WEATHERSHIFT: TRANSITIONAL WEATHER VIDEO GENERATIVE MODELING FOR AUTONOMOUS DRIVING PERCEPTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Robust perception for autonomous driving requires generalizing beyond discrete weather categories to the continuously evolving conditions encountered on the road. Existing autonomous-driving datasets and perception methods predominantly represent weather as fixed states such as clear, rain, fog, or snow, ignoring the transitions between them, yet real driving sequences pass through intermediate regimes, such as sunny to rainy or rainy to sunny, with continuously varying levels of adversity. Recent weather-transition generation methods primarily operate at the image level and typically assume a static background, limiting their ability to capture the temporal evolution of weather alongside the dynamic scene content and ego-motion inherent to real-world driving videos. To address these limitations, we propose **WeatherShift**, a video-based framework for generating temporally coherent weather transitions in autonomous-driving scenes. Starting from real-world driving videos in nuScenes, we construct extreme-weather counterparts, namely **Rainy-nuScenes**, **Foggy-nuScenes**, and **Snowy-nuScenes**, by applying weather-specific scene transformations and atmospheric effects. We then generate intermediate weather-transition videos in which weather attributes progressively evolve while scene content, ego-motion, and background elements dynamically change across frames. Building on this framework, we introduce **nuScenes-WT** (Nuscenes Weather Transition Video Dataset), comprising twenty directional transition categories spanning sunny, rainy, foggy, snowy, and cloudy conditions. We further evaluate the generated videos for downstream object detection and semantic segmentation, establishing a unified generative-to-perceptual benchmark for transitional weather. Experimental results demonstrate that **WeatherShift** generates visually coherent weather transitions while preserving temporal scene dynamics, and that nuScenes-WT provides a challenging testbed for evaluating perception under continuously changing weather conditions.

1 INTRODUCTION

Autonomous driving has advanced substantially over the past decade, progressing from controlled demonstrations toward increasingly capable real-world systems (Zheng et al., 2026; Xiong et al., 2026). This progress has been driven by advances in learned perception, including multi-camera fusion, bird's-eye-view representations, 3D scene understanding, and increasingly integrated perception-and-planning models (Huang et al., 2026; Xiong et al., 2026). Large-scale benchmarks such as nuScenes (Caesar et al., 2020), Waymo (Sun et al., 2020), and BDD100K (Yu et al.) have further enabled this development by providing diverse driving scenes and annotations for object detection, segmentation, tracking, and scene understanding. However, these datasets do not fully capture the range and temporal variability of real-world driving conditions, with adverse weather remaining a major challenge for reliable visual perception (Qian et al., 2026; Song et al., 2026a).

Rain, fog, and snow affect the visual appearance and physical characteristics of driving scenes in fundamentally different ways (Cao et al., 2026; Wu et al., 2026). Rain introduces droplets and streaks with reduced visibility, fog attenuates scene contrast in a depth-dependent manner, and snow introduces moving particles while altering road-surface appearance and structure. These effects modify the image statistics and spatial cues used by perception models, degrading tasks such as object de-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tection and semantic segmentation (Song et al., 2026b; Pan et al., 2025). Prior work has addressed these challenges through adverse-weather benchmarks (Sakaridis et al., 2021; Bijelic et al., 2020; Zhou et al., 2026a), synthetic weather augmentation and physically based rendering (Sakaridis et al., 2018; Tremblay et al., 2021), and image restoration and domain-adaptation techniques (Liu et al., 2026; Rajagopalan & Patel, 2025). However, adverse weather is typically formulated as discrete conditions, with scenes categorized as clear, rainy, foggy, or snowy and perception evaluated independently at these fixed states (Lu et al., 2025a; Kulkarni et al., 2026).

Real-world weather is rarely static: a vehicle may transition from clear to rain, experience increasing precipitation, or return to clear conditions, while fog can similarly develop or dissipate as the vehicle moves through different regions (Madhavi et al., 2026a). We refer to these evolving conditions as transitional weather (Sun et al., 2022). Unlike isolated adverse-weather conditions, transitional weather involves continuously changing weather severity alongside evolving vehicle and scene dynamics (Kondapally et al., 2024b; 2026). Visibility, contrast, illumination, and weather artifacts vary across frames, while ego-motion and object motion alter the scene configuration (Madhavi et al., 2026a). These coupled changes challenge object consistency, temporal correspondence, tracking, motion estimation, and multi-frame fusion, making fixed-weather endpoint evaluation insufficient for characterizing video-based perception under changing weather. A key challenge in studying transitional weather is the lack of large-scale data capturing realistic weather evolution. Collecting real-world driving sequences spanning different weather types, transition directions, and intermediate severity levels is difficult because weather evolution cannot be readily controlled or reproduced. Simulation provides an alternative by enabling explicit control over weather parameters (Dosovitskiy et al., 2017; Sun et al., 2022); however, synthetic scenes may differ from real driving environments in geometry, materials, appearance, and sensor characteristics. These challenges motivate data-generation approaches that create controllable weather transitions directly from real driving sequences while preserving their underlying scene and motion dynamics.

Recent generative approaches have explored this direction by synthesizing intermediate weather states from clear and adverse versions of real driving imagery. In particular, image-level weather-transition generation has shown that intermediate conditions can be constructed through controlled interpolation, enabling datasets with multiple transition categories (Kondapally et al., 2024b; Madhavi et al., 2026b). However, these approaches typically operate on individual images or a static scene, varying weather severity while keeping the underlying scene content fixed. They therefore do not capture the coupled temporal dynamics of real driving sequences, where weather evolves concurrently with ego-motion, object motion, and changes in the surrounding environment. Consequently, image-level transitions provide limited representation of how perception systems respond when both weather and scene dynamics evolve simultaneously.

Table 1 highlights the limitations of existing approaches. To address these limitations, we introduce **WeatherShift**, a two-stage framework that generates temporally coherent weather transitions from real autonomous-driving videos. First, we construct frame-aligned weather-specific counterparts of real nuScenes sequences, using depth-guided atmospheric modeling for fog and appearance editing with particle-based effects for rainy and snowy conditions. These sequences provide adverse-weather endpoints while preserving the temporal structure of the original driving videos. Second, WeatherShift generates intermediate weather states using a temporally varying weather-severity schedule, allowing weather to evolve continuously while the underlying scene, ego-motion, and surrounding agents continue to change. Using this framework, we construct **nuScenes-WT**, a video benchmark comprising 20 directional transitions among five weather states: Sunny, Cloudy, Rainy, Foggy, and Snowy. Unlike earlier image-level transitional-weather datasets, nuScenes-WT captures weather evolution within moving driving scenes, providing temporally varying environmental conditions for evaluating both generation quality and downstream perception. Finally, we evaluate object detection, and semantic segmentation on the generated transitional-weather sequences. These experiments assess both the fidelity of the generated transitions and their ability to reveal perception challenges that are difficult to characterize using fixed-weather or static-scene datasets.

Our main contributions are summarized as follows:

1. **WeatherShift:** We propose a two-stage video generation framework for synthesizing temporally coherent weather transitions from real autonomous-driving videos, where weather severity evolves jointly with dynamic scene content and ego-motion.

Table 1: Comparison with representative works in driving scene generation, weather modeling, and transitional-weather perception. R→Rainy, C→Cloudy, S→Sunny, F→Foggy, ✓→Present, ✗→Not present, Tasks: Det→detection, Seg→segmentation, Rec→weather recognition, Fcst→weather forecasting, Cap→captioning, Res→restoration, Gen→generation only, Sim→simulated.

