https://doi.org/10.31436/iiumej.v27i3.4689

Compute-Aware Evaluation of Severity Conditioning and Lightweight Architectures for Real-World Image Deraining

TEDDY SURYA GUNAWAN[1*], MIRA KARTIWI[2]

[1]*Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*

[2]*Department of Information Systems, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia*

**Corresponding author: tsgunawan@iium.edu.my*

(Received: 28 July 2026; Accepted: 29 August 2026; Published online: 12 September 2026)

ABSTRACT: Image deraining for embedded perception must balance restoration quality against complete inference cost. Yet density-conditioned methods require annotations unavailable for captured rain, while efficiency studies often omit conditioning-signal cost. This study examines whether label-free severity conditioning justifies its end-to-end cost and whether complexity measures predict realized latency. Under a common L_1 protocol, we compare a PReNet baseline with scalar feature-wise linear modulation, spatial severity modulation (SSM), a matched-capacity constant-map control, and two lightweight architectures. Zero-, shuffled-, and constant-prior interventions test SSM's dependence on prior presence, image correspondence, and spatial structure. Six models were trained for 100 epochs on 10,000 LHP-Rain patches using three seeds and evaluated on the 1,000-image test split. SSM achieved 31.775±0.109 dB, compared with 31.741±0.095 dB for Base and 31.754±0.102 dB under zero-prior replacement. A seed-level equivalence analysis based on paired image differences supported practical equivalence between SSM and Base (95% CI: −0.026 to 0.094 dB; margin fixed before primary analysis: ±0.1 dB), consistent with the half-scale experiment. The two-stage model reduced computation from 265.0 to 44.6 GFLOPs and model latency from 20.43 to 4.50 ms, but the prior generation increased total latency to 24.74 ms. CBAM-light required 14.1 GFLOPs yet recorded 33.68 ms. The label-free statistic achieved a cross-scene AUC of 0.772 on RealRain-1k, indicating moderate agreement with density labels. Thus, severity-conditioned and lightweight deraining systems should be assessed through verified prior use and measured end-to-end latency rather than accuracy, parameters, or GFLOPs alone.

ABSTRAK: Penyahhujanan untuk sistem terbenam perlu mengimbangi kualiti pemulihan dengan kos inferens menyeluruh. Namun, kaedah yang bergantung pada kekerapan memerlukan anotasi yang tidak tersedia untuk hujan yang dirakam, manakala kajian kecekapan sering kali mengabaikan kos isyarat pengkondisian. Kajian ini meneliti sama ada penentu tahap tanpa label membenarkan kos hujung-ke-hujung dan sama ada ukuran kerumitan meramalkan kelewatan sebenar. Di bawah protokol L1, garis dasar PReNet dibandingkan dengan modulasi linear mengikut ciri skalar, modulasi tahap spatial (SSM), kawalan peta tetap berkapasiti sepadan, dan dua seni bina ringan. Intervensi sifar-, diacak-, dan malar menguji kebergantungan SSM pada kewujudan, kesepadanan imej, dan struktur spatial. Enam model dilatih untuk 100 epok menggunakan 10,000 tampalan LHP-Rain dengan tiga benih dan dinilai pada 1,000 imej ujian. SSM mencapai 31.775±0.109 dB, berbanding 31.741±0.095 dB bagi Base dan 31.754±0.102 dB di bawah penggantian sebelum sifar. Analisis kesetaraan tahap biji berdasarkan perbezaan imej berpasangan menyokong kesetaraan praktikal antara SSM dan Base (95% CI: −0.026 hingga 0.094 dB; margin ditetap sebelum analisis utama: ±0.1 dB), ianya konsisten dengan eksperimen separuh skala. Model

dua peringkat mengurangkan pengiraan daripada 265.0 kepada 44.6 GFLOP dan model pendaman daripada 20.43 kepada 4.50 ms, tetapi bagi penjanaan sebelum, meningkatkan jumlah pendaman kepada 24.74 ms. CBAM-light memerlukan 14.1 GFLOPs tetapi merekodkan 33.68 ms. Statistik tanpa label mencapai AUC rentas senarai 0.772 pada RealRain-1k, menunjukkan persetujuan sederhana pada label ketumpatan. Jadi, sistem penyahhujanan ringan dan bersyarat teruk perlu dinilai melalui penggunaan yang disahkan sebelumnya dan ukuran kelewatan hujung-ke-hujung, bukan pada ketepatan, parameter, atau GFLOPs.

KEYWORDS: *Image Deraining; Severity Conditioning; Lightweight Architectures; End-to-End Latency; Real-World Rain*

1. INTRODUCTION

Rain introduces streaks, veiling, splash, and contrast loss into images acquired by mobile robots, unmanned aerial vehicles, and advanced driver-assistance systems. These degradations can obscure scene structure and compromise downstream perception precisely when reliable visual information is most needed. Because restoration may need to run on the sensing platform, practical deraining requires more than high peak signal-to-noise ratio (PSNR) or structural similarity (SSIM). Moreover, it must also satisfy constraints on latency, computation, and memory [1]. Nevertheless, deraining studies commonly use parameter count and floating-point operations (FLOPs) as efficiency surrogates without measuring the full inference path. This convention is particularly problematic for conditional models, whose reported cost may omit generation of the auxiliary signal that drives the restoration network.

Deep image deraining has progressed from convolutional and recurrent networks to transformer- and attention-based restoration systems. PReNet established a compact recurrent formulation that progressively refines the restored image through repeated stages [2]. Density-aware deraining further showed that restoration can be adapted to different rain regimes by predicting a density category and routing features through specialized streams [3]. More recent high-capacity architectures, including Restormer and frequency-enhanced state-space models, have expanded the accuracy-oriented design space through long-range and frequency-aware modeling [4, 5]. In parallel, lightweight approaches have pursued recursive pyramids, recursive window transformers, and saliency-guided convolutional attention to reduce model size and arithmetic cost [6-8]. These developments demonstrate that both conditional adaptation and architectural compression can improve the accuracy–complexity tradeoff, but they do not establish whether low parameter counts or FLOPs translate into low realized latency under a common implementation and hardware configuration.

The shift from synthetic benchmarks to captured rain creates a second challenge. Real-rain datasets better represent the interaction of streaks, accumulated rain, splash, illumination, and scene content, but constructing aligned clean references is intrinsically difficult. GT-RAIN acquires real rainy and clean observations through carefully controlled capture, whereas LHP-Rain derives pseudo-ground-truth references from video through robust low-rank tensor recovery [9, 10]. These strategies improve realism but can retain alignment errors or residual degradation in the reference images. Moreover, LHP-Rain does not provide rain-density annotations. Such annotations are generally unavailable in uncontrolled capture because density is neither prescribed during acquisition nor readily recoverable afterward as an objective ground-truth quantity. Consequently, a density-conditioned model developed for synthetic or controlled data cannot be transferred directly to LHP-Rain without constructing a label-free conditioning signal.

Conditioning mechanisms also differ in the information they communicate to a restoration network. Feature-wise linear modulation (FiLM) converts a global conditioning vector into channel-wise affine transformations, whereas spatial feature transformation produces location-dependent modulation from a spatial prior [11, 12]. Global conditioning can represent image-level degradation, but it cannot indicate where individual degradation structures occur. Spatial conditioning preserves this location information, but its larger conditioning head and map-generation procedure add computational cost. Related work in adverse-weather restoration has applied color-imbalance conditioning to a compact transformer for sandstorm restoration and stability-aware physics-prior conditioning to lightweight image dehazing [13, 14]. Although these studies support the broader use of degradation-aware priors, they do not establish whether a label-free spatial rain prior is informative for real-world image deraining or whether any resulting improvement justifies its end-to-end cost. The present study addresses this question by coupling a deterministic, training-free rain-severity estimator with an identity-initialized spatial severity modulation (SSM) head and evaluating the resulting system using matched-capacity conditioning alternatives and prior-replacement controls.

These limitations define the research gap this study addresses. Existing work has not jointly isolated the contribution of conditioning granularity, controlled for conditioning-head capacity, challenged a trained spatially conditioned model using falsification controls, and charged prior generation to the end-to-end inference cost. Likewise, lightweight deraining architectures are often compared through published parameter or FLOP values obtained under different protocols rather than through measured latency under a common evaluation. The present study is therefore positioned as a controlled evaluation rather than a claim of a universally superior architecture. Its objective is to determine whether a label-free severity prior provides a repeatable restoration advantage on real captured rain commensurate with its full computational cost, and whether conventional complexity measures reliably predict realized latency.

