TruRescue: A Federated Learning Based Drone Driven Road Emergency Assistant with Generative AI

Jiaxuan He[1], Chen-Yeou Yu[2], Olivia Lewis[3] and Wensheng Zhang[4]

[1, 2, 3] Department of Computer and Data Science, Truman State University, Kirksville MO 63501, USA
fo35563@truman.edu, cyyu@truman.edu, we53156@truman.edu
[4] Department of Computer Science, Iowa State University, Ames IA 50011, USA
wzhang@iastate.edu

Abstract. Timely and accurate assessment of road traffic accidents is critical for effective emergency response. However, aerial imagery captured by UAVs is often degraded by motion blur, occlusion, and suboptimal visibility—conditions that impair detection accuracy and delay dispatch decisions. This paper presents **TruRescue**, a decentralized UAV-based framework that integrates **Federated Learning (FL)** and **Generative AI (GenAI)** to enhance image quality, support privacy-preserving accident detection, and enable real-time risk estimation. Each UAV locally trains a detection model while contributing updates to a global model via FL. Simultaneously, a generative module performs multi-view fusion and reconstruction to recover high-fidelity accident scenes. A novel **CDV-based risk index**—incorporating Clarity gain (C), Damage severity (D), and Visibility level (V)—quantifies the scene risk to guide emergency dispatch.
We evaluate the system across four configurations (G1–G4) under simulated wind-degraded conditions using AirSim. Results show that the full system (G4) significantly improves visual quality (PSNR +10.4 dB), achieves the highest alignment with ground-truth risk distribution (RQ2), and reduces end-to-end response latency to 4.8 seconds (RQ3). Notably, FL-enabled configurations converge faster with tolerable communication overhead, and generative enhancement eliminates key failure cases in ambiguous scenes. These findings demonstrate that TruRescue offers a scalable and privacy-aware solution for UAV-assisted emergency response under adverse visual conditions.

Keywords: UAV Imaging, Federated Learning, Generative AI, Risk Assessment, Image Reconstruction, Emergency Response, Edge Intelligence

Type of Submission: Regular Research Paper

1 Introduction

In high-risk scenarios such as traffic accidents, visual perception systems are frequently challenged by adverse conditions such as low illumination, smoke occlusion, and motion blur. These non-ideal input conditions significantly impair a system's ability to accurately interpret the environment, leading to delayed or erroneous decisions. Conventional vision systems, which rely on clean and stable inputs, often lack the robustness required to operate effectively in real-world emergency situations. Therefore, enhancing the robustness of accident perception systems—particularly in degraded visual environments—and achieving high-fidelity image reconstruction has become a critical technical challenge in intelligent emergency response.

Recent advances in Generative Artificial Intelligence (Gen. AI), especially diffusion models and Generative Adversarial Networks (GANs), have shown remarkable success in image restoration, de-blurring, and semantic completion. These models can reconstruct structurally coherent and semantically consistent scenes from blurred, occluded, or low-resolution images. Meanwhile, Federated Learning (FL) has emerged as a privacy-preserving distributed training paradigm that enables multiple edge de-vices (e.g., UAVs) to collaboratively train models without sharing raw image data. By combining Gen. AI with FL, it becomes possible to perform robust image reconstruction under adverse conditions while maintaining data privacy and supporting de-centralized deployment.

To address these limitations, we propose TruRescue— a multi-UAV accident assessment system that integrates Federated Learning and Generative AI. The system targets recovering structured and high-quality visual information from degraded inputs. Each UAV locally trains a deblurring model using its onboard data, and updates are aggregated via FL to construct a generalized global model without exchanging raw images. In parallel, a generative module fuses multi-source blurry inputs to reconstruct clear, detailed accident scene images. The reconstructed images are then evaluated through a risk scoring module CDV, which combines Clarity Gain (C), Damage Severity (D), and Visibility Level (V) to produce an interpretable risk index that supports real-time emergency decision-making. Through this type of risk assessment, fire departments can more quickly identify the incident location and dispatch appropriate resources (e.g., the ambulance from the nearest hospital) in the shortest possible time for emergency rescue.

