https://doi.org/10.65646/3cp3es20dpa7t232

A GAN-Based Inpainting Framework for Smoke Removal in Critical Safety Imagery

Isuru Atthanayake
Dept. of Computer Science
Informatics Institute of Technology
in collaboration with University of Westminster
Colombo,Srilanka.
isuru3.atthanayake@gmail.com

Siyad Najmudeen
Dept.of Electrical and Computer Engineering
The Open University Of Srilanka
Colombo,Srilanka
nmsiy@ou.ac.lk

***Abstract*— Smoke generated during fire and other critical incidents severely degrades the visual quality of captured imagery, hindering safety monitoring, evacuation planning, and firefighting decision-making. Existing smoke- and haze-removal techniques, ranging from prior-based dehazing to single-branch generative models, struggle with thick, structureless smoke and often leave residual artefacts, colour distortion, or require large paired real-world datasets. Across the reviewed literature, prior-based dehazing such as the Dark Channel Prior typically reports Peak Signal-to-Noise Ratio (PSNR) values of only around 13-17 dB under dense haze or smoke, while GAN-based desmoking methods raise this to roughly 18-25 dB yet still report Structural Similarity Index (SSIM) scores below 0.85 and visible residual smoke once density is high, and multi-model pipelines such as DHSGAN incur added computational cost from training two networks separately. This paper proposes a two-stage Generative Adversarial Network (GAN) framework for blind-inpainting-based smoke removal. The framework decouples smoke localization from image reconstruction through two cooperating sub-networks: a Mask Prediction Network (MPN) that identifies smoke-affected regions, and a Robust Inpainting Network (RIN) that reconstructs the identified regions using adversarial and VGG16-based perceptual losses. This method reviews the relevant literature, presents GAN fundamentals for completeness, detail the proposed system architecture, network designs, dataset considerations, design goals, and loss formulation, and position the framework against existing smoke- and haze-removal approaches identified through a structured literature review. The proposed design is intended as the basis for a subsequent implementation and empirical evaluation.**

Keywords —blind image inpainting, deep learning, disaster imagery, Generative Adversarial Networks (GANs), image restoration, residual networks, smoke removal

I. INTRODUCTION

The presence of smoke during a critical fire incident introduces severe noise into captured imagery, obscuring structural and human detail that safety-monitoring and rescue personnel [9]. Reduced visibility directly obstructs evacuation planning and increases risk during firefighting operations [8]. This visibility reduction negatively impacts change detection mapping, climate change monitoring and disaster monitoring[16].Prior image-restoration research has largely targeted generic corruptions such as noise, low resolution, and poorly defined boundaries [10], [13]; smoke, however, is dynamic and structureless, varying unpredictably in density, shape, and colour cast, which makes it substantially harder to remove than haze or fog of a fixed physical model [5].

Beyond fire safety, reliable smoke removal has applications across autonomous vehicles, where visual obstruction from smoke can degrade the performance of perception systems, and disaster management more broadly, where restored imagery supports emergency response and post-incident safety monitoring. These applied domains motivate a restoration approach that is both accurate and generalizable across varying capture conditions, rather than tuned narrowly to a single dataset or lighting scenario.

Recent work has applied GANs to desmoking, but existing approaches leave residual smoke under thick, high-opacity conditions [3], depend on synthetic or narrow-domain datasets [6], or require multiple independently trained models that increase computational cost [7], [14]. This paper addresses these gaps by proposing a framework, rather than an implementation: Blind-inpainting along with GAN architecture that separates the smoke-affected pixels from the images and reconstructs them. The framework is derived from a structured review of existing dehazing, desmoking, and inpainting literature, and is intended as the design basis for a subsequent implementation and empirical evaluation, which is reported as a separate study.

The remainder of this paper is organized as follows. Section II related work and presents background on GANs for completeness. Section III presents the proposed framework and methodology, including its architecture, sub-networks, dataset considerations, and loss formulation. Section IV states the framework's design goals. Section V positions the framework against existing approaches. Section VI reports results,

Section VII discusses the framework and its anticipated limitations, and Section VIII concludes the paper.