To achieve this objective, the study uses PReNet as a fixed backbone and compares an unconditioned baseline with global FiLM, spatial SSM, and a matched-capacity SSM supplied with a constant map. We further evaluate each trained SSM checkpoint after replacing its estimated prior with zero, shuffled, and constant alternatives, testing dependence on the prior's presence, image correspondence, and spatial structure. We then compare the matched-backbone family with a prior-gated two-stage network and an image-only CBAM-light network under the same reconstruction objective and test protocol [7, 15]. A deterministic estimator supplies the label-free severity map, and we assess its cross-scene agreement with low- and high-density labels independently on RealRain-1k rather than assuming it represents ground-truth density [16]. Evaluation combines PSNR, SSIM, and learned perceptual image patch similarity with parameter count, GFLOPs, model latency, prior-generation latency, and total sequential latency. Severity-stratified and upper-versus-lower regional analyses examine whether aggregate results conceal condition- or region-specific behavior.

This study makes four contributions. First, it establishes a label-free difficulty-stratification protocol and quantifies its agreement with labeled real-rain data while explicitly limiting its interpretation as a density measure. Second, it provides a matched-backbone comparison that separates global conditioning, spatial conditioning, and added conditioning-head capacity, supplemented by three prior-replacement controls. Third, it evaluates two lightweight alternatives against the PReNet family while accounting for the cost of every auxiliary signal consumed during inference. Fourth, it introduces severity-stratified and regional analyses to test whether spatial conditioning reduces performance disparities associated with difficult and ground-dominant image content. Together, these contributions

https://doi.org/10.31436/iiumej.v27i3.4689

provide an evidence-based assessment of when severity conditioning and lightweight architectural design translate into practical end-to-end efficiency for real-world image deraining.

2. Label-Free Severity Conditioning and Lightweight Deraining Architectures

This section develops the label-free conditioning framework and comparison architectures needed to determine whether spatial rain information improves restoration without confounding signal granularity with model capacity.

2.1. Comparative Framework and PReNet Backbone

The comparison is organized around four conditions that share a common PReNet backbone: an unconditioned baseline (Base), global feature-wise linear modulation (FiLM), spatial severity modulation (SSM), and a matched-capacity constant-map condition (SSM-const). PReNet provides a compact recurrent architecture that progressively refines the restored image through repeated stages [2]. Its recurrent depth, feature width, and main restoration pathway are held fixed across the four conditions. Consequently, differences within this family reflect the conditioning mechanism rather than changes in backbone capacity.

This comparison distinguishes three factors often conflated in conditional restoration: the presence of a conditioning signal, the spatial information in that signal, and the additional capacity introduced by the conditioning head. Base establishes the unconditioned reference. FiLM tests whether an image-level severity statistic is sufficient. SSM retains the spatial distribution of the estimated rain response, whereas SSM-const preserves the SSM head but removes spatial variation from its input. Two additional lightweight models, including a prior-gated two-stage network and an image-only CBAM-light network, are evaluated as independent architectures rather than PReNet variants.

2.2. Label-Free Rain-Severity Estimator

A deterministic, training-free procedure derives a per-pixel response map $S(x)$ and an image-level summary statistic $\bar{s}$ directly from the rainy input, as shown in Fig. 1. The procedure is designed for real-rain datasets that do not provide density annotations. It identifies bright, narrow, predominantly near-vertical structures consistent with rain streaks, but it is not intended to recover a physically calibrated rain density.

Let $Y(x)$ denote the luminance of the rainy image at spatial location x. A guided self-filter $G(\cdot)$, with radius 4 and regularization parameter $\varepsilon_g = 0.08^2$, estimates the locally smooth component of the luminance image [17]. The positive high-frequency residual is

$$R(x) = \max(Y(x) - G(Y)(x), 0). \quad (1)$$

Retaining only the positive residual suppresses slowly varying illumination and dark edges while emphasizing locally bright structures. Because scene textures can produce similar responses, $R(x)$ is treated as an intermediate streak-sensitive representation rather than a rain mask.

The residual is then convolved with a bank of zero-mean line kernels $K_{\theta_k,l}$, indexed by orientation θ_k and length l. The strongest absolute response across the bank is

$$E(x) = \max_{\theta_k,l} \left| \left(K_{\theta_k,l} * R\right)(x) \right|. \quad (2)$$

https://doi.org/10.31436/iiumej.v27i3.4689

The kernel bank spans five near-vertical orientations from 60° to 120° in 15° increments and uses lengths of 7 and 11 pixels. Taking the maximum across orientations and lengths permits the estimator to respond to streaks with modest variations in direction and scale without introducing learned parameters.

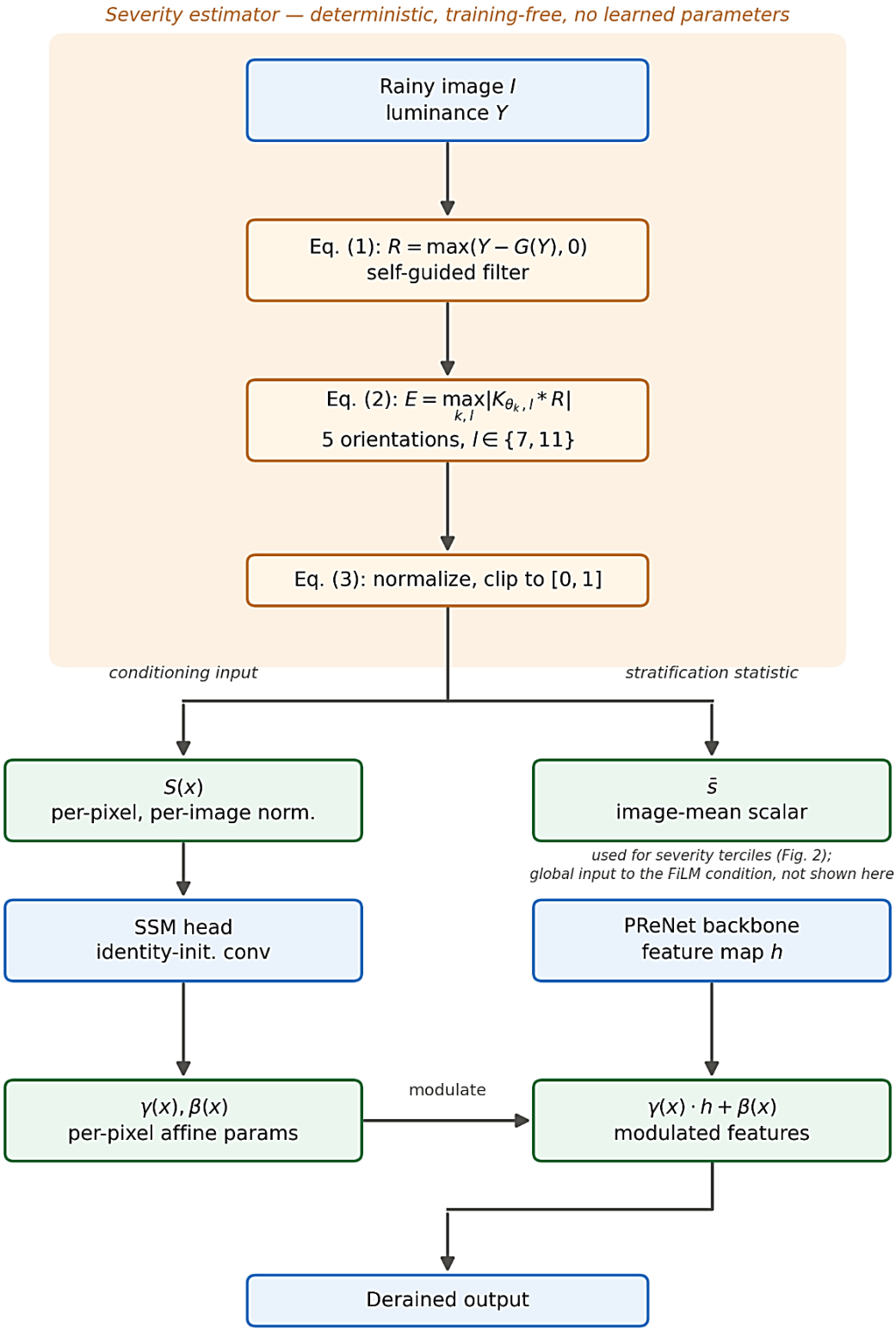

Figure 1. Construction of the label-free severity map through Eqs. (1)–(3) and its integration into the SSM conditioning head.

The response is normalized within each image using its 5th and 95th percentiles:

$$S(x) = clip\left(\frac{E(x)-q_{0.05}(E)}{q_{0.95}(E)-q_{0.05}(E)+\varepsilon}, 0,1\right), \quad \bar{s} = \frac{1}{HW}\sum_x S(x), \tag{3}$$

where H and W are the image height and width, respectively, and ε prevents division by zero. The normalized map $S(x)$ serves as the spatial conditioning input, while $\bar{s}$ summarizes the distribution of the normalized response at the image level. Because the normalization is performed independently for each image, $\bar{s}$ is not interpreted as an absolute rain-density measurement. It is used only as a reproducible difficulty statistic, and its agreement with externally provided low- and high-density labels is evaluated separately.

