remote sensing

MDPI

Article

Deep Learning Style Transfer for Enhanced Smoke Plume Visibility: A Standardized False Color Composite (SFCC) in GEMS Satellite Imagery

Yemin Jeong [1], Seung Hee Kim [2], Menas Kafatos [2], Jeong-Ah Yu [3], Kyoung-Hee Sung [3], Sang-Min Kim [3], Seung-Yeon Kim [3], Goo Kim [3], Jae-Jin Kim [4] and Yangwon Lee [1,*]

1 Major of Geomatics Engineering, Division of Earth Environmental System Sciences, Pukyong National University, Busan 48513, Republic of Korea; jym6677@pukyong.ac.kr
2 Institute for Earth, Computing, Human and Observing (ECHO), Chapman University, Orange, CA 92866, USA; sekim@chapman.edu (S.H.K.); kafatos@chapman.edu (M.K.)
3 Environmental Satellite Center, National Institute of Environmental Research, Incheon 22689, Republic of Korea; jayu@korea.kr (J.-A.Y.); yamacco9@korea.kr (K.-H.S.); smkim91@korea.kr (S.-M.K.); sseung@korea.kr (S.-Y.K.); kimgoo9@korea.kr (G.K.)
4 Major of Environmental Atmospheric Sciences, Division of Earth Environmental System Sciences, Pukyong National University, Busan 48513, Republic of Korea; jjkim@pknu.ac.kr
* Correspondence: modconfi@pknu.ac.kr

Highlights

What are the main findings?

- The creation of SFCC using style transfer models substantially improves the detection and clarity of smoke plumes within Geostationary Environment Monitoring Spectrometer (GEMS) imagery.
- Key factors for achieving standardized color tones in time-series GEMS imagery are the optimal combination of spectral channels and the employment of state-of-the-art (SOTA) image fusion models.

What are the implication of the main findings?

- GEMS imagery acquired as a time series often exhibits variability in visual brightness and color tone, primarily due to factors such as acquisition time, local atmospheric conditions, and solar altitude variations. While these inconsistencies can compromise the consistency and reliability of smoke detection, the application of our SFCC, which uses style transfer models, substantially improves the detection and clarity of smoke plumes within the GEMS data.
- We sought to maximize the benefits of deep learning-based style transfer for improving smoke plume visibility by identifying the optimal spectral channel combination for the RGB color composite. This optimal combination utilizes the ΔUV (Radiance (380 nm)–Radiance (340 nm)) to the red channel, Radiance (380 nm) to the green, and Radiance (340 nm) to the blue. Moreover, through a comparison of multiple statistical and deep learning models, we established the Style Injection Diffusion Model (SI-DM) as the optimal choice for SFCC creation.

Check for updates

Academic Editors: Hao Zhu and Biao Hou

Received: 13 December 2025
Revised: 21 January 2026
Accepted: 29 January 2026
Published: 2 February 2026

Abstract

Wildfire smoke visualization using geostationary satellite imagery is essential for real-time monitoring and atmospheric analysis; however, inconsistencies in color tone across Geostationary Environment Monitoring Spectrometer (GEMS) images hinder reliable interpretation and model training. This study proposes a Standardized False Color Composite (SFCC) framework based on deep learning style transfer to enhance the visual consistency and interpretability of wildfire smoke scenes. Four tone-standardization methods were compared: the statistical Empirical Cumulative Distribution Function (ECDF) correction and three neural approaches—ReHistoGAN, StyTr^2, and Style Injection Diffusion Model (SI-DM). Each model was evaluated visually and quantitatively using six metrics (SSIM, LPIPS, FID, histogram similarity, ArtFID, and LSCI) and validated on three major wildfire events in Korea (2022–2025). Among the tested models, SI-DM achieved the most balanced performance, preserving structural features while ensuring consistent color-tone alignment (ArtFID = 1.620; LSCI mean = 0.894). Qualitative assessments further confirmed that SI-DM effectively delineated smoke boundaries and maintained natural background tones under complex atmospheric conditions. Additional analysis using GEMS UVAI, VISAI, and CHOCHO demonstrated that the styled composites partially reflect the optical and chemical characteristics distinguishing wildfire smoke from dust aerosols. The proposed SFCC framework establishes a foundation for visually standardized satellite smoke imagery and provides potential for future aerosol-type classification and automated detection applications.

Keywords: geostationary environment monitoring spectrometer; wildfire smoke visualization; deep learning; style transfer; color-tone standardization

1. Introduction

Wildfires are disasters that pose a serious threat to natural ecosystems and human societies worldwide, and in recent years, they have shown an increasing trend in both frequency and intensity. In particular, as climate change, the persistence of hot and dry environments, and anthropogenic factors act in combination, the likelihood of large wildfires has become much higher than in the past, and in fact, in the United States, Australia, Canada, and Russia, megafires in which more than hundreds of thousands of hectares of forest are lost have been frequently reported [1–3]. Such wildfires go beyond localized damage and release massive amounts of smoke and delicate particulate matter into the atmosphere, causing large-scale air pollution that can lead to a variety of secondary impacts, such as respiratory diseases, aviation disruptions, and disturbances to the climate system [4,5]. Recent studies have further shown that exposure to wildfire smoke has increased substantially over the past decade, exacerbating its impacts on air quality, public health, and climate systems amid ongoing climate change [6,7]. Therefore, technologies that can precisely detect and monitor how smoke spreads in space and time generated by wildfires have become essential in fields such as air quality forecasting, disaster response, and climate change research.

For these reasons, satellite-based smoke detection has been employed as an essential tool to minimize air pollution and secondary damage from wildfires. Early studies on smoke detection primarily relied on rule-based classification using multi-threshold algorithms in the visible and near-infrared bands, which distinguished smoke pixels based on brightness or color differences [8,9]. Chrysoulakis et al. [10] detected smoke-affected areas based on the Normalized Difference Vegetation Index (NDVI) and reflectance differences, and then estimated the entire smoke plume through spatial expansion. Wang et al. [11] identified smoke pixels in Moderate Resolution Imaging Spectroradiometer (MODIS) imagery by applying various band combinations in a physics-based approach. They employed K-means clustering and thresholding to extract smoke regions in a physically informed manner. While these studies demonstrated the feasibility of quantifying smoke detection, they also revealed structural limitations, as consistent rules are complex to apply across varying atmospheric background conditions and satellite image quality.

To address these challenges, deep learning-based approaches have emerged that leverage large training datasets to improve generalization performance. Through deep learning-based image recognition techniques, various methods have been proposed to integrate spatial and spectral contextual information for smoke detection. These studies can broadly be categorized into (1) scene-level classification approaches, which determine whether smoke is present in an image, and (2) pixel-level segmentation approaches, which distinguish smoke pixels from background pixels to represent the spatial distribution of smoke. First, Ma et al. [12] proposed a CNN-based smoke-detection method that uses MODIS RGB satellite imagery as input. Their model incorporated both spatial and channel attention modules, forming a scene-level classification structure designed to determine the presence of smoke at the image level. This approach evaluated the generalizability of smoke detection across visually similar atmospheric conditions, including smoke, clouds, and dust. Zhao et al. [13] introduced the Variant Input Bands for Smoke Detection (VIB_SD) model based on Landsat-5/8 imagery, which exploited the characteristic that smoke tends to absorb more and reflect less in the near-infrared (NIR) and shortwave infrared (SWIR) bands. Zhao et al. [14] investigated cross-sensor transfer learning for fire-smoke scene detection using variable-band multi-spectral satellite imagery, demonstrating enhanced generalization when deploying a model trained on one sensor to data from another. By learning spatial contextual information from the contrasting spectral characteristics of smoke and background pixels, their method determined the presence of smoke in the imagery.

More recently, pixel-level semantic segmentation studies have focused on delineating the spatial boundaries and distribution of wildfire smoke using deep learning models trained on satellite imagery. These approaches aim to explicitly represent smoke morphology, plume structure, and spatial distribution in addition to simple presence detection. Liu et al. [15] proposed a transformer-boosted U-Net architecture for smoke segmentation in complex backgrounds using multispectral Landsat imagery, demonstrating superior delineation performance by capturing long-range contextual features relative to standard convolutional models. Similarly, Andrade et al. [16] presented a real-time satellite image segmentation framework for classifying smoke and cloud pixels from Geostationary Operational Environmental Satellites (GOES) imagery, achieving high accuracy and intersection-over-union (IoU) scores while enhancing the robustness of smoke–cloud separation in operational monitoring scenarios. Despite these advances, segmentation accuracy remains highly sensitive to radiometric inconsistencies and visual ambiguities among smoke, clouds, and haze. To address the challenge of indistinct smoke boundaries in complex background conditions, Wang et al. [17] proposed the Attention-Guided Optical Satellite Video Smoke Segmentation Network (AOSVSSNet), which leverages spatiotemporal patterns from satellite video data to enhance robustness against illumination changes and noise. Nonetheless, these pixel-level segmentation methods remain highly dependent on the quality and consistency of input imagery, and still face limitations in reliably distinguishing smoke from visually similar atmospheric features such as clouds and fog.

So far, studies have made meaningful progress in improving the accuracy and automation of wildfire smoke detection using satellite imagery. Traditional threshold-based methods relied on simple rules to visually detect smoke distributions, whereas deep learning techniques have achieved high classification accuracy and fine-grained segmentation. Nevertheless, challenges remain: distinguishing smoke from other atmospheric phenomena is difficult, and limitations in image quality—such as inconsistent colors across different times and regions—impose constraints. These issues negatively affect the generalization performance of deep learning models [18–20]. To enhance the reliability and usability of

https://doi.org/10.3390/rs18030483

satellite-based smoke detection, such problems must be resolved. However, classical statistical methods such as the Empirical Cumulative Distribution Function (ECDF) alone are insufficient to achieve high-quality imagery. Therefore, it is essential to obtain high-quality imagery from a sequence of time-series satellite images that preserves visual structures while maintaining a consistent color tone [21].

In addition, because wildfire smoke spreads rapidly, high-temporal-resolution imagery is required for its effective detection and tracking. To this end, we employed data from the Geostationary Environment Monitoring Spectrometer (GEMS), which provides hourly daytime images. Owing to its ability to offer high-precision hourly observations over East Asia, GEMS has been widely employed in various environmental monitoring studies, such as hazardous gas surveillance and air quality analysis [22,23]. However, GEMS imagery acquired as a time series tends to exhibit inconsistencies in visual brightness and color tone due to factors such as acquisition time, atmospheric conditions, and variations in solar altitude. These inconsistencies can affect the consistency and reliability of smoke detection.

To address these challenges and improve the visual quality of GEMS-based smoke visualization, this study proposes a Standardized False Color Composite (SFCC) generation method using style transfer techniques. We designed an image standardization approach that applies various style transfer models to normalize tonal variations while preserving structural visual features. In this study, style transfer does not refer to the modification of artistic properties, such as texture or aesthetic patterns, commonly found in computer vision. Instead, it specifically denotes the alignment of image representations—including color distribution and luminance balance—across diverse observational and solar conditions. Consequently, the proposed framework aims to normalize representational variations, ensuring consistent visual interpretation across different temporal and spatial contexts. Its performance is evaluated using six objective metrics. This enables a quantitative assessment of visual consistency across different times and observational conditions, as well as an examination of the applicability and limitations of style transfer techniques for satellite imagery. We show that SFCC reduces inconsistencies caused by the spectral similarity between smoke and clouds and mitigates analytical discrepancies arising from tonal differences across time-series imagery.

2. Technical Background

2.1. Evolution of Style Transfer

Style transfer was initially developed to apply the visual style of one image to another. Early approaches mainly performed simple color transformations, but with advances in deep learning, the technique has expanded to transfer higher-level visual characteristics. In the standard formulation, two inputs are used—a content image and a reference style image (RSI)—and the model synthesizes a new image that preserves the content's structural features while imposing the style's color tone and texture.

Deep learning-based style transfer began with the Neural Style Transfer (NST) proposed by Gatys et al. [24]. Their method, built on a pre-trained VGG-19 network, represents image content using activation features from intermediate layers, and represents style using Gram-matrix statistics aggregated across multiple layers. The final image is produced by an iterative optimization that minimizes a weighted sum of content and style losses, demonstrating for the first time the practical application of artistic styles to high-resolution natural images. While impressive results were shown, the approach faces practical constraints: tens of seconds to minutes of computation per image, substantial computational cost, and the potential for high-frequency artifacts.

To address computational inefficiencies, Huang and Belongie [25] introduced Adaptive Instance Normalization (AdaIN). AdaIN aligns the mean and variance of the content features to match the statistical moments of the style features, enabling real-time style transfer through a single feed-forward network without requiring explicit optimization of the loss function. The method supports arbitrary style transfer and achieves a significant improvement in processing speed. In experiments on the COCO dataset and various artistic painting styles, AdaIN achieved high perceptual quality and intense user satisfaction. However, because it relies only on first-order statistics (mean and variance), it is limited in its ability to represent fine-grained, spatially localized style details.

Subsequently, StyleGAN expanded the concept of style transfer to the broader field of image generation, integrating a style control mechanism into the traditional GAN framework [26]. This model maps the latent vector to an intermediate latent space and applies AdaIN operations at each convolutional layer, allowing multi-scale modulation of style attributes. However, because the model was primarily designed for image generation rather than direct style transfer, it has structural limitations when applied to existing content images.

To address these limitations, Liu et al. [27] proposed Adaptive Attention Normalization (AdaAttN). This method generates an attention map to assign spatially varying weights during style normalization, enabling the model to integrate both global and local features and thereby overcome the shortcomings of AdaIN. In particular, the introduction of a local feature loss strengthened the representation of regional visual quality. It confirmed the perceptual quality and stable performance across diverse style images. However, the attention mechanism inherently increases memory consumption and slows inference speed, which remain practical challenges. Although the attention mechanism requires significant computational resources, this is being addressed by the rapid advancements in hardware.

2.2. Applications of Style Transfer in Satellite Imagery

Satellite imagery can exhibit significant variations in color and visual quality for the same region, depending on observation time, sensor characteristics, atmospheric conditions, and solar altitude. Such visual inconsistencies can negatively affect the reliability of image interpretation and deep learning-based analysis. In this context, style transfer techniques have gained attention as an effective means of correcting visual discrepancies and reconciling representational differences between domains. Recently, several studies have explored the application of style transfer to remote sensing imagery.

Karatzoglidi et al. [28] first demonstrated the potential applicability of style transfer to satellite imagery through an experiment that applied artistic styles to remote sensing images. Using Sentinel-2 data, they performed semantic segmentation to separate water and land classes, and then applied distinct style images to each class to produce a composite collage. This approach meaningfully expanded the visual diversity and potential uses of satellite imagery. However, the results exhibited inconsistent representations across surface types and visual confusion in unsegmented areas, revealing limitations in semantic coherence. Consequently, their study highlighted that semantic consistency and region-specific style correspondence are critical when applying style transfer to satellite imagery.

Sokolov et al. [29] proposed the HR-SemI2I framework, based on semantically consistent image-to-image translation (SemI2I), to reduce style discrepancies between different satellite images while maintaining semantic consistency. Their approach focused on aligning the visual representations of WorldView-2 and SPOT-6 imagery through style transfer, without degrading segmentation accuracy. During transformation, the framework combined an AdaIN-based architecture with cross-reconstruction, self-reconstruction, and gradient loss terms to strengthen structural preservation. Experimental results showed improved semantic segmentation accuracy compared to conventional methods. They demonstrated that style transfer can effectively support domain alignment and semantic preservation across heterogeneous satellite datasets, establishing a foundation for subsequent semantics-aware image translation research in remote sensing.

Sun and Wu [30] presented an application of style transfer for generating map-style images from satellite imagery. They extracted color and texture features from reference imagery across different regions. They aligned them with semantic segmentation results using a cross-attention mechanism, enabling class-specific transfers for categories such as roads, water bodies, and vegetation. Furthermore, they employed a Latent Diffusion-based architecture that enabled style layering and the preservation of semantic hierarchy, both of which are suitable for cartographic visualization. In experiments, their method achieved higher structural preservation and visual consistency compared with conventional GAN-based approaches. They demonstrated a specialized application of style transfer for map-style rendering and experimentally verified that the technique can achieve a balance between maintaining semantic hierarchy and maintaining visual coherence in remote sensing imagery.

Within the context of satellite-based environmental monitoring, systematic evaluations and practical implementations of style transfer techniques remain scarce. In particular, color-tone consistency and visual quality enhancement in time-series satellite imagery are crucial for reliable interpretation and automated analysis, yet have been scarcely investigated. Consequently, a technical gap remains in methodologies that can effectively apply style transfer to real-world remote sensing applications, such as smoke visualization and image standardization. Our study will be a significant attempt to resolve this.

3. Materials and Methods

3.1. GEMS Datasets

Operational GEMS Level-1/Level-2 products undergo periodic updates as retrieval algorithms, calibration parameters, and quality-screening procedures are refined. In accordance with the GEMS processing system's data policy, major algorithm updates typically trigger a full reprocessing of the mission archive—dating back to the commencement of routine observations in 2021—using the updated processing chain. These reprocessed products are subsequently redistributed via the archive. For this study, GEMS L1C radiance and L2 products were acquired within a single period, utilizing the latest reprocessed release available at the time of analysis.

3.1.1. L1C Radiance Data

The wildfire smoke visualization was conducted using the GEMS sensor onboard the GK-2B satellite. GEMS is a geostationary hyperspectral environmental satellite that observes the entire East Asian region approximately 8 times per day during daylight hours. It provides fine spectral sampling between 300 and 500 nm at 0.6 nm intervals, enabling real-time monitoring of atmospheric pollutants over East Asia by measuring ultraviolet radiative energy. Ultraviolet spectral measurements are particularly effective for aerosol detection over both land and ocean surfaces because they are less affected by surface reflectance. Such hyperspectral satellite sensors provide continuous, narrow-band spectral information, enabling precise identification of the unique spectral signatures of atmospheric pollutants. In particular, the ultraviolet range can effectively distinguish absorbing aerosols (e.g., smoke, soot, and desert dust) from non-absorbing aerosols (e.g., clouds, haze). Wildfire smoke is characterized by a significant difference between the 340 nm and the 380 nm radiance.