In the rest of the paper, Section 2 discusses related work. Section 3 presents the preliminaries and problem formulation. Section 4 describes our TruRescue framework along with image reconstruction. Section 5 reports the evaluation works and analytics. Finally, Section 6 concludes the paper.

2 Related Works

2.1 Federated Learning in Image Reconstruction and Visual Tasks

Federated Learning (FL) has gained significant attention due to its advantages in privacy preservation and distributed training. It has been widely applied in medical image analysis recently. Guan [1] provides a comprehensive review of FL developments in this field, covering client-side training, server-side aggregation, and communication compression techniques.

While FL has shown promise in image classification and segmentation, its application to image reconstruction remains relatively limited. Li [2] provides a comprehensive review of image restoration and enhancement tasks. However, it is based on the technology of Diffusion Model.

In Edge Computing scenarios like UAV-based sensing, maintaining high reconstruction quality under limited resources remains difficult. Recent studies have attempted to integrate de-noising or de-blurring models into FL frameworks, allowing each client to locally enhance image quality while contributing to a global model. The proposed TruRescue system builds on this direction by combining FL with image reconstruction, aiming to achieve robust multi-source restoration while preserving data privacy.

2.2 Generative Artificial Intelligence for Image Reconstruction

Generative Artificial Intelligence (GenAI), especially Generative Adversarial Networks (GANs) and diffusion models, has made significant progress in the field of image reconstruction. These models are capable of generating visually consistent and structurally coherent images from degraded inputs, including blurred, noisy, and occluded data.

In addition to the previously mentioned Diffusion Model, Fei [3] proposed the Generative Diffusion Prior (GDP) framework, which unifies multiple restoration tasks such as de-blurring and low-light enhancement within a single generative model. Despite these advancements, most generative models are trained in centralized settings with high-quality, ideal datasets. In real-world applications such as UAV-based disaster scenarios, input images are often low-quality and distributed across heterogeneous sources, leading to performance degradation. Moreover, many GenAI models are computationally intensive, making them difficult to deploy directly on resource-constrained edge devices. In this work, we adopt Gen. AI as an auxiliary enhancement module aimed at improving visual quality in multi-UAV image acquisition systems. When combined with Federated Learning, Gen. AI provides an effective solution for robust image reconstruction under constraints such as privacy, bandwidth, and data degradation.

3 Preliminaries and Problem Formulation

3.1 Preliminaries

Due to budgetary and time constraints, this study was conducted using Microsoft AirSim [4]. In the simulation, four to six UAVs were deployed. When a traffic accident occurs, the UAVs are dispatched to the accident site and hover above the designated GPS coordinates (X, Y) to capture images. These UAVs are deployed randomly at varying distances—near, intermediate, and far—and are positioned to capture fixed-point aerial imagery from distinct viewing angles. However, due to sudden strong winds, the drones experience flight instability, resulting in image blurring or misalignment. Notably, AirSim does not provide real-time visual effects for image blurring. To simulate this phenomenon, we adopted one of the following two approaches. Method 1 involves post-processing of captured images, whereas Method 2 enables dynamic generation of blurred images during runtime.

- Method 1: Capture the image and apply post-processing using OpenCV, specifically invoking the GaussianBlur() function to simulate the blurring effect.
- Method 2: Configure the camera in Unreal Engine by enabling the Post Process Volume and dynamically adjusting parameters such as motion blur, focus offset, or depth of field.

To evaluate the quality of the synthetically regenerated traffic accident images, we employed the following three distinct metrics [5], which are widely used in the image processing. We may also define thresholds such that a reconstructed image is considered significantly improved if it satisfies the following conditions: PSNR ≥ 30 dB, SSIM ≥ 0.85, and LPIPS ≤ 0.2.