Fig. 1 summarizes the conceptual landscape surrounding smoke removal in images: the common corruption types it is related to, the technologies applied to the problem, the evaluation methods used, applied domains and limitations of this research.



Figure 1:Concept graph of smoke removal in images

II. RELATED WORK

1) *Prior-Based and CNN-Based Methods*

Early restoration approaches treated smoke as a variant of haze and applied prior-based dehazing methods such as the Dark Channel Prior, which estimates a transmission map from statistical regularities in haze-free images [5]. These methods perform poorly on thick smoke, where the underlying physical assumptions break down and object detection within the recovered image becomes unreliable. CNN-based approaches later improved on hand-crafted priors by learning restoration mappings directly from data, but single-branch architectures still combine the tasks of detecting corrupted regions and reconstructing them, limiting output quality when corruption is dense or irregular.

2) *GAN-Based Desmoking Methods*

GAN-based desmoking methods followed the similar way as haze removal. DHSGAN pairs two separate models for fog and smoke removal, at the cost of increased computational requirements [7]. The Desmoke GAN framework uses an attention mechanism with a PatchGAN classifier and removes smoke to a useful degree, but leaves visible residue when input smoke is thick [3]. Generative Smoke Removal applies a Pix2Pix-style architecture to laparoscopic imagery, but relies heavily on synthetic training data, limiting generalization to real-world scenes [6]. MSDGAN further extends this direction with a multi-scale dilated architecture, but benchmarking against AOD-Net, CAP, DehazeNet, and the Dark Channel Prior indicates that restoration quality is still highly sensitive to the density and structure of the smoke present in the input image [14].

3) *Blind Image Inpainting*

Separately, the image-inpainting literature has established blind inpainting, reconstructing missing or corrupted regions without a supplied ground-truth mask, as a viable strategy for irregular, unknown corruptions [12]. Decoupled probabilistic modelling with a Style-GAN backbone has been used to reconstruct missing image regions, but tends to apply only minor adjustments rather than substantial reconstruction, and struggles when contextual indicators are sparse [4]. Broader surveys of GAN-based computer vision [11] and of deep-learning image inpainting [10] confirm that image inpainting produces more realistic restorations than denoising, super-resolution, or colourization alone, but note that few inpainting methods have been purpose-built for the specific structure of smoke corruption. This gap, a blind-inpainting GAN designed specifically around smoke's dynamic, structureless character, motivates the framework proposed in Section III.

A. *Background: Generative Adversarial Networks*

GANs are composed of two cooperating sub-models: a generator and a discriminator. The generator is trained to produce synthetic data that closely resembles real instances, while the discriminator is trained to distinguish generated data from real data. Training proceeds as a minimax game in which the generator tries to imitate the statistical properties of real data closely enough to trick the discriminator, while the discriminator tries to remain accurate; the resulting competition allows GANs to produce highly realistic output and has made them a core technique across computer vision, artistic synthesis, and generation of synthetic data [1].

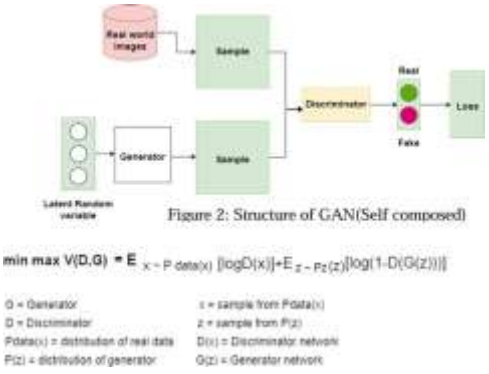


Figure 2: Structure of GAN(Self composed)

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z)))]$$

G = Generator — x = sample from $P_{data}(x)$
D = Discriminator — z = sample from $P(z)$
$P_{data}(x)$ = distribution of real data — D(x) = Discriminator network
P(z) = distribution of generator — G(z) = Generator network

Figure 2. General structure of a Generative Adversarial Network

III. PROPOSED FRAMEWORK AND METHODOLOGY

A. *Design Rationale and Overview*

The framework uses the image inpainting technique to remove smoke from images through the approach of conditional GANs. Smoke-filled images and smoke-free images from the dataset are used as the training data. A GAN architecture, comprising a generator and

a discriminator, is then used to remove the smoke in images. The model is trained on the dataset to recognize and remove the smoke in the images, filling the smoke-covered area using the inpainting technique, and its performance is finally evaluated.