3.1.2. L2 Ultraviolet Aerosol Index (UVAI)

The Ultraviolet Aerosol Index (UVAI) is a dimensionless index expressed on a logarithmic scale that represents the spectral contrast between measured reflectance at 354 nm and 388 nm and the corresponding normalized radiance calculated for a Rayleigh atmosphere under aerosol-free conditions (Equation (1)).

$$UVAI = -100\left[log\left(\frac{N_{\lambda 1}}{N_{\lambda 2}}\right) - log\left(\frac{N_{\lambda 1}(A_{LER})}{N_{\lambda 2}(A_{LER})}\right)_{Rayleigh}\right],\ (\lambda 1 = 354\ \text{nm},\ \lambda 2 = 380\ \text{nm}) \quad (1)$$

The UVAI utilizes the difference in shortwave radiance to indicate the degree of ultraviolet absorption by aerosols. A higher (or more positive) UVAI value suggests stronger absorption at shorter wavelengths, a condition typical when absorbing fine aerosols, such as carbonaceous particles from wildfire smoke, are present. Additionally, for the same aerosol loading, UVAI tends to increase when aerosols are located at higher altitudes, resulting from enhanced background contrast. Because UVAI is more sensitive to absorption strength, layer height, and background brightness than to aerosol optical depth (AOD), it was employed as a supplementary indicator to confirm the presence of absorbing aerosols and to evaluate the consistency between physical absorption features and the color-tone variations observed in our SFCC images. UVAI products are accompanied by quality and validity flags generated by the operational retrieval algorithms. Accordingly, we utilized this official GEMS quality-screening information to exclude invalid or low-confidence retrievals, such as those affected by extreme viewing geometries, sun glint, or optically thick clouds.

3.1.3. L2 Visible Aerosol Index (VISAI)

The Visible Aerosol Index (VISAI) is a dimensionless index expressed on a logarithmic scale that represents the spectral contrast between the measured reflectance at 477 nm and 490 nm and the corresponding normalized radiance calculated for a Rayleigh atmosphere under aerosol-free conditions (Equation (2)).

$$VISAI = -100\left[log\left(\frac{N_{\lambda 1}}{N_{\lambda 2}}\right) - log\left(\frac{N_{\lambda 1}(A_{LER})}{N_{\lambda 2}(A_{LER})}\right)_{Rayleigh}\right],\ (\lambda 1 = 477\ \text{nm},\ \lambda 2 = 490\ \text{nm}) \quad (2)$$

The index is derived by matching the Lambertian Equivalent Reflectance (LER) at 490 nm and applying the same LER to 477 nm, thereby quantifying the discrepancy between observed and Rayleigh-modeled reflectance. Because absorption effects are weaker in the visible region than in the ultraviolet, VISAI primarily reflects Ångström-type variations related to particle size. In general, fine-mode-dominated plumes (e.g., wildfire smoke, urban pollution) exhibit high VISAI values, whereas coarse-mode-dominated plumes (e.g., dust) tend to have low or near-zero VISAI values. VISAI is more sensitive to factors such as the fine-to-coarse mode ratio, surface Bidirectional Reflectance Distribution Function (BRDF), and sea-surface glint than to total AOD. VISAI products are provided with quality and validity flags generated by the operational retrieval algorithms. Leveraging this official GEMS quality-screening information, we excluded invalid or low-confidence retrievals, particularly those associated with extreme viewing geometries, sun glint, and optically thick clouds. VISAI was used as auxiliary validation data to verify particle-size differences between wildfire smoke and yellow dust identified in our SFCC images. The classification was complemented with glyoxal (CHOCHO) data, in which VISAI served as an indicator of particle size, and CHOCHO provided information on combustion-related characteristics of smoke aerosols.

3.1.4. L2 Glyoxal (CHOCHO)

Glyoxal (CHOCHO) is a trace gas pollutant produced by biomass burning (e.g., forest and grassland fires) or through the oxidation of volatile organic compounds (VOCs). The GEMS sensor retrieves CHOCHO concentrations using its absorption features between approximately 433 and 461 nm, first deriving the Slant Column Density (SCD) and then converting it to the Vertical Column Density (VCD) via the Air Mass Factor (AMF) obtained from the observed spectra. Previous studies have reported that CHOCHO VCD increases significantly during wildfire events, with smoke plumes persisting for several tens of hours and being transported over long distances—up to approximately 1500 km downstream. These findings support CHOCHO as a reliable satellite indicator for detecting and tracking wildfire smoke [31].

CHOCHO was also used as auxiliary validation data to verify whether smoke plumes observed in our SFCC images could be physically and chemically distinguished. Specifically, VISAI reflects the particle-size tendency (fine vs. coarse mode), while CHOCHO serves as an indicator of combustion origin. The simultaneous variation of these two indices was examined to assess the feasibility of distinguishing wildfire smoke from yellow dust using the SFCC-based visualization. CHOCHO VCD products are provided with quality and validity flags generated by the operational retrieval algorithm. Accordingly, we applied the official GEMS quality-screening criteria to exclude invalid or low-confidence retrievals.

3.2. Initial FCC Image

Preprocessing of the L1C radiance data includes removing outliers caused by sun-glint, excessive solar zenith angles (>70°), and viewing zenith angles (>70°). The initial False Color Composite (FCC) was created using the preprocessed L1C radiance at 340 nm and 380 nm. The radiance values were normalized to the 0–255 range for convenient color scaling. The initial FCC images were created using the radiance difference between 340 nm and 380 nm, by applying ΔUV (Radiance (380 nm)–Radiance (340 nm)) to the red channel, Radiance (380 nm) to the green channel, and Radiance (340 nm) to the blue channel (Table 1). Three major wildfire cases (Table 2 and Figure 1) were analyzed: Case 1 (2022): A large-scale wildfire in the Uljin–Samcheok region (Gangwon Province) that spread rapidly under strong winds; Case 2 (2023): An inland wildfire in Hongseong (Chungcheongnam-do); Case 3 (2025): An extensive wildfire event across Uiseong and Andong (Gyeongsangbuk-do) with significant burned areas.

Table 1. Main channels and their wildfire smoke-related features.

Channel	GEMS Radiance	Description
Red	380 nm–340 nm (ΔUV)	Highlights aerosol contrast; smoke appears strongly in the difference band.
Green	380 nm	Higher reflectance for UV-absorbing aerosols.
Blue	340 nm	Lower reflectance for smoke.

Table 2. Overview of Korean wildfire events (2022–2025) used in this study.

Case	Region	Wildfire Period	Burned Area (ha)
Case 1 (2022)	Uljin–Samcheok, Gangwon	3–13 March 2022	27,995
Case 2 (2023)	Hongseong	2–4 April 2023	1337
Case 3 (2025)	Uiseong, Andong	22–30 March 2025	26,708

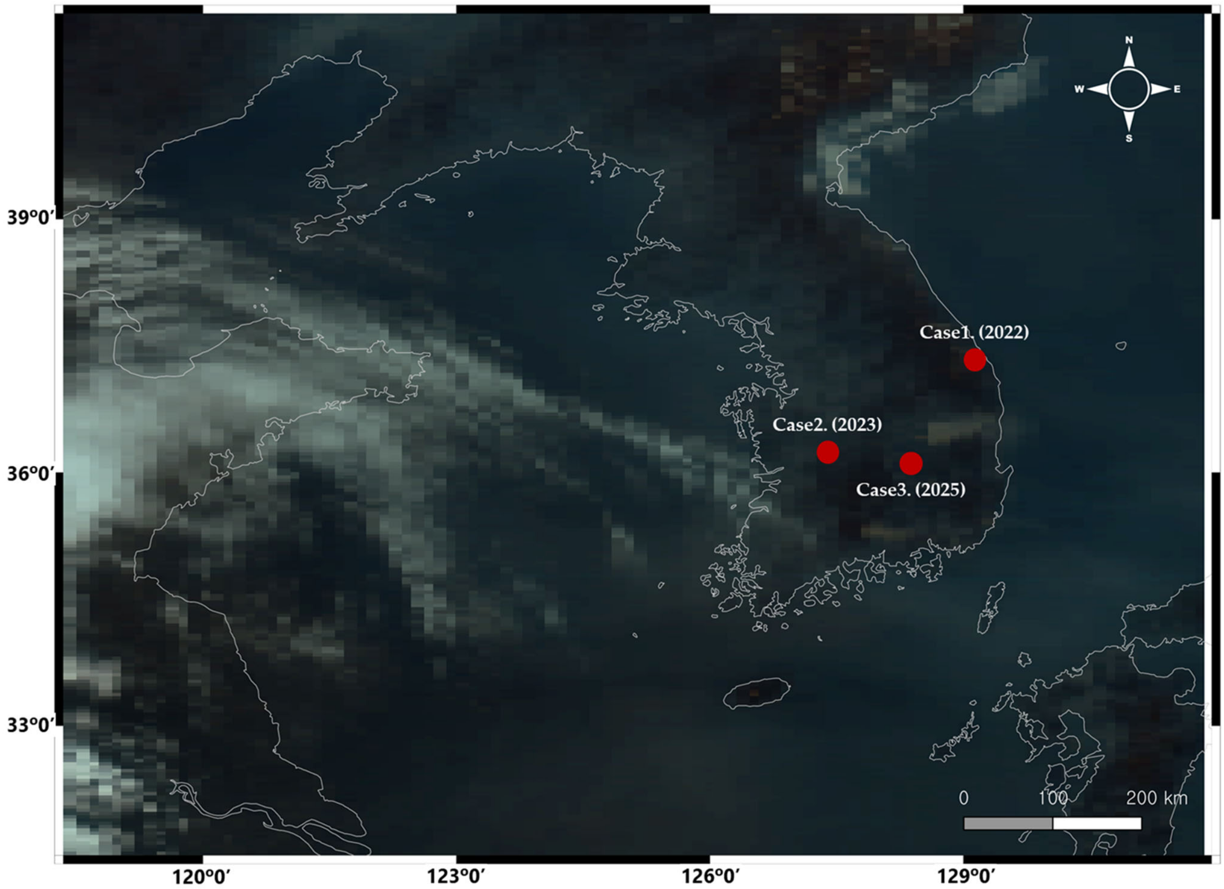

Figure 1. Locations of Korean wildfire events (2022–2025) used in this study.

However, the initial FCC images exhibited color-tone inconsistencies across observation times, attributed to differences in illumination intensity, scattering conditions, and changes in the solar zenith angle (SZA). Because GEMS has a spectral resolution of 0.6 nm, it is highly susceptible to even subtle variations in radiance values. In the ultraviolet region, the Top-of-Atmosphere (TOA) reflectance is known to fluctuate significantly with SZA, aerosol loading, cloud distribution, and meteorological conditions. Previous studies using TROPOMI data have reported that UV reflectance can vary markedly over time and across atmospheric states, even within the same region, with these effects becoming more pronounced in high-resolution imagery [31]. Similarly, the TOMS algorithm indicated that variations in Ultraviolet (UV) reflectance are strongly influenced by solar radiation geometry and aerosol scattering properties [32]. To mitigate these color-tone inconsistency issues, we create new SFCC images that ensure color-tone consistency across time and observation conditions.

Figure 2 presents the initial FCC wildfire smoke imagery generated using GEMS data. It demonstrates that, even on the same day, the image color tone varies with observation time, leading to inconsistent visual outputs. Similarly, Figure 3 shows that when wildfire smoke is present, the color tone of the initial FCC image continues to change over time. These temporal color inconsistencies cause the model to perceive different patterns even under identical surface and atmospheric conditions, thereby reducing training efficiency and generalization performance. To mitigate these issues, we propose generating SFCC imagery to ensure tonal consistency across time and observational conditions. The overall workflow of the proposed approach is illustrated in Figure 4.

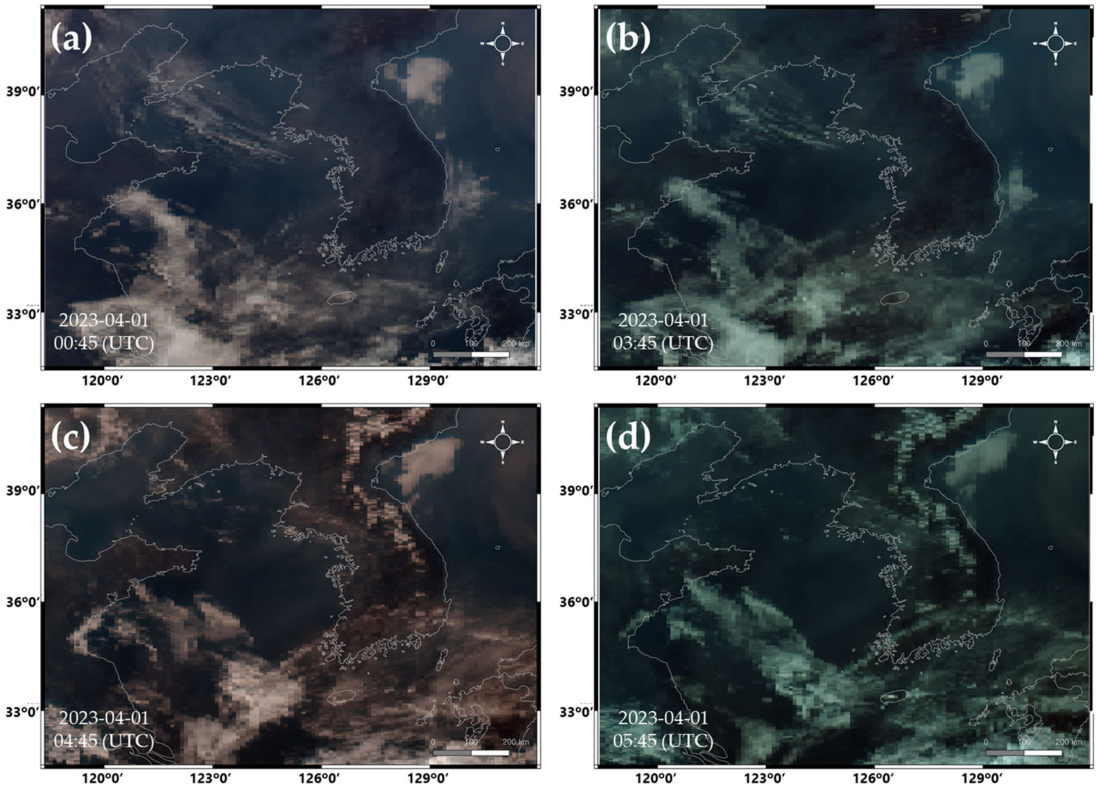

Figure 2. Initial FCC images generated from GEMS radiance data on 1 April 2023 show color-tone variation on the same day under clear-sky conditions without wildfire smoke. (**a**) 00:45 UTC, (**b**) 03:45 UTC, (**c**) 04:45 UTC, (**d**) 05:45 UTC.

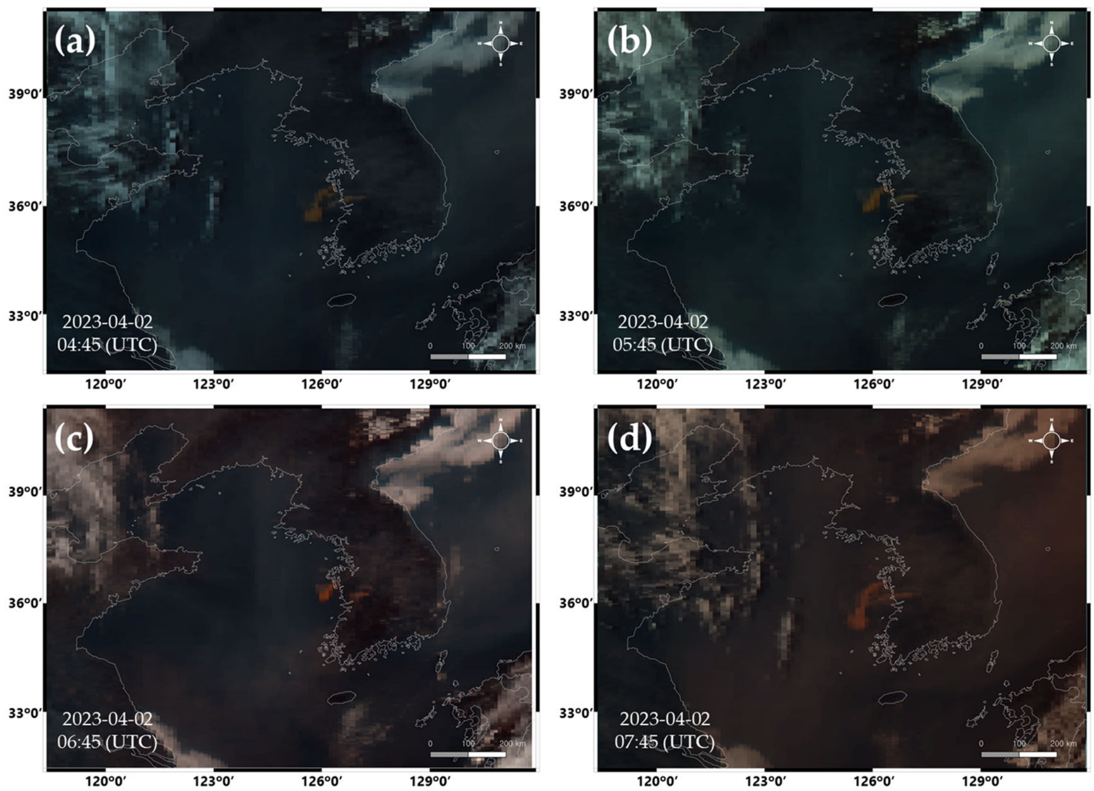

Figure 3. Initial FCC images generated from GEMS radiance data on 2 April 2023, showing color-tone variation on the same day in the presence of wildfire smoke. (**a**) 04:45 UTC, (**b**) 05:45 UTC, (**c**) 06:45 UTC, (**d**) 07:45 UTC.