2.3. Global and Spatial Severity Conditioning

The four PReNet-family conditions differ only in the signal supplied to the backbone and the mechanism used to transform that signal, as shown in Table 1.

Table 1. Comparison of Conditioning Mechanisms.

Condition	Conditioning signal	Purpose
Base	None	Establishes the unconditioned reference
FiLM	Image-level statistic $\bar{s}$	Tests whether global severity information is sufficient
SSM	Spatial map $S(x)$	Tests whether location-dependent severity information is beneficial
SSM-const	Constant map with value $\bar{s}$	Controls for SSM-head capacity while removing spatial variation

FiLM and SSM apply the same general affine transformation to an intermediate recurrent feature tensor h:

$$h' = h \odot [1 + \gamma(c) + \beta(c)], \quad c = \begin{cases} \bar{s}, & FiLM, \\ S(x), & SSM. \end{cases} \tag{4}$$

FiLM converts the scalar $\bar{s}$ into channel-wise modulation parameters γ and β, which are broadcast across spatial positions. This design follows the principle of feature-wise affine conditioning [11]. SSM instead derives channel- and position-dependent modulation maps from $S(x)$, following the broader principle of spatial feature transformation [12]. This modulation is applied once within each of PReNet's six recurrent iterations, immediately after the convolutional-LSTM gate update and before the five residual blocks, using the same conditioning head (i.e., shared, not iteration-indexed, parameters) at every iteration.

The final affine layer of each conditioning head is initialized to zero. Therefore, $\gamma(x) = 0$ and $\beta(c) = 0$ at initialization, and Eq. (4) initially reduces to $h' = h$. This identity initialization lets each conditioned model start from the unmodulated feature pathway and learn only the degree of conditioning supported by the training data. SSM-const uses the same architecture and parameter count as SSM but replaces $S(x)$ with a spatially constant map whose value is $\bar{s}$. It therefore preserves the conditioning head and image-level response magnitude while removing location-dependent structure. The contrast between SSM and SSM-const isolates the contribution of spatial variation more directly than a comparison between SSM and Base alone.

2.4. Lightweight Comparison Architectures

Two independent lightweight architectures broaden the comparison beyond alternative conditioning heads on PReNet. They use the same reconstruction objective but are not exact reproductions of previously published models.

The first is a prior-gated two-stage architecture organized as local repair followed by global correction. A depthwise-separable repair network receives the concatenation of the rainy image I and severity map S. Its residual output produces an intermediate image I_o:

$$I_o = I + g_\phi([I, S]). \tag{5}$$

An AOD-Net-style module then estimates a spatial correction factor $K_\psi(I_o)$ and produces the restored image [18]:

$$\hat{J} = clip\big(K_\psi(I_o) \odot I_o - K_\psi(I_o) + b, 0,1\big). \tag{6}$$

The first stage is termed de-occlusion because its local prior emphasizes bright, oriented streak structures that may obscure scene details. This term describes the intended computational function rather than asserting that the stage identifies every form of rain occlusion or splash. The second stage is termed de-veiling because it applies the AOD-style reformulation to correct broader contrast and visibility degradation without separately estimating transmission and atmospheric light. The de-occlusion-first ordering was fixed before the main comparison based on preliminary architectural screening; that screening motivates the configuration but is not treated as causal evidence for its physical interpretation.

The second architecture, CBAM-light, is an image-only model that combines dilated convolution with channel and spatial attention. It follows the general lightweight visual-saliency pattern used in LDVS and implements attention through convolutional block attention modules [7, 15]. CBAM-light is a study-specific reimplementation, not an exact reproduction of LDVS. Its default width yields 54,998 trainable parameters, close to the approximately 50,703 parameters reported for LDVS. Unlike FiLM, SSM, SSM-const, and the two-stage architecture, CBAM-light receives no externally computed severity signal. It therefore provides a lightweight, image-only comparison to determine whether reduced arithmetic complexity translates into reduced realized latency.

3. Experimental Design

This section establishes a common and reproducible protocol for testing restoration quality, dependence on the conditioning prior, and realized end-to-end efficiency under controlled training and evaluation conditions.

3.1. Datasets and Data Splits

LHP-Rain is the primary training and restoration-evaluation dataset [10]. It contains rain captured under real outdoor conditions across multiple cities and provides pseudo-ground-truth clean references recovered from video through robust low-rank tensor modeling. Although this construction offers more realistic degradation than synthetic rain generation, residual rain and imperfect alignment may remain in the recovered references. Reference-based metrics are therefore interpreted as measures of agreement with the released pseudo-ground truth rather than as complete measures of perceptual restoration quality. LHP-Rain does not provide rain-density annotations.

The experiments use the released train-simple and local test configurations rather than the full-resolution benchmark protocol. The train-simple split contains 75,838 pre-cropped 256×256 patches, representing approximately 10% of the complete training material. The primary configuration uses a fixed-seed sample of 10,000 patches generated with seed 20260814, together with 1,000 additional patches reserved for validation. The test set contains 1,000 paired 512×512 crops, all of which are included in the reported evaluation. By contrast,

https://doi.org/10.31436/iiumej.v27i3.4689

the published LHP-Rain benchmark evaluates full-resolution 1920×1080 frames. Published benchmark values are therefore included only as contextual anchors and not as like-for-like comparisons.

The 10,000-patch sample represents 13.2% of train-simple. However, because train-simple consists of cropped patches rather than independently captured scenes, the nominal patch count may overstate the diversity of the training data. The initial experiment screened all six conditions using 5,000 patches for 50 epochs. To determine whether its principal finding persisted under a materially larger training budget, we subsequently doubled the number of unique training patches to 10,000 and increased the training duration to 100 epochs. This change increased the total number of sample presentations and optimizer updates by approximately fourfold. We adopted the larger configuration as the primary training and reporting scale, while retaining the original 5,000-patch experiment as a training-budget sensitivity check. Because we increased patch count and training duration simultaneously, this comparison evaluates robustness to a larger overall training budget rather than isolating the effect of data quantity alone.

A filename-level audit of the released manifests showed that the train-simple subsample and local test split are not scene-disjoint. Using the `<scene>_<frame>` naming convention of the LHP-Rain release, 446 of the test set's 463 distinct source scenes (96.3%) also contributed patches to the 10,000-patch training sample. The corresponding overlap was 399 of 463 scenes (86.2%) for the earlier 5,000-patch sample and 454 of 463 scenes (98.1%) for the complete 75,838-patch train-simple release. By contrast, the training and validation subsamples shared no scene identifiers under the same audit procedure.

Matching filename identifiers across the training and test sets were comparatively uncommon: 7 of 10,000 training images at the primary scale, 5 of 5,000 at the screening scale, and 101 of 75,838 in the complete train-simple release. Because the released training and test images can differ in resolution and crop offset, a matching identifier does not necessarily indicate pixel-identical content. Nevertheless, the extensive scene-level overlap shows that most test scenes originate from source videos also represented in the training material. This overlap is a property of the officially released train-simple and local test configurations, not an artifact introduced by the present subsampling procedure.

All evaluated conditions use identical training, validation, and test lists; therefore, the overlap does not create an explicit condition-specific procedural advantage. It may nevertheless elevate absolute PSNR and SSIM values by exposing models to scene-related content and could interact differently with model capacity or conditioning. Accordingly, interpret both the absolute restoration scores and comparisons with published anchors as results on the official overlapping configuration rather than as evidence of generalization to unseen scenes.

RealRain-1k is used only to assess whether the label-free statistic agrees with externally assigned low- and high-density conditions [16]. The dataset contains the same 1,120 scenes represented at two rain densities through density-controllable temporal filtering. The present validation uses the released 224-scene test split from RealRain-1k-L and RealRain-1k-H, producing 448 images. RealRain-1k is not used to train or evaluate the restoration architectures.

3.2. Training Objective and Protocol

All six trained conditions, including Base, FiLM, SSM, SSM-const, two-stage, and CBAM-light, are optimized using the same reconstruction objective and data protocol. This common training design isolates the effects of conditioning and architecture from differences introduced by model-specific loss functions.

https://doi.org/10.31436/iiumej.v27i3.4689

Let $I \in [0,1]^{3\times H\times W}$ denote the rainy input, J its pseudo-ground-truth reference, and z the model-specific conditioning input. The restored image is expressed as $\hat{J} = f_\theta(I, z)$, where θ denotes the trainable model parameters. The conditioning input is defined as

$$z = \begin{cases} \emptyset & \text{Base and CBAM-light} \\ \bar{s} & \text{FiLM} \\ S(x) & \text{SSM and two-stage} \\ \bar{s}\mathbf{1}_{H\times W} & \text{SSM-const} \end{cases} \tag{7}$$

where $\emptyset$ denotes the absence of an auxiliary conditioning input, $\bar{s}$ is the image-level severity statistic, $S(x)$ is the estimated spatial severity map, and $\mathbf{1}_{H\times W}$ is an all-one map. Thus, $\bar{s}\mathbf{1}_{H\times W}$ supplies SSM-const with the same image-level statistic as FiLM while removing spatial variation.