The framework's specific novelty and research contribution, relative to the existing GAN-based desmoking and image-restoration frameworks reviewed in Section II, lies in this decoupling itself. Existing GAN-based desmoking methods typically use a single generator to perform detection and reconstruction jointly [3], [6], [7], [14], which conflates two objectives, locating corrupted pixels and synthesizing reasonable content, that benefit from different network behaviour. The proposed framework instead separates these into two cooperating sub-networks, the MPN and the RIN, each optimized for its own objective, and is intended to generalize more robustly across varying smoke density than single-branch architectures, since reconstruction quality no longer depends on a single network it correctly performs in both tasks at once.

The relationship between smoke removal and image inpainting is direct: rather than modelling the physical process by which smoke scatters or weakens light, as prior-based dehazing does [5], the framework treats smoke-affected pixels as missing or corrupted content and reconstructs them using image inpainting based on the surrounding visual information.Specifically, the MPN first processes the smoke-filled input image and predicts a spatial mask that identifies the pixels affected by smoke; this predicted mask, together with the original image, is then passed to the RIN, which reconstructs plausible underlying content within the masked region while leaving the unmasked, smoke-free pixels unchanged. Because the reconstruction stage depends directly on the localization stage, the two networks are trained jointly within a single adversarial objective, so that reconstruction quality provides a signal that shapes the predicted mask, and mask quality in turn determines which regions the reconstruction stage is asked to fill.

The framework was developed through the structured literature review summarized in Section II, which identified three recurring limitations in existing desmoking approaches: residual smoke under thick conditions, reliance on synthetic or narrow-domain training data, and the computational overhead of independently trained models. Each architectural decision responds directly to one of these limitations: the two-network MPN-RIN split addresses the residual-smoke limitation by separating localization from reconstruction the choice of the Smoke100k dataset, which pairs smoke-filled and smoke-free real images rather than purely synthetic ones, addresses the synthetic-data limitation; and combining both networks within a single adversarial objective, rather than training them independently as in DHSGAN [7], addresses the computational-cost limitation. The five design goals stated formally in Section IV were derived from this same review and used throughout development to check candidate architectural choices.

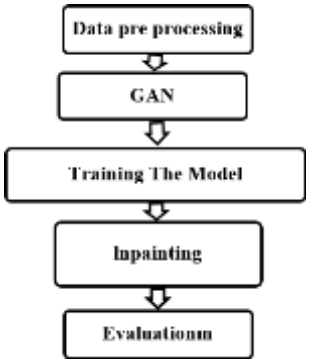


Figure 3. High-level solution methodology for the proposed framework.

B. System-Level Architecture

Beyond the model architecture, the framework is intended to sit within a three-tier system architecture, as shown in Fig. 4: a presentation tier through which an operator supplies an input image, a logic tier that performs model loading and restoration, and a data tier that manages dataset, trained-model, and output-image storage.

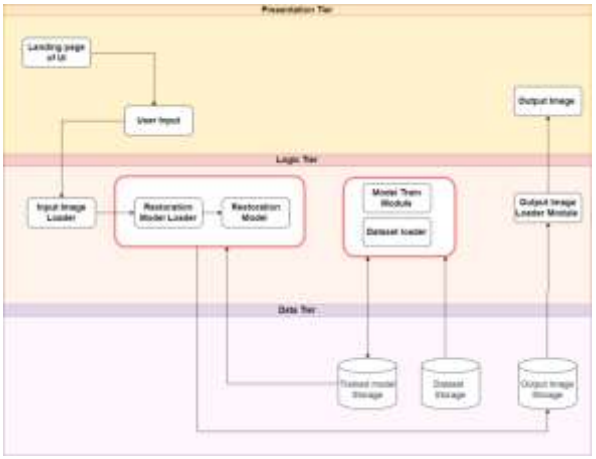


Figure 4. Proposed three-tier system architecture for deploying the framework.

