.5(SD 3.5) [28] and OmniGen2 [17] serve as foundational generative baselines. ICEdit [29] and InstructPix2Pix(IP2P) [18] represent instruction-based editing models. WeatherDG [27] is included as a meteorological GAN. Additionally, we selected Stable Diffusion 2.1 trained with ControlNet [19] to represent methods with strong structural control. Notably, we trained the ControlNet model on the Weath-

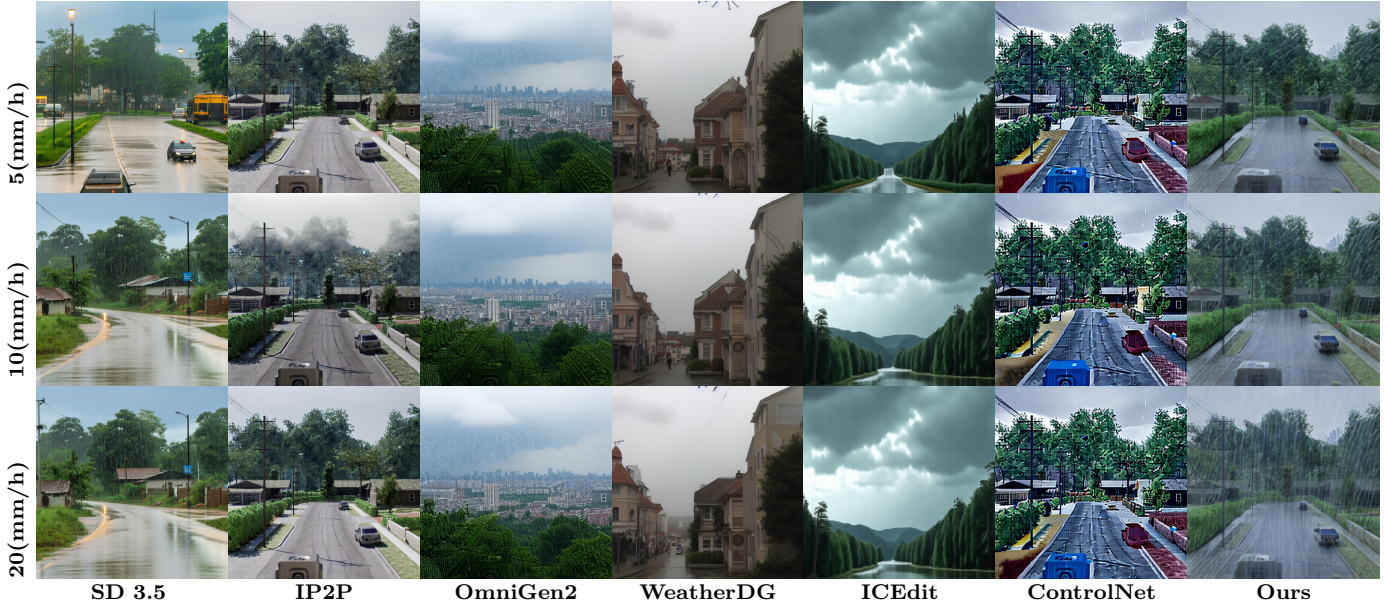


Fig. 6: **Qualitative analysis of controllability under varying precipitation intensities.** Responding to fine-grained intensity changes from 5(mm/h) to 20(mm/h), our method exhibits a precise monotonic visual response with linearly expanding rain streak density and surface wetness. Conversely, baseline methods fail to reflect numerical changes, showing either negligible visual variation or severe structural collapse under higher intensities.

erImage dataset until convergence. We employ FID [30] to evaluate distribution realism and CLIPScore [31] to assess semantic consistency (Table III).

Overall, existing methods trade off between semantic generation and image fidelity. Our method mitigates the conflict between high-fidelity structure preservation and fine-grained intensity control. It achieves the best balance across all weather categories.

**Rain and Snow Domain:** Baselines like ICEdit [29] suffer severe structural degradation due to high-frequency noise injection. In contrast, our method maintains competitive FID, proving it synthesizes meteorological particles rather than merely overlaying superficial noise.

**Fog and Overcast Domain:** While ControlNet [19] preserves structure, its conservative generation results in low semantic alignment. Our method effectively renders depth-dependent visibility, ensuring high consistency with meteorological descriptions.