Every condition minimizes the mean absolute reconstruction error:

$$\mathcal{L}_1(\theta) = \frac{1}{3HW}\sum_{c=1}^{3}\sum_{x}|[f_\theta(I,z)]_c(x) - J_c(x)|. \tag{8}$$

Using the same $\mathcal{L}_1$ objective prevents the architecture comparison from being confounded by differences in loss formulation. The retrained Base condition is not intended to reproduce the originally published PReNet score because the official PReNet implementation uses mean squared error or negative SSIM, whereas the present study applies $\mathcal{L}_1$ uniformly to all conditions [2]. Severity-weighted and detector-based auxiliary losses are excluded from the validated comparison because they would introduce additional objectives requiring independent weight selection and controlled ablation.

Each condition is trained for 100 epochs with a batch size of 16 using random seeds 42, 1337, and 2024. The same training, validation, and test partitions are maintained across conditions. Reported results are calculated as the arithmetic mean and sample standard deviation of the three seed-level measurements. For reproducibility, each run retains its resolved configuration, training log, metric output, final checkpoint, latency measurements, and source-revision identifier. The three prior-replacement controls are not separately trained models; they are post-training interventions applied to each retained SSM checkpoint, as described in Section 3.3.

All conditions were optimized with Adam (learning rate 1×10^{-4}, no learning-rate schedule) and a batch size of 16. The only data augmentation applied during training was random 256×256 cropping of each rainy/clean pair; no flipping, rotation, or color jittering was used. Evaluation uses each run's final-epoch checkpoint for every condition, with no best-on-validation selection, so no condition benefits from selective early stopping. All experiments used PyTorch 2.5.1 (CUDA 12.1, cuDNN 9.1.0) on a single NVIDIA GeForce RTX 4090 GPU under Python 3.10.20; LPIPS was computed with the lpips package (v0.1) using its AlexNet backbone.

3.3. Prior-Replacement Falsification Controls

Three post-training interventions test whether the trained SSM model depends on the information carried by its estimated prior. These interventions are applied to every trained SSM checkpoint without retraining or changing any model weights.

In the **zero-prior control**, $S(x)$ is replaced with an all-zero map. This intervention tests whether the model depends on a nonzero conditioning signal. In the **shuffled-prior control**, $S(x)$ is replaced with a severity map obtained from another image in the same evaluation batch. This preserves the broad distribution of prior values while disrupting correspondence between

https://doi.org/10.31436/iiumej.v27i3.4689

the rainy image and its prior. In the **constant-prior control**, $S(x)$ is replaced with a spatially constant map whose value equals the image's $\bar{s}$. This retains its image-level statistic while removing spatial variation.

The input tensors used for the estimated, shuffled, zero, and constant conditions are verified to be distinct where required. Together, these controls test three increasingly specific forms of prior use: sensitivity to the prior's presence, dependence on image–prior correspondence, and dependence on its spatial structure. They complement SSM-const, which tests the effect of training a matched-capacity model from the outset with a spatially invariant input.

3.4. Restoration-Quality Metrics

PSNR and SSIM are computed on the luminance channel of YCbCr, whereas learned perceptual image patch similarity (LPIPS) is computed on RGB. This combination captures pixel fidelity, structural similarity, and learned perceptual distance.

For images normalized to [0,1], luminance mean squared error and PSNR are

$$MSE_Y = \frac{1}{HW}\sum_x \left[\hat{Y}(x) - Y_J(x)\right]^2, \quad PSNR_Y = 10\log_{10}\left(\frac{1}{MSE_Y}\right) \tag{9}$$

where $\hat{Y}$ and Y_J denote the luminance channels of the restored image and reference image, respectively.

SSIM measures agreement in local luminance, contrast, and structure [19]:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)}. \tag{10}$$

LPIPS measures the distance between normalized deep representations extracted at multiple feature levels [20]:

$$d(a,b) = \sum_l \frac{1}{H_l W_l}\sum_x \left\| w_l \odot \left[\hat{\phi}_l(a)_x - \hat{\phi}_l(b)_x\right]\right\|_2^2. \tag{11}$$

Higher PSNR and SSIM indicate better agreement with the reference, whereas lower LPIPS indicates greater perceptual similarity. We first compute metrics for each test image and checkpoint, then aggregate across the full test split. The three seed-level results are summarized using their mean and sample standard deviation. Because three seeds provide limited power for formal significance testing, the experiment emphasizes the magnitude and consistency of observed differences rather than claiming statistical significance from seed counts alone.

To determine whether the small mean difference reported in Section 4.2 represents practical equivalence rather than merely an absence of statistical significance, we compute paired per-image PSNR differences between SSM and Base on the shared 1,000-image test split for each seed. Each seed's mean paired difference is treated as one independent training replicate (n=3), and the mean across replicates is reported with its 95% confidence interval. We evaluate equivalence using a project-defined margin of ±0.1 dB, selected after the 5,000-patch screening experiment but fixed before analysis of the primary 10,000-patch configuration as the smallest PSNR difference considered operationally meaningful. The margin was not adjusted after observing the primary results. Equivalence is assessed using two one-sided tests (TOST) at $\alpha = 0.05$ [21]. Because the margin is project-defined and only three independent training replicates are available, the result is interpreted as equivalence under the stated experimental criterion rather than as evidence of exact equality.

https://doi.org/10.31436/iiumej.v27i3.4689

3.5. Severity Validation and Stratified Analyses

The label-free statistic $\bar{s}$ is validated independently on RealRain-1k through within-scene and cross-scene analyses. Within-scene ordering examines whether the high-density version of a scene receives a larger value than its corresponding low-density version. Cross-scene discrimination is assessed using the area under the receiver operating characteristic curve (AUC) and point-biserial correlation between $\bar{s}$ and the binary low/high label. These analyses distinguish paired sensitivity to changes in rain conditions from general discrimination across scenes with different textures and content.

For LHP-Rain, tercile boundaries are computed from the $\bar{s}$ distribution of the 1,000-image test split. The resulting low-, medium-, and high-statistic groups provide a deterministic difficulty partition for evaluating whether aggregate performance conceals condition-specific behavior. Because $\bar{s}$ is affected by both streak-like responses and scene texture, the groups are described as severity-statistic or difficulty strata rather than ground-truth rain-density classes.

A complementary regional analysis divides each 512×512 test image into an upper half (rows 0–255) and a lower half (rows 256–511). We compute regional PSNR and SSIM for Base and SSM using the same checkpoints as the whole-image evaluation. The division is motivated by the ground-dominant degradation reported for LHP-Rain, but the two regions are geometric proxies rather than semantic segmentations of sky, road, splash, or other scene components. The analysis therefore tests whether SSM changes a reproducible regional performance gap without assigning that gap to a specific physical degradation.

3.6. Complexity and Latency Evaluation

Computational efficiency is assessed using trainable parameter count, GFLOPs, model latency, prior-generation latency, and total sequential latency. Parameter count measures model storage requirements, while GFLOPs estimate the arithmetic operations required for a single 512×512 forward pass. Neither measure substitutes for realized latency because execution time also depends on memory traffic, kernel-launch overhead, operator implementation, and hardware utilization.

Latency is measured at batch size 1 on an NVIDIA GeForce RTX 4090. Each model is evaluated over 50 timed repetitions following five warm-up passes, with CUDA synchronization before and after timing to prevent asynchronous execution from biasing the measurements. Model latency covers the restoration-network forward pass. The deterministic prior is computed on the host CPU, and its measured generation time is charged only to conditions that consume $\bar{s}$ or $S(x)$.

Total latency is defined as the sum of prior-generation latency and model latency under sequential execution:

$$T_{total} = T_{prior} + T_{model}. \tag{12}$$

This definition does not assume asynchronous overlap between CPU prior generation and GPU inference. It therefore provides a transparent end-to-end estimate for the tested implementation rather than a hardware-independent claim about deployment speed.