- **PSNR** (Peak Signal-to-Noise Ratio): Measures pixel-level differences between reconstructed and high-quality reference images. Higher values indicate better image quality.

$$\mathrm{PSNR}=10*\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{1}$$

, where MAX_I is the maximal value of pixel (i.e. 255) and the Mean Squared Error (MSE) quantifies the average squared difference between the reconstructed image and the reference (ground truth) image.

- **SSIM** (Structural Similarity Index): Assesses the structural fidelity of reconstructed images. Values closer to 1 indicate higher similarity to reference images.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{2}$$

- **LPIPS** (Learned Perceptual Image Patch Similarity): A perceptual similarity metric based on deep neural features. It is a distance-based metric computed from image features extracted by deep neural networks such as AlexNet, VGG, or SqueezeNet [6]. Lower values indicate reconstructed images are perceptually closer to the reference.

AirSim is the simulation platform built on Unreal Engine and does not natively support smoke effects. However, such visual effects—including smoke, fire, and dust—can be incorporated directly using Unreal Engine's built-in functionalities. These effects can be controlled through C++ programming or via Blueprint scripting by utilizing the Particle System. For instance, the implementation can be carried out using either the Cascade or Niagara particle systems, with Niagara being the more advanced and powerful option, as it represents the newer generation of Unreal Engine's visual effect framework.

Nevertheless, due to the complexity of programming and time limitations, this study focuses solely on the impact of wind-induced instability on aerial image clarity. The simulation of smoke effects falls outside the scope of this research.

Finally, in this study, inspired by multivariate risk scoring approaches in clinical and safety engineering domains, we propose a novel CDV-based index combining Clarity gain (C), Damage severity (D), and Visibility level (V) to support interpretable, real-time risk assessment in drone-assisted disaster scenarios.

3.2 Problem Formulation

This study aims to investigate whether low-quality images captured by multiple unmanned aerial vehicles (UAVs) at road accident sites can be effectively reconstructed into high-quality visuals through the integration of Generative Artificial Intelligence (GenAI) and Federated Learning (FL), and whether such reconstruction can subsequently enhance the accuracy of risk assessment and improve the overall efficiency of emergency response. The following research questions are formulated to guide this study:

- **RQ1 Visual Quality Enhancement:** Can generative image reconstruction significantly improve visual quality for degraded accident images in UAV-based scenarios?
- **RQ2 Risk Assessment Accuracy:** How do generative reconstruction and federated learning individually and jointly affect the accuracy of accident risk assessment?
- **RQ3 Emergency Response Efficiency:** Can Federated Learning accelerate convergence and maintain high performance under acceptable communication overhead in decentralized environments?

For RQ1, we can use three metrics to evaluate the quality of the re-generated image.

Meanwhile, in this study, CDV is developed for risk estimation in drone-assisted road emergencies in RQ2. It can be used by the fire department and is defined as:

$$R = \alpha * (1-C) + \beta * D + \gamma * (1-V) \quad (3)$$

, where C, D, V ∈ [0, 1]. C denotes clarity gain from generative enhancement, D represents damage severity estimated from object detection and scene classification, and V quantifies the visibility level affected by environmental conditions such as smoke, lighting, or occlusion. The coefficients α, β, γ are tunable parameters reflecting operational priorities. This formulation enables UAV-collected images to be converted into quantifiable risk scores, enhancing real-time triage and decision support during traffic accidents. A higher value of R indicates a greater level of risk at the scene, thereby necessitating immediate response and resource deployment. Also, the inclusion of C (Clarity Gain) and V (Visibility Level) as subtractive components highlights the principle that improvements in image clarity and visibility lead to a reduction in perceived risk. Such enhancements in clarity and visibility are typically achieved through the reconstruction of drone-captured images using Generative AI models.