In the data tier, dataset storage holds the images used to train and test the model, trained-model storage holds the pre-trained MPN/RIN weights, and output-image storage holds the restored images produced by the model. In the logic tier, an input-image loader supplies images to the model, either user-supplied or drawn from the stored dataset; a restoration-model loader retrieves the trained model; and the restoration model itself, the MPN–RIN described in the following subsections, performs the reconstruction. A model-train module and dataset loader support (re-)training the model against the stored dataset, and an output-image loader module returns the restored image to the presentation tier for the operator to view.

C. Mask Prediction Network(MPN)

The MPN takes the smoke-obscured image as input and predicts a spatial mask indicating the regions affected

by smoke. It consists of an input convolutional layer that projects the image into a base feature width, a sequence of four residual blocks that progressively expand the feature width from the base channel count toward a wider bottleneck ("neck") channel count, a bilinear up-sampling layer, and an output convolutional stage of three convolutional layers with intermediate ReLU activations and a final Tanh activation that bounds the predicted mask to a fixed output range. Because the network is trained without direct mask supervision, relying instead on smoke-filled/smoke-free image pairs and the adversarial and reconstruction signal propagated from the RIN stage, it must learn to distinguish smoke-affected regions from clean background purely from that indirect signal. Ground-truth masks included in the training dataset are deliberately withheld from this loss and are used only after training, to evaluate localization quality (Section VI), which makes mask quality central to the overall framework's performance.

D. Robust Inpainting Network(RIN)

The RIN receives the MPN's predicted mask together with the original image and reconstructs the masked, smoke-affected region. It mirrors the MPN's layer composition, an input convolutional layer, residual blocks, bilinear up-sampling, and a multi-layer output convolutional stage, but is optimized specifically for photorealistic reconstruction rather than localization. The two networks are combined within a single GAN, referred to here as the SmokeRemovalGAN, in which the MPN acts as the generator, the RIN acts as the inpainter refining the generator's output, and a discriminator provides the adversarial training signal for the combined pipeline.

E. Discriminator Network

The discriminator is a sequential convolutional network of two 3×3 convolutional layers with stride 1 and padding 1, separated by a LeakyReLU activation and finished with a sigmoid output, classifying image regions as real (smoke-free ground truth) or fake (reconstructed by the RIN). Operating on local patches rather than the whole image at once, this discriminator design encourages locally coherent, realistic textures in the reconstructed region rather than blurred or averaged content, which is a known weakness of purely regression-based inpainting.

F. Loss Formulation

Reconstruction quality is measured with a feature (perceptual) loss computed using a pre-trained VGG16 network. Because VGG16 is pre-trained rather than trained from scratch, it supplies stable, meaningful feature representations at negligible additional training cost. The reconstructed and ground-truth smoke-free images are resized to a common resolution and compared in VGG16 feature space using a mean-squared-error criterion, which is combined with the discriminator's adversarial loss so that the MPN and RIN are jointly optimized to produce realistic reconstructed regions while preserving the overall visual features of the original image

G. Dataset and Implementation Consideration

The framework is designed around the Smoke100k dataset [15], a computer-vision image dataset providing smoke-filled images, corresponding smoke-free ground-truth images, and smoke masks, together with smoke-filled and smoke-free image pairs for testing. Only the smoke-filled and smoke-free image pairs are used to train the MPN-RIN pipeline, consistent with the blind, mask-free formulation described in Section III-C; Smoke100k's provided masks are withheld from training and reserved for evaluating the MPN's predicted-mask quality after training (Section VI). Because the framework performs blind inpainting, the trained model does not require mask input at inference time either; end users supply only a smoke-filled image, which matches the practical constraint that smoke masks are not available when the system is deployed in the field.

At the implementation level, the framework is intended to be realized in Python using PyTorch, chosen for its dynamic computation graph, GPU acceleration, and strong ecosystem support for custom GAN architectures. This choice of tooling is noted here for completeness; implementation details, training configuration, and empirical results fall outside the scope of this design-stage paper and are addressed in a companion implementation study.