4. RESULTS

This section evaluates the severity statistic, restoration accuracy, dependence on the estimated prior, regional behavior, and the relationship between theoretical complexity and measured end-to-end latency.

https://doi.org/10.31436/iiumej.v27i3.4689

4.1. Agreement of the Label-Free Statistic With Rain-Density Labels

The label-free statistic $\bar{s}$ was first evaluated on the 448-image RealRain-1k test set, comprising 224 scenes under each low- and high-density condition. Terciles calculated from the RealRain-1k distribution had boundaries of 0.287 and 0.329. Among the low-density images, 48.2% were assigned to the low tercile, while 48.7% of the high-density images were assigned to the high tercile. Both proportions exceeded the 33.3% expected under equal random assignment, but neither indicated clean separation.

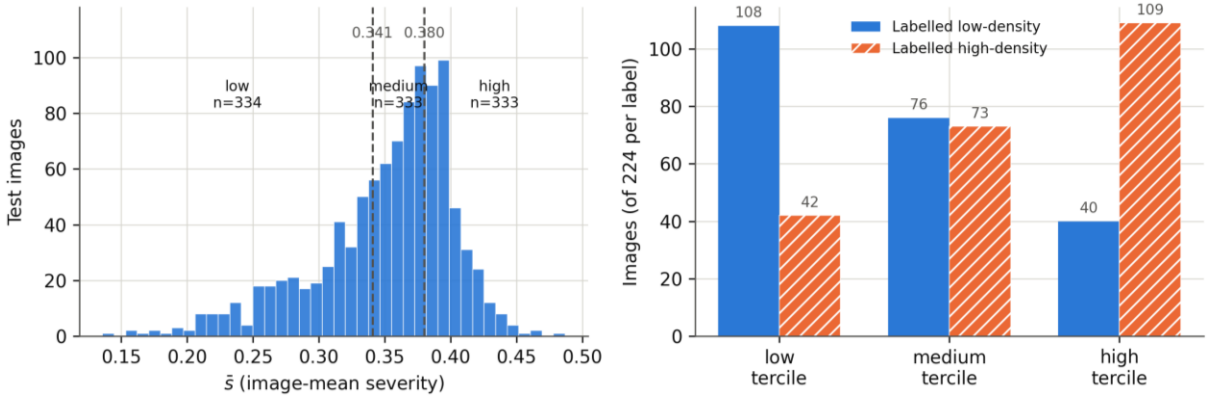

Figure 2. Distribution and external validation of the label-free statistic. (a) Distribution of $\bar{s}$ on the LHP-Rain test set ($n = 1{,}000$), with tercile boundaries at 0.341 and 0.380. (b) RealRain-1k tercile assignments by low- and high-density label ($n = 448$).

Across scenes, $\bar{s}$ achieved an area under the receiver operating characteristic curve of 0.772 and a point-biserial correlation of r=0.455 with the binary density label. These results indicate moderate agreement, not a direct measurement of rain density. Accordingly, $\bar{s}$ is used as a reproducible difficulty statistic, not as a substitute for a validated density annotation.

On the 1,000-image LHP-Rain test split, $\bar{s}$ had a mean of 0.351 and a standard deviation of 0.053. Tercile boundaries of 0.341 and 0.380 divided the test set into 334 low-, 333 medium-, and 333 high-statistic images. Figure 2 summarizes the LHP-Rain distribution and the relationship between RealRain-1k tercile assignments and the externally assigned density labels.

4.2. Restoration Accuracy and Dependence on the Prior

Table 2 examines whether the principal finding changes under a larger training budget. Increasing the configuration from 5,000 patches and 50 epochs to 10,000 patches and 100 epochs improved Base by 0.099 dB in PSNR and 0.0054 in SSIM, while SSM improved by 0.116 dB and 0.0059, respectively. However, the SSM–Base PSNR difference increased only from 0.017 to 0.034 dB. Because the number of patches and training epochs were increased simultaneously, this comparison cannot separate the effects of additional data from those of additional optimization. It nevertheless shows that the small advantage of SSM persisted but did not become practically substantial under the larger training budget.

Table 3 reports restoration quality on the complete LHP-Rain test split under the primary configuration. Among the trained PReNet-family conditions, SSM attained the highest mean PSNR and SSIM and the lowest LPIPS. Its improvements over Base were nevertheless limited to 0.034 dB in PSNR, 0.0011 in SSIM, and 0.0016 in LPIPS. FiLM and SSM-const also differed only marginally from Base. The numerical ranking therefore does not establish a practically meaningful advantage for spatial severity conditioning.

Table 2. Training-budget sensitivity of Base and SSM under the 5,000-patch/50-epoch screening configuration and the primary 10,000-patch/100-epoch configuration; values are mean ± sample standard deviation across three seeds.

Condition	Training configuration	PSNR ↑ (dB)	SSIM ↑
Base	5,000 patches/50 epochs	31.642 ± 0.047	0.8975 ± 0.0012
Base	10,000 patches/100 epochs (primary)	31.741 ± 0.095	0.9029 ± 0.0030
SSM	5,000 patches/50 epochs	31.659 ± 0.034	0.8981 ± 0.0011
SSM	10,000 patches/100 epochs (primary)	31.775 ± 0.109	0.9040 ± 0.0043

Table 3. Restoration quality on the LHP-Rain test split ($n = 1{,}000$). Values are mean ± sample standard deviation across three seeds. The zero-, shuffled-, and constant-prior rows are post hoc interventions applied to the trained SSM checkpoints without retraining.

Condition	PSNR ↑ (dB)	SSIM ↑	LPIPS ↓
Base	31.741 ± 0.095	0.9029 ± 0.0030	0.1112 ± 0.0026
FiLM	31.758 ± 0.104	0.9034 ± 0.0030	0.1108 ± 0.0033
SSM with estimated $S(x)$	31.775 ± 0.109	0.9040 ± 0.0043	0.1096 ± 0.0036
SSM with zero prior	31.754 ± 0.102	0.9014 ± 0.0037	0.1107 ± 0.0038
SSM with shuffled prior	31.739 ± 0.110	0.9028 ± 0.0044	0.1108 ± 0.0037
SSM with constant prior	31.753 ± 0.121	0.9038 ± 0.0044	0.1093 ± 0.0034
SSM-const, trained	31.765 ± 0.100	0.9030 ± 0.0031	0.1106 ± 0.0027
Two-stage	31.626 ± 0.032	0.8964 ± 0.0003	0.1180 ± 0.0010
CBAM-light	31.652 ± 0.025	0.8974 ± 0.0002	0.1148 ± 0.0003

The prior-replacement interventions test whether the trained SSM model uses the information contained in the estimated spatial map. Relative to the estimated prior, replacing $S(x)$ with zeros reduced mean PSNR by 0.021 dB, shuffling priors across images reduced it by 0.036 dB, and replacing each prior with its corresponding constant map reduced it by 0.022 dB. These effects were small and were not consistently ordered across PSNR, SSIM, and LPIPS. In particular, the constant-prior intervention produced marginally better LPIPS than the estimated prior. The separately trained SSM-const condition was also only 0.010 dB below SSM in PSNR. The controls therefore provide limited evidence that prior correspondence or spatial variation materially contributes to restoration performance.

The seed-level equivalence analysis supports this interpretation. Based on the mean paired image difference calculated independently for each seed, the SSM–Base PSNR difference was +0.034 dB, with a 95% confidence interval of [−0.026, +0.094] dB. The interval lies within the project-defined practical-equivalence margin of ±0.1 dB, which was fixed after the screening experiment but before analysis of the primary configuration. The two one-sided tests support equivalence under this margin (p=0.021). This result indicates practical equivalence within the stated experimental framework rather than exact equality between the models. Its interpretation remains bounded by the project-defined margin and the availability of only three independent seed-level replicates.