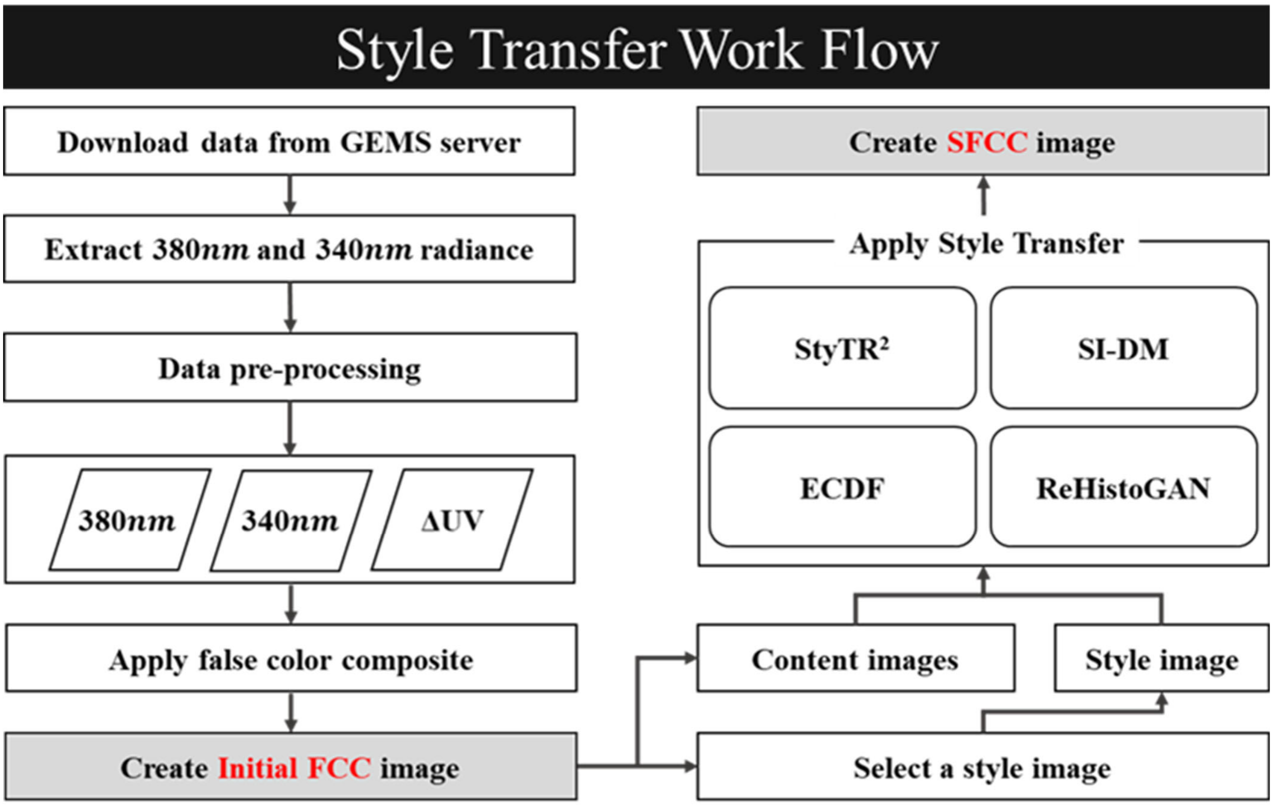

Figure 4. Workflow of initial FCC and the SFCC creation using GEMS radiance of 340 nm and 380 nm.

3.3. Style Transfer Models for SFCC

To generate the SFCC images, we used one statistical correction method and three deep learning-based style transfer models. Each method performs content structure preservation and color distribution adjustment separately, and their effectiveness is expected to vary with the unique characteristics of satellite imagery.

3.3.1. Empirical Cumulative Distribution Function (ECDF) Correction

The ECDF is a statistical image correction method. Using the ECDF for a target image and the ECDF for the reference image, the corrected pixel values are adjusted like Figure 5. In Equation (3), x represents the pixel value of the target image, x' denotes the corrected pixel value, $F_I(x)$ indicates the ECDF value of the target image, and F_{ref}^{-1} refers to the inverse ECDF of the reference image.

$$x' = F_{ref}^{-1}(F_I(x)) \tag{3}$$

We adopted an ECDF-based correction to address temporal reflectance variation and sensor-induced quality imbalance arising from the GEMS satellite's repeated observations. A representative cumulative distribution was first computed from all GEMS images, and then each image's pixel values were aligned with this reference distribution to perform color standardization. Unlike learning-based models, this statistical approach requires no additional labels or external data, relying solely on the inherent statistical properties of the imagery for normalization. This helps improve the overall stability and visual consistency of satellite imagery products.

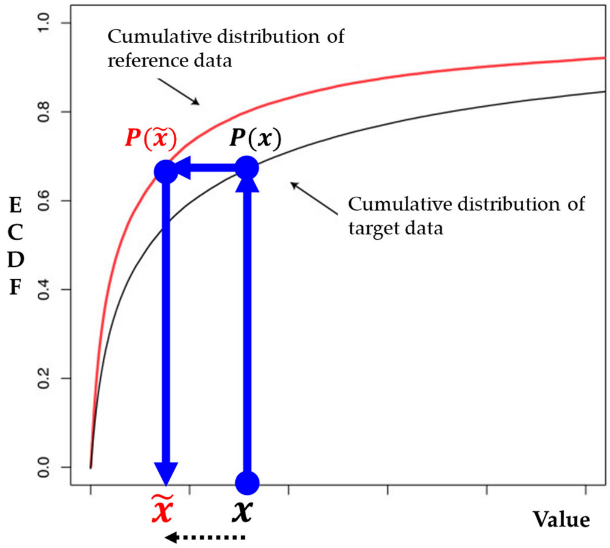

Figure 5. Color-tone correction procedure using ECDF. The original value x is transformed into $\tilde{x}$ by referencing the cumulative distribution of reference data, as indicated by the dotted arrow.

3.3.2. ReHistoGAN

ReHistoGAN is a GAN-based image recoloring framework that extends HistoGAN to process real images in an unsupervised manner, offering explicit control over color distributions while preserving structural integrity [33]. The model is designed to mitigate the structural degradation commonly observed in color and style transfer tasks by combining histogram-based color alignment with structure-preserving constraints. Architecturally, ReHistoGAN consists of two primary components: (1) a StyleGAN-based generator head conditioned on differentiable color histogram features, and (2) an encoder–decoder network that maps a real input image into the generator's latent space. Color information is represented via a differentiable histogram in log-chroma space, facilitating robust color control while minimizing sensitivity to illumination variations (Figure 6). This design enables the model to align global color statistics without requiring semantic correspondence between the input and reference images. ReHistoGAN is optimized using a composite objective that balances color consistency with structural preservation. Histogram alignment is formulated as a distribution mapping between the output image and the RSI, expressed as follows:

$$I_{out}(p) = H_s^{-1}(H_c(I_c(p))) \tag{4}$$

Here, H_c denotes the cumulative distribution of the content image, and H_s^{-1} represents the inverse cumulative distribution of the RSI. In addition, ReHistoGAN incorporates a Structural Similarity Index Measure (SSIM)-based loss to preserve structural similarity and an edge-preservation loss to maintain fine spatial details, such as terrain boundaries and smoke contours, in satellite imagery (Equation (5)).

$$L_{structure} = 1 - SSIM(I_c, I_{out}) \tag{5}$$

As described in the original study, ReHistoGAN is trained in a fully unsupervised manner on large-scale natural image datasets across diverse domains, including human faces, animals, landscapes, and aerial imagery [33]. The training procedure initializes the generator head with pre-trained HistoGAN weights and jointly optimizes the encoder–decoder network, enabling the model to learn a content-preserving mapping that matches target color distributions. In this work, we employ ReHistoGAN in an inference-only configuration using the official pre-trained weights, foregoing additional fine-tuning on satellite imagery. This implementation choice ensures a controlled comparison across models and facilitates an evaluation of ReHistoGAN's intrinsic ability to harmonize color tones and preserve structures in GEMS-based false color composites.

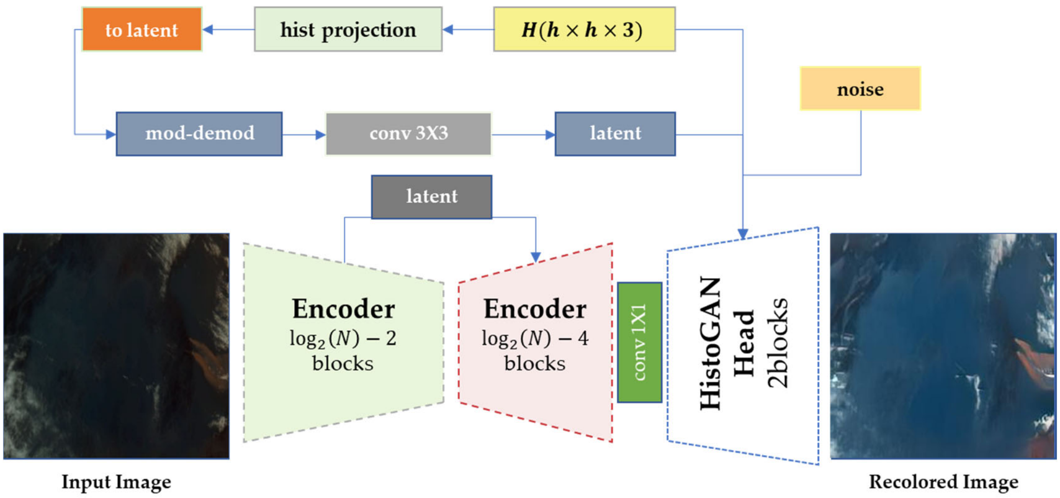

Figure 6. Overall architecture of the ReHistoGAN model.

3.3.3. StyTr^2

StyTr^2 is a transformer-based arbitrary style transfer network designed to explicitly model long-range dependencies between content and style via attention mechanisms [34]. The framework employs two separate transformer encoders—dedicated to the content image and the RSI, respectively—and aligns their representations within a multi-layer transformer decoder through cross-attention. This architecture is intended to preserve structural details in the content while capturing style-related appearance cues from the RSI more effectively than conventional, purely convolutional pipelines.

The model encodes the content image I_c and the RSI I_s using independent transformer encoders E_c and E_s.

$$F_c = E_c(I_c), \quad F_s = E_s(I_s) \tag{6}$$

Here, F_c and F_s represent the feature embeddings of the content and RSI, respectively. During decoding, color-tone harmonization is achieved through a cross-attention mechanism, where the style tokens influence the content tokens as follows:

$$O = Decoder(Attention(F_c, F_s)) \tag{7}$$

The style tokens are directly injected into the decoder to guide the style representation, thereby achieving style alignment while simultaneously preserving structural integrity (Figure 7). StyTr^2 additionally introduces Content-Aware Positional Encoding (CAPE) to enhance scale-invariant stylization and stabilize structural preservation across varying image resolutions. Following the transformer decoder, a lightweight CNN-based decoder progressively upsamples and refines the patch-level outputs to produce the final stylized image.

As described in the original study, StyTr^2 is trained on large-scale natural image datasets, utilizing MS-COCO for content and WikiArt for style. Training images are randomly cropped to a fixed resolution of 256 × 256 pixels. The network is optimized using the Adam optimizer with an initial learning rate of 5×10^{-4} and a warm-up adjustment strategy, employing a batch size of 8 over 160,000 iterations [34]. The objective function combines perceptual content and style losses derived from a pre-trained VGG-19 network, alongside identity losses to enhance representation fidelity and stabilize the stylization process.

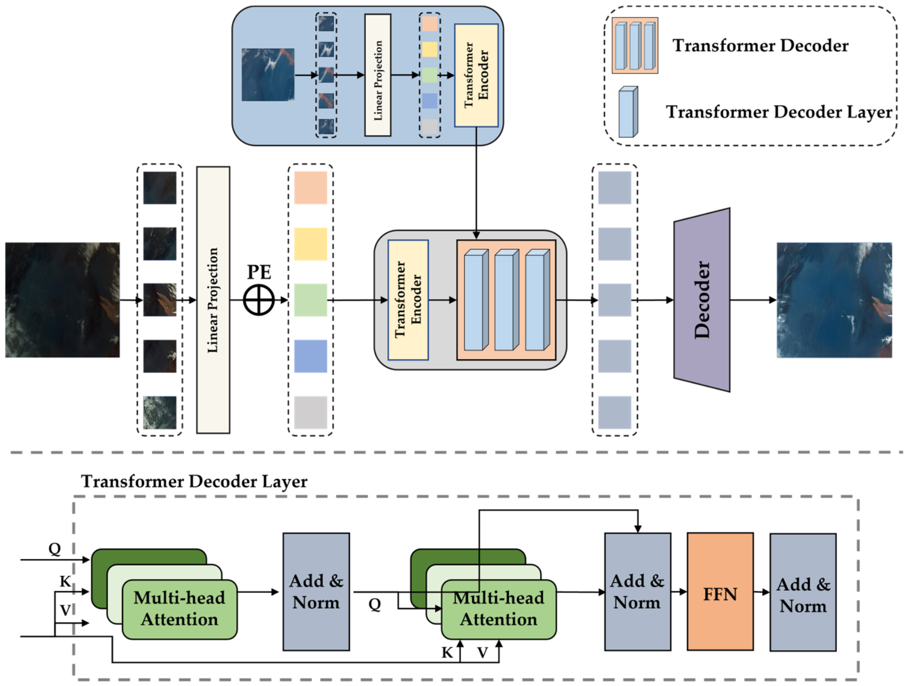

Figure 7. Network architecture of the Style Transfer 2 (StyTr^2) model.

3.3.4. Style Injection Diffusion Model (SI-DM)

The Style Injection Diffusion Model (SI-DM) leverages a pre-trained diffusion model architecture to achieve both structural preservation and color-tone standardization by injecting style information at intermediate denoising stages. The overall architecture and style injection mechanism of SI-DM are illustrated in Figure 8. The fundamental reverse denoising process of a diffusion model is an iterative procedure that progressively reconstructs the clean image from a noisy input [35].

$$x_{t-1} = f_\theta(x_t,\ t) \tag{8}$$

Here, x_t represents the noisy image at timestep t, and f_θ denotes the learned denoising function. In SI-DM, style injection is implemented through an attention mechanism applied at intermediate stages. Specifically, the Query (Q) is derived from the content image's latent features, while the Key (K) and Value (V) are derived from the RSI's latent features. This process forms a self-attention-based structure, allowing the diffusion model to incorporate style information dynamically while maintaining the spatial and structural integrity of the content representation.

$$Attention(Q, K, V) = softmax\left(\frac{QK^\tau}{\sqrt{d_k}}\right)V \tag{9}$$

SI-DM is built upon a pre-trained latent diffusion foundation model and performs example-based style transfer by injecting style features during the reverse diffusion process [35]. Instead of updating the model weights, this method intervenes directly in the self-attention layers of the denoising U-Net. Specifically, content-derived features serve as queries, while style-derived keys and values are extracted from the RSI. This design facilitates the transfer of color-tone characteristics while strictly maintaining the spatial structure.

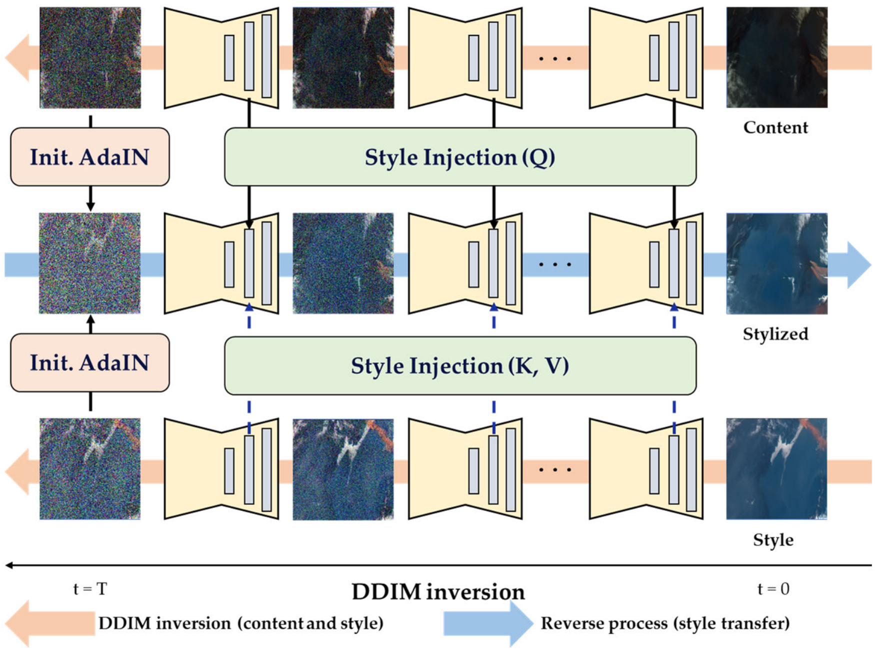

Figure 8. Overall framework of the Style Injection Diffusion Model (SI-DM).

In the original study, SI-DM utilizes Stable Diffusion—trained on the large-scale LAION dataset—as its backbone. The training configurations (e.g., optimizer, learning rate schedule, and batch size) align with those of the released foundation model. Style transfer is subsequently executed in a training-free manner using DDIM inversion and sampling. Fixed inference hyperparameters, such as the number of diffusion steps and injection strength, are used to modulate the content–style trade-off. We employ the official pre-trained weights and apply SI-DM to GEMS false color composites in an inference-only configuration, without additional fine-tuning on satellite imagery.

3.3.5. Implementation Details

All deep learning-based style transfer models used in this study—ReHistoGAN, StyTr^2, and SI-DM—were implemented using pre-trained weights released by their original authors. All implementations were based on the PyTorch framework (ver. 1.11), and both the publicly available source codes and trained parameters were obtained from the respective official repositories. No fine-tuning or additional training was performed on GEMS data. This design choice was intentional, as all models were evaluated under identical inference settings to assess their intrinsic generalization capability when directly applied to GEMS FCC imagery, without introducing task- or sensor-specific adaptation. Despite the absence of fine-tuning, the proposed SFCC framework demonstrates stable performance across multiple wildfire cases under varying observational conditions, supporting its applicability to satellite imagery without task-specific retraining.

For SI-DM, inference was conducted using a latent diffusion backbone initialized from Stable Diffusion v1.4, with style information injected into intermediate self-attention layers during the reverse diffusion process. The primary inference hyperparameters—including attention injection layers, the temperature parameter, and the style strength γ—aligned with the recommended settings reported in the original work, as detailed in Table 3. For ReHistoGAN and StyTr^2, inference was performed using the officially released pre-trained networks with fixed weights. The corresponding histogram-matching parameters (ReHistoGAN) and transformer/VGG-based encoder–decoder components (StyTr^2) were utilized as provided by their respective original implementations. All experiments were executed on a workstation equipped with an NVIDIA RTX 3090 Ti GPU (24 GB VRAM) within a unified software environment. To ensure a consistent evaluation, all style transfer experiments were conducted on the same SFCC images. The RSI was selected as the uniform reference across all models to maintain rigorous comparability.

Table 3. Implementation summary of the used style transfer models.

Model	Pre-Trained	Fine-Tuning	Inference Configuration	Framework	Hardware
ReHistoGAN	Yes (official)	No	network capacity = 18 pyramid levels = 8 histogram σ = 0.04 γ = 4.0 α = 16 β = 3.0	PyTorch	RTX 3090 Ti
StyTr^2	Yes (official)	No	fixed pre-trained VGG, transformer, decoder weights	PyTorch	RTX 3090 Ti
SI-DM	Yes (Stable Diffusion v1.4)	No	attention injection layers = 9, 10, 11 γ = 0.3 temperature T = 1.5	PyTorch	RTX 3090 Ti

3.4. Selection of Reference Style Image (RSI)

A single RSI was selected to serve as the basis for the style transfer process. In style transfer, the RSI defines the color tone and texture characteristics to be applied to all content images; therefore, its selection was primarily guided by image quality and visual representativeness.