Method	Driving	Video	Weather	Transition	Temporal	Generated	Real/Sim	Tasks	Temporal Weather & Scene Evaluation
nuScenes	✓	✓	✓	✗	✓	✗	Real	Det, Track	✗
ACDC	✓	✗	✓	✗	✗	✗	Real	Seg	✗
Seeing Through Fog	✓	✗	✓	✗	✗	✗	Real	Det	✗
AWRaCLe	✗	✗	✓	✗	✗	✗	Sim	Res	✗
Foggy Cityscapes	✓	✗	✓	✗	✗	✓	Real	Seg	✗
Rain rendering	✓	✗	✓	✗	✗	✓	Real	Det, Seg	✗
SHIFT	✓	✓	✓	✓	✓	✓	Sim	Det, Seg	✗
WeatherEdit	✓	✓	✓	✗	✓	✓	Real	Gen	✗
WeatherCity	✓	✓	✓	✗	✓	✓	Real	Gen	✗
AIWD6	✓	✗	✓	✓	✗	✓	Real	Rec	✗
mSAT	✓	✗	✓	✓	✗	✓	Real	Det	✗
CaRS	✓	✗	✓	✓	✗	✓	Real	Seg	✗
TWFNet	✓	✓	✓	✓	✗	✓	Real	Fcst	✗
VLAAD-TW	✓	✗	✓	✓	✗	✓	Real	Cap	✗
Panacea	✓	✓	✓	✗	✓	✓	Real	Gen, Det	✗
MagicDrive-V2	✓	✓	✓	✗	✓	✓	Real	Gen	✗
Video-As-Prompt	✗	✓	✗	✗	✓	✓	Real	Gen	✗
WeatherShift (ours)	✓	✓	✓	✓	✓	✓	**Real**	**Det, Seg**	✓

2. **Extreme-weather video construction:** We develop *Rainy-nuScenes*, *Foggy-nuScenes*, and *Snowy-nuScenes* by applying weather-specific scene and atmospheric transformations to real nuScenes sequences, providing adverse-weather endpoints for controllable transition generation.
3. **nuScenes-WT:** We introduce a video benchmark comprising twenty directional weather-transition categories, extending transitional-weather modeling from static images to dynamically evolving driving scenes.
4. **Generative-to-perceptual evaluation:** We systematically evaluate the generated transitional-weather videos using object detection, and semantic segmentation, establishing a unified framework for assessing perception under continuously changing weather conditions.

2 Proposed Method

2.1 Overview

Let $V = \{I_t\}_{t=1}^{N}$ denote a clear-weather driving clip, where the scene content, object configurations, and camera viewpoint evolve over time. Our objective is to generate a transition clip $\tilde{V} = \{\tilde{I}_t\}_{t=1}^{N}$ in which the weather condition changes progressively from a source state to a target state while retaining the temporal dynamics of the original driving sequence. The proposed *WeatherShift* framework consists of two stages. First, for each extreme weather condition, we generate a frame-aligned reference clip from the original clear-weather (Sunny) sequence (Sec. 2.2). These reference clips provide the target appearance required for transition generation. Second, we derive a spatial weather mask by comparing the original and corresponding extreme-weather frames and progressively modulate its strength according to a temporal severity schedule (Sec. 2.3–2.4). When neither endpoint is the sunny state, the transition is constructed through an intermediate Sunny state by concatenating two transitions (Sec. 2.5). This formulation allows the same transition operator to generate multiple weather directions while maintaining the motion and scene evolution of the source driving clip.

2.2 Extreme-weather Video Generation

Transition generation requires a frame-aligned extreme-weather counterpart for each source clip (Yang et al., 2026a). We therefore construct weather-specific reference videos using the two branches illustrated in Figure 1. Fog is generated using depth-guided atmospheric scattering because its appearance depends strongly on scene depth, whereas rain, snow, and cloud are generated through prompt-driven appearance editing followed, where applicable, by weather-particle synthesis. Both branches operate on the complete driving sequence rather than independently editing isolated frames. We employ a background editing model with temporal consistency alignment (Qian et al., 2026), allowing weather-induced appearance changes to follow the underlying scene motion across frames. The resulting reference clip remains temporally aligned with the source sequence and can therefore be used for subsequent mask estimation and transition synthesis.

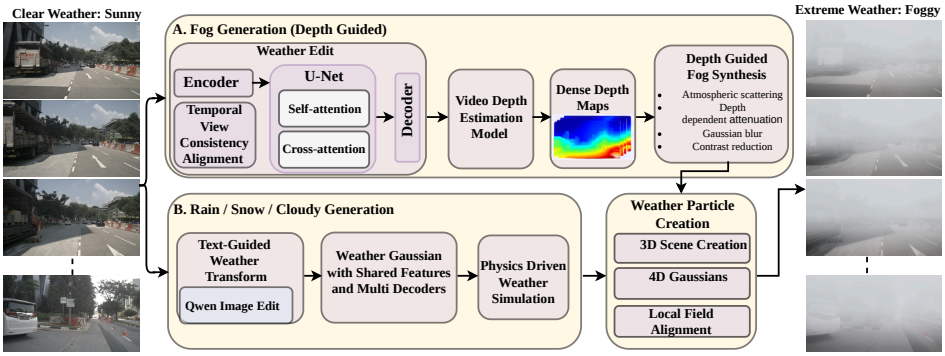

Figure 1: Overview of the proposed extreme-weather video generation framework. A clear-weather driving video is processed through two parallel branches: depth-guided fog synthesis using monocular depth and atmospheric scattering, and prompt-guided editing for rain, snow, and cloud, with particle generation for rain and snow. Both branches produce temporally aligned extreme-weather videos for subsequent transition generation while preserving consistency with the evolving scene.

Depth-guided fog synthesis. For each frame I_t, the background scene editor first produces an initial fog-conditioned frame J_t, modifying scene illumination, sky appearance, and global colour characteristics. In parallel, a video depth model estimates a dense depth map $d_t(x)$, and the atmospheric light A is estimated from the scene. We then model the depth-dependent attenuation of fog using the Beer–Lambert scattering formulation:

$$t_{\mathrm{r}}(x) = \exp!\left(-\beta, \hat{d} * t(x)\right), \qquad I^{\mathrm{fog}} * t(x) = t * \mathrm{r}(x) J_t(x) + (1 - t * \mathrm{r}(x))\, A, \tag{1}$$

where $\hat{d}_t(x)$ denotes the normalized depth and β controls the atmospheric extinction strength. Consequently, nearby regions with smaller depth retain more of their original appearance, whereas distant regions experience stronger attenuation toward the atmospheric light. This produces the characteristic depth-dependent visibility degradation of fog. The corresponding contrast attenuation can be expressed as $C_{\mathrm{obs}}(x) = C_0(x) \exp!\left(-\beta, \hat{d}_t(x)\right)$, where $C_0(x)$ and $C_{\mathrm{obs}}(x)$ denote the original and observed contrast, respectively. Thus, increasing β progressively reduces the visibility of distant scene structures.

Rain, snow, and cloud synthesis. For rain, snow, and cloud, we apply a text-guided image editing model (Wu et al., 2026) to each source frame using weather-specific prompts. The editing process modifies weather-dependent visual attributes while preserving the scene layout, objects, and geometric structure. In particular, rain and snow alter road and surface appearance and introduce weather-specific visual effects, whereas cloud primarily modifies sky appearance, illumination, and overall scene tone. For rain and snow, appearance editing alone is insufficient because these conditions introduce dynamic particles that partially occlude the scene. We therefore employ the particle-generation component of the weather editing model (Qian et al., 2026). The particles are represented in a reconstructed 3D scene and temporally evolved as a 4D Gaussian field under a physically motivated simulation before being rendered according to the scene geometry. The rendered particles are composited with the edited frame using depth-ordered alpha blending:

$$I_t^w(p) = \sum_{i \in \mathcal{N}(p)} c_i \alpha_i \prod_{j<i} (1 - \alpha_j) + \left[\prod_{i \in \mathcal{N}(p)} (1 - \alpha_i) \right] \tilde{J}_t(p), \tag{2}$$

where $\tilde{J}_t$ denotes the edited frame, c_i and α_i are the colour and opacity of particle i, and $\mathcal{N}(p)$ represents the particles projected onto pixel p, ordered according to depth. Rain and snow use their respective particle representations, while cloud requires no particle layer because its primary effect is a change in sky and illumination. This process produces an extreme-weather reference clip

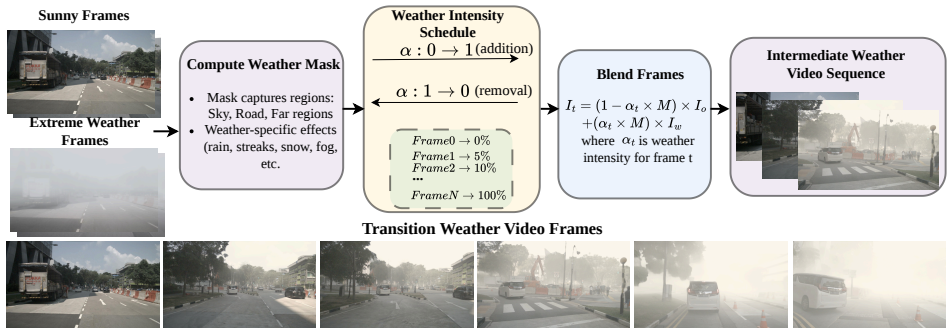

Figure 2: Architecture of transition generation between Sunny and an extreme-weather. A weather mask, computed by comparing each original frame with its corresponding extreme-weather frame, localizes weather effects to relevant regions rather than uniformly modifying the frame. A temporal severity schedule controls the blending between the two states: increasing severity produces forward transitions (Sunny→Rainy), while decreasing severity produces reverse transitions (Rainy→Sunny).