Figure 3 tests whether aggregate averaging conceals a greater benefit on more difficult images. PSNR decreases from the low- to high-statistic tercile for every condition, confirming that the label-free statistic separates images by restoration difficulty. However, the differences among Base, FiLM, SSM, and SSM-const remain below 0.1 dB across the three terciles. Replacing the estimated SSM prior with zero, shuffled, or constant maps likewise produces no consistent deterioration, including within the high-statistic tercile where severity conditioning would be expected to offer its greatest benefit. The statistic therefore identifies difficult images more clearly than conditioning improves their restoration.

https://doi.org/10.31436/iiumej.v27i3.4689

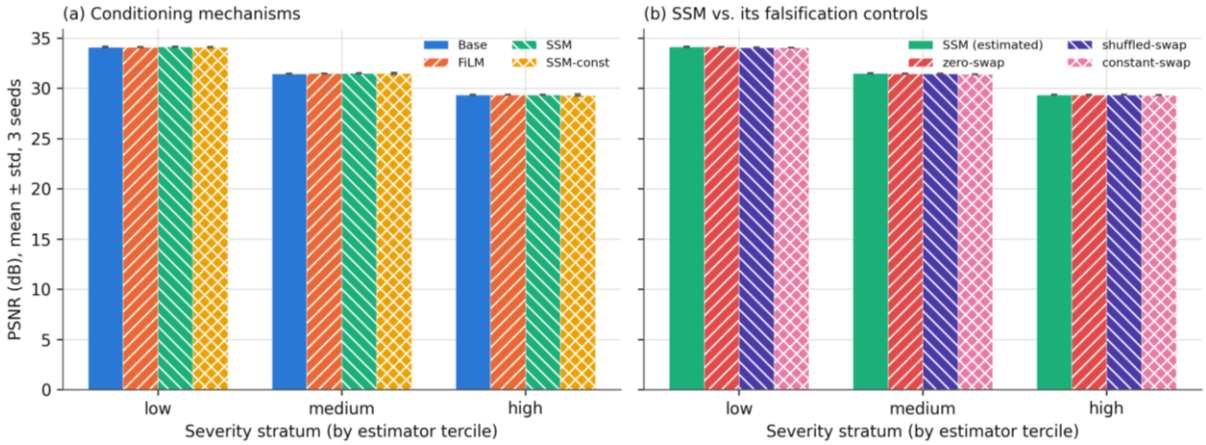

Figure 3. Restoration PSNR across LHP-Rain severity-statistic terciles under the primary training configuration. (a) Comparison of Base, FiLM, SSM, and SSM-const. (b) Comparison of the estimated SSM prior with zero-, shuffled-, and constant-prior interventions.

The lightweight architectures incurred modest accuracy penalties relative to Base. Two-stage was lower by 0.115 dB in PSNR, 0.0065 in SSIM, and 0.0068 in LPIPS, whereas CBAM-light was lower by 0.089 dB, 0.0055, and 0.0036, respectively. Relative to the mean PSNR of the four trained PReNet-family conditions, the deficits were 0.134 dB for Two-stage and 0.108 dB for CBAM-light. These accuracy reductions must be considered together with the computational and latency results rather than interpreted independently.

4.3. Computational Complexity and Realized Latency

Table 4 compares model size, arithmetic cost, and measured latency at 512×512 resolution. SSM increased the PReNet parameter count and GFLOPs by approximately 11.1%. Its model-only latency increased from 20.43 to 23.79 ms, while prior generation added another 19.73 ms. The resulting sequential total of 43.52 ms was more than twice Base's latency.

Table 4. Model complexity and latency at batch size 1 and 512×512 resolution on an NVIDIA GeForce RTX 4090.

Condition	Parameters	GFLOPs	Model (ms)	Prior (ms)	Total (ms)
Base	168,963	265.0	20.43	0	20.43
FiLM	171,139	265.0	20.90	20.11	41.01
SSM	187,779	294.4	23.79	19.73	43.52
SSM-const	187,779	294.4	23.79	19.73	43.52
Two-stage	170,446	44.6	4.50	20.23	24.74
CBAM-light	54,998	14.1	33.68	0	33.68

The two-stage model reduced restoration-network computation from 265.0 to 44.6 GFLOPs and model-only latency from 20.43 to 4.50 ms. However, generating its spatial prior increased total latency to 24.74 ms, which was 21% higher than Base. The approximately 20 ms prior-generation time therefore became a dominant fixed cost once the restoration network itself was made inexpensive.

CBAM-light exhibited the opposite discrepancy. It used approximately 3.1 times fewer parameters and 18.8 times fewer GFLOPs than Base, yet its measured latency was 33.68 ms, or 1.6 times higher. Because we did not conduct operator-level profiling, Table 4 shows an implementation- and hardware-specific mismatch between arithmetic complexity and realized latency rather than a universal limitation of CBAM-based architectures.

https://doi.org/10.31436/iiumej.v27i3.4689

4.4. Regional Performance

Table 5 presents the upper- and lower-half results for Base and SSM. Both models performed less favorably in the lower half. For Base, the upper-to-lower difference was 0.179 dB in PSNR and 0.0156 in SSIM. For SSM, the corresponding differences were 0.165 dB and 0.0152. SSM improved lower-half PSNR over Base by only 0.041 dB and therefore did not materially reduce the regional disparity.

Table 5. Upper- and lower-half restoration quality on LHP-Rain; values are mean ± sample standard deviation across three seeds.

Condition	Region	PSNR ↑ (dB)	SSIM ↑
Base	Upper half	32.074 ± 0.074	0.9106 ± 0.0022
Base	Lower half	31.895 ± 0.109	0.8950 ± 0.0038
SSM	Upper half	32.101 ± 0.073	0.9116 ± 0.0032
SSM	Lower half	31.936 ± 0.128	0.8964 ± 0.0055

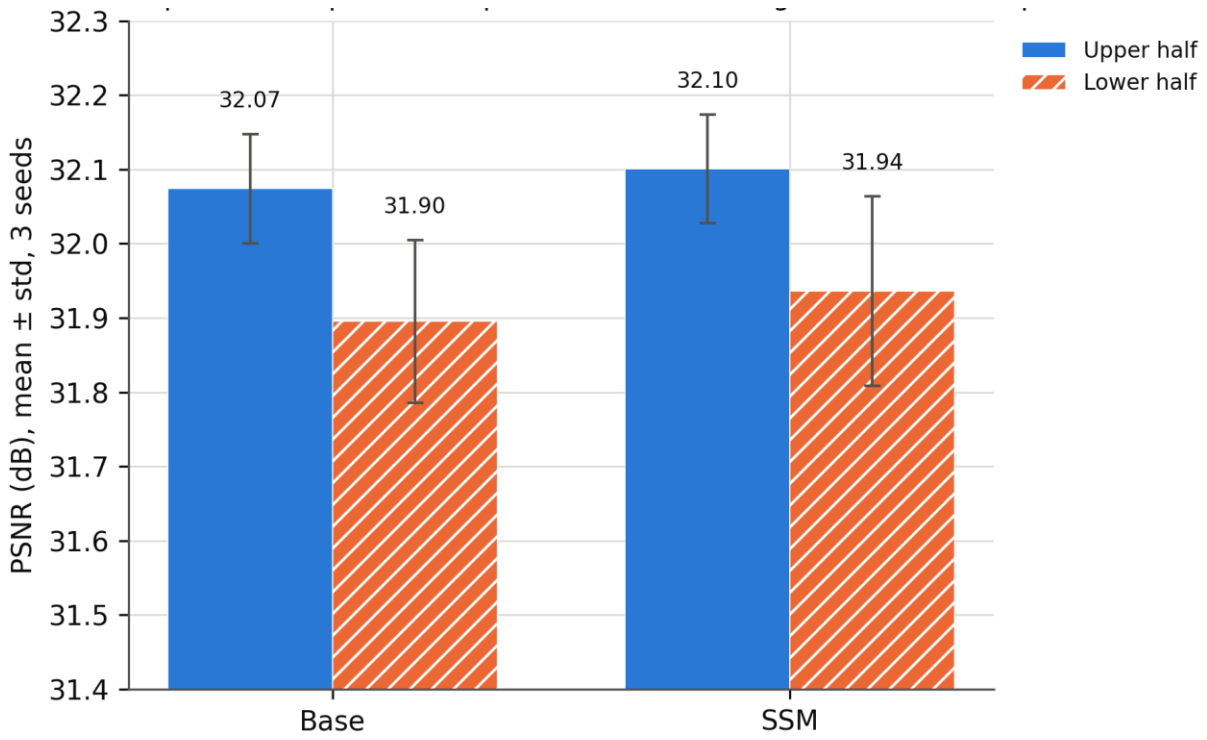

Figure 4. Upper- and lower-half PSNR for Base and SSM on LHP-Rain. Bars show means across three seeds, and error bars show sample standard deviations. The truncated vertical axis highlights a regional difference below 0.3 dB; the regions are geometric proxies rather than semantic masks.

The arithmetic mean of upper- and lower-half PSNR is not expected to equal whole-image PSNR in Table 3 because PSNR is a logarithmic transformation of MSE. Averaging separately calculated log-domain values differs from calculating PSNR from the combined image-level MSE. The corresponding SSIM averages do not show this offset.

Figure 4 visualizes the regional difference. The geometric division does not semantically identify sky, ground, road, or splash; therefore, the lower performance cannot be attributed to a specific rain phenomenon. It nevertheless shows that the tested spatial prior did not eliminate the reproducible lower-region deficit.

4.5. Contextual Comparison With Published LHP-Rain Results

Table 6 provides published LHP-Rain values only to contextualize the scale of the present results [10]. The protocols are not directly comparable. The published values use the full-resolution configuration, different training losses, and 1920×1080 test frames, whereas the present study uses 10,000 training patches, a common $\mathcal{L}_1$ objective, and 512×512 test crops. The published PReNet value exceeds the present Base result by 0.389 dB, which is plausible given its larger training configuration and different loss design. Table 6 should therefore not be interpreted as a ranking of the evaluated methods.