The selection process consisted of two main steps. First, images were screened to ensure that at least one pixel had an Aerosol Index (AI) value $\geq$ 6 and that the cloud coverage ratio was between 30% and 60%. These criteria ensured that the RSI included wildfire smoke (orange tone in SFCC), clouds (white), and clear-sky or ocean areas simultaneously—conditions under which global color-tone alignment could be achieved across diverse atmospheric components. Second, the visual quality of the filtered candidates was examined. Through visual inspection, the final RSI was selected to minimize banding artifacts, radiometric distortions, and abnormal brightness, while maintaining stable color balance and overall tonal harmony.

The selected style reference image is presented in Figure 9. The upper plume (approximately 37°N, 130°E) corresponds to Case 1 and appears deep orange. This coloration occurs because strong carbon absorption at 340 nm markedly reduces Top-of-Atmosphere (TOA) reflectance, while the spectral contrast between 380 nm and 340 nm increases, emphasizing the red channel. In contrast, the lower plume (approximately 32.7°N, 122°E) corresponds to a dust region observed on the same date, when the Korea Meteorological Administration issued a high-concentration dust advisory. This region appears light brown due to a gentle spectral slope and weaker UV absorption caused by coarse mineral particles, which limit the difference in the (380–340 nm) channel and produce a lower-saturation brown hue. These tonal characteristics provide only qualitative cues in the SFCC images. Therefore, additional validation using auxiliary datasets—such as UVAI/VISAI, and Ångström Exponent (AE) with Fine Mode Fraction (FMF)—is essential to ensure the robustness of aerosol-type identification.

The selected RSI is fed into each model's style encoder or attention block, allowing the color tone and brightness information to be harmoniously transferred across all content images. Ultimately, all style transfer results were generated using the same style reference image, ensuring experimental consistency and objective comparability among model performances.

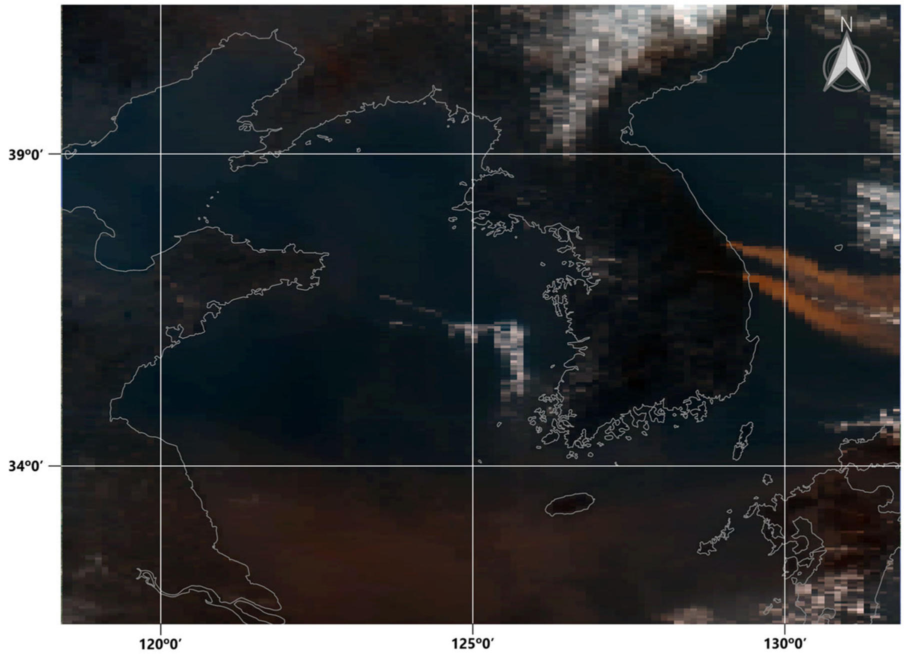

Figure 9. A reference style image (RSI) is selected to standardize color tone.

3.5. Evaluation Metrics

We employed six primary image quality assessment metrics: Histogram Similarity, Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), Fréchet Inception Distance (FID), Artistic Fréchet Inception Distance (ArtFID), and the Latent Style Consistency Index (LSCI). Each of these metrics measures a different aspect of image quality—visual fidelity, perceptual similarity, structural preservation, and color distribution standardization—allowing for an objective, comprehensive comparison of style transfer models' performance.

3.5.1. Histogram Similarity

Histogram similarity is a traditional method for evaluating the similarity of pixel value distributions between images. It is primarily used to analyze the correspondence of brightness, color tone, and channel-wise distributions (Equation (10)). In style transfer, it serves as a key metric for assessing color-tone harmonization performance.

$$BC(P,Q) = \sum_{i=1}^{N} \sqrt{P(i) \cdot Q(i)} \tag{10}$$

We employed histogram similarity based on the Bhattacharyya Coefficient, where $P(i)$ and $Q(i)$ represent the probability values of the i-th bin in two histograms, and N denotes the total number of bins (typically 256). The Bhattacharyya Coefficient quantifies the similarity between the color distributions of two images, with values closer to 1 indicating greater overlap. This metric was used to evaluate how well the style-transferred images preserved or reproduced the global color-tone distribution of the style reference image. Because the Bhattacharyya Coefficient is based on the morphological similarity of probability distributions, it responds sensitively even to global changes in color representation that may occur during style transfer. However, it does not account for structural variations or texture-related differences in visual style. Therefore, this metric should be interpreted in conjunction with qualitative assessments of visual characteristics for a more comprehensive evaluation.

3.5.2. Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is a representative traditional metric for evaluating image quality. It quantifies structural similarity between two images by simultaneously considering three key components: luminance, contrast, and structure. Unlike pixel-wise comparison metrics, SSIM is based on the statistical characteristics of local image patches, enabling a more perceptually meaningful assessment of image similarity. The general form of the SSIM equation is defined as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{11}$$

Here, $\mu_x,\ \mu_y$ represent the mean luminance of each image, $\sigma_x^2,\ \sigma_y^2$ denote the variance (contrast information), and σ_{xy} is the covariance reflecting structural similarity between the two images. $C_1,\ C_2$ are small constants introduced to ensure numerical stability in the computation. The SSIM value ranges from -1 to 1, where values closer to 1 indicate greater structural similarity between the images. SSIM was used to evaluate how effectively structural information—such as the shape of smoke plumes, terrain boundaries, and atmospheric patterns—was preserved after applying style transfer.

3.5.3. Learned Perceptual Image Patch Similarity (LPIPS)

The Learned Perceptual Image Patch Similarity (LPIPS) is a representative deep learning-based image similarity metric designed to reflect human perceptual similarity better. Unlike traditional evaluation metrics such as L2 distance or SSIM, which often fail to align with human visual perception, LPIPS measures the difference in deep feature representations between two images [36]. This metric computes perceptual distance using feature maps extracted from pre-trained neural networks (e.g., AlexNet, VGG). Specifically, it calculates the L2 distance between feature activations of image patches, which is mathematically defined as follows:

$$LPIPS(x,y) = \sum_l \frac{1}{H_l W_l} \sum_l ||\omega_l \cdot f_l^x(h,w) - f_l^y(h,w)||_2^2 \tag{12}$$

Here, $f_l^x,\ f_l^y$ denote the feature maps of images x, y at the l-th layer, respectively; ω_l represents the channel weighting factor, and $H_l,\ W_l$ indicate the height and width of the feature map. This equation computes the Euclidean distance between the feature activations extracted from each layer and aggregates them using a weighted sum to quantify the perceptual distance. Formally, for a set of original content images $X_c = \left\{X_c^{(1)}, \ldots, X_c^{(N)}\right\}$ and a corresponding set of style-transferred images $X_g = \left\{X_g^{(1)}, \ldots, X_g^{(N)}\right\}$, the perceptual distance $d\left(X_c^{(i)}, X_g^{(i)}\right)$ is calculated for each image pair $X_c^{(i)},\ X_g^{(i)}$. The mean of these distances across all samples is then used to evaluate the overall content preservation performance of the model.

LPIPS quantifies feature-level variations perceptually salient to human vision, effectively capturing differences in texture, structural details, and stylistic appearance in a manner consistent with human perceptual judgments. LPIPS was employed as a key metric to assess the visual standardization quality and perceptual fidelity of the images generated after style transfer.

3.5.4. Fréchet Inception Distance (FID)

The Fréchet Inception Distance (FID) is a widely used metric for evaluating the quality of generated images by measuring the distance between the feature distributions of real and generated image sets. It assumes that the embedding vectors extracted from the intermediate layers of a specific neural network—typically InceptionV3—follow a multivariate Gaussian distribution, and computes the Fréchet distance between these two distributions [37].

$$FID(x,\ y) = ||\mu_x - \mu_y||^2 + Tr\left(\Sigma_x + \Sigma_y - 2\left(\Sigma_x \Sigma_y\right)^{\frac{1}{2}}\right) \tag{13}$$

Here, $\mu_X,\ \mu_Y$ represent the mean feature vectors of the real and generated image sets, respectively, while $\Sigma_X,\ \Sigma_Y$ denote their corresponding covariance matrices, and $Tr(\cdot)$ indicates the trace operation. The first term measures the difference in content information between the two distributions using the Euclidean distance between their mean vectors. In contrast, the second term quantifies the differences in style information and structural variation by comparing their covariance matrices.

FID is computed as the sum of these two terms and thus measures the discrepancy between high-dimensional feature distributions of real and generated image sets. A lower FID value indicates that the two distributions are more similar, implying that the generated images are visually closer to real images. Consequently, the FID approaches zero as the generated imagery becomes indistinguishable from real observations. This demonstrates that FID effectively evaluates not only pairwise visual similarity but also the global statistical consistency of texture, tone, and structural coherence across image sets. Therefore, FID was used to assess overall style consistency and to evaluate improvements in visual quality in satellite image style transfer.

3.5.5. Artistic Fréchet Inception Distance (ArtFID)

Conventional evaluations of style transfer models have primarily relied on subjective human assessments, limiting the objectivity and reproducibility of results and hindering fair performance comparisons among models. To address these limitations, Wright and Ommer [38] proposed the ArtFID methodology, a quantitative evaluation metric that integrates the two core components of style transfer: content preservation and style matching.

The content preservation term measures how accurately the original image's semantic structure and content are retained in the style-transferred result. As described in Section 3.5.3, this is quantified using the LPIPS metric [36]. The style-matching term, on the other hand, evaluates how faithfully the style-transferred image reproduces the visual characteristics of the original RSI, and this is quantified using the FID, as explained in Section 3.5.4.

Finally, ArtFID is computed by combining these two key evaluation components—content preservation and style matching—using a multiplicative formulation that balances their relative importance (Equation (14)). The resulting metric enables an objective, comprehensive quantitative assessment of style transfer quality across models.

$$ArtFID(X_g, X_c, X_s) = \left(1 + \frac{1}{N}\sum_{i=1}^{N} d(X_c^{(i)}, X_g^{(i)})\right) \cdot \left(1 + FID\left(X_s, X_g\right)\right) \tag{14}$$

3.5.6. Latent Style Consistency Index (LSCI)

The Latent Style Consistency Index (LSCI) is a quantitative metric proposed to evaluate the style consistency and cohesiveness of images generated by a style transfer model. This approach leverages the latent space representations of the large-scale image–language model Contrastive Language-Image Pre-training (CLIP) to measure how uniformly the generated images share common visual style characteristics. Unlike metrics such as the FID or Inception Score (IS), which primarily assess overall image quality and diversity, LSCI directly evaluates one of the core objectives of style transfer—the consistent application of style across an entire dataset.

Given a set of style-transferred images $X_g = \left\{ X_g^{(1)}, X_g^{(2)}, \ldots, X_g^{(N)} \right\}$ the vision backbone of a pre-trained CLIP model is used to extract the latent feature embeddings $v_g^{(i)}$ for each image. These embeddings encapsulate high-dimensional visual and semantic information, effectively capturing style-related characteristics. Next, cosine similarity is computed between all pairs of embeddings, measuring the directional alignment between two vectors—i.e., the degree to which two images share similar styles within the latent feature space. The resulting set of similarity values for all image pairs $\left(X_g^{(i)}, X_g^{(j)} \right)$ is expressed in Equation (12).

$$S = \{cos(\theta_{ij}) = \frac{v_g^{(i)} \cdot v_g^{(j)}}{||v_g^{(i)}|| \cdot ||v_g^{(j)}||} \quad | \quad i, j \in \{1, \ldots, N\}, i \neq j\} \tag{15}$$

The LSCI metric is defined as the mean and standard deviation of all computed cosine similarities. The mean value represents the average level of style agreement across the dataset, while the standard deviation quantifies the variance of style consistency. This metric extends beyond evaluating the style quality of individual images, providing a measure of how consistently a model applies a uniform style across an entire dataset. LSCI is particularly useful for assessing style degradation that can occur when using the same style to diverse image content. Unlike qualitative human evaluations, LSCI provides objective numerical indicators—the mean and standard deviation—allowing fair and efficient comparisons of performance across different models.

4. Results

4.1. Evaluation Results of Style Transfer Models

4.1.1. Quantitative Evaluation Using Common Image Generation Metrics

Table 4 presents a quantitative comparison of the performance of each style transfer model based on three evaluation metrics: SSIM, LPIPS, and FID. SSIM represents structural similarity, LPIPS measures perceptual distance, and FID quantifies the distance between feature distributions. Generally, a higher SSIM indicates better structural preservation, while lower LPIPS and FID values correspond to higher visual quality in the generated images.

Table 4. Quantitative evaluation results of style transfer performance by the statistical and deep learning models used in this study (↑: higher is better; ↓: lower is better).

Metric	ECDF	ReHistoGAN	StyTr^2	SI-DM
SSIM (↑)	0.845	0.884	0.840	0.792
LPIPS (↓)	0.261	0.200	0.259	0.292
FID (↓)	196	207	201	195

ReHistoGAN achieved the highest SSIM (0.884) and LPIPS (0.200) scores, indicating the most substantial structural similarity and perceptual closeness to the RSI. In contrast, SI-DM recorded the lowest FID (195), suggesting the closest overall distribution to the style reference. ECDF showed stable performance in SSIM (0.845) and FID (196), while StyTr^2 exhibited generally average results across all metrics.

However, these quantitative metrics have inherent limitations in fully capturing the visual characteristics and semantic consistency of satellite imagery. Despite its high SSIM and LPIPS scores, ReHistoGAN produced over-saturated colors and blurred boundaries. In contrast, SI-DM, despite having a relatively lower SSIM, preserved smoke boundaries and terrain contours more effectively, yielding visually superior results. This discrepancy highlights the gap between conventional quantitative metrics and actual perceptual quality in satellite image processing. It underscores the need for domain-specific evaluation criteria tailored to remote sensing imagery.

Table 5 presents the histogram similarity between the style and generated images for each model, evaluated using the Bhattacharyya Coefficient. The similarity values across color channels ranged from 0.7 to 0.99, with SI-DM showing the most stable color alignment, achieving values around 0.99 across all channels.

Table 5. Histogram similarity (in terms of Bhattacharyya Coefficient) between the RSI and the generated images by the statistical and deep learning models used in this study.

Metric	ECDF	ReHistoGAN	StyTr^2	SI-DM
Red channel	0.909	0.953	0.988	0.992
Green channel	0.773	0.798	0.977	0.993
Blue channel	0.732	0.756	0.955	0.991

First, the ECDF method, which is based on statistical cumulative distribution matching, achieved similarity scores of 0.909, 0.773, and 0.732 for the red, green, and blue channels, respectively. These relatively lower values indicate that ECDF, while directly normalizing the cumulative distribution of pixel intensities, does not adequately capture spatial or semantic structures. Consequently, even though its histogram-level similarity may appear high, the resulting images often exhibit unnatural color tones or artifacts. This suggests that a comprehensive evaluation of style transfer performance should consider not only statistical metrics but also spatial consistency and perceptual realism.

Following the ECDF, the ReHistoGAN and StyTr^2 models achieved higher overall similarity but exhibited noticeable variability across color channels. ReHistoGAN yielded the lowest similarity in the blue channel (0.756), while StyTr^2 showed a discrepancy between the red (0.988) and blue (0.955) channels. In contrast, SI-DM consistently maintained high similarity across all channels—0.992 (red), 0.993 (green), and 0.991 (blue)—demonstrating the most balanced and stable color alignment. A more detailed qualitative comparison is discussed in Section 5.

4.1.2. Extended Evaluation with Style-Transfer-Specific Metrics

To evaluate the models more comprehensively, we additionally employed two style-transfer-specific metrics—ArtFID and LSCI. Unlike general image generation metrics, these indicators directly measure the core aspects of style transfer: content preservation, style matching, and dataset-level style consistency.

Table 6 reports the ArtFID scores of the three models. ArtFID integrates content preservation and style matching into a single metric, with lower values indicating better performance. SI-DM achieved the lowest score (1.620), marking the highest overall style-transfer quality among the models. ReHistoGAN ranked second (1.701), while StyTr^2 obtained a relatively higher value (1.760). These results suggest that SI-DM provides a more balanced performance—maintaining semantic content structure while faithfully reproducing the visual characteristics of the style reference.

Table 6. ArtFID-based evaluation results of style transfer quality across deep learning models (↓: lower is better).

Metric	ReHistoGAN	StyTr^2	SI-DM
ArtFID (↓)	1.701	1.760	1.620

Table 7 presents the LSCI results for ECDF correction and the three style-transfer models. LSCI quantifies the consistency of styled outputs across the dataset; higher means and lower standard deviations indicate better performance. SI-DM achieved the highest style consistency (mean 0.894, std 0.042). ReHistoGAN followed (mean 0.867, std 0.046), and StyTr^2 ranked third (mean 0.857, std 0.046). The ECDF correction yielded the lowest consistency (mean 0.846, std 0.060). A detailed analysis and qualitative discussion of these findings are provided in Section 5.

Table 7. LSCI-based evaluation results of style consistency across models.

LSCI	ECDF	ReHistoGAN	StyTr^2	SI-DM
Mean	0.846	0.867	0.857	0.894
Std.	0.060	0.046	0.046	0.042

4.2. Model-Specific Characteristics Under Spatial and Temporal Variations

4.2.1. Qualitative Analysis of Boundary Preservation and Artifact Formation

To enable a more detailed comparison of style transfer performance across models, Figure 10 presents zoomed-in local regions centered on smoke-affected areas. These regions were selected as representative sites where boundary sharpness and visibility, the key indicators of smoke-visualization quality, can be assessed intuitively.