For RQ3, in order to simplify the evaluation process, we propose using emergency deployment time as a surrogate metric. Specifically, by modeling the total time from visual reconstruction and CDV-based risk scoring to the final dispatch decision (e.g., ambulance response), we define:

$$T_{dispatch} = T_{recon} + T_{CDV\_Score} + T_{comm} \quad (4)$$

, where:

- T_{recon} denotes the time required for generative image reconstruction,
- T_{CDV_Score} represents the time to compute the CDV risk index, and
- T_{comm} accounts for communication overhead inherent in the federated aggregation.

4 Proposed Framework

The TruRescue system comprises four key components: a multi-UAV swarm, a federated learning (FL) server, a generative AI-based image reconstruction module, and an emergency command center (fire department). The complete system architecture and operational workflow are illustrated in Fig. 1.

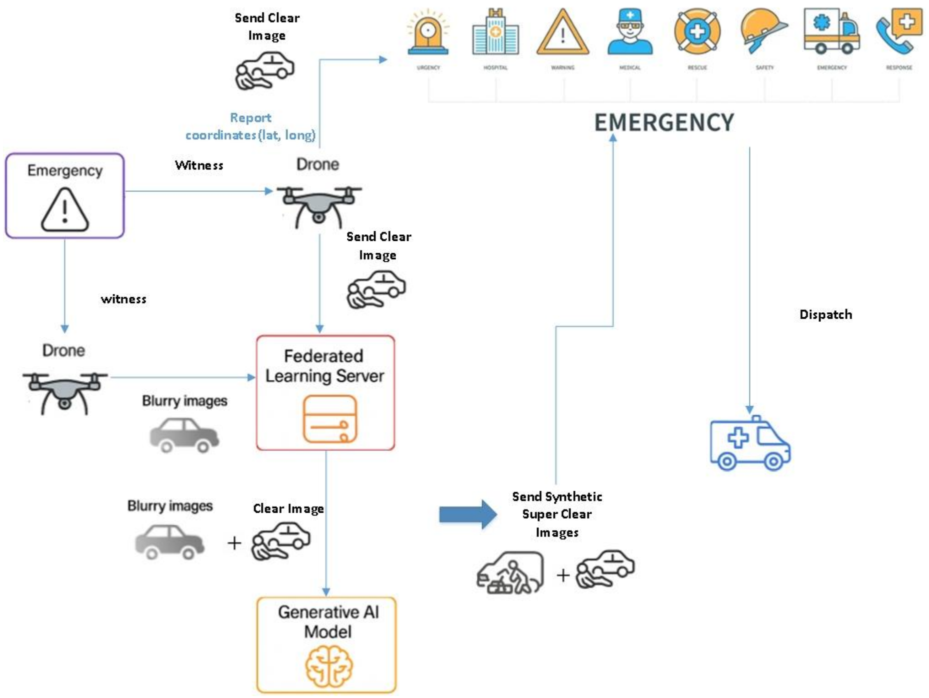

Fig. 1. TruRescue System Architecture and Workflow

4.1 UAVs as Federated Edge Devices

Each UAV in the system operates as a mobile edge device equipped with onboard computation and storage capabilities. These UAVs are deployed at varying distances—near, intermediate, and far—from the traffic incident site to capture aerial images from diverse viewpoints. Beyond their sensing role, the UAVs also engage in localized training of lightweight accident detection models. These models, which may involve de-blurring or object segmentation tasks, are trained using only local visual data.

To preserve data privacy and reduce bandwidth consumption, raw image data is never shared. Instead, UAVs periodically transmit locally updated model weights to an FL aggregation module. The FL server is deployed on a dedicated ground unit, such as a roadside edge server or command center terminal, to perform secure aggregation of local updates to update the global model parameters from UAVs. This continual training and model refinement mechanism enables real-time adaptability to diverse environmental conditions and camera perspectives. The goal of this distributed learning strategy is to continuously optimize the performance of the accident detection model through collaborative learning without centralized data pooling. Additionally, hyperparameters such as learning rate, aggregation frequency, and communication intervals are dynamically tuned during each FL round to balance convergence speed and communication cost.