**Sandstorm Extreme Domain:** IP2P [18] and OmniGen2 [17] suffer content collapse. Ours achieves the experiment’s highest CLIPScore with stable fidelity, demonstrating exceptional robustness in reconstructing environmental tones while strictly preserving geometric semantics.

### C. Qualitative Analysis

We visually evaluate the generation quality across five typical weather conditions (Figs. 3 to 5). As observed, baseline methods struggle to balance structural fidelity and physical realism. Generative editors often suffer from severe structural collapse, hallucinating content unrelated to the input layout. Conversely, structure-guided methods preserve geometry but fail to render complex physical

effects, often producing texture mapping artifacts without volumetric depth or correct lighting. In contrast, our method achieves orthogonal decoupling: it strictly preserves the geometric topology while accurately rendering high-frequency meteorological details governed by physical laws, such as depth-dependent visibility in fog and rain streaks in rain.

Rain Effect Coverage is quantified via the absolute pixel-wise difference between generated images and clear-weather references. As precipitation intensity increases across 5, 10, and 20 (mm/h), the measured coverage scales to 56.40%, 59.73%, and 69.03%. This trend corroborates a monotonic visual response to linear parameter changes (Fig. 6), a fine-grained capability absent in baseline methods. Fig. 7 confirms robust generalization to diverse real-world scenes.

### D. Ablation Study

To validate the effectiveness of the proposed decoupled architecture and optimization strategy, we conduct a progressive ablation study involving four configurations, Table IV and Fig. 8 present respectively the quantitative and qualitative results:

- **Baseline:** The original pre-trained SD-Turbo model without any architectural modifications.
- **Config A:** The Baseline equipped with the Structure-Preserving Backbone to lock the geometric anchor, but without meteorological semantic injection.
- **Config B:** Config A + Meteorological Semantic Domain Adaptation Module and Cross-Modal Injection, trained with fixed loss weights.



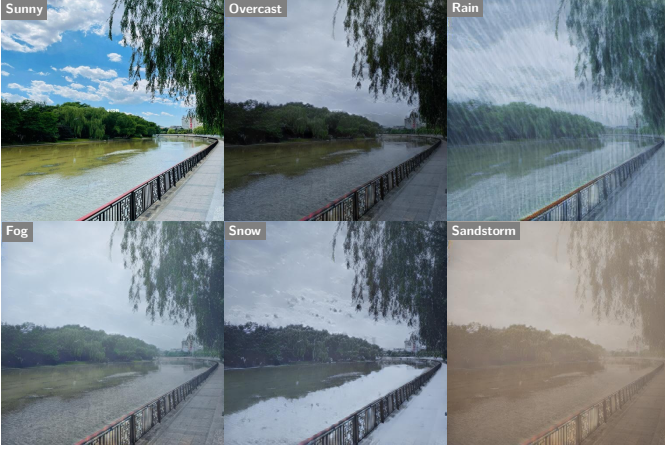


Fig. 7: **Generalization to real-world scenes.** The model successfully applies complex weather effects to diverse real-world photographs. This validates that our method bridges the synthetic-to-real domain gap, demonstrating strong robustness in real-world applications.

- **Config C:** The full WeatherDiff framework integrating the Dynamic Balance Optimization strategy.

1) *Impact of Structure-Preserving Backbone:* As shown in Fig. 8, the Baseline suffers from severe *Structural Drift*. Config A introduces the Structure-Preserving Backbone. This effectively locks the low-frequency topology, stabilizing the global structure. This is reflected in the table as a decrease in FID [30]. However, although scene geometry is well preserved, the model fails to manifest meteorological attributes. Specifically, precipitation-related prompts do not result in visible rain streaks or wet textures. This confirms that the structural anchor alone is insufficient to enable fine-grained meteorological control, and the CLIPScore [31] is still relatively low.

2) *Impact of Meteorological Semantic Adapter:* Config B further incorporates the Meteorological Semantic Domain Adaptation Module to activate the physics-to-texture mapping. Compared to Config A, precipitation cues are triggered. Nevertheless, under the Static Injection strategy, we observe failing to render rain streaks. We attribute this to Gradient Competition between the objectives of structure preservation and texture injection, lack spatial coherence and physical plausibility. The CLIPScore [31] increases but the FID [30] does not decrease.