$V^w = \{I_t^w\}_{t=1}^N$. that is spatially and temporally aligned with the original sequence. Since the scene geometry and camera motion originate from the same driving clip, the reference preserves the temporal evolution of vehicles, roads, and other scene elements required for transition generation. The procedure for generating the extreme-weather reference videos is summarized in Algorithm 1.

2.3 Weather Mask Estimation

Weather effects are spatially non-uniform. For example, rain and snow may affect the sky, road, and image foreground through particles, fog progressively attenuates distant structures, and cloud primarily modifies the sky and illumination. At the same time, many scene regions, such as nearby vehicle surfaces and static structures, should remain largely unchanged. To localize these weather-induced changes, we compute a soft spatial mask by comparing each source frame with its aligned extreme-weather counterpart: $M_t^w = \phi!\left(I_t, I_t^w\right) \in [0,1]^{H\times W}$, where $\phi(\cdot)$ measures the normalized per-pixel appearance difference between the two aligned frames. A larger value of $M_t^w(x)$ indicates that pixel x is strongly affected by the target weather, whereas values close to zero indicate regions that should remain close to the original appearance. The use of an aligned reference is important because the mask represents weather-induced appearance differences at corresponding scene locations rather than differences caused by object motion or camera displacement. This spatially adaptive mask allows weather effects to develop selectively instead of uniformly modifying the entire frame.

2.4 Weather Severity Scheduling and Transition Synthesis

Given the weather mask M_t^w, we define a temporal severity function $\alpha : 1, \ldots, N \to [0,1]$, where α_t specifies the target-weather strength at frame t. The transition frame is synthesized as

$$\tilde{I}_t = (1 - \alpha_t M_t^w) \odot I_t + (\alpha_t M_t^w) \odot I_t^w, \tag{3}$$

where $\odot$ denotes element-wise multiplication, broadcast across colour channels. For $\alpha_t = 0$, the synthesized frame is identical to the original clear-weather frame, whereas $\alpha_t = 1$ produces the corresponding extreme-weather reference. Intermediate values progressively introduce the weather appearance according to the spatial mask. Consequently, a clear-to-adverse transition is obtained by increasing α_t, while the reverse transition is obtained by decreasing it. Unless otherwise specified, we use a linear schedule, $\alpha_t = \frac{t-1}{N-1}$, for a forward transition and $1 - \alpha_t$ for the reverse direction. More generally, the framework supports any monotonic severity schedule, allowing different rates of weather development to be modeled without changing the underlying transition operator. An important property of this formulation is that the weather severity is explicitly known for every generated frame. Hence, each frame can be associated with a continuous severity value α_t, enabling downstream perception performance to be analyzed as a function of weather intensity rather than only at discrete weather endpoints.

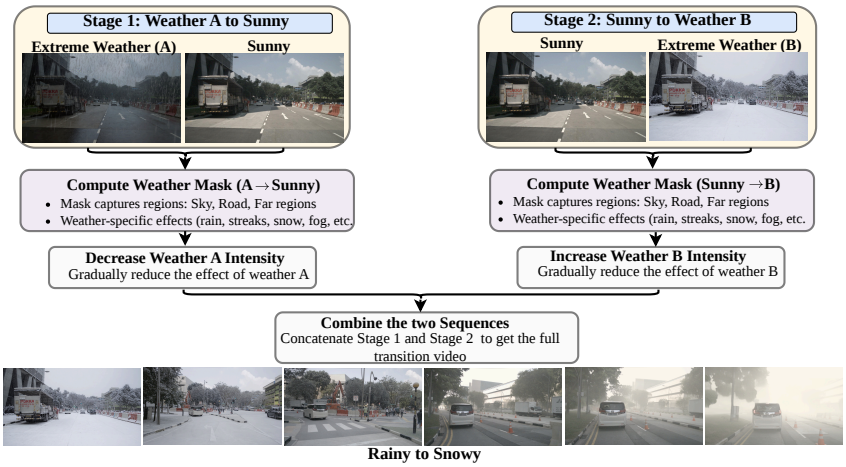

Figure 3: Architecture for transition generation between two non-Sunny weather conditions using Sunny as an intermediate state, following the sequence Weather A → Sunny → Weather B.

2.5 Composing Transitions Through the Clear State

The direct transition formulation naturally handles cases in which Sunny is one of the endpoints. However, directly interpolating between two adverse-weather reference clips is undesirable because the two weather effects may be visually and physically incompatible. For example, directly blending rain and snow can result in simultaneous or unnatural weather artefacts rather than a meaningful temporal transition. We therefore use Sunny as a common intermediate state for transitions between two non-Sunny weather conditions. Given source weather A and target weather B, the transition is decomposed as $A \rightarrow \textit{Sunny} \rightarrow B$. The first stage gradually decreases the severity of weather A from 1 to 0, thereby returning the sequence to its clear-weather state. The second stage starts from this clear state and progressively increases the severity of weather B from 0 to 1. Both stages use the same mask-based transition operator defined in Eq. 3. For example, *Snowy* → *Sunny* → *Rainy* is obtained by first removing the snowy appearance and subsequently introducing rain. Similarly, *Rainy* → *Sunny* → *Foggy* first reduces rain severity and then progressively increases fog severity. Because both stages pass through the same clear-weather state, the transition remains continuous at the junction. More importantly, this strategy avoids directly mixing unrelated weather effects and provides a unified mechanism for generating arbitrary weather-direction combinations. The procedure for generating the transitional-weather videos is summarized in Algorithm 2.

Nuscenes-WT. Applying the proposed transition strategy to the selected weather conditions produces the transition categories considered in our dataset. We distinguish between two types of transitions: (i) direct transitions in which Sunny is one of the endpoints, and (ii) composed transitions between non-Sunny weather conditions through the intermediate Sunny state. For each source driving clip, we retain the original clear-weather sequence, the corresponding extreme-weather sequence, and the generated transition sequence. All three sequences originate from the same underlying driving scene and preserve its temporal evolution. This design enables controlled evaluation of perception robustness under progressively changing weather conditions.

2.6 Downstream Perception Evaluation

We evaluate transitional-weather videos on downstream autonomous-driving tasks, including object detection and semantic segmentation. Since WeatherShift modifies weather appearance while preserving scene layout and object identities, source annotations can be retained where the task and transformation permit. Beyond final-weather endpoints, we evaluate performance across the severity parameter α_t to obtain degradation curves that characterize task behavior as weather progressively intensifies, providing a finer assessment than isolated clear- and extreme-weather evaluation.

Table 2: Nuscenes-WT dataset details

No. of transitions	No of clips in transition	Total no of Transition videos	Video length (sec)	FPS	No. of videos annotated	Source Dataset	Weather conditions Covered
20	39	20000	4	10	1000	Nuscenes	Sunny, cloudy, foggy, rainy, snowy

3 EXPERIMENTS

3.1 NUSCENES-WT DATASET

To generate nuScenes-WT, we construct a five-condition weather transition benchmark for AVs, built over real driving video rather than static images. We start from $1{,}000$ scenes of the nuScenes dataset (Caesar et al., 2020), drawn from the front camera, each of which comprises 39 frames; converting these to a frame rate of 10 fps yields clips of approximately 4 *seconds*. Using WeatherShift, every clear scene is first extended into its extreme-weather counterparts, namely Cloudy-, Rainy-, Foggy-, and Snowy-nuScenes. Traversing the severity schedule α_t (Eq. 3) between every ordered pair of these five conditions, either directly for the eight sunny-anchored pairs or through the clear state for the twelve adverse-to-adverse pairs, yields *twenty* weather transition states, namely *sunny to cloudy*, *cloudy to sunny*, *sunny to rainy*, *rainy to sunny*, *sunny to foggy*, *foggy to sunny*, *sunny to snowy*, *snowy to sunny*, together with the twelve transitions among *cloudy*, *rainy*, *foggy*, and *snowy*. Figure 6 from the Appendix shows the qualitative analysis of these transition states. Applying this construction to all $1{,}000$ scenes gives $20{,}000$ transition videos, each 39 frames long at 10 fps, for a total of $1{,}000 \times 20 \times 39 = 7{,}80{,}000$ frames across nuScenes-WT. Table 2 summarizes **nuScenes-WT** in terms of its weather conditions, transition states, scene and video statistics, frame rate, total frames, and dynamic scene content. Further, Table 4 from Appendix describes the extreme and transition weather implementation details, covering backbone, approach, epochs, and hyperparameters.