Table 6. Contextual LHP-Rain results reported under different training and evaluation protocols.

Method	PSNR (dB)	SSIM
Rainy input, as reported	29.97	0.8497
PReNet, as reported	32.13	0.9177
SCD-Former, as reported	34.33	0.9468
Base, present study	31.741 ± 0.095	0.9029 ± 0.0030

4.6. Qualitative Comparison

The qualitative examples in Figure 5 reinforce the small quantitative differences reported in Table 3. Base and SSM produce nearly indistinguishable outputs across the low-, medium-, and high-severity-statistic terciles. Both models generally preserve major edges, colors, and scene structures, but SSM shows no obvious advantage in suppressing rain or recovering detail. This similarity is especially important in the high-severity example: although restoration becomes more difficult as the severity statistic increases, the spatial prior does not visibly redirect the model toward more effective correction of the prominent rain-affected regions. This observation is consistent with Figure 3, where the SSM advantage remains below 0.1 dB across all terciles.

The final row presents a more demanding regional case, selected because Base produced its largest upper-minus-lower PSNR difference among the high-tercile images. Nevertheless, Base and SSM again produce visually similar lower-region reconstructions. This observation supports the regional results in Table 5: SSM improves lower-half PSNR by only 0.041 dB and does not materially reduce the upper-to-lower performance gap. However, the image cannot establish that the remaining degradation is caused by splash. The geometric regional criterion does not provide a semantic label, and the estimator primarily responds to bright, near-vertical, high-frequency structures. The example therefore indicates a limitation in coverage rather than demonstrating a specific physical failure mechanism.

The practical implication is that the spatial prior does not deliver an observable improvement commensurate with its computational cost. When Base and SSM produce comparable quantitative and visual outcomes, Base is the more defensible deployment choice because it avoids the conditioning head and approximately 20 ms of prior-generation latency reported in Table 4. This negative result is operationally meaningful: adding a physically plausible map is insufficient unless the model demonstrably uses its spatial information to improve difficult regions.

The qualitative evidence must nevertheless be interpreted conservatively. Figure 5 contains only four examples, uses one seed, and was not evaluated through blinded human assessment, magnified residual views, or downstream perception tasks. Moreover, the reference images are video-derived pseudo-ground truth and may contain residual degradation or smoothing, particularly in difficult regions. Figure 5 therefore corroborates the aggregate

findings but does not independently establish perceptual equivalence. Stronger perceptual evaluation would require enlarged crops, Base–SSM difference maps, multiple checkpoints, human ratings, or task-level measures such as detection accuracy.

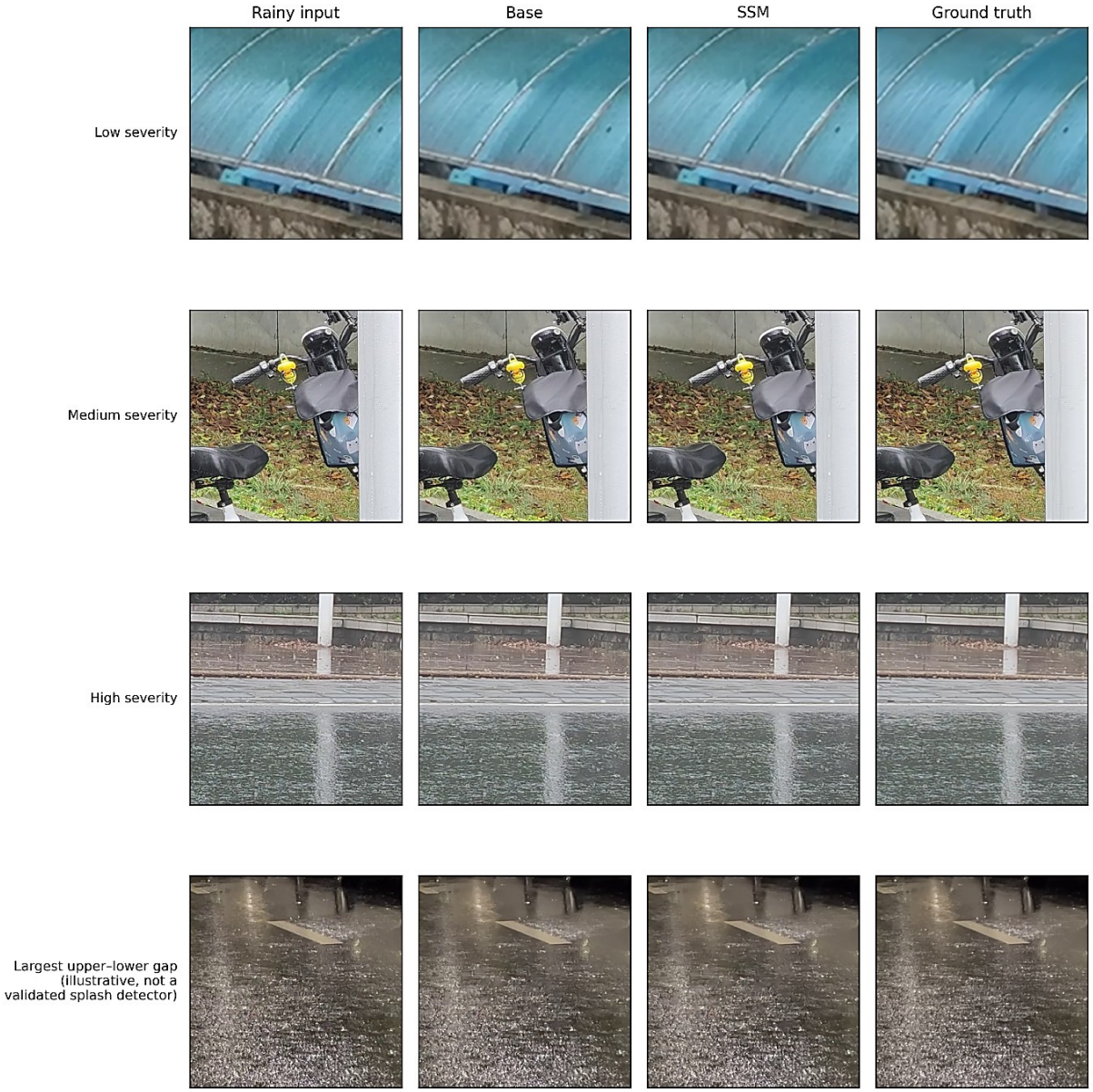

Figure 5. Qualitative comparison of rainy inputs, Base outputs, SSM outputs, and pseudo-ground-truth references. The first three rows represent the low-, medium-, and high-severity-statistic terciles. The final row shows the preselected high-tercile image with the largest Base upper-minus-lower PSNR difference. All outputs use the seed-42 checkpoint.

5. DISCUSSION

This section interprets the results by determining what the limited conditioning gains and mismatched complexity–latency relationships imply for designing and deploying severity-aware, lightweight real-world image deraining systems.

5.1. Utility of Spatial Severity Conditioning

The matched-backbone experiment provides no evidence that the tested spatial severity prior produces a practically meaningful restoration advantage at the present scale. As shown in Table 3, SSM improved mean PSNR over Base by only 0.034 dB and over its zero-prior intervention by 0.021 dB. Figure 3 further shows that this limited aggregate improvement does not conceal a substantial benefit in any severity-statistic tercile. The seed-level equivalence analysis described in Section 3.4 and reported in Section 4.2 strengthens this interpretation: the SSM–Base difference had a 95% confidence interval of [−0.026, +0.094] dB, which lies within the project-defined ±0.1-dB equivalence margin (TOST p=0.021). This result supports practical equivalence under the stated criterion rather than exact equality between the models. Its interpretation remains bounded by the project-defined margin and the availability of only three independent training replicates.

Two mechanisms may account for the limited effect. First, $S(x)$ is a deterministic transformation of the rainy image already available to PReNet. It therefore supplies an inductive bias rather than independent information, and the recurrent convolutional backbone may learn comparable local evidence directly from I. Second, the per-image percentile normalization in Eq. (3) suppresses absolute response magnitude. Although this stabilizes the spatial map, it also makes $\bar{s}$ sensitive to the distribution of oriented scene textures such as foliage, brickwork, and railings. The moderate agreement observed in Figure 2 is consistent with this interpretation.

The prior-replacement controls nevertheless suggest that the trained SSM is not entirely insensitive to its conditioning input. Shuffling or spatially flattening the map reduced performance more than replacing it with zeros, but the ordering was not consistent across PSNR, SSIM, and LPIPS. The current evidence therefore supports a narrow conclusion: this particular estimator–head combination does not provide sufficient benefit to justify its cost. It does not establish that all physical or degradation-aware priors are ineffective.