3) *Effectiveness of Dynamic Balance Optimization:* Config C resolves the aforementioned conflict through the Dynamic Balance Optimization strategy. This dynamic scheduling enables the synthesis of natural, spatially consistent rain streaks while maintaining the strict geometric anchor provided by the backbone. Compared to Config B, our full approach significantly reduces texture artifacts and improves the coherence between scene structure and weather semantics, as shown in Fig. 8.

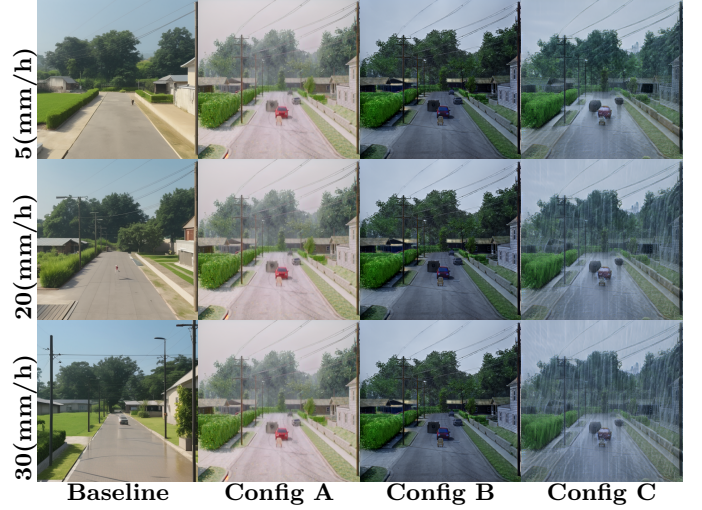


Fig. 8: **Qualitative Ablation Analysis of Controllability.**

TABLE IV: Quantitative Comparison of Different Methods Under Various Rain Intensities

| Method   | Rain 5mm/h   |              | Rain 20mm/h  |              | Rain 30mm/h  |              |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|
|          | CLIP↑        | FID↓         | CLIP↑        | FID↓         | CLIP↑        | FID↓         |
| Baseline | 15.79        | 128.36       | 15.91        | 168.01       | 15.69        | 235.17       |
| Config A | 19.14        | 87.49        | 19.77        | 120.62       | 19.73        | 184.34       |
| Config B | 27.94        | 86.55        | 28.42        | 111.70       | 28.19        | 183.49       |
| Config C | <b>29.17</b> | <b>84.08</b> | <b>31.29</b> | <b>92.39</b> | <b>31.34</b> | <b>94.64</b> |

#### E. Inference Efficiency Analysis

We evaluate the computational efficiency of our framework. Baseline models typically require 28 denoising steps, whereas our method utilizes a one-step configuration, achieving an inference time of only 0.41s. This significant speedup stems from leveraging the one-step capability of SD-Turbo, completing structure preservation and texture synthesis in a single forward pass. This efficiency demonstrates high potential for real-time meteorological visualization.

## V. CONCLUSION

This paper addresses the lack of physical controllability and structural fidelity in meteorological scene generation. We propose a collaborative data-model solution. We constructed WeatherImage. It is the first large-scale paired dataset supporting continuous intensity supervision. We also proposed the WeatherDiff framework. We established a precise mapping from continuous physical parameters to visual representations. Experimental results demonstrate the effectiveness of our parameterized adapter and dynamic strategy. They achieve orthogonal decoupling of texture injection and structure preservation. Compared to existing methods, our framework achieves optimal FID and CLIPScore. It also realizes a continuous monotonic response to physical quantities. This study provides a standardized data foundation and a universal generation paradigm.



However, our method has certain limitations. First, the strict geometric constraint restricts the synthesis of weather phenomena that physically alter scene topology. Second, due to the distribution bias in the WeatherImage dataset towards common conditions, the model’s responsiveness to extreme or rare weather scenarios remains limited.

Future work involves strengthening physical weather controllability and expanding datasets to include rare meteorological events.

#### ACKNOWLEDGMENT

This work was supported by the Sichuan Science and Technology Program (No. 2024NSFTD0043), the Key Technology Research and Development Project of Sichuan Province (Grant Nos. 2025YFNH0008 and 2024ZDZX0020), and the National Natural Science Foundation of China (Grant No. 42375148).