Evaluation Metrics. We evaluate scene and object preservation using CLIP-S, DINO-S, Object Preservation Recall, Matched Object IoU, and Object Count Error. CLIP-S measures the semantic similarity between corresponding original and generated frames using CLIP embeddings $\text{CLIP-S} = \frac{1}{N}\sum_{i=1}^{N}\cos(\mathbf{e}_i^{orig}, \mathbf{e}_i^{gen})$. DINO-S measures visual and structural similarity using DINO embeddings: $\text{DINO-S} = \frac{1}{N}\sum_{i=1}^{N}\cos(\mathbf{d}_i^{orig}, \mathbf{d}_i^{gen})$. Object Preservation Recall measures the fraction of original objects that are successfully preserved in the generated frames: $\text{OPR} = \frac{N_{\text{matched}}}{N_{\text{original}}}$. Matched Object IoU measures the spatial overlap between the bounding boxes of matched objects: $\text{IoU} = \frac{|B^{orig} \cap B^{gen}|}{|B^{orig} \cup B^{gen}|}$. Object Count Error measures the relative difference between the number of objects in the original and generated frames: $\text{OCE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_i^{orig} - C_i^{gen}|}{C_i^{orig}}$. Higher CLIP-S, DINO-S, Object Preservation Recall, and Matched Object IoU indicate better preservation, while lower Object Count Error indicates better consistency.

3.2 COMPARISON WITH BASELINES

Table 3 presents the quantiative evaluation of WeatherShift against three strong general-purpose video generation and editing baselines, OmniFlash 1.1 (OmniFlash, 2026), Seedance 2.0 (ByteDance, 2026), and Kling 3.0 Mini (Kuaishou Technology, 2026), prompted with the same weather-specific instructions used by our appearance-editing stage, and report object preservation, CLIP-S, DINO-S, matched-object IoU, and object count error. These metrics disentangle weather transformation quality from scene preservation: CLIP-S measures alignment with the target weather description, DINO-S measures background feature similarity outside the weather-affected regions, object preservation is the fraction of source-frame objects still detected after editing, matched-object IoU measures how much a preserved object's box drifts, and object count error catches insertions or removals that per-object matching can miss. WeatherShift achieves the best score across all metrics, while existing prompt-based editors prioritize prompt fidelity and per-frame quality without explicitly preserving source objects. In contrast, WeatherShift preserves scene content by attenuating radiance along existing scene rays and depth-compositing weather particles, avoiding object identity, position, and count drift. Further, Figure 4 presents a qualitative comparison between WeatherShift



Figure 4: Qualitative comparison with OmniFlash 1.1 on Nuscenes-WT where WeatherShift preserves object identity, position, and count across the edit, whereas OmniFlash regenerates scene content and drops or displaces objects under the same weather prompt.

Table 3: Comparison of WeatherShift with general-purpose video generation baselines on object preservation, CLIP-S, DINO-S, matched-object IoU, and object count error (↓ unless noted), averaged across all the transitions.

Models	CLIP-S (mean)	DINO-S (mean)	Object Preservation Recall	Matched Object IoU	Object Count Error
OmniFlash-1.1	0.81	0.64	30.15%	0.56	0.71
Seedance 2.0	0.82	0.70	35.20%	0.75	0.74
Kling 3.0 Omni	0.81	0.71	33.30%	0.78	0.75
Ours	**0.85**	**0.76**	**81.53%**	**0.94**	**0.20**

and OmniFlash 1.1 on the nuScenes-WT dataset under the same weather prompts. WeatherShift preserves the identity, position, and count of objects across the weather transition, whereas OmniFlash regenerates scene content, resulting in missing or displaced objects. This highlights the advantage of WeatherShift's source-guided formulation, which applies weather effects to the existing scene rather than regenerating scene content, thereby maintaining geometric and semantic consistency in autonomous-driving sequences. Figure 5 presents the object detection and segmentation results of WeatherShift on nuScenes-WT. The results illustrate how the predicted objects and semantic regions evolve as the weather condition changes throughout the driving sequence. Despite the continuously changing weather, WeatherShift maintains the underlying scene structure and object locations, enabling the perception models to track consistent scene content across transition states.

4 Limitations and Future Work

While **WeatherShift** provides a unified framework for generating temporally evolving weather transitions, several aspects offer opportunities for future development:

1. **Camera coverage.** Transitions are generated for the front camera, which carries most of the evidence used by driving perception. Extending them to the six-camera surround view is our main

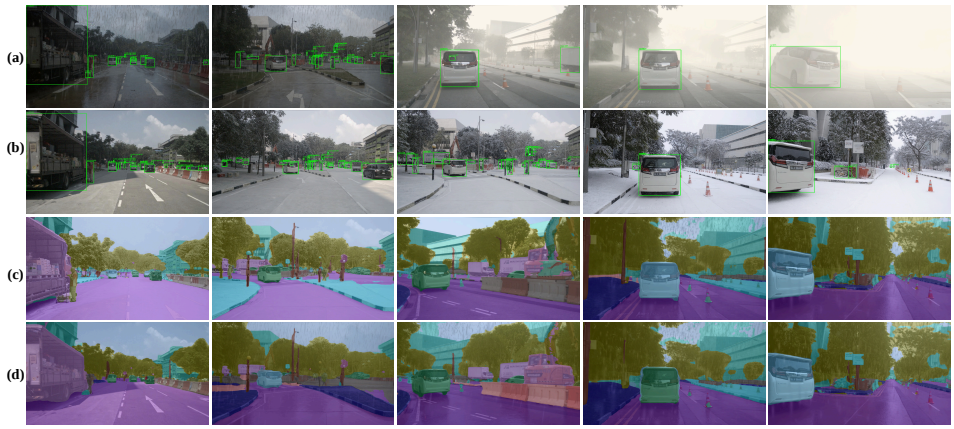

Figure 5: Object detection and semantic segmentation performance of WeatherShift on Nuscenes-WT. Results demonstrate perception performance under progressively changing weather conditions. (a) Rainy to Foggy (b)Sunny to Snowy (c) Snowy to Rainy (d) Sunny to Rainy transitions.

next step. The same schedule drives every camera, and enforcing agreement between overlapping views turns temporal consistency into spatio-temporal consistency, as surround-view perception.

2. **Sensing and task coverage.** We evaluate camera-based detection, and segmentation, the tasks most affected by appearance shift. Since the schedule is modality-agnostic and the pipeline needs only monocular video, driving lidar and radar degradation with the same α_t, applying it to further compound shifts such as weather with day-to-night, and using the per-frame severity label for tracking, planning, and test-time adaptation all follow from the same formulation.
3. **Calibration of severity.** Fog severity is physically parameterized by extinction and visibility, whereas rain, snow, and cloud use the schedule α_t. Extending these conditions to physical precipitation rates would enable meteorological calibration and comparison with field measurements.

5 Conclusion

In this paper, we presented **WeatherShift**, a framework for generating temporally evolving weather transitions in real-world driving videos, together with **nuScenes-WT**, the resulting benchmark. Starting from nuScenes driving clips, we construct frame-aligned extreme-weather counterparts using depth-guided Beer–Lambert scattering for fog and prompt-driven appearance editing with particle compositing for rain and snow, while cloud conditions are modeled through appearance transformation. A spatial weather mask and temporal severity schedule α_t then enable weather intensity to evolve continuously throughout the clip while preserving the underlying scene dynamics. Transitions between two adverse conditions are further constructed through the clear-weather state, providing a unified mechanism for generating diverse weather directions. **WeatherShift** provides three key advantages over fixed-condition and static-background weather data. First, weather severity is continuous and explicitly known for every frame, enabling perception robustness to be analyzed as a function of weather intensity rather than only across discrete weather categories. Second, scene content, object motion, and ego-motion evolve naturally throughout the transition, providing a more realistic setting for evaluating perception under dynamically changing conditions. Third, the generation process preserves the underlying scene geometry and temporal structure of the source driving sequences, allowing the resulting data to support both training and evaluation. Experiments across object detection, and semantic segmentation demonstrate that perception performance can progressively degrade during moving-scene weather transitions, revealing robustness characteristics that fixed-endpoint evaluation may overlook. Furthermore, training with **nuScenes-WT** improves performance over the corresponding baselines, with gains of 0.03 in CLIP-S, 0.05 in DINO-S, 46.33 in object-preservation recall, and 0.16 in matched-object IoU, while reducing object count error by 0.51, demonstrating the effectiveness of **WeatherShift** for generating faithful and scene-consistent weather transitions. Future work will extend the framework toward multi-camera weather transition generation with cross-view consistency.