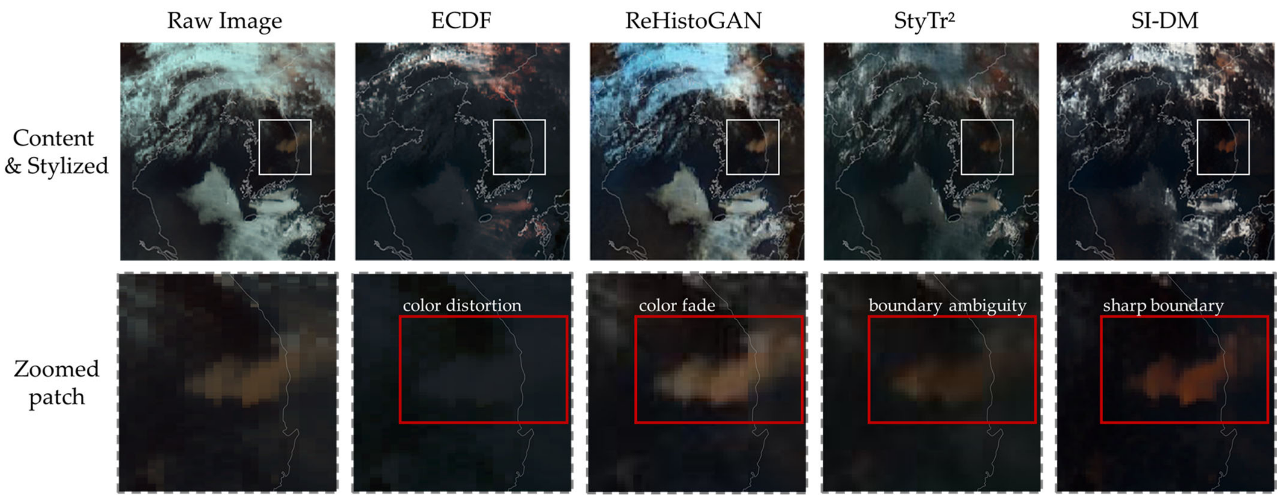

Figure 10. Qualitative visual comparisons between the style transfer model results, focusing on the wildfire smoke pixels (white and red rectangles).

In Figure 10, the ECDF method exhibits severe tonal distortion and uninterpretable outputs in the smoke regions, resulting in a complete loss of smoke structure. In contrast, ReHistoGAN and StyTr^2 partially preserve the smoke contours, but the images exhibit patch-level discontinuities and disrupted texture coherence. This artifact reflects a structural limitation of GAN-based decoders: color statistics are applied globally, while local texture continuity is not adequately maintained.

By comparison, SI-DM restores the smoke structures at high resolution while producing balanced color-tone adaptation without excessive color enhancement. This can be attributed to its self-attention-based style injection mechanism, which preserves content-query information while inserting style key and value features at intermediate stages, thereby achieving both spatial alignment and texture consistency. Moreover, the iterative reconstruction inherent in the diffusion process enables progressive refinement of smoke representation, preserving structure without spatial distortion—unlike GAN-based methods. As a result, SI-DM maintains clear boundary delineation even in scenes where the smoke–background separation is visually ambiguous. These structural properties enhance smoke visibility and demonstrate that SI-DM provides the most consistent, visually stable style transfer among the evaluated models.

Figure 11 focuses on comparing the overall color-tone alignment of background regions and the continuity between image patches, rather than the smoke itself. $StyTr^2$ and ReHistoGAN effectively apply their respective global color-tone styles, but both suffer from inconsistent color-tone alignment between patches, resulting in an unnatural overall appearance. In addition, both models exhibit visible block artifacts, with local tonal discontinuities between adjacent patches, further degrading the imagery's visual coherence. In particular, $StyTr^2$, which adopts a Transformer-based encoder–decoder architecture, exhibits color-tone inconsistency due to domain mismatch with the training dataset and is sensitive to brightness variations and atmospheric scattering.

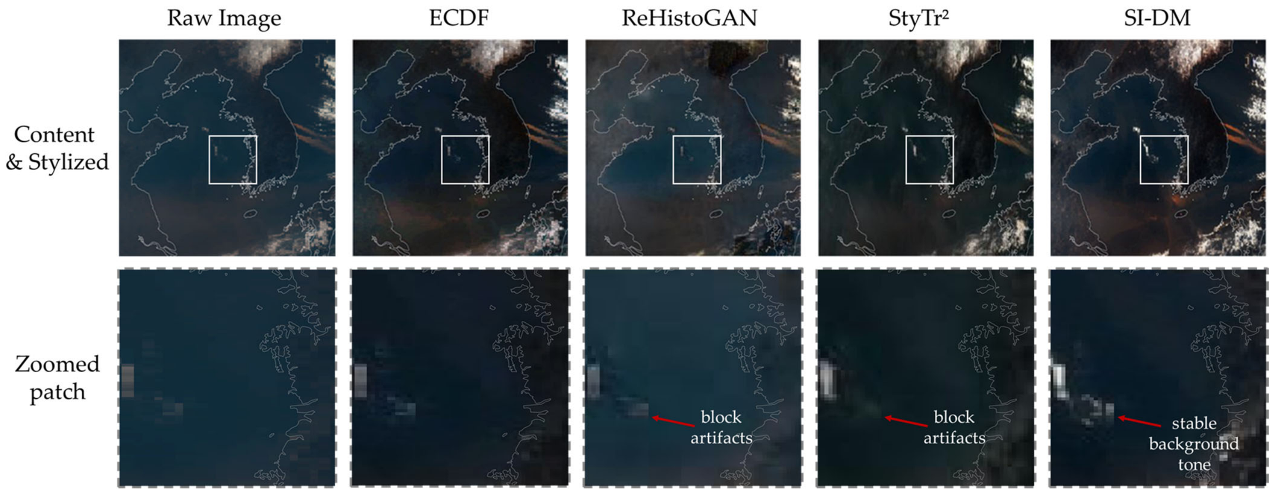

Figure 11. Qualitative visual comparisons between the style transfer model results, focusing on the block artifact phenomenon.

In contrast, SI-DM employs an initial latent AdaIN mechanism that softly aligns color-tone statistics (mean and variance) at the early latent stage, combined with attention-based style injection that ensures local alignment across the image. Consequently, the overall color-tone distribution naturally harmonizes, spatial continuity is maintained, and the resulting imagery appears visually stable and free of block artifacts, providing the most coherent results among the compared models.

4.2.2. Sub-Regional Assessment of Time-Adjacent Consistency

The SFCC framework supports the consistent visual interpretation of GEMS false color composite imagery by stabilizing color tones while preserving scene structures critical for smoke analysis, such as smoke plume boundaries and adjacent background regions. SFCC is applied at the spatial scale illustrated in Figure 1, which serves as the standard processing unit for normalization. For evaluation, we analyzed specific sub-regions within these patches where short-term color variations most directly influence visual interpretation. The evaluation ROIs were defined as localized patches centered on visually discernible smoke plume boundaries to assess boundary delineation and interpretability across temporally adjacent observations. This sub-regional evaluation provides a practical basis for assessing temporal consistency in time-series observations.

The primary objective of SFCC is to reduce visual inconsistency arising from observation-induced factors rather than to modify physically meaningful plume dynamics.

In geostationary satellite observations, image appearance is strongly influenced by changes in observational geometry and illumination conditions—including variations in the solar zenith angle—even over short temporal intervals. To assess whether SFCC interferes with genuine smoke plume evolution under such conditions, we conducted a temporal stability analysis using adjacent observations acquired at 2 h intervals.

To evaluate temporal proximity consistency in time-series GEMS observations, we conducted a sub-regional analysis using temporally adjacent images acquired at 2 h intervals. This interval was selected to capture short-term observation-induced variations—such as changes in illumination geometry and atmospheric scattering—while minimizing the influence of large-scale physical plume evolution. For this analysis, subregional regions of interest (ROIs) were defined as localized image patches centered on visually discernible smoke plume boundaries. These boundary-centered regions represent areas where small temporal inconsistencies most directly affect visual interpretation, boundary delineation, and perceived plume continuity. By focusing on plume-edge regions rather than homogeneous background areas, the evaluation more sensitively reflects the practical interpretability of smoke features under temporally adjacent observations. The same ROI definition strategy was applied consistently across all wildfire cases and time steps. For each ROI, histogram similarity, Structural Similarity Index Measure (SSIM), and Latent Style Consistency Index (LSCI) were computed between adjacent time pairs (t vs. t + 2 h and t + 2 h vs. t + 4 h), and the results are summarized in Table 8.

Table 8. Quantitative evaluation of temporal proximity consistency for subregional smoke plume ROIs across temporally adjacent observations ($\Delta t = 2$ h). ROIs were defined as localized patches centered on visually identifiable smoke plume boundaries and were applied consistently across all cases. Metrics include histogram similarity, Latent Style Consistency Index (LSCI; mean $\pm$ standard deviation), and Structural Similarity Index (SSIM), computed for consecutive time pairs (t vs. t + 2 h and t + 2 h vs. t + 4 h).

Model	**Histogram Similarity**		**LSCI Mean (±std)**	**SSIM**	
Δt	t vs. t + 2	t + 2 vs. t + 4	-	t vs. t + 2	t + 2 vs. t + 4
ECDF	0.782	0.814	-	0.896	0.857
ReHistoGAN	0.687	0.962	0.916 (±0.035)	0.720	0.686
StyTr^2	0.956	0.956	0.936 (±0.020)	0.802	0.789
SI-DM	0.920	0.963	0.936 (±0.022)	0.750	0.796

Importantly, visual inspection confirms that the spatial extent, shape, and displacement of the smoke plume are maintained across all time steps (Figure 12). While global color appearance is stabilized, plume transport remains clearly observable. Among the evaluated methods, SI-DM provides the most stable temporal harmonization without introducing blocking artifacts, yielding consistent tone alignment while preserving plume boundaries and internal texture across consecutive frames. In contrast, ECDF excessively flattens smoke-related contrast, rendering the plume region visually indistinguishable from the surrounding background and thereby reducing interpretability. StyTr^2 exhibits limited separability between the plume and background, with weakened edge definition that makes boundary delineation ambiguous. ReHistoGAN shows insufficient color-tone harmonization, leading to noticeable temporal inconsistencies in both background tone and plume appearance. Overall, these qualitative observations indicate that SFCC, particularly when implemented via SI-DM, selectively reduces observation-induced variability while avoiding the suppression or distortion of physically meaningful plume evolution.

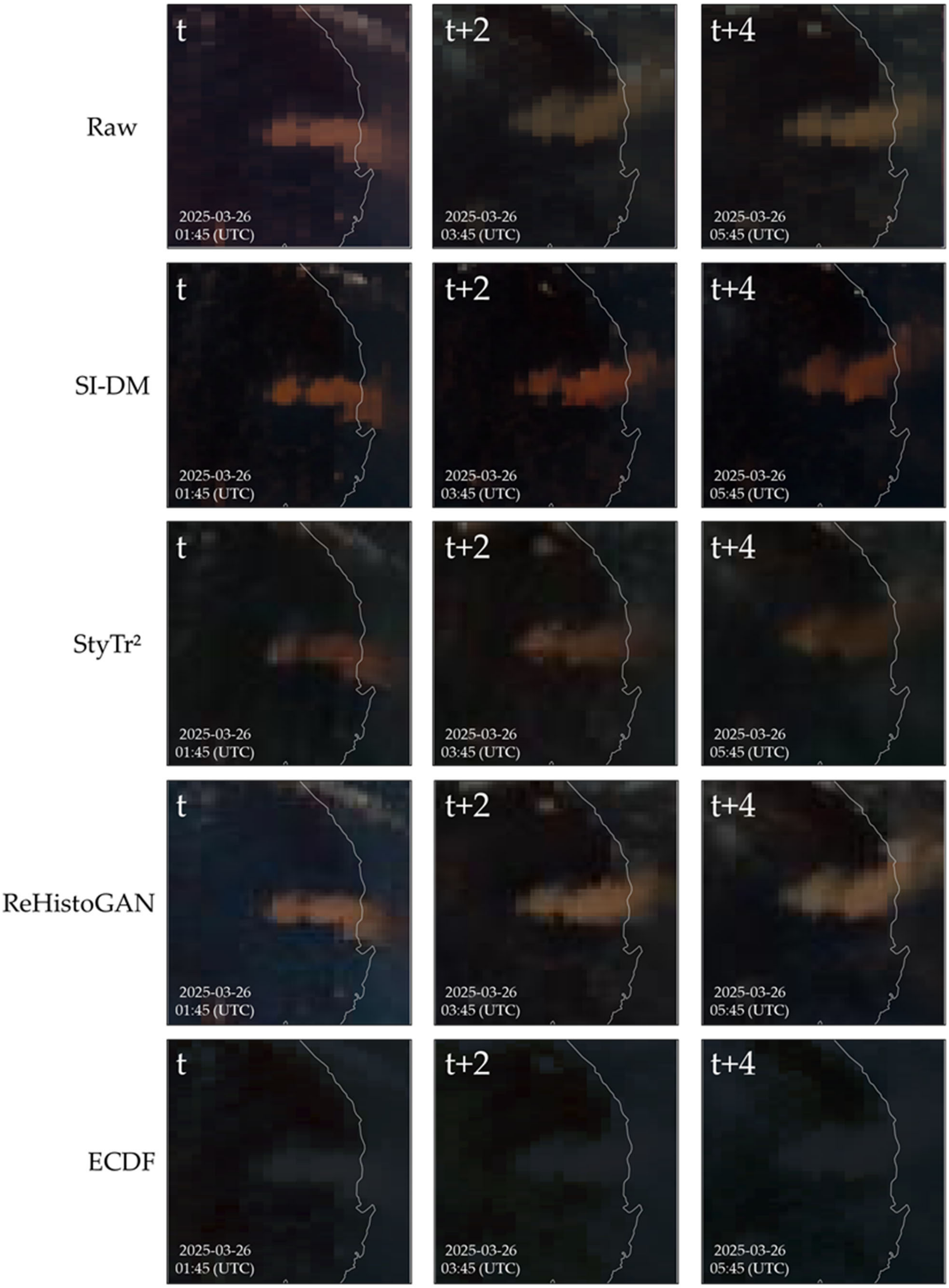

Figure 12. Visual comparison of sub-regional smoke plume ROIs across consecutive observation time steps (t, t + 2, t + 4).

These findings demonstrate that SFCC operates as an effective visualization standardization mechanism at localized spatial and short-term temporal scales. By stabilizing background tone and color alignment without altering plume morphology, the framework supports more reliable manual interpretation and provides temporally consistent inputs for downstream analyses—such as change detection and data-driven smoke monitoring—while preserving the physical integrity of aerosol evolution.

4.2.3. Qualitative Assessment of Clear-Sky Scenes

To further evaluate potential artifact generation or unintended distortions from the style transfer process, we present SFCC results for both smoke-affected and representative clear-sky scenes. As illustrated in Figure 13, the proposed SFCC enhances tonal consistency in smoke-laden regions while maintaining the fidelity of clear-sky areas, without introducing noticeable artifacts or chromatic distortions in the examined cases. For brevity, Figure 13 displays the SFCC outputs generated by SI-DM as a representative example, demonstrating its robust performance across diverse atmospheric conditions.

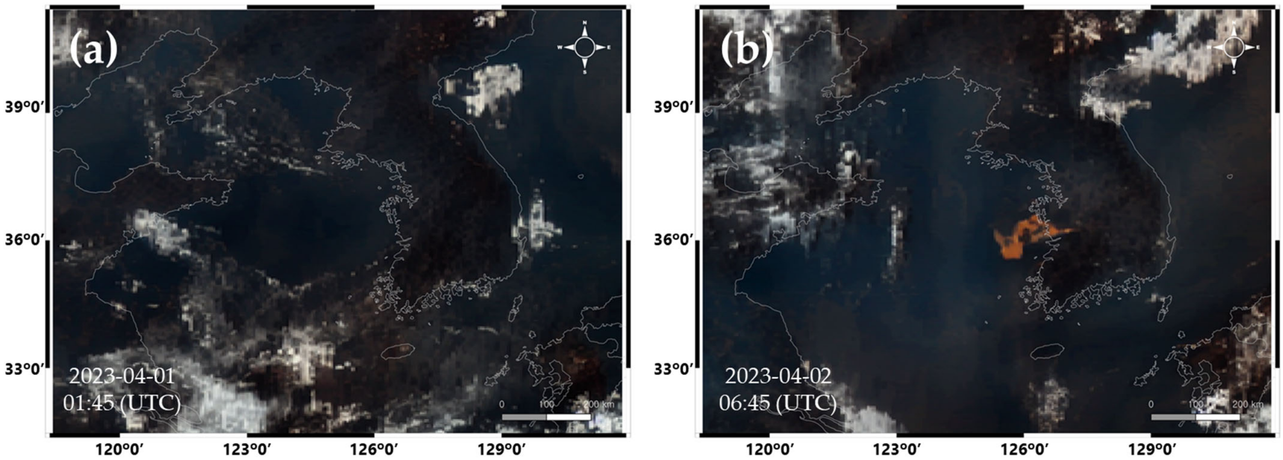

Figure 13. Standardized False Color Composite (SFCC) results for representative clear-sky and smoke-affected scenarios. (**a**) A clear-sky scene acquired on 1 April 2023 (01:45 UTC), showing a consistent and natural background tone. (**b**) A smoke-affected scene from 2 April 2023 (06:45 UTC), where the smoke plume (highlighted in orange) is clearly distinguishable from the standardized background. The framework ensures visual consistency across different dates and atmospheric conditions.

4.2.4. Qualitative Comparison with Traditional Color Enhancement Methods

Figure 14 presents a qualitative comparison of FCC images processed using various color enhancement approaches for the same observation time. Figure 14a displays the original FCC image without normalization. In contrast, the proposed SI-DM SFCC (Figure 14b) yields a more uniform color tone while faithfully preserving the structural features of both smoke-affected and clear-sky regions. Conversely, the linear contrast stretching result in Figure 14c enhances local contrast but introduces scene-dependent color shifts, leading to an inconsistent visual appearance across different regions. Histogram equalization (Figure 14d) further exaggerates contrast, producing severe color distortion and artificial artifacts that obscure the physical interpretation of atmospheric features. These results indicate that conventional color enhancement techniques, while effective for improving local contrast, are ill-suited for the standardized visualization of time-series satellite imagery. In comparison, SFCC provides a more stable and visually consistent representation, underscoring its effectiveness in harmonizing FCC images under varying observational conditions.