IV. DESIGN GOALS

Table I states the five design goals that guided the proposed framework, derived from the target use case of safety-critical imagery restoration.

Table I:DESIGN GOALS OF THE PROPOSED FRAMEWORK

Design Goal	Description
Performance	Reconstruct realistic, high-quality smoke-free images without losing background detail, using blind inpainting, and perform efficiently within an acceptable time frame.
Adaptability	Restore images captured under varying lighting conditions, according to whatever input image is supplied.
Usability	Remain operable by non-technical users, requiring only minimal guidance.
Scalability	Work with large datasets exhibiting visual degradations such as challenging lighting conditions and varied backgrounds.
Correctness	Effectively eliminate smoke for safety monitoring during rescue operations, prioritizing precision and reliability.

V. COMPARATIVE POSITIONING

Table II summarizes how the proposed MPN–RIN framework is intended to address limitations reported for representative existing approaches identified in Section II.

TABLE II. COMPARISON WITH EXISTING SMOKE- AND HAZE-REMOVAL APPROACHES

Approach	***Method***	***Key Limitation***	***How the Proposed Framework Responds***
[5]	Dark Channel Prior dehazing	Struggles with thick smoke; colour distortion	MPN explicitly localizes smoke regions before reconstruction, avoiding global colour shift
[6]	GAN for laparoscopic smoke removal	Relies on synthetic training data	RIN trained with VGG16 perceptual loss for realistic, generalizable texture
[7]	DHSGAN (fog + smoke)	Two separate models; high compute cost	Single unified MPN + RIN pipeline reduces redundant computation
[3]	Desmoke GAN (single image)	Residual smoke left when density is high	Blind mask targets exactly the smoke-affected pixels regardless of density
[4]	Decoupled probabilistic inpainting	Minor adjustments only; weak on missing detail	RIN performs full region reconstruction rather than incremental correction
[14]	MSDGAN multi-scale dilated GAN	Restoration quality still sensitive to smoke density/structure	MPN–RIN decoupling targets density variation explicitly via localization before reconstruction

VI. RESULTS

This paper presents a proposed framework rather than a trained system, and no implementation has yet been built; consequently, no image results or quantitative metrics can be reported at this stage, and none are fabricated here. The companion implementation study is intended to instantiate the MPN and RIN architecture described in Section III and evaluate it using standard image-restoration metrics: the Structural Similarity Index (SSIM), which measures perceived structural fidelity between the reconstructed and ground-truth images; Peak Signal-to-Noise Ratio (PSNR), which measures pixel-level reconstruction fidelity; and binary cross-entropy loss on the predicted masks, which measures MPN localization quality. These metrics would be benchmarked against the approaches summarized in Table II, and the resulting preliminary results will be reported once training and evaluation are complete.

VII. DISCUSSION

The MPN-RIN approach is expected to perform better than the existing desmoking methods reviewed in Section II for reasons tied directly to each method's reported limitation, summarized in Table II. Against prior-based dehazing such as the Dark Channel Prior, which estimates a single global transmission map and consequently distorts colour once its physical assumptions break down under thick smoke [5], the MPN's explicit, learned localization step is expected to confine reconstruction to the affected region and so avoid a global colour shift. Against single-branch GAN methods that conflate detection and reconstruction in one generator [3], [6], decoupling these tasks is expected to let the RIN specialize in photorealistic reconstruction, supported by the VGG16 perceptual loss, rather than dividing network capacity between localization and synthesis. Against two-model pipelines such as DHSGAN, which train separate networks for fog and smoke at additional computational cost [7], the shared adversarial objective linking the MPN and RIN is expected to reduce redundant computation while still allowing each sub-network to specialize. Finally, because the framework performs blind inpainting rather than supervised localization, it is expected to remain applicable across the varying, unlabelled smoke conditions reported as a weakness of density-sensitive methods such as MSDGAN [14], though this claim, like the others in this section, is a design expectation to be confirmed empirically rather than a demonstrated result.