AI Use Statement. Generative AI tools were used during the preparation of this manuscript for language refinement, grammar correction, and improving the clarity and readability of the text.

Ethics Statement. NuScenes-WT is derived entirely from the publicly released nuScenes (Caesar et al., 2020) autonomous-driving dataset and introduces no new recordings or collection of personal data. We inherit the privacy safeguards and anonymization of the original dataset, including the blurring of faces and license plates. Our weather-generation pipeline modifies the appearance of existing driving scenes without introducing new identifiable individuals, vehicles, or locations. WeatherShift and nuScenes-WT are intended for research on improving autonomous-driving perception under adverse and transitional weather conditions. As with any synthetic benchmark, models trained on nuScenes-WT should undergo additional validation and safety testing before real-world deployment.

Reproducibility Statement. We provide detailed descriptions of the WeatherShift framework, including extreme-weather video generation, weather-mask estimation, severity scheduling, transition synthesis, and the construction of nuScenes-WT. We report the dataset composition, transition settings, evaluation metrics, and experimental protocols to facilitate reproduction. We plan to release the implementation, configuration files, and generated dataset or the corresponding generation pipeline, subject to the licensing and usage conditions of the underlying nuScenes data and the third-party models used in the framework.

REFERENCES

Yuxuan Bian, Xin Chen, Zenan Li, Tiancheng Zhi, Shen Sang, Linjie Luo, and Qiang Xu. Video-as-prompt: Unified semantic control for video generation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 114558–114583, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/ba8d1b46292c5e82cbfb3b3dc3b968af-Paper-Conference.pdf.

Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11682–11692, June 2020.

ByteDance. Seedance 2.0: Official launch. https://seed.bytedance.com/en/blog/official-launch-of-seedance-2-0, 2026. Accessed 2026-09-25.

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11618–11628, 2020. doi: 10.1109/CVPR42600.2020.01164.

Shilei Cao, Hehai Lin, Jiashun Cheng, Yang Liu, Guowen Li, Xuehe Wang, Juepeng Zheng, Haoyuan Liang, Meng Jin, Chengwei Qin, et al. Task-adaptive parameter-efficient fine-tuning for weather foundation models. In *International Conference on Learning Representations*, volume 2026, pp. 47078–47116, 2026.

Rui Chen, Zehuan Wu, Yichen Liu, Yuxin Guo, Jingcheng Ni, Haifeng Xia, and Siyu Xia. Unimlvg: Unified framework for multi-view long video generation with comprehensive control capabilities for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 25453–25463, October 2025.

Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pp. 1–16. PMLR, 2017.

Ruiyuan Gao, Kai Chen, Bo Xiao, Lanqing Hong, Zhenguo Li, and Qiang Xu. Magicdrive-v2: High-resolution long video generation for autonomous driving with adaptive control. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 28135–28144, October 2025.

Junhao Ge, Zuhong Liu, Longteng Fan, Yifan Jiang, Jiaqi Su, Yiming Li, Zhejun Zhang, and Siheng Chen. Unraveling the effects of synthetic data on end-to-end autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 28859–28869, October 2025.

Zhao Huang, Bowen Zhang, Zhongzhu Li, and Di Lin. Occdriver: Future occupancy guided dual-branch trajectory planner in autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 141940–141962, 2026.

Madhavi Kondapally, K Naveen Kumar, and C Krishna Mohan. Object detection in transitional weather conditions for autonomous vehicles. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2024a.

Madhavi Kondapally, K. Naveen Kumar, Chalavadi Vishnu, and C. Krishna Mohan. Towards a transitional weather scene recognition approach for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 25(6):5201–5210, 2024b. doi: 10.1109/TITS.2023.3331882.

Madhavi Kondapally, K Naveen Kumar, and C Krishna Mohan. Eyes on the road, words in the changing skies: Vision-language assistance for autonomous driving in transitional weather. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=PCEDvdVJon.

Kuaishou Technology. Kling 3.0. https://picsart.com/ai-models/kling-3-0/, 2026. Accessed 2026-09-25. evaluates the "3.0 Mini" variant; no technical report was found for this specific version at the time of writing.

Ashutosh Kulkarni, Prashant W. Patil, Santosh Kumar Vipparthi, Subrahmanyam Murala, and Balasubramanian Raman. Dtmir-pro: Domain translation with prompt-based latent-space generalization for multi-weather image restoration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3846–3856, March 2026.

Ao Liang, Lingdong Kong, Tianyi Yan, Hongsi Liu, Yu Yang, Ziqi Huang, Wei Yin, Jialong Zuo, Yixuan Hu, Dekai Zhu, Dongyue Lu, Youquan Liu, Guangfeng Jiang, Linfeng Li, Xiangtai Li, Long Zhuo, Lai Xing Ng, Benoit R. Cottereau, Changxin Gao, Liang Pan, Wei Tsang Ooi, and Ziwei Liu. Worldlens: Full-spectrum evaluations of driving world models in real world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 36385–36399, June 2026.

Beibei Lin, Yeying Jin, Wending Yan, Wei Ye, Yuan Yuan, Shunli Zhang, and Robby T Tan. Nighttrain: Nighttime video deraining via adaptive-rain-removal and adaptive-correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 3378–3385, 2024.

Yuetong Liu, Yunqiu Xu, Yang Wei, Xiuli Bi, and Bin Xiao. Clear nights ahead: Towards multi-weather nighttime image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7395–7403, 2026.

Xin Lu, Jie Xiao, Yurui Zhu, and Xueyang Fu. Continuous adverse weather removal via degradation-aware distillation. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 28113–28123, June 2025a.

Yifan Lu, Xuanchi Ren, Jiawei Yang, Tianchang Shen, Zhangjie Wu, Jun Gao, Yue Wang, Siheng Chen, Mike Chen, Sanja Fidler, and Jiahui Huang. Infinicube: Unbounded and controllable dynamic 3d driving scene generation with world-guided video models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27272–27283, October 2025b.

Kondapally Madhavi, K Naveen Kumar, C Krishna Mohan, and Sobhan Babu. Cars: A causal intervention segmentation framework and benchmark dataset for autonomous driving under transitional weather conditions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4046–4056, March 2026a.

Kondapally Madhavi, K. Naveen Kumar, and C. Gayathri. Twfnet: Introducing transitional weather conditions for autonomous driving with a spatio-temporal forecasting network. *Pattern Recognition*, 171:112154, 2026b. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2025.112154.

Valentina Muşat, Ivan Fursa, Paul Newman, Fabio Cuzzolin, and Andrew Bradley. Multi-weather city: Adverse weather stacking for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2906–2915, 2021.

OmniFlash. Omniflash 1.1. https://omniflash.ai/omni-1-1-flash, 2026. Accessed 2026-09-25. Product/version identity not independently verified; confirm against the actual API or service used for evaluation.

Yuwen Pan, Rui Sun, Wangkai Li, and Tianzhu Zhang. Exploring weather-aware aggregation and adaptation for semantic segmentation under adverse conditions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13952–13962, October 2025.

Chenghao Qian, Wenjing Li, Yuhu Guo, and Gustav Markkula. Weatheredit: Controllable weather editing with 4d gaussian field. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 8511–8519, 2026.

Sudarshan Rajagopalan and Vishal M Patel. Awracle: All-weather image restoration using visual in-context learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 6675–6683, 2025.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10765–10775, 2021.

Khaled M. Seyam, Julian Wiederer, Markus Braun, and Bin Yang. Svs-gan for semantic synthesis of traffic videos for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 8650–8659, March 2026.

Yifan Shen, Peiyuan Zhu, Zijian Li, Shaoan Xie, Namrata Deka, Zongfang Liu, Zeyu Tang, Guangyi Chen, and Kun Zhang. Controllable video generation with provable disentanglement. In *International Conference on Learning Representations*, volume 2026, pp. 139053–139080, 2026.

Chen Shi, Shaoshuai Shi, Xiaoyang Lyu, Chunyang Liu, Kehua Sheng, Bo Zhang, and Li Jiang. Unisplat: Unified spatio-temporal fusion via 3d latent scaffolds for dynamic driving scene reconstruction. In *International Conference on Learning Representations*, volume 2026, pp. 125218–125235, 2026.

Changhoon Song, Teng-Yuan Chang, and Youngjoon Hong. Extreme weather nowcasting via local precipitation pattern prediction. In *International Conference on Learning Representations*, volume 2026, pp. 109146–109177, 2026a.

Youngjun Song, Hyeongyu Kim, and Dosik Hwang. Cd-buffer: Complementary dual-buffer framework for test-time adaptation in adverse weather object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15050–15059, June 2026b.

Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2443–2451. IEEE, 2020.

Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. Shift: a synthetic driving dataset for continuous multi-task domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Maxime Tremblay, Shirsendu Sukanta Halder, Raoul de Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. *International Journal of Computer Vision*, 129(2):341–360, 2021. doi: 10.1007/s11263-020-01366-3.

Yuqing Wen, Yucheng Zhao, Yingfei Liu, Fan Jia, Yanhui Wang, Chong Luo, Chi Zhang, Tiancai Wang, Xiaoyan Sun, and Xiangyu Zhang. Panacea: Panoramic and controllable video generation for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6902–6912, 2024.

Wenhua Wu, Huai Guan, Zhe Liu, and Hesheng Wang. Weathercity: Urban scene reconstruction with controllable multi-weather transformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 40949–40958, June 2026.

Tianze Xia, Yongkang Li, Lijun Zhou, Jingfeng Yao, Kaixin Xiong, Haiyang Sun, Bing Wang, Kun Ma, Guang Chen, Hangjun Ye, et al. Drivelaw: Unifying planning and video generation in a latent driving world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 39701–39712, 2026.

Kaixin Xiong, Xiangyu Guo, Fang Li, Sixu Yan, Gangwei Xu, Lijun Zhou, Long Chen, Haiyang Sun, Bing Wang, Kun Ma, et al. Recogdrive: A reinforced cognitive framework for end-to-end autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 157518–157556, 2026.

Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. Language-driven all-in-one adverse weather removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24902–24912, June 2024.

Xuemeng Yang, Licheng Wen, Tiantian Wei, Yukai Ma, Jianbiao Mei, Xin Li, Wenjie Lei, Daocheng Fu, Pinlong Cai, Min Dou, Liang He, Yong Liu, Botian Shi, and Yu Qiao. Drivearena: A closed-loop generative simulation platform for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26933–26943, October 2025a.

Yongqi Yang, Huayang Huang, Xu Peng, Xiaobin Hu, Donghao Luo, Jiangning Zhang, Chengjie Wang, and Yu Wu. Towards one-step causal video generation via adversarial self-distillation. In *International Conference on Learning Representations*, volume 2026, pp. 34858–34876, 2026a.

Zhenya Yang, Zhe Liu, Yuxiang Lu, Liping Hou, Chenxuan Miao, Siyi Peng, Bailan Feng, Xiang Bai, and Hengshuang Zhao. Geniedrive: Towards physics-aware driving world model with 4d occupancy guided video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35680–35690, June 2026b.

Zhuoran Yang and Yanyong Zhang. Consisdrive: Identity-preserving driving world models for video generation by instance mask. *arXiv preprint arXiv:2602.03213*, 2026.

Zhuoran Yang, Xi Guo, Chenjing Ding, Chiyu Wang, Wei Wu, and Yanyong Zhang. Instadrive: Instance-aware driving world models for realistic and consistent video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 25410–25420, October 2025b.

Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. *The International Journal of Robotics Research*, 39(1):1–18.

Kaiwen Zhang, Zhenyu Tang, Xiaotao Hu, Xingang Pan, Xiaoyang Guo, Yuan Liu, Jingwei Huang, Li Yuan, Qian Zhang, Xiao-Xiao Long, Xun Cao, and Wei Yin. Epona: Autoregressive diffusion world model for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27220–27230, October 2025.

Weicheng Zheng, Xiaofei Mao, Nanfei Ye, Pengxiang Li, Kun Zhan, Xianpeng Lang, and Hang Zhao. Driveagent-r1: Advancing vlm-based autonomous driving with active perception and hybrid thinking. In *International Conference on Learning Representations*, volume 2026, pp. 125576–125610, 2026.

Xin Zhou, Dingkang Liang, Sifan Tu, Xiwu Chen, Yikang Ding, Dingyuan Zhang, Feiyang Tan, Hengshuang Zhao, and Xiang Bai. Hermes: A unified self-driving world model for simultaneous 3d scene understanding and generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27817–27827, October 2025.

Xingcheng Zhou, Xuyuan Han, Feng Yang, Yunpu Ma, Volker Tresp, and Alois Knoll. Opendrivevla: Towards end-to-end autonomous driving with large vision language action model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13782–13790, 2026a.

Yang Zhou, Hao Shao, Letian Wang, Zhuofan Zong, Hongsheng Li, and Steven Waslander. Driving gen: A comprehensive benchmark for generative video world models in autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 103502–103524, 2026b.

Benjin Zhu, Xiaogang Wang, and Hongsheng Li. Consistentcity: Semantic flow-guided occupancy dit for temporally consistent driving scene synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26382–26392, October 2025.

Ziyue Zhu, Zhanqian Wu, Zhenxin Zhu, Lijun Zhou, Haiyang Sun, Bing Wang, Kun Ma, Guang Chen, Hangjun Ye, Jin Xie, et al. Worldsplat: Gaussian-centric feed-forward 4d scene generation for autonomous driving. In *International Conference on Learning Representations*, volume 2026, pp. 68258–68277, 2026.

A Appendix

A.1 Related Work

Perception under Adverse Weather. Adverse-weather perception has traditionally been studied using real and synthetic datasets containing discrete weather conditions. Large driving benchmarks such as BDD100K (Yu et al.), nuScenes (Caesar et al., 2020), and Waymo Open Dataset (Sun et al., 2020) provide large-scale driving data covering diverse environmental conditions, but are dominated by clear conditions. Dedicated benchmarks fill this gap with real adverse weather capture. ACDC (Sakaridis et al., 2021) pairs fog, night, rain, and snow images with clear-weather correspondences for semantic segmentation, while Seeing Through Fog dataset (Bijelic et al., 2020) records multimodal sensor data in heavy fog, snow, and rain for object detection. Recent work adapts segmentation models with weather-aware feature aggregation (Pan et al., 2025), adapts detectors at test time under adverse weather (Song et al., 2026b), and restores degraded inputs with all-in-one models that remove several weather types with a single network (Yang et al., 2024; Rajagopalan & Patel, 2025; Kulkarni et al., 2026; Liu et al., 2026), including video deraining (Lin et al., 2024). Across these datasets and methods, each image or sequence carries a single, fixed weather condition. None of them evaluates or trains on weather whose severity changes within a sequence, which is the setting we target. Image-level weather synthesis cannot capture temporal phenomena such as gradually emerging precipitation, changing atmospheric visibility, or evolving weather appearance while maintaining consistent object motion and scene dynamics.

Synthetic adverse-weather data. Because real adverse-weather data is scarce, many works synthesize it. Physics-based rendering adds fog to Cityscapes using estimated depth and the atmospheric scattering model (Sakaridis et al., 2018), and renders rain and fog onto Cityscapes, KITTI, and nuScenes, releasing Rainy- and Foggy-nuScenes (Tremblay et al., 2021). Both apply weather to individual frames at a small set of fixed intensities. Simulators give full parametric control: CARLA (Dosovitskiy et al., 2017) exposes weather settings directly, and SHIFT (Sun et al., 2022) uses this control to render driving sequences in which cloudiness, rain, fog, and time of day shift continuously, the closest existing setting to ours. Its scenes, however, are rendered, so models evaluated on SHIFT face a sim-to-real gap. More recently, reconstruction-based methods edit weather on real scenes through a 3D or 4D representation, such as a 4D Gaussian field (Qian et al., 2026) or a reconstructed urban scene with controllable multi-weather transformation (Wu et al., 2026). These methods produce view-consistent weather but target fixed conditions rather than transitions, and they are not evaluated as training data for perception. We use WeatherEdit (Qian et al., 2026) to restyle backgrounds when constructing our extreme endpoints, and build transitions on top of it.

Transitional Weather for Autonomous Driving. A growing line of work treats transitions between weather states as a perception problem. It has introduced transitional weather scene recog-

nition (Kondapally et al., 2024b; Sun et al., 2022), object detection (Kondapally et al., 2024a), semantic and instance segmentation with a sixteen-category benchmark generated by interpolating the latent representations of stable-weather images (Madhavi et al., 2026a), forecasting of upcoming weather states from image sequences (Madhavi et al., 2026b), and vision-language assistance that describes scenes under changing weather (Kondapally et al., 2026). These works establish that intermediate weather states degrade perception and that they can be generated at scale. Their data, however, is built on a *fixed background*: each transition modifies weather-related appearance over a single static scene, including when the transition is presented as an image sequence (Madhavi et al., 2026b). Scene content, surrounding agents, and ego-motion therefore do not change during a transition. Our proposed WeatherShift method removes this restriction by generating transitions over real driving video, so the weather and the scene evolve together.