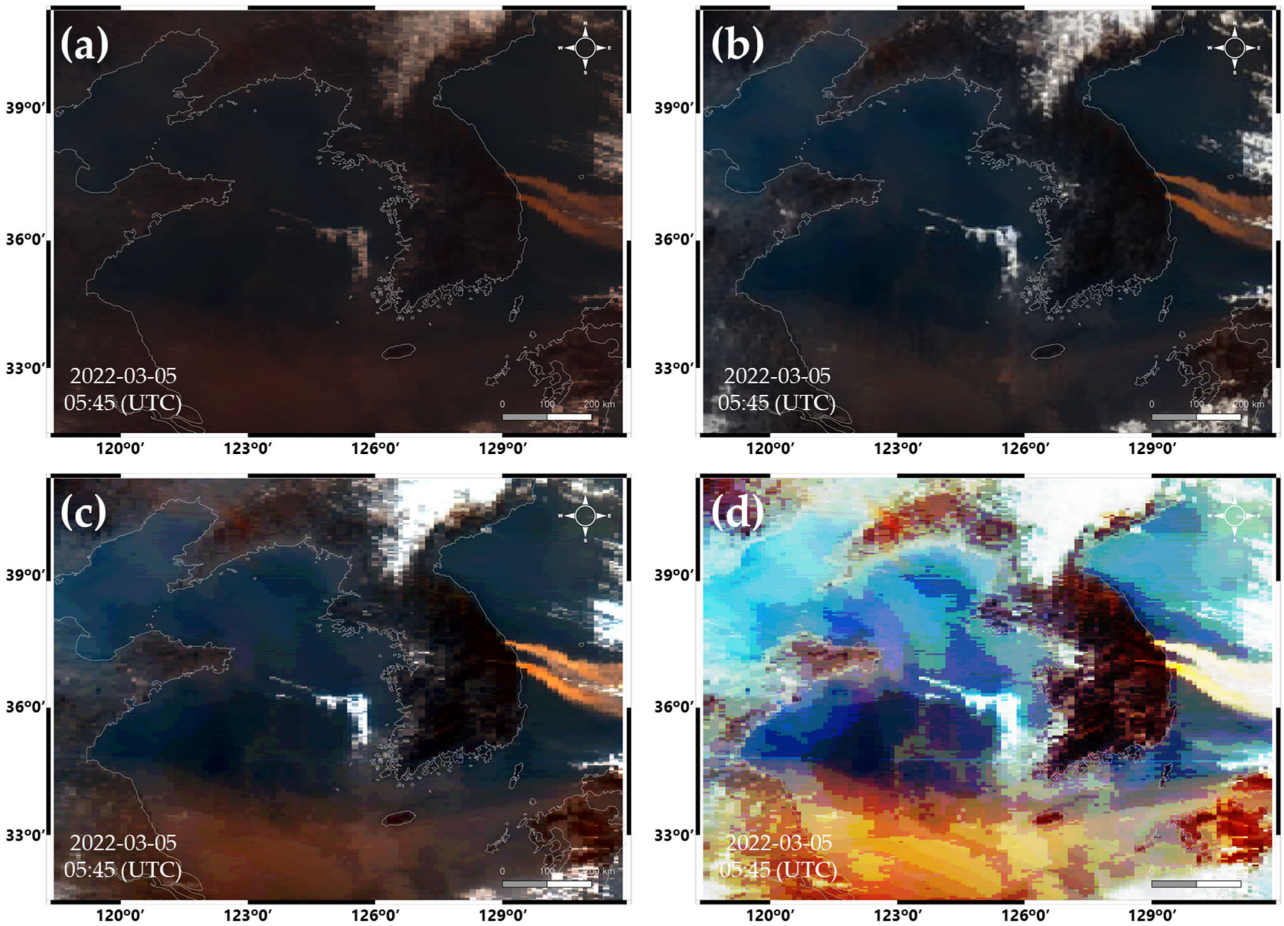

Figure 14. Qualitative comparison of color enhancement methods for a representative GEMS false color composite (FCC) scene (5 March 2022, at 05:45 UTC). (**a**) Original FCC without normalization, displaying substantial observation-induced color shifts. (**b**) The proposed SI-DM-based SFCC, providing a standardized and uniform color tone. (**c**) Linear contrast stretching, which improves local contrast but introduces scene-dependent color inconsistencies. (**d**) Histogram equalization, exhibiting severe color distortion and artificial artifacts that obscure atmospheric features.

4.3. Visual Validation Using Major Wildfire Cases (2022–2025)

In this section, we compared the style transfer models using major wildfire events in Korea (2022, 2023, and 2025) and evaluated the reproducibility and stability of the SI-DM model. Despite considerable variations in regional conditions, atmospheric states, and solar altitude across years, SI-DM consistently demonstrated stable performance in both smoke structure preservation and background color-tone alignment.

Figure 15 presents the SFCC result for Case 1 (the 2022 Uljin–Samcheok wildfire). This case featured weak smoke intensity and complex illumination conditions, under which ECDF and ReHistoGAN successfully corrected background color tone but showed limited improvement in smoke contrast and visibility. In comparison, SI-DM effectively preserved distinct smoke patterns while achieving stable style transfer without introducing excessive color distortion.

Figure 15. Style transfer results on the 2022 Uljin-Samcheok wildfire case. The yellow box indicates the dust area. The SI-DM result is indicated by the red box.

This case coincided with a dust episode, and the SFCC imagery also revealed aerosols, visualized as dust plumes (yellow box in Figure 15). To verify these observations, a cross-comparison with GEMS UVAI data for the same period was conducted (Figure 16). Dust plumes, dominated by coarse mineral particles, typically exhibit a gentle spectral slope and weaker UV absorption at 380 and 354 nm, resulting in relatively low UVAI values due to background contrast effects in the lower atmosphere. In this case, dust showed UVAI values of approximately 1–2.2, whereas wildfire smoke exhibited higher UVAI values, allowing clear differentiation between the two aerosol types. In summary, regardless of aerosol type (smoke or dust), SI-DM maintained structural consistency between smoke boundaries and background tone, thereby enhancing the visual interpretability of wildfire scenes in the SFCC imagery.

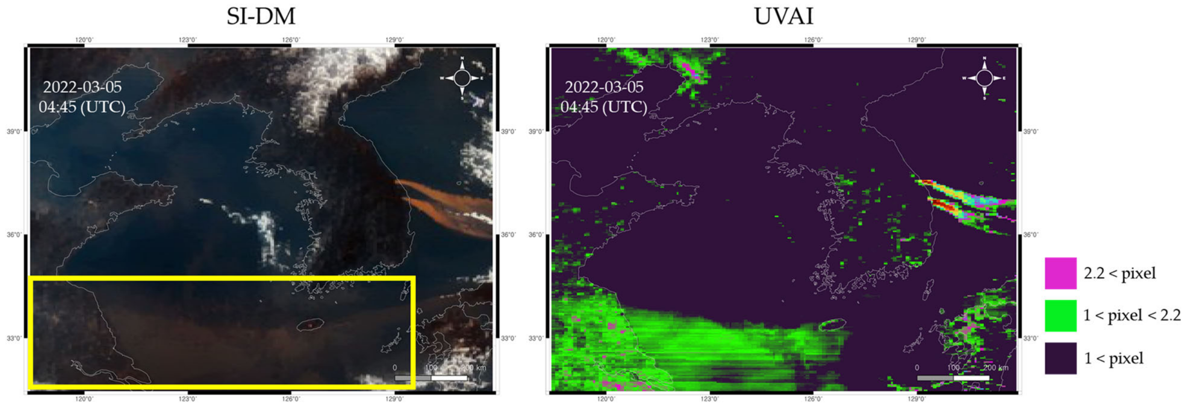

Figure 16. Comparison between the SI-DM result and GEMS UVAI for the same observation period (5 March 2022 04:45 (UTC)). The yellow box indicates the dust area.

Figure 17 presents the SFCC result for Case 2 (the 2023 Hongseong wildfire). This case represents a low-contrast, highly complex scene in which the boundaries between the smoke plume and the surrounding terrain were visually intertwined, making interpretation difficult. Under these challenging conditions, StyTr2 tended to blur the visual distinction between smoke and the background, while ReHistoGAN caused a global color-tonal shift that altered the overall color balance. In contrast, SI-DM successfully preserved the shape of the smoke plume while maintaining consistent tonal alignment across regions and effectively retaining complex structural patterns within the scene.

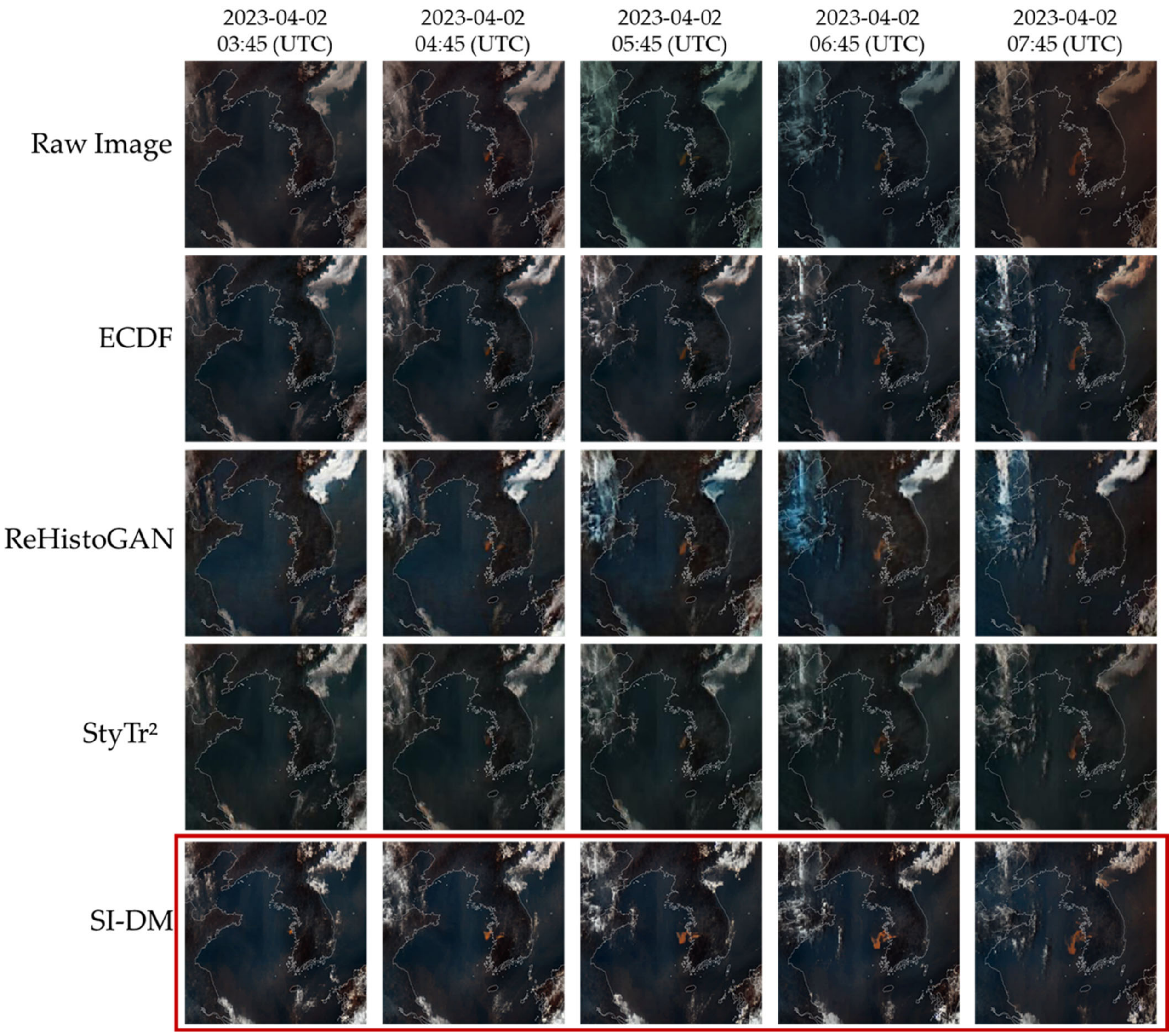

Figure 17. Style transfer results on the 2023 Hongseong wildfire case. The SI-DM result is indicated by the red box.

Figure 18 shows the SFCC result for Case 3 (the 2025 nationwide concurrent wildfire event). This scene involved both high smoke concentration and strong atmospheric scattering, representing one of the most challenging observation conditions. In this case, ECDF excessively reduced the contrast of the smoke regions, resulting in poor detectability, and StyTr^2 introduced color-tone inconsistencies between clouds and smoke. Conversely, SI-DM preserved sharp smoke boundaries and consistent background tone, ensuring high visual interpretability even under such complex atmospheric conditions.

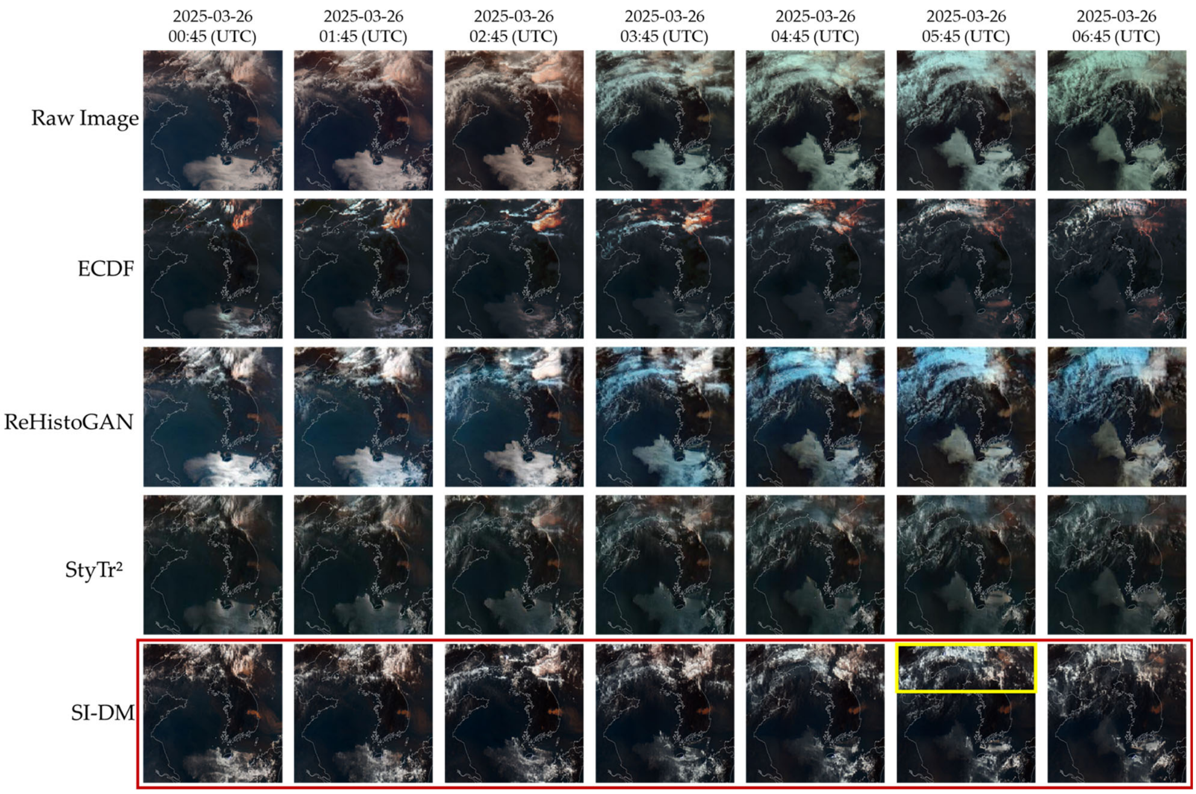

Figure 18. Style transfer results on the 2025 nationwide wildfire event in Korea. The yellow box indicates the dust area. The SI-DM result is indicated by the red box.

In addition, the SFCC image revealed an aerosol feature over the northeastern part of the Korean Peninsula (upper-right area), in addition to the main wildfire smoke plume. To determine whether this feature originated from absorbing aerosols or was simply due to atmospheric scattering, a comparison with GEMS UVAI data for the same observation time was conducted (Figure 19). The region emphasized by SI-DM corresponded spatially to the high UVAI region, indicating the presence of absorbing aerosols. Notably, in the northeastern Korean Peninsula, this aerosol feature was obscured by clouds in the raw imagery but was clearly recovered in the SI-DM result. These findings demonstrate that SI-DM produces results consistent with the physically based UVAI indicator, accurately reproducing both smoke boundaries and tonal variations in a stable and physically meaningful manner.

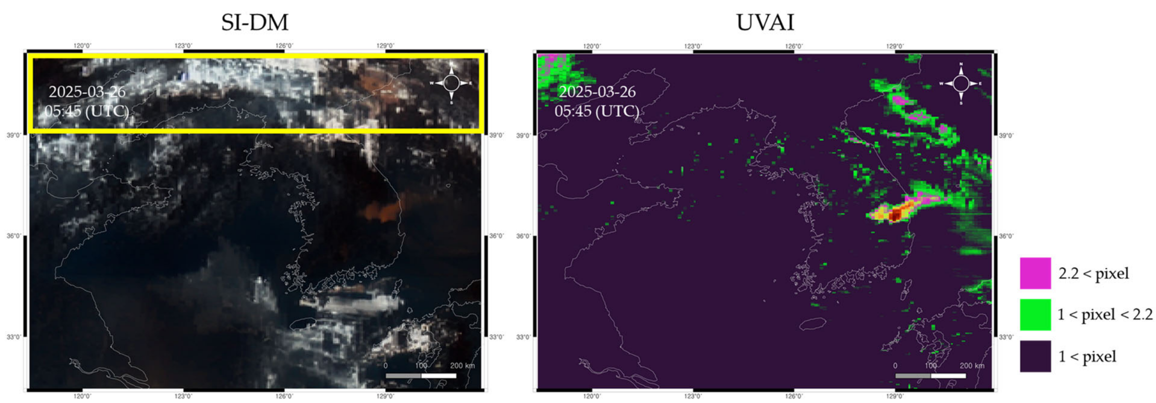

Figure 19. Comparison between the SI-DM result and GEMS UVAI for the same observation period (26 March 2025 05:45 (UTC)). The yellow box indicates the dust area.

5. Discussion

5.1. Histogram Similarity Evaluation

As shown in Table 5, SI-DM yields consistently high histogram similarity across all color channels, indicating stable and well-aligned color-tone reproduction with respect to the reference style image. Figure 20 compares the histogram similarity (Bhattacharyya Coefficient) of the R, G, and B channels between the style reference image and each style-transferred result on a one-to-one basis. In Figure 20a (ECDF), the ECDF correction does not sufficiently reproduce the histogram distribution of the RSI. In particular, in the G and B channels, the orange dashed lines exhibit repeated fine peaks and discontinuous distributions. This occurs because ECDF matching aligns the cumulative distribution function (CDF) rather than directly preserving local density variations across individual pixel intervals. In other words, while CDF-based mapping ensures agreement in the global cumulative probability, it does not capture the local structure or continuous color-tone transitions of the histogram. As a result, even if the corrected image shares a similar global statistical distribution with the RSI, it tends to exhibit overemphasized contrast or discontinuous color variations. Consequently, although ECDF performs reasonably well as a statistical similarity metric, it has limited ability to achieve visual naturalness and structural consistency.