Several limitations of the design are anticipated. Training two cooperating networks under a shared adversarial objective is more computationally demanding than a single-branch model, and blind mask prediction, while removing the need for annotated masks, may be less accurate than supervised localization under very high smoke density. The framework's reliance on a diverse training dataset for adaptability across lighting conditions is also untested at this design stage and forms a key subject of the planned empirical evaluation. Finally, because Smoke100k is the primary dataset considered, generalization to smoke captured by different camera sensors or in substantially different environments remains an open question for future work.

VIII. CONCLUSION

This paper proposed a two-stage GAN-based inpainting framework for smoke removal in critical safety imagery, decoupling smoke localization (MPN) from image reconstruction (RIN) and combining adversarial and VGG16-based perceptual objectives within a three-tier system architecture. The design was derived from a structured review of dehazing, desmoking, and blind-inpainting literature, grounded in GAN fundamentals, and positioned against representative existing approaches using five explicit design goals. It should be stated clearly that the contribution presented in this paper is a framework and

design proposal, not a trained or benchmarked system: no image results, PSNR/SSIM scores, or other quantitative metrics have been produced, and the expected advantages argued for in Section VII remain design expectations rather than demonstrated outcomes. Its practical effectiveness therefore remains to be validated empirically; future work will implement the proposed architecture and evaluate it against these design goals and comparative baselines in the companion implementation study.

ACKNOWLEDGMENT

The author gratefully acknowledges the guidance and support provided by his supervisor throughout the development of the proposed framework.

REFERENCES

[1] Y. Weng, S. Ding, and T. Zhou, "A survey on improved GAN based image inpainting," in Proc. 2022 2nd Int. Conf. Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2022, pp. 319–322, doi: 10.1109/ICCECE54139.2022.9712740.

[2] W.-T. Chen et al., "Color channel-based smoke removal algorithm using machine learning for static images," in Proc. 2018 25th IEEE Int. Conf. Image Processing (ICIP), Athens, Greece, 2018, pp. 2855–2859.

[3] Y. Huang et al., "Single image desmoking via attentive generative adversarial network for smoke detection process," Fire Technology, vol. 57, no. 6, pp. 3021–3040, 2021.

[4] W. Li et al., "Image inpainting via iteratively decoupled probabilistic modeling," arXiv:2212.02963, 2023.

[5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Miami, FL, 2009, pp. 1956–1963.

[6] O. Sidorov, C. Wang, and F. A. Cheikh, "Generative smoke removal," arXiv:1902.00311, 2019.

[7] R. Malav et al., "DHSGAN: An end to end dehazing network for fog and smoke," in Computer Vision, Pattern Recognition, Image Processing, and Graphics, 2019, pp. 593–608.

[8] A. A. Khan et al., "A review of critical fire event library for buildings and safety framework for smart firefighting," Int. J. Disaster Risk Reduction, vol. 83, p. 103412, 2022.

[9] Z. Wang et al., "Predicting transient building fire based on external smoke images and deep learning," J. Building Engineering, vol. 47, p. 103823, 2022.

[10] H. Xiang et al., "Deep learning for image inpainting: A survey," Pattern Recognition, vol. 134, p. 109046, 2023.

[11] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy," Computer Science Review, 2023.

[12] N. Cai et al., "Blind inpainting using the fully convolutional neural network," The Visual Computer, vol. 33, no. 2, pp. 249–261, 2017.

[13] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," Applied Soft Computing, vol. 138, p. 110176, 2023.

[14] K. Gwak and J. R. Young, "MSDGAN: Multi-scale dilated generative adversarial network for smoke removal and image restoration," 2023. [Online]. Available: https://doi.org/10.2139/ssrn.4429184

[15] H.-Y. Cheng et al., "Smoke 100k: A database for smoke detection," in Proc. 2019 IEEE 8th Global Conf. Consumer Electronics (GCCE), 2019, pp. 596–597.

[16] Neagoe, I.C., Vaduva, C. and Datcu, M. (2021). Haze and Smoke Removal for Visualization of Multispectral Images: A DNN Physics Aware Architecture. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. 11 July 2021. Brussels, Belgium: IEEE, 2102–2105. Available from https://doi.org/10.1109/IGARSS47720.2021.9553735