Driving video generation and world models. Generative models of driving video have advanced quickly. Controllable generators produce multi-view or panoramic video conditioned on layouts, trajectories, and text (Wen et al., 2024; Gao et al., 2025; Chen et al., 2025), improve the consistency of individual instances across frames (Yang et al., 2025b; Yang & Zhang, 2026), and synthesize video from semantic or occupancy representations (Seyam et al., 2026; Zhu et al., 2025). World models couple generation with prediction, understanding, or planning (Zhang et al., 2025; Zhou et al., 2025; Yang et al., 2026b; Xia et al., 2026), support closed-loop simulation (Yang et al., 2025a), and extend to controllable 4D scenes (Lu et al., 2025b; Zhu et al., 2026; Shi et al., 2026). Dedicated benchmarks now evaluate these models for visual quality and downstream use (Zhou et al., 2026b; Liang et al., 2026), and synthetic data from such models has been shown to affect end-to-end driving in measurable ways (Ge et al., 2025). Beyond autonomous driving, video-generation methods provide semantic and disentangled control (Bian et al., 2026; Shen et al., 2026) and efficient causal generation (Yang et al., 2026a). However, these methods typically treat weather as a clip-level condition, fixed through a prompt or style input, rather than modeling its severity as a temporal trajectory within a clip. In contrast, WeatherShift edits recorded nuScenes sequences rather than generating new scenes, preserving scene geometry and alignment with the original annotations.

Limitations of existing works. Despite progress in driving world models, adverse-weather synthesis, controllable video generation, and weather-robust perception, several gaps remain. *(i) Discrete weather evolution:* Existing weather benchmarks, restoration, and rendering methods typically model only discrete weather conditions or severity levels (Sakaridis et al., 2021; Yang et al., 2024; Sakaridis et al., 2018; Tremblay et al., 2021). *(ii) Simulation dependence:* Continuous weather variation along driving trajectories often relies on simulated data, introducing a potential sim-to-real gap (Sun et al., 2022). *(iii) Static scene transitions:* Existing transitional-weather methods vary weather severity on real imagery but largely retain a fixed scene, limiting simultaneous changes in ego-motion and scene content (Madhavi et al., 2026a;b). *(iv) Weather-static video generation:* Existing driving-video generators control scene layout and motion but generally maintain a fixed weather condition within a clip, rather than modeling a temporal transition between source and target weather states (Wen et al., 2024; Gao et al., 2025). Our work addresses the intersection of these areas by generating *temporally evolving weather transitions in driving videos*. Rather than imposing a fixed background or independently modifying individual frames, we aim to preserve the temporal structure of driving scenes while enabling the weather condition to progressively evolve between source and target states. Furthermore, we evaluated downstream detection, and semantic segmentation on the resulting benchmark nuscenes-WT.

A.2 Preliminaries

A.2.1 Transitional Weather Conditions

Weather transitions are common in the real world and pose a challenge for autonomous vehicle perception (AVT), particularly in rain, clouds, and fog. Existing datasets typically capture stable weather conditions (e.g., cloudy or rainy), but real-world transitions, such as from sunny to rainy, involve varying intensities. For example, transitioning from sunny to rainy involves intensities ranging from light to heavy rain (Sun et al., 2022). Recent studies highlight challenges posed by continuous weather shifts for learning systems, underscoring the need for datasets that capture dynamic weather changes (Sun et al., 2022).

Importance of addressing weather transitions in autonomous vehicles. Transitional weather conditions significantly impact scene perception performance in autonomous driving. Although existing datasets such as Foggy Cityscapes (Sakaridis et al., 2018) and MultiWeatherCity (Muşat et al., 2021) capture adverse weather scenarios, they mainly focus on discrete weather conditions such as heavy rain or intense fog. On the other hand, real-world driving environments often involve gradual weather transitions (for example, foggy to rainy or rainy to sunny). These transitions introduce additional complexities that go beyond those encountered in extreme weather conditions.

Unique challenges posed by transitional weather. Existing AVT datasets and models inadequately address critical challenges associated with weather transitions: *(i)* unlike static weather states, transitions involve dynamic changes in visibility and illumination, which can disrupt perception models trained on fixed conditions; *(ii)* autonomous systems rely on stable environmental features for perception, however weather transitions introduce inconsistencies in visual cues, leading to misclassification or unreliable predictions.

A.3 Video Interpolation and Dynamic Scene Navigation

A separate line of challenge concerns *how* intermediate weather states should be produced, and over what kind of scene. Prior transitional-weather datasets synthesize intermediate states through image-level interpolation, generating a sequence of increasingly severe frames over a single, frozen background (Sun et al., 2022; Madhavi et al., 2026a). This is a natural starting point, but it factors weather transition and scene motion as though they were independent, when in a moving vehicle they are not: the ego-vehicle continuously *navigates* the scene while weather evolves, so surrounding agents, road geometry, and viewpoint change frame to frame at the same time as visibility and illumination do.

From image interpolation to video-level transitions. Interpolating between two encoded images produces a smooth path in appearance space, but decoding each step independently gives no guarantee that adjacent frames remain consistent with one another; extending this recipe to video by decoding many such steps in sequence compounds the problem, since nothing in the interpolation ties one decoded frame to the next. A dynamic driving scene additionally requires that this consistency hold jointly with motion: a transition sequence must look like a single continuous drive, not a slide show of independently plausible frames. This motivates treating weather transition as a *video* generation problem, in which an entire clip is edited as a temporally coherent unit, rather than as repeated single-image generation.

Dynamic scene navigation. Because the ego-vehicle is moving, a valid transition must also preserve *scene dynamics*: the same road layout, the same set of surrounding vehicles and pedestrians, and the same camera trajectory that the original clip contains, with only the weather changing on top of them. A transition built over a frozen background cannot express this, since the vehicle never proceeds anywhere; it can only show a static viewpoint growing foggier or wetter. Modeling weather transitions over real, moving driving footage instead requires every synthesized frame to remain aligned with the scene the vehicle is actually navigating, so that changes in visibility or precipitation are attributable to weather and not to an inconsistent or frozen background.

Limitations of existing models in handling transitional weather. Most AVT models are trained on datasets with extreme or static weather conditions, making them ill-equipped to handle gradual transitions over moving scenes. Their primary limitations include: *(i)* models trained on clear or extreme weather conditions lack the adaptability to generalize across intermediate weather changes; *(ii)* existing models often rely on static images, or short sequences over a fixed background, missing the gradual evolution of visual features together with scene motion over time; *(iii)* during transitions, environmental features may not match any single weather category, causing prediction errors; *(iv)* because scene content and viewpoint are held fixed in existing constructions, none of these models are ever exposed to weather changing while the vehicle navigates, which is the setting perception failures in deployment actually occur in. These limitations reduce model robustness in real-world deployments, where smooth weather transitions over a continuously changing scene are the norm rather than the exception. Addressing these gaps, jointly, is what motivates WeatherShift and nuscenes-WT in the main text.

A.4 Additional Methodology Details

The proposed WeatherShift framework consists of two main stages: extreme weather reference generation and transitional weather generation. The complete procedure for generating extreme weather references, including the different processing strategies for sunny, cloudy, rainy, foggy, and snowy conditions, is provided in Algorithm 1. The procedure for generating temporally smooth weather transitions between two weather conditions is described in Algorithm 2. It covers the use of a weather change mask and temporal severity scheduling to progressively reduce or increase the target weather intensity. When Sunny is not one of the endpoint conditions, the transition is generated through Sunny as an intermediate weather state, as detailed in the algorithm.