In contrast, the SI-DM distribution closely follows that of the RSI while maintaining a smooth and evenly spread color-tone distribution. This behavior is associated with the model's structural characteristics, particularly its iterative denoising process and attention-based style injection, which enable localized color-tone adjustment while preserving spatial coherence. Both ReHistoGAN and StyTr^2 employ a single forward-pass mapping for style transfer, which inherently restricts the flexibility of color-tone adaptation. ReHistoGAN reduces the style representation into an RGB–UV color space and regulates the color distribution by minimizing the Hellinger distance between histograms. However, projecting the three-dimensional RGB space onto a two-dimensional chromatic space inevitably results in color information loss, as noted by the authors [33]. StyTr^2, on the other hand, utilizes multi-head self-attention within a Transformer-based architecture to model long-range dependencies between content and style features. Through the cross-attention mechanism, similarities between content queries and style keys are computed, and the corresponding style features are fused via weighted summation. During Softmax normalization of the attention weights, higher weights are assigned to highly similar style features, causing the histogram of the target image to converge to a limited range of values and produce distinct peaks [34].

In contrast, SI-DM has a structure that progressively restores the image through a multi-stage denoising process. The reverse process of the diffusion model starts from an initial pure noise state and removes noise step by step at each timestep while injecting style information. In this process, SI-DM uses a self-attention mechanism that references the RSI key and value, thereby inheriting a structural advantage that maintains spatially diverse pixel-value distributions. Unlike the single-mapping method, which minimizes the global loss at once, the multi-stage process learns local distributions at each step, enabling a more detailed reproduction of the histogram of the RSI.

However, the spectral bias, a well-known phenomenon in deep neural networks, is an inherent limitation common to all models. Spectral bias refers to the phenomenon in which a neural network prioritizes learning low-frequency components during training, while learning of high-frequency components is relatively slow or incomplete [39]. This is due to the formation of a broader basin for low-frequency functions in parameter space and to the smoothness of activation functions. Models based on single mapping, such as ReHistoGAN and StyTr^2, cannot fully learn sharp contrast changes or detailed color-tone distributions due to this spectral bias, leading to histogram peaks converging on specific color values.

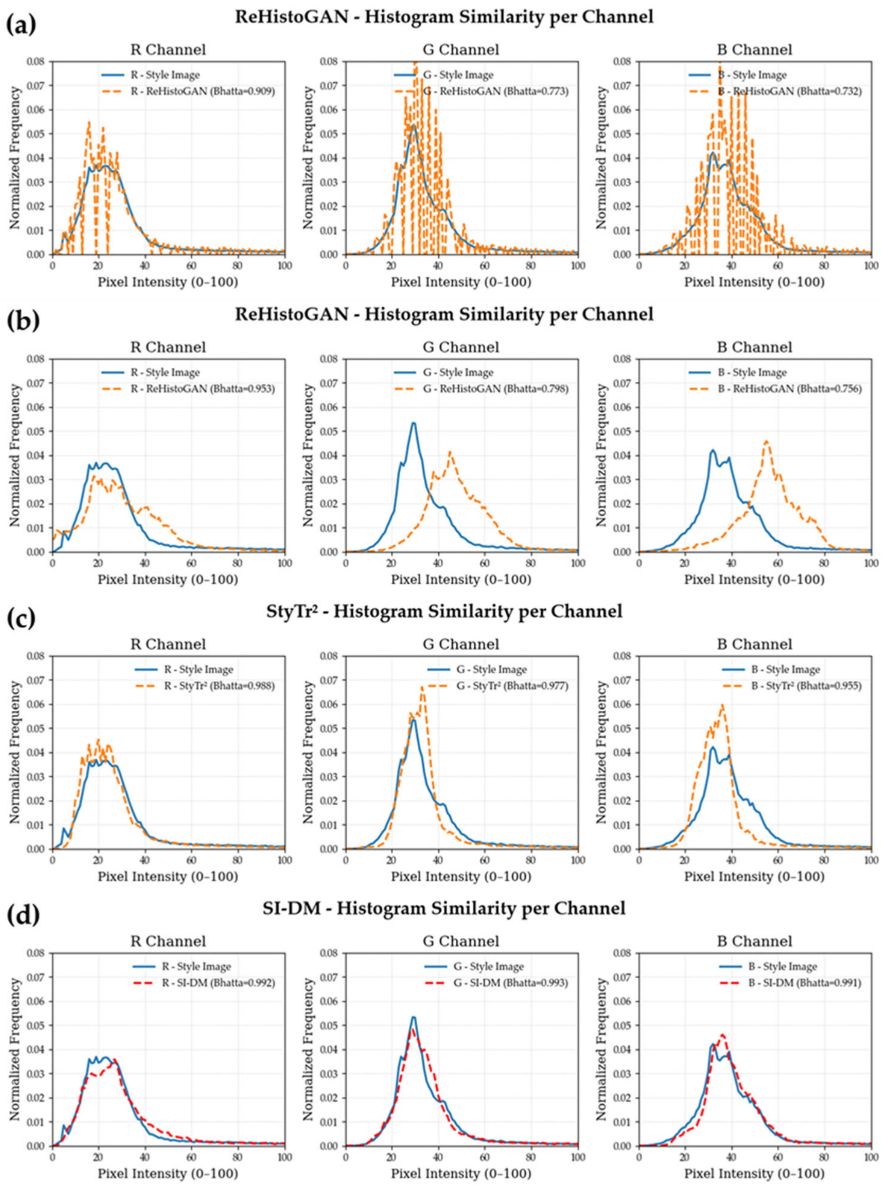

Figure 20. Histogram similarity evaluation results using the Bhattacharyya Coefficient. (**a**) ECDF, (**b**) ReHistoGAN, (**c**) StyTr^2, (**d**) SI-DM.

SI-DM is similarly affected by the aforementioned spectral bias; however, it partially compensates for this limitation through its style injection mechanism. Diffusion models exhibit a hierarchical generation property, whereby low-frequency global structures are established during early timesteps, followed by the addition of high-frequency details at later stages [40]. This multi-stage refinement allows high-frequency components to be reintroduced and adjusted across successive iterations, rather than being determined in a single forward pass. In the proposed SI-DM framework, the attention features of the RSI are continuously injected at intermediate stages of the diffusion process, which preserves the target style distribution more effectively than single-pass methods. Although SI-DM still faces challenges in fully restoring high-frequency components, the repeated style referencing across multiple timesteps helps prevent premature convergence to overly smoothed representations. Consequently, this process maintains a broader and more stable histogram distribution that remains consistent with the RSI.

Ultimately, the SI-DM model showed the highest histogram similarity with the RSI. These results are interpreted as due to the SI-DM's characteristic of gradually integrating content and style information. Therefore, the high Bhattacharyya Coefficient indicates excellent global color alignment, meaning that SI-DM performs the transfer while maintaining the overall color tone of the RSI, suppressing unnecessary color noise, and emphasizing key visual elements.

5.2. Enhanced Evaluation with ArtFID and LSCI

We evaluated the quantitative performance of the style transfer results using SSIM, LPIPS, FID, and channel-wise histogram similarity (Bhattacharyya Coefficient). These metrics are widely used as standard benchmarks for comparing image performance in terms of structural similarity, perceptual distance, and distribution-based statistics. However, as shown in the results, there was an apparent discrepancy between the quantitative scores and the visually perceived image quality.

Traditional quantitative metrics such as SSIM, LPIPS, and FID were initially developed to evaluate image restoration, compression, or generation quality, focusing on measuring the similarity between generated and original images. Consequently, these metrics are designed to answer "How similar is the result to the original?" rather than "How effectively has the style been transferred?", highlighting a fundamental difference between image similarity and style transfer quality.

First, SSIM evaluates local similarity based on luminance, contrast, and structural features within patches. As a result, even if the image is globally blurred or exhibits color-tone shifts, SSIM can still yield high scores when local patch distributions remain relatively similar. Next, LPIPS is a representative metric for evaluating content preservation. It measures perceptual distortion by computing profound differences in features across color, texture, and contrast [41]. Owing to this property, LPIPS effectively evaluates how well the structural elements of the original content are maintained, but is limited in its ability to quantitatively assess how successfully the style characteristics have been applied.

In our experiments, ReHistoGAN achieved the highest SSIM (0.884) and LPIPS (0.200) scores; however, qualitative inspection revealed significant visual degradation, including blurred smoke boundaries, reduced local sharpness, and blocking artifacts. This discrepancy highlights a critical limitation of traditional image quality metrics in the context of remote sensing smoke visualization. SSIM and LPIPS primarily reward global content similarity and may remain high even when perceptually vital cues for smoke interpretation—such as boundary sharpness, edge fidelity, and inter-patch continuity—are compromised. Consequently, high SSIM and LPIPS values do not necessarily translate into enhanced interpretability for SFCC-based smoke scenes.

In contrast, SI-DM recorded lower scores (SSIM = 0.792 and LPIPS = 0.292) yet produced superior visual results, effectively preserving smoke contours, maintaining color-tone alignment with the background, and improving interpretability. These findings indicate that conventional metrics do not adequately reflect the analytical suitability of style-transferred imagery. In SFCC images, critical information, such as smoke visibility, depends not only on structural preservation but also on a combination of perceptually integrated factors—edge clarity, local contrast, and global color-tone consistency. Current quantitative metrics, however, evaluate these aspects independently rather than holistically. To address these limitations, we employed newly designed style transfer-specific metrics to achieve a more comprehensive evaluation.

Accordingly, we used ArtFID and LSCI to assess the quality of the SFCC imagery. As shown in Tables 6 and 7, SI-DM achieves the lowest ArtFID (1.620) and the highest LSCI with low variability (0.894 $\pm$ 0.042), indicating a well-balanced performance in content preservation and style consistency. ArtFID, which integrates content preservation and style matching, was explicitly proposed for evaluating the performance of style transfer. The results showed that SI-DM achieved the lowest ArtFID value (1.620), indicating the most balanced performance in maintaining both structural fidelity and style alignment. In contrast, ReHistoGAN and StyTr^2 partially reproduced the RSI's color distribution but exhibited structural distortions and detail loss, resulting in higher ArtFID scores. This demonstrates that SI-DM successfully fuses structural and stylistic information, moving beyond simple color-matching to achieve perceptually coherent, visually balanced results.

We excluded the SFCC dataset generated by ECDF correction from the ArtFID evaluation. ECDF correction is a statistical approach that forces the pixel distributions of other images to match the cumulative distribution function (CDF) of the RSI. As a result, when evaluated using ArtFID, simple distribution alignment can artificially lower the FID component, leading to overestimated performance. Moreover, ArtFID is specifically designed to assess "how effectively a model preserves content while learning and applying style information." In contrast, ECDF directly injects the style distribution via a non-learning-based preprocessing method and thus does not reflect the learned model's performance. Therefore, ECDF was excluded from the ArtFID evaluation and instead assessed separately using the histogram similarity metric and LSCI to validate its effectiveness.

LSCI uses CLIP-based vision embeddings to evaluate the overall style consistency of the generated SFCC dataset. By computing the mean cosine similarity among all pairs of style vectors within the dataset, LSCI measures how consistently a model applies the same style across all outputs. The mean represents the average style similarity across all image pairs, while the standard deviation (std) reflects the variance of this similarity, indicating the stability of style application. The SI-DM achieved the highest mean (0.894) and the lowest standard deviation (0.042) among all models, indicating highly consistent style reproduction across different regions, time periods, and smoke patterns. The low std value suggests that the model maintains uniform style adaptation across varying environmental conditions, a critical attribute for satellite image applications, where spatial and temporal tonal uniformity is essential for reliable analysis.

While the quantitative comparison demonstrates that SI-DM achieves the highest mean LSCI score (0.894), the practical significance of this result merits further discussion. LSCI evaluates the consistency of high-level visual representations by measuring their proximity within a shared embedding space. In the context of SFCC, a high LSCI score indicates that smoke imagery acquired on different dates—and under varying atmospheric and illumination conditions—is mapped to highly comparable tonal representations. This consistency directly enhances the continuity and reliability of manual interpretation, as it allows analysts to more effectively distinguish genuine physical changes in smoke structures from observation-induced variability. Furthermore, such representational stability provides a robust foundation for constructing standardized, high-quality training datasets for data-driven smoke detection and segmentation models, thereby supporting downstream applications beyond visual analysis.

It is also crucial to acknowledge the operational boundaries of LSCI. Since the metric is built upon the CLIP visual encoder—trained primarily on large-scale natural image datasets—a domain gap inevitably exists when applied to satellite remote sensing imagery, particularly in scenes dominated by complex atmospheric effects or low-texture backgrounds. This discrepancy may impose a practical upper bound on the discriminative capability of LSCI in remote sensing contexts. Accordingly, the consistently high LSCI values achieved by SI-DM, which approach this perceived upper bound, should be interpreted as strong evidence of the model's ability to produce stable and semantically coherent SFCC representations, rather than as an indication that the metric is universally optimal or saturated.

Nevertheless, both ArtFID and LSCI have limitations. ArtFID is computed as the product of LPIPS and FID, implicitly assuming equal importance for both components. However, depending on the application, content preservation may be more critical than style matching, or vice versa. Thus, a parameterized evaluation framework that allows the weighting of these components would be desirable. Similarly, LSCI relies on CLIP's pre-trained visual–language embeddings, which are trained primarily on natural image domains. Consequently, in remote sensing imagery, where visual structures differ from natural images, the accuracy of style representation may be inherently constrained.

5.3. Comparative Analysis of Wildfire Smoke and Yellow Dust

As described in Figure 9, both wildfire smoke and yellow dust appear as aerosols in the initial FCC image. However, because the two aerosol types share similar color tones, they cannot be clearly distinguished from the initial FCC image alone. Therefore, additional satellite-derived variables were employed to examine the physical differences between the two aerosols.

The UVAI serves as an indicator of ultraviolet-absorbing aerosols and can exhibit high values for both dust and smoke. In contrast, the VISAI reflects particle size and scattering characteristics in the visible spectrum, typically showing higher responses for coarse particles such as dust. CHOCHO, a photochemical oxidation product of aromatic compounds and VOCs, is known to occur at higher concentrations in combustion-origin aerosols, such as wildfire smoke. Based on these relationships, Case 1, in which wildfire smoke and dust co-occurred, was analyzed by delineating the wildfire smoke area (in red) and the yellow dust area (in brown) (Figure 21).

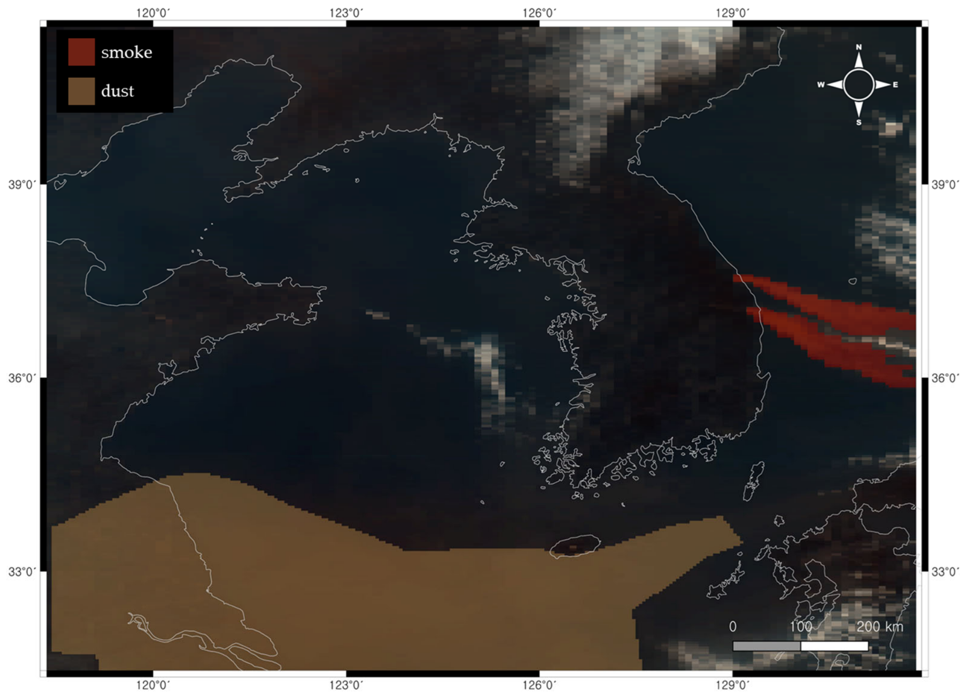

Figure 21. Delineation of the wildfire smoke area (in red) and the yellow dust area (in brown) in the SFCC image for Case 1 (5 March 2022, 04:45 (UTC)).

In Figure 22, red points represent wildfire pixels, and yellow points represent dust pixels. The scatter plot reveals two distinct clusters: dust forms a group characterized by higher VISAI and lower CHOCHO values, whereas wildfire smoke shows higher CHOCHO and lower VISAI values.

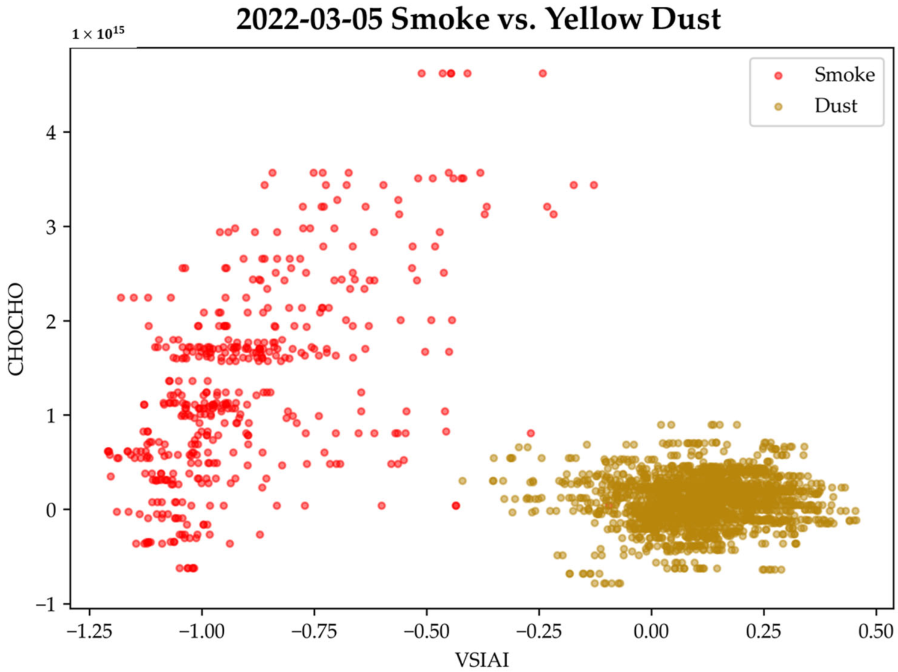

Figure 22. Separation of wildfire smoke and yellow dust using GEMS VISAI and CHOCHO for Case 1 (5 March 2022, 04:45 (UTC)).