5.2. Arithmetic Efficiency Versus End-to-End Efficiency

Table 4 demonstrates that model-only complexity can misrepresent operational efficiency in two distinct ways. For the two-stage architecture, network compression was successful: the restoration model required 5.9 times fewer GFLOPs and ran 2.7 times faster than Base. However, prior generation eliminated much of that advantage and raised total latency above the unconditioned baseline. The prior therefore becomes the principal optimization target when the restoration network is already lightweight.

CBAM-light presents the converse result. Despite having the fewest parameters and GFLOPs, it produced the highest image-only latency. Parameters and GFLOPs do not account for kernel-launch overhead, memory access, operator fusion, or hardware utilization. Consequently, the paper's compute-aware contribution lies not in declaring one architecture universally faster, but in showing that efficiency claims should include measured latency for the complete inference pathway.

These latency rankings remain specific to the tested implementation and RTX 4090 platform. Different results may arise with TensorRT, operator fusion, mixed precision, CPU vectorization, asynchronous prior generation, or deployment-class accelerators. Nevertheless, the comparison establishes that theoretical arithmetic savings cannot be assumed to translate directly into end-to-end speed.

https://doi.org/10.31436/iiumej.v27i3.4689

5.3. Interpretation of Severity and Regional Behavior

The severity statistic is useful primarily as a difficulty partition. Figure 2 shows moderate agreement with the RealRain-1k density labels, while Figure 3 shows a substantial decline in restoration quality from the low to high tercile. However, these two observations do not imply that $\bar{s}$ measures rain density alone. Scene texture, illumination, residual degradation in the pseudo-ground truth, and other sources of restoration difficulty may also influence the statistic.

The regional findings reinforce this cautious interpretation. Table 5 and Figure 4 identify a consistent lower-half performance deficit, but the geometric split cannot determine whether the difference arises from splash, road texture, reflections, accumulated rain, or reference quality. SSM does not materially reduce the gap, which is compatible with an estimator designed around bright, near-vertical streaks rather than low-frequency veiling or isotropic splash. Figure 5 similarly shows little visible separation between Base and SSM, but the examples cannot establish the physical cause of the regional result.

5.4. Limitations and Scope of Inference

The primary limitation is the training scale. The models use 10,000 of the 75,838 released train-simple patches, and these cropped patches may not represent independent scenes. The numerical ranking could change with a larger training sample or longer optimization. The difference between the present results and the published values in Table 6 further indicates that the findings should not be generalized to the full-resolution LHP-Rain benchmark.

The scene-overlap audit in Section 3.1 further shows that the training samples and test split are not scene-disjoint, with 86–98% of test-scene identifiers shared with the training data depending on the training scale. The training and validation subsamples, by contrast, share no scene identifiers. Because every condition uses the same training and test lists, the overlap does not create an explicit condition-specific split advantage; however, it could interact differently with model capacity or conditioning. Consequently, the absolute PSNR and SSIM values, including the contextual comparison with published LHP-Rain results in Table 6, cannot be assumed to generalize to unseen scenes. A released scene-disjoint partition was not available for the simple splits, and constructing one remains necessary for stronger generalization claims.

The equivalence analysis is also conditional on its specified criterion. The ±0.1-dB margin used in Sections 3.4 and 4.2 was selected after the screening experiment but fixed before analysis of the primary configuration as the smallest PSNR difference considered operationally meaningful. Because this threshold is project-defined rather than externally established or preregistered, the equivalence result should be interpreted as evidence under the stated margin, not as an unconditional demonstration that SSM and Base are identical. The availability of only three independent training replicates further limits the precision and generalizability of the equivalence estimate.

Reference quality also bounds the interpretation. LHP-Rain uses video-derived pseudo-ground truth; consequently, residual rain and alignment errors can affect L_1 training and all reference-based metrics. This limitation may be most consequential for difficult images, where reference recovery itself may be less reliable.

The estimator is intentionally limited to bright, near-vertical, high-frequency structures. It does not explicitly model low-frequency veiling, oblique rain outside the kernel bank, or ground splash. Likewise, the upper/lower analysis uses fixed geometric regions rather than semantic labels. The reported regional gap is therefore descriptive rather than causal.

Finally, conditioning is evaluated on one common backbone, external validation is limited to RealRain-1k, and latency is measured using one software–hardware stack. Three seeds characterize training variability but limit the precision of the estimated effects. The findings should therefore be interpreted as a controlled assessment of the tested configurations rather than a universal rejection of severity-aware deraining.

6. CONCLUSIONS

This study establishes that severity conditioning must be justified by verified information use and complete inference cost, not architectural plausibility alone. Under a common 10,000-patch, 100-epoch, three-seed protocol, spatial severity modulation improved mean PSNR over the unconditioned PReNet baseline by only 0.034 dB and over its zero-prior intervention by 0.021 dB, while increasing parameters, computation, model latency, and prior-generation time. The matched-capacity control, prior-replacement interventions, seed-level equivalence analysis, and half-scale sensitivity experiment consistently showed no practically meaningful restoration advantage from the tested spatial prior. The lightweight architectures exposed a related deployment limitation: the two-stage network reduced model computation from 265.0 to 44.6 GFLOPs but became slower than Base when prior generation was included, whereas CBAM-light had the fewest parameters and GFLOPs yet the highest image-only latency. Model size and arithmetic complexity are therefore insufficient indicators of operational efficiency; the entire inference pathway must be measured on the intended platform. The label-free statistic remains useful for reproducible difficulty stratification, but its cross-scene AUC of 0.772 indicates only moderate agreement with rain-density labels and precludes its interpretation as ground-truth density. A severity prior is operationally valuable only when its restoration benefit survives controlled falsification and outweighs its end-to-end computational cost. Future work should evaluate larger, scene-disjoint training configurations, introduce semantically annotated weather regions, and benchmark optimized end-to-end inference on deployment-class accelerators.

ACKNOWLEDGEMENT

This work is supported by the AOARD project FA2386-25-1-4061 and recorded at IIUM SPI25-295-0295. The authors used an AI coding assistant for clarity and proofreading, reviewed and verified all AI-assisted content, and take full responsibility for the accuracy, originality, and integrity of the published work.

REFERENCES

[1] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 196, pp. 146-177, 2023.

[2] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: IEEE, pp. 3937-3946.

[3] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: IEEE, pp. 695-704.

[4] S. Yamashita and M. Ikehara, "Image deraining with frequency-enhanced state space model," in *Asian Conference on Computer Vision*, 2024: Springer, pp. 318-334.

[5] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5728-5739.

[6] B. Xiaojun, C. Zheng, Y. Jianyu, and W. Haibo, "LRP-net: A lightweight recursive pyramid network for single image deraining," *Neurocomputing,* vol. 490, pp. 181-192, 2022.

[7] M. Hu, J. Yang, N. Ling, Y. Liu, and J. Fan, "Lightweight single image deraining algorithm incorporating visual saliency," *IET Image Processing,* vol. 16, no. 12, pp. 3190-3200, 2022.

[8] Y. Liang, S. Anwar, and Y. Liu, "Drt: A lightweight single image deraining recursive transformer," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022: IEEE, pp. 589-598.

[9] Y. Ba *et al.*, "Not just streaks: Towards ground truth for single image deraining," in *European Conference on Computer Vision*, 2022: Springer, pp. 723-740.

[10] Y. Guo, X. Xiao, Y. Chang, S. Deng, and L. Yan, "From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: IEEE, pp. 12097-12107.

[11] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32, no. 1.

[12] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: IEEE, pp. 606-615.

[13] T. S. Gunawan and M. Kartiwi, "Stability-aware physics-prior conditioning for lightweight image dehazing," in *The 12th IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) 2026*, Kuching, Sarawak, 29-30 July 2026, pp. 210-215.

[14] T. S. Gunawan and M. Kartiwi, "CIART: A compact color-imbalance-conditioned transformer for sandstorm image restoration," in *The 12th IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) 2026*, Kuching, Sarawak, 29-30 July 2026, pp. 221-226.

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *European Conference on Computer Vision*, 2018: Springer, pp. 3-19.

[16] W. Li, Q. Zhang, J. Zhang, Z. Huang, X. Tian, and D. Tao, "Toward real-world single image deraining: A new benchmark and beyond," *arXiv preprint arXiv:2206.05514,* 2022.

[17] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, no. 6, pp. 1397-1409, 2013.

[18] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-net: All-in-one dehazing network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017: IEEE, pp. 4770-4778.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, pp. 600-612, 2004.

[20] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: IEEE, pp. 586-595.

[21] D. Lakens, "Equivalence tests: A practical primer for t tests, correlations, and meta-analyses," *Social psychological and personality science,* vol. 8, no. 4, pp. 355-362, 2017.