#### REFERENCES

- [1] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [2] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, “Benchmarking robustness in object detection: Autonomous driving when winter is coming,” *arXiv preprint arXiv:2107.07484*, 2019.
- [3] M. Rautenhaus, M. Böttinger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. Röber, and R. Westermann, “Visualization in meteorology—a survey of techniques and tools for data analysis tasks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 12, pp. 3268–3296, 2017.
- [4] X. Li, K. Kou, and B. Zhao, “Weather gan: Multi-domain weather translation using generative adversarial networks,” *arXiv preprint arXiv:2103.05422*, 2021.
- [5] S. Zheng, C. Lu, and S. G. Narasimhan, “Tpsence: Towards artifact-free realistic rain generation for deraining and object detection in rain,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 5394–5403.
- [6] R. Gong, D. Dai, Y. Chen, W. Li, D. P. Paudel, and L. Van Gool, “Analogical image translation for fog generation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 2, 2021, pp. 1433–1441.
- [7] Z. Liu, Y. Li, and M. Okutomi, “Cfdnet: A generalizable foggy stereo matching network with contrastive feature distillation,” 2024.
- [8] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.
- [9] C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman *et al.*, “Laion-5b: An open large-scale dataset for training next generation image-text models,” *Advances in neural information processing systems*, vol. 35, pp. 25 278–25 294, 2022.
- [10] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “Bdd100k: A diverse driving dataset for heterogeneous multitask learning,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.
- [11] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, “Desnownet: Context-aware deep network for snow removal,” *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3064–3073, 2018.
- [12] R. T. Tan, “Visibility in bad weather from a single image,” in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.
- [13] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “Airsim: High-fidelity visual and physical simulation for autonomous vehicles,” in *Field and service robotics: Results of the 11th international conference*. Springer, 2017, pp. 621–635.
- [14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [16] V. Schmidt, A. S. Luccioni, M. Teng, T. Zhang, A. Reynaud, S. Raghupathi, G. Cosne, A. Juraver, V. Vardanyan, A. Hernandez-Garcia *et al.*, “Climategan: Raising climate change awareness by generating images of floods,” *arXiv preprint arXiv:2110.02871*, 2021.
- [17] S. Xiao, Y. Wang, J. Zhou, H. Yuan, X. Xing, R. Yan, C. Li, S. Wang, T. Huang, and Z. Liu, “Omnigen: Unified image generation,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 13 294–13 304.
- [18] T. Brooks, A. Holynski, and A. A. Efros, “Instructpix2pix: Learning to follow image editing instructions,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 18 392–18 402.
- [19] L. Zhang, A. Rao, and M. Agrawala, “Adding conditional control to text-to-image diffusion models,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 3836–3847.
- [20] R. Li, R. T. Tan, and L.-F. Cheong, “All in one bad weather removal using architectural search,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3175–3185.
- [21] H. Wang, Z. Yue, Q. Xie, Q. Zhao, Y. Zheng, and D. Meng, “From rain generation to rain removal,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 791–14 801.
- [22] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, “Removing rain from single images via a deep detail network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3855–3863.
- [23] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen *et al.*, “Lora: Low-rank adaptation of large language models,” *ICLR*, vol. 1, no. 2, p. 3, 2022.
- [24] A. Sauer, D. Lorenz, A. Blattmann, and R. Rombach, “Adversarial diffusion distillation,” in *European Conference on Computer Vision*. Springer, 2024, pp. 87–103.
- [25] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [26] G. Parmar, T. Park, S. Narasimhan, and J.-Y. Zhu, “One-step image translation with text-to-image models,” *arXiv preprint arXiv:2403.12036*, 2024.
- [27] C. Qian, Y. Guo, Y. Mo, and W. Li, “Weatherdgc: Llm-assisted procedural weather generation for domain-generalized semantic segmentation,” *IEEE Robotics and Automation Letters*, 2025.
- [28] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer, F. Boesel *et al.*, “Scaling rectified flow transformers for high-resolution image synthesis,” in *Forty-first international conference on machine learning*, 2024.
- [29] Z. Zhang, J. Xie, Y. Lu, Z. Yang, and Y. Yang, “Enabling instructional image editing with in-context generation in large scale diffusion transformer,” in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
- [30] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.
- [31] J. Hessel, A. Holtzman, M. Forbes, R. Le Bras, and Y. Choi, “Clipscore: A reference-free evaluation metric for image captioning,” in *Proceedings of the 2021 conference on empirical methods in natural language processing*, 2021, pp. 7514–7528.