These results suggest that style transfer-based visualization not only improves color-tone adjustment but also partially reflects the chemical and optical distinctions between aerosol types. Specifically, wildfire smoke exhibits increased CHOCHO concentrations indicative of combustion origin, while dust shows elevated VISAI values corresponding to coarse-particle dominance. Future studies may leverage these relationships to integrate style-transfer-based visualization with GEMS multi-wavelength indices (UVAI, VISAI, CHOCHO, etc.) to develop comprehensive aerosol classification methods.

5.4. Sensitivity to Reference Style Image (RSI) Selection

We also evaluated the robustness of the methods with respect to variations in the RSI. Figure 23 illustrates the different reference style images utilized in this analysis. RSI A, which was previously assessed in Figure 9, serves as the baseline reference; it was selected from scenes satisfying the following criteria: an Aerosol Index (AI) peak > 6, a cloud fraction between 30% and 60%, and a Solar Zenith Angle (SZA) in the range of 50–60°. These criteria represent moderate atmospheric conditions with clear visual separability between the smoke plume and the background.

To determine whether the observed performance trends remain consistent under diverse reference conditions, two additional RSIs (RSI B and RSI C) were introduced. RSI B was selected from scenes with an AI peak > 6, a high cloud fraction (70–90%), and a low SZA (10–30°). This RSI represents cloud-dominant conditions where the spectral and visual characteristics of smoke and clouds are heavily intertwined, increasing the ambiguity of the reference style statistics. In contrast, RSI C was selected from scenes with an AI peak > 3, a lower cloud fraction (10–30%), and a moderate SZA (30–50°). RSI C deviates from both RSI A and RSI B in terms of atmospheric composition and illumination geometry, thereby enabling a comprehensive assessment of model behavior across varying cloud and solar conditions.

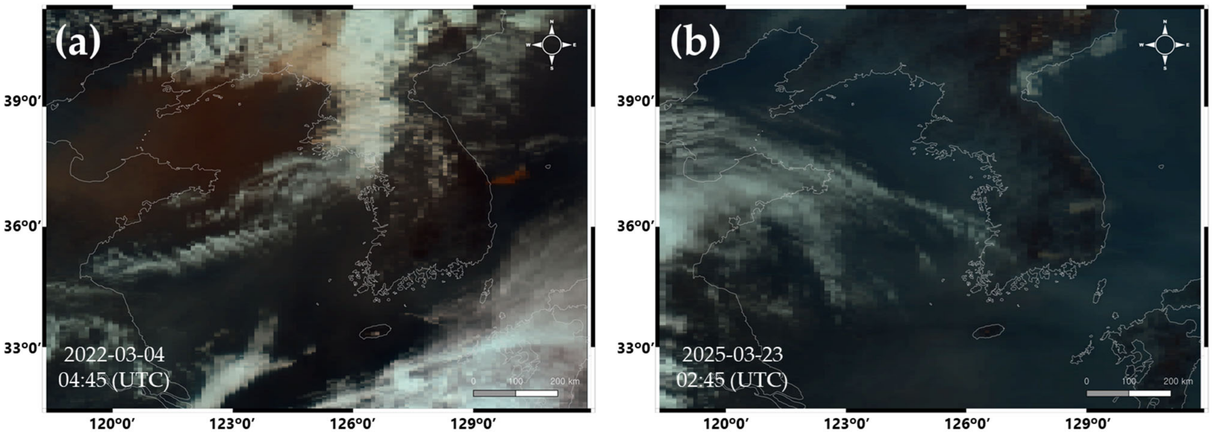

Figure 23. Reference style images (RSI B and RSI C) used to assess sensitivity to RSI selection. (**a**) RSI B, (**b**) RSI C.

The quantitative evaluation results for RSI B and RSI C are summarized in Table 9. Under the RSI B configuration, SI-DM achieves a mean LSCI of 0.900 ± 0.035, indicating high style consistency and minimal inter-scene variability. The small standard deviation suggests that SI-DM maintains stable performance even when the reference style is derived from cloud-dominant scenes. In comparison, ReHistoGAN and StyTr^2 yield lower mean LSCI values (0.836 and 0.861, respectively) with substantially larger standard deviations (0.081 and 0.070), reflecting an increased sensitivity to reference variability. ART-FID results show comparable performance for SI-DM and ReHistoGAN (both 1.670), while StyTr^2 exhibits a higher ART-FID (1.742), indicating a greater deviation from the reference distribution.

Table 9. Stability of evaluation metrics under different RSIs (↓: lower is better).

Model	**RSI A**		**RSI B**		**RSI C**	
	LSCI (Mean ± std)	**ArtFID (↓)**	**LSCI (Mean ± std)**	**ArtFID (↓)**	**LSCI (Mean ± std)**	**ArtFID (↓)**
SI-DM	0.894 (±0.042)	1.620	0.900 (±0.035)	1.670	0.894 (±0.044)	1.546
StyTr^2	0.857 (±0.046)	1.760	0.861 (±0.070)	1.742	0.867 (±0.053)	1.710
ReHistoGAN	0.867 (±0.046)	1.701	0.836 (±0.081)	1.670	0.832 (±0.059)	1.717

A consistent pattern is observed for RSI C. SI-DM records a mean LSCI of 0.894 ± 0.044 and an ART-FID of 1.546, demonstrating robust performance under differing illumination conditions. Again, ReHistoGAN and StyTr^2 show lower mean LSCI values (0.832 and 0.867) and higher ART-FID values (1.717 and 1.710, respectively), indicating that their outputs are more susceptible to fluctuations in reference style statistics associated with cloud fraction and SZA variations.

Across both RSI B and RSI C, although absolute metric values vary depending on the atmospheric and illumination characteristics of the reference image, the relative performance ranking remains consistent. In particular, SI-DM consistently exhibits the highest style consistency with the lowest variability, confirming its reduced sensitivity to reference selection. These findings suggest that SI-DM's performance is not contingent upon a specific reference image but is instead attributable to its inherent ability to maintain stable color and structural statistics across diverse reference conditions.

6. Conclusions

This study aimed to ensure visual consistency and enhance the perceptual clarity of smoke in imagery generated by the GEMS geostationary environmental satellite by comparing state-of-the-art deep learning-based style transfer methods. To our knowledge, this is the first study to systematically apply style transfer techniques to satellite-borne wildfire-smoke imagery, to conduct visual validation using real wildfire cases from Korea (2018–2025), and to quantitatively and qualitatively assess how style-transfer models influence the visibility and quality of smoke.

In detailed comparative experiments involving ECDF correction, the transformer-based StyTr^2, ReHistoGAN, and SI-DM, the proposed SI-DM-based SFCC framework consistently demonstrated superior performance. SI-DM effectively preserved fine-scale smoke structures, reduced inter-patch discontinuities, and achieved more natural background color-tone alignment across time. In particular, its self-attention-based style-injection mechanism enabled clearer delineation of smoke–background boundaries, while the iterative refinement inherent to the diffusion process contributed to improved perceptual quality and visual stability.

Beyond visual enhancement, these properties translate into distinct practical application pathways. By mitigating observation-induced color variability while preserving intrinsic structural characteristics, standardized SFCC imagery can significantly enhance the robustness of data-driven smoke detection and segmentation models across varying illumination and viewing geometries. Furthermore, the framework facilitates the generation of temporally consistent monitoring products from high-frequency geostationary observations, thereby ensuring reliable time-series interpretation and seamless integration into operational wildfire and air-quality monitoring workflows.

By explicitly accounting for the domain-specific characteristics of satellite imagery and the distinct visual signatures of wildfire smoke, this work differs fundamentally from prior true-color-oriented studies. Under the objective of standardizing GEMS-based smoke imagery, the proposed approach—focused on improving tonal consistency and visual discriminability—provides a practical and extensible foundation for earlier detection and more reliable operational response to wildfire events.

This study also underscores the limitation that conventional quantitative metrics (e.g., SSIM, LPIPS) do not fully reflect the perceptual quality of style-transferred imagery, as they primarily assess similarity to the original rather than the fidelity of style application. To mitigate these gaps, we employed ArtFID and LSCI, which complement traditional metrics and provide more reliable, task-aligned assessments for style-transfer evaluation. Consequently, future research should further develop and adopt evaluation protocols tailored to satellite-image style transfer, integrating task-relevant indicators to better capture both content preservation and style consistency.

In summary, this work originates from the problem of standardizing satellite-based smoke visualization and represents an initial, comprehensive application of modern deep learning to smoke-focused style transfer. It highlights promising avenues toward automated wildfire detection, enhanced early-warning systems, and data-driven aerosol characterization in real-world settings.

Author Contributions: Conceptualization, Y.J. and Y.L.; methodology, Y.J. and Y.L.; formal analysis, Y.J.; data curation, Y.J.; writing—original draft preparation, Y.J.; writing—review and editing, S.H.K., M.K., J.-A.Y., K.-H.S., S.-M.K., S.-Y.K., G.K., J.-J.K. and Y.L.; supervision, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Institute of Environmental Research (NIER), funded by the Ministry of Climate, Energy and Environment (MCEE) of the Republic of Korea (NIER-2023-01-02-096). This research was supported by a grant (2021-MOIS37-002) of Intelligent Technology Development Program on Disaster Response and Emergency Management funded by Ministry of Interior and Safety (MOIS, Korea). This study was carried out with the support of R&D Program for Forest Science Technology (Project No. RS-2025-25404070) provided by Korea Forest Service (Korea Forestry Promotion Institute). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2025-25431264). This work was supported by the National Research Foundation (NRF), Korea, under project BK21 FOUR.

Data Availability Statement: The data will be made available upon reasonable request. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Potapov, P.; Tyukavina, A.; Turubanova, S.; Hansen, M.C.; Giglio, L.; Hernandez-Serna, A.; Lima, A.; Harris, N.; Stolle, F. Unprecedentedly high global forest disturbance due to fire in 2023 and 2024. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2505418122. [CrossRef]
2. Jolly, W.M.; Cochrane, M.A.; Freeborn, P.H.; Holden, Z.A.; Brown, T.J.; Williamson, G.J.; Bowman, D.M. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* **2015**, *6*, 7537. [CrossRef]
3. Abatzoglou, J.T.; Williams, A.P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11770–11775. [CrossRef] [PubMed]
4. Hadiwijoyo, E.; Rijal, H.B.; Abdullah, N. Cascading Impacts of Wildfire Emissions on Air Quality, Human Health, and Climate Change Based on Literature Review. *Fire* **2025**, *8*, 471. [CrossRef]
5. Johnston, F.H.; Henderson, S.B.; Chen, Y.; Randerson, J.T.; Marlier, M.; DeFries, R.S.; Kinney, P.; Bowman, D.M.; Brauer, M. Estimated global mortality attributable to smoke from landscape fires. *Environ. Health Perspect.* **2012**, *120*, 695–701. [CrossRef] [PubMed]
6. Lopes, M.; Monteiro, A.; Mouzourides, P.; Kouis, P. The health burden of wildfire smoke in a changing climate: Exposure, risks, and strategies for mitigation. *Curr. Opin. Environ. Sci. Health* **2025**, *46*, 100631. [CrossRef]
7. Chowdhury, S.; Hänninen, R.; Uppstu, A.; Sofiev, M.; Aunan, K. Global health burden from acute exposure to fine particles emitted by fires. *npj Clean Air* **2025**, *1*, 24. [CrossRef]
8. Eck, T.F.; Holben, B.; Reid, J.; Dubovik, O.; Smirnov, A.; O'neill, N.; Slutsker, I.; Kinne, S. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res. Atmos.* **1999**, *104*, 31333–31349. [CrossRef]
9. Zhao, L.; Liu, J.; Peters, S.; Li, J.; Mueller, N.; Oliver, S. Learning class-specific spectral patterns to improve deep learning-based scene-level fire smoke detection from multi-spectral satellite imagery. *Remote Sens. Appl. Soc. Environ.* **2024**, *34*, 101152. [CrossRef]
10. Chrysoulakis, N.; Herlin, I.; Prastacos, P.; Yahia, H.; Grazzini, J.; Cartalis, C. An improved algorithm for the detection of plumes caused by natural or technological hazards using AVHRR imagery. *Remote Sens. Environ.* **2007**, *108*, 393–406. [CrossRef]
11. Wang, J.; Song, W.; Wang, W.; Zhang, Y.; Liu, S. A new algorithm for forest fire smoke detection based on MODIS data in Heilongjiang province. In Proceedings of the 2011 International Conference on Remote Sensing, Environment and Transportation Engineering, Nanjing, China, 24–26 June 2011; pp. 5–8.
12. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention. *Remote Sens.* **2019**, *11*, 1702. [CrossRef]
13. Zhao, L.; Liu, J.; Peters, S.; Li, J.; Oliver, S.; Mueller, N. Investigating the impact of using IR bands on early fire smoke detection from landsat imagery with a lightweight CNN model. *Remote Sens.* **2022**, *14*, 3047. [CrossRef]
14. Zhao, L.; Liu, J.; Peters, S.; Li, J.; Mueller, N.; Oliver, S. Cross-sensor transfer learning for fire smoke scene detection using variable-bands multi-spectral satellite imagery aided by spectral patterns. *Int. J. Remote Sens.* **2024**, *45*, 3332–3348. [CrossRef]
15. Liu, J.; Li, J.; Peters, S.; Zhao, L. A transformer boosted UNet for smoke segmentation in complex backgrounds in multispectral LandSat imagery. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101283. [CrossRef]
16. Andrade, S.; Shafiei, N.; Mehlitz, P. Smoke or cloud: Real-time satellite image segmentation in a wildfire data integration application. *Comput. Geosci.* **2025**, *204*, 105960. [CrossRef]

17. Wang, T.; Hong, J.; Han, Y.; Zhang, G.; Chen, S.; Dong, T.; Yang, Y.; Ruan, H. AOSVSSNet: Attention-guided optical satellite video smoke segmentation network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8552–8566. [CrossRef]
18. Pinto, M.M.; Libonati, R.; Trigo, R.M.; Trigo, I.F.; DaCamara, C.C. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 260–274. [CrossRef]
19. Yin, Z.; Zou, L.; Sun, J.; Zhang, H.; Zhang, W.; Shen, X. A Nonlinear Radiometric Normalization Model for Satellite Imgaes Time Series Based on Artificial Neural Networks and Greedy Algroithm. *Remote Sens.* **2021**, *13*, 933. [CrossRef]
20. Su, H.; Hu, Y.; Lu, H.; Sun, W.; Du, Q. Diversity-driven multikernel collaborative representation ensemble for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2861–2876. [CrossRef]
21. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
22. Choi, W.J.; Moon, K.-J.; Kim, G.; Lee, D. Reliability analysis based on air quality characteristics in East Asia using primary data from the test operation of geostationary environment monitoring spectrometer (GEMS). *Atmosphere* **2023**, *14*, 1458. [CrossRef]
23. Choi, H.; Liu, X.; Jeong, U.; Chong, H.; Kim, J.; Ahn, M.H.; Ko, D.H.; Lee, D.-W.; Moon, K.-J.; Lee, K.-M. Geostationary Environment Monitoring Spectrometer (GEMS) polarization characteristics and correction algorithm. *Atmos. Meas. Tech.* **2024**, *17*, 145–164. [CrossRef]
24. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
25. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
26. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
27. Liu, S.; Lin, T.; He, D.; Li, F.; Wang, M.; Li, X.; Sun, Z.; Li, Q.; Ding, E. Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 6649–6658.
28. Karatzoglidi, M.; Felekis, G.; Charou, E. Neural Style Transfer for Remote Sensing. *arXiv* **2020**, arXiv:2007.15920. [CrossRef]
29. Sokolov, M.; Henry, C.; Storie, J.; Storie, C.; Alhassan, V.; Turgeon-Pelchat, M. High-Resolution Semantically Consistent Image-to-Image Translation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 482–492. [CrossRef]
30. Sun, Y.; Wu, M. Flexible style transfer from remote sensing images to maps. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104566. [CrossRef]
31. Alvarado, L.M.; Richter, A.; Vrekoussis, M.; Hilboll, A.; Kalisz Hedegaard, A.B.; Schneising, O.; Burrows, J.P. Unexpected long-range transport of glyoxal and formaldehyde observed from the Copernicus Sentinel-5 Precursor satellite during the 2018 Canadian wildfires. *Atmos. Chem. Phys.* **2020**, *20*, 2057–2072. [CrossRef]
32. Sorenson, B.T.; Zhang, J.; Reid, J.S.; Xian, P.; Jaker, S.L. Ozone Monitoring Instrument (OMI) UV aerosol index data analysis over the Arctic region for future data assimilation and climate forcing applications. *Atmos. Chem. Phys.* **2023**, *23*, 7161–7175. [CrossRef]
33. Afifi, M.; Brubaker, M.A.; Brown, M.S. Histogan: Controlling colors of gan-generated and real images via color histograms. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7941–7950.
34. Deng, Y.; Tang, F.; Dong, W.; Ma, C.; Pan, X.; Wang, L.; Xu, C. Stytr2: Image style transfer with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11326–11336.
35. Chung, J.; Hyun, S.; Heo, J.-P. Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 8795–8805.
36. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
37. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
38. Wright, M.; Ommer, B. Artfid: Quantitative evaluation of neural style transfer. In Proceedings of the DAGM German Conference on Pattern Recognition, Konstanz, Germany, 27–30 September 2022; pp. 560–576.
39. Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; Courville, A. On the spectral bias of neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 5301–5310.

40. Falck, F.; Pandeva, T.; Zahirnia, K.; Lawrence, R.; Turner, R.; Meeds, E.; Zazo, J.; Karmalkar, S. A Fourier Space Perspective on Diffusion Models. *arXiv* **2025**, arXiv:2505.11278. [CrossRef]
41. Sjögren, O.; Pihlgren, G.G.; Sandin, F.; Liwicki, M. Identifying and mitigating flaws of deep perceptual similarity metrics. *arXiv* **2022**, arXiv:2207.02512. [CrossRef]