Algorithm 1: Extreme Weather Reference Generation

Require: Original image/frame I, weather condition W
Ensure: Extreme weather image I_{ext}

$W \in \{\text{Sunny, Cloudy, Rainy, Foggy, Snowy}\}$
if $W =$ Foggy **then**
  Generate initial foggy image using WeatherEdit Background Scene Editing Model
  Generate corresponding depth map D for I
  Estimate atmospheric light A
  Set high fog parameter β
  Compute transmission:

$$T(x, y) = \exp(-\beta D(x, y))$$

  Generate depth-aware foggy image:

$$I_{\text{ext}}(x, y) = T(x, y)I(x, y) + (1 - T(x, y))A$$

else
  if $W =$ Cloudy **then**
    Use Cloudy weather prompt
  else if $W =$ Rainy **then**
    Use Rainy weather prompt
  else if $W =$ Snowy **then**
    Use Snowy weather prompt
  else if $W =$ Sunny **then**
    Use original Sunny image as the reference
  end if
  Generate weather-transformed image using Qwen Image Edit
  Pass the generated image to the WeatherEdit Particle Addition Model
  Add the corresponding weather particles
  Obtain the extreme weather image I_{ext}
end if
return I_{ext}

A.4.1 Rationale of the Proposed Approach

Why not latent interpolation? Generating weather transitions for driving videos requires more than a continuous change in weather appearance. The transition should preserve temporal consistency, ego-motion, moving objects, changing viewpoints, and scene structure, while localizing weather effects to relevant regions and providing an explicit measure of weather severity. These requirements are difficult to guarantee through latent interpolation between clear- and adverse-weather representations of a single image (Madhavi et al., 2026a; Kondapally et al., 2026). Independently decoded interpolation states may introduce temporal inconsistencies and flicker, while image-level interpolation does not capture the evolving scene dynamics of a driving sequence. Moreover, latent interpolation does not explicitly localize weather effects and its interpolation coefficient represents position in latent space rather than an interpretable weather severity. These requirements motivate WeatherShift, which operates on temporally coherent, frame-aligned clear and extreme-weather videos (Sec. 2.2). Corresponding frames are blended using the spatial weather mask M_t^w. and an explicit severity parameter α_t, preserving the source scene dynamics while localizing weather changes

Algorithm 2: Weather Transition Generation

Require: Source weather W_A, target weather W_B, source frame I_A, extreme weather references
Ensure: Transitional sequence $\mathcal{T}$

$W_A, W_B \in \{\text{Sunny, Cloudy, Rainy, Foggy, Snowy}\}$
if $W_A = \text{Sunny} \vee W_B = \text{Sunny}$ **then**
    Use the corresponding source and target weather references
    Compute the weather-change mask M between the source frame and target-weather reference
    if $W_A = \text{Sunny}$ **then**
        Gradually increase target-weather intensity
    else
        Gradually decrease source-weather intensity
    end if
    Generate transitional frames using the weather mask M
else
    Set intermediate weather $W_I = \text{Sunny}$
    Stage 1: $W_A \rightarrow \text{Sunny}$
    Compute weather mask M_A between W_A and Sunny
    Gradually decrease the intensity of W_A
    Generate intermediate frames until the Sunny state is reached
    Stage 2: $\text{Sunny} \rightarrow W_B$
    Compute weather mask M_B between Sunny and W_B
    Gradually increase the intensity of W_B
    Generate intermediate frames until W_B is reached
    Concatenate both stages to form the complete transition

$$W_A \rightarrow \text{Sunny} \rightarrow W_B$$

end if
Apply temporal intensity scheduling to ensure smooth weather change
return transitional sequence $\mathcal{T}$

Table 4: Implementation details of the proposed WeatherShift approach.

Extreme				Transition		
Weather Category	**Backbone**	**Method**	**Hyper params**	**Backbone**	**Method**	**Hyper params**
Foggy	WeatherEdit	Depth guided fog synthesis	$\beta_{\text{start}} : 0$ $\beta_{\text{end}} : 0.09$ $T = \exp(-\beta d)$	Video-Depth-Anything + OpenCV	Weather-mask-guided severity scheduling and frame blending	Smoothstep scheduling Diff. threshold: 0.015 Weather-specific mask gain Mask power: 0.45 Depth-aware atmosphere
Rainy, Snowy Cloudy	Qwen Image Edit + WeatherEdit	Text guided transformation + particle addition	CFG: 3.5 Steps: 20 Seed: 0 BF16			

Table 5: Hardware and system configuration used for WeatherShift.

Component	**Specification**
GPU	NVIDIA Quadro RTX 6000 (24 GB VRAM)
CPU	Intel Xeon E5-1650 v4 (12 cores)
System Memory	32 GB RAM
Operating System	Ubuntu 24.04
NVIDIA Driver	595.84
CUDA Toolkit	12.0

to affected regions. The monotonic α_t provides a directly controllable severity progression, further grounded in the depth-dependent atmospheric model for fog through the extinction coefficient β (Sec. 2.2). Thus, WeatherShift explicitly controls **where, when, and how strongly** weather evolves while retaining the underlying dynamic scene.

A.5 Experimental Setup

Table 4 summarizes the implementation details of WeatherShift for extreme-weather and transitional-weather generation, including the backbones, methods, and key hyperparameters.

Extreme-weather generation. Fog references are produced with WeatherEdit as the depth-guided-fog-synthesis backbone, ramping the extinction coefficient linearly from β_{start} to $\beta_{\text{end}} = 0.09$ and compositing with $T = \exp(-\beta d)$ (Eq. 1). Rain, snow, and cloudy references are produced with Qwen-Image-Edit for text-guided appearance transformation followed by WeatherEdit's particle addition stage (Eq. 2), using classifier-free guidance scale 3.5, 20 diffusion steps, a fixed seed of 0, and BF16 precision, so that generation is deterministic and reproducible across runs.

Transition synthesis. The severity schedule and frame blending stage (Sec. 2.4) uses Video-Depth-Anything for weather-mask-guided scheduling: severity follows a smoothstep schedule rather than a linear ramp, the weather mask M_t^w is thresholded at a frame-difference value of 0.015, and the mask is further shaped by a condition-specific gain and a mask power of 0.45, with depth-aware atmosphere applied so that masking strength additionally accounts for scene depth rather than appearance difference alone.

We implement WeatherShift in Python 3.10 using PyTorch and We split nuScenes-WT into 70% training, 10% validation, and 20% testing, matching the protocol of our prior work, with the split performed at the scene level so that all transitions and extreme-weather counterparts of a given scene fall in the same split and do not leak across train/test.

A.5.1 Computational Setup

Table 5 presents the hardware setup for conducting experiments. All experiments were conducted on a workstation equipped with an NVIDIA Quadro RTX 6000 GPU with 24 GB of VRAM. The system uses an Intel Xeon E5-1650 v4 CPU with 12 CPU cores and 32 GB of system memory. The experiments were conducted on Ubuntu 24.04. The NVIDIA driver version was 595.84, while the installed CUDA toolkit version was 12.0.

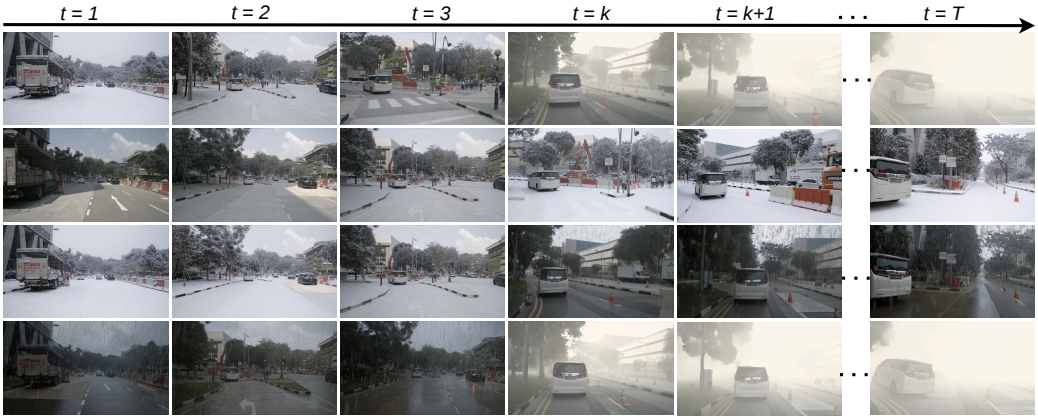

Figure 6: Nuscenes-WT dataset transitional weather video sequences (e.g., (a) snowy to foggy, (b) sunny to snowy, (c) snowy to rainy, and (d) rainy to foggy). For a given length T, the WeatherShift model produces smooth and realistic transitions between weather conditions.

A.6 Additional qualitative analysis

Figure 6 presents representative transitional-weather video sequences from the nuScenes-WT dataset generated by WeatherShift, including (a) Snowy $\rightarrow$ Foggy, (b) Sunny $\rightarrow$ Snowy, (c) Snowy $\rightarrow$ Rainy, and (d) Rainy $\rightarrow$ Foggy transitions. For a given sequence length T, each example consists of temporally ordered frames in which the weather condition progressively evolves from the source state toward the target state. Importantly, both the weather and the driving scene evolve simultaneously: ego-motion changes the camera viewpoint, surrounding vehicles and other agents move, and scene content changes as the vehicle progresses through the environment. WeatherShift preserves this natural scene evolution while progressively modifying the weather appearance, resulting in smooth transitions without freezing the underlying scene or treating it as a static background. The examples demonstrate consistent evolution of precipitation, fog, visibility, and illumination alongside changing road layouts, object configurations, and viewpoints. These qualitative results illustrate the ability of WeatherShift to generate diverse directional weather transitions while maintaining temporal coherence with the dynamic driving scene.