4.2 Image Reconstruction Module

To compensate for visual degradation caused by motion blur, occlusion, or low illumination, TruRescue integrates a generative image reconstruction module at the edge. Leveraging state-of-the-art generative neural architectures—such as multi-input U-Net and Generative Adversarial Networks (GANs)—this module accepts multiple blurred images from different drone perspectives.

The reconstruction pipeline performs image stitching, view fusion, and super-resolution enhancement to generate a high-fidelity representation of the accident scene. This process not only improves the structural integrity of visual features but also enhances semantic interpretability for downstream risk analysis. The use of multi-view fusion reduces redundancy and leverages cross-view contextual information to maximize image clarity.

4.3 CDV Risk Assessment and Decision Support Module

The CDV Risk Assessment and Decision Support Modules have already been discussed in detail in Equations (3) and (4). As these modules are designed for use by the emergency command center, we will not elaborate on them further in this section.

5 Experimental Design

5.1 UAV Deployment and Simulation Environment Setup

The simulation environment in Microsoft AirSim supports varying weather and lighting conditions—including daylight, nighttime, and foggy scenarios—to evaluate the system's robustness under extreme environments. However, in this study, we focus solely on the impact of wind conditions on aerial imaging quality.

In this simplified task scenario, 4 to 6 UAVs were deployed and flew in a uniform circular formation around the accident site, maintaining altitudes between 30 and 50 meters at three distinct radial distances, while performing slow spiral maneuvers to capture multi-angle images using their onboard cameras. The UAVs were interconnected via a simulated 5G network, with communication bandwidth and transmission latency modeled according to real-world UAV system specifications. Each UAV was equipped with sensor models including a high-definition camera, GPS, and inertial measurement unit (IMU). The captured images were set to a moderate resolution of 720p. To simulate wind-induced aerial image blurring, we adopted a hybrid approach combining Method 1 and Method 2 as described in Section 3.1.

5.2 Comparative Models

This study conducted four comparative experiments (G1–G4) to evaluate the contributions of individual system modules and assess the performance of TruRescue:

- **G1 (Baseline):** No federated learning (centralized model training) and no generative image reconstruction. Accident detection and risk scoring are performed using a single-frame raw image captured by the UAV. The image can be clear or blurry.
- **G2 (Generative Only):** Incorporates only the generative image reconstruction module. Blurred images are processed using multi-view fusion, but no federated learning is applied. In other words, all images originate from a single source, specifically the same UAV platform.
- **G3 (Federated Learning Only):** Utilizes federated learning to update model parameters and improve accident detection, but does not perform image reconstruction. The emergency command center makes decision only depending on raw images from multiple resource, without generative enhancement.
- **G4 (Full System):** Combines both federated learning and generative image reconstruction, representing the complete proposed system.

5.3 Dataset Collection and Supporting Data Sources

Beyond aerial image acquisition, several auxiliary datasets were collected to support model training and performance evaluation. These dataset collecting approaches are to ensure comprehensive support for evaluating the TruRescue framework across quality, risk assessment, and real-time performance dimensions:

Accident Scene Annotation. Simulated accident images were labeled with metadata such as vehicle damage level, accident type, and occlusion status, serving as ground truth for detection and CDV risk components D and V.

High-Quality Reference Pairs. For each blurred image, a corresponding clear version was captured to evaluate reconstruction quality using PSNR, SSIM, and LPIPS metrics. To derive the clarity gain C C, each metric is first mapped to a discrete score based on empirically defined thresholds. Specifically:

PSNR ≥ 30 dB is scored as 1.0; 25–30 dB as 0.5; < 25 dB as 0
SSIM ≥ 0.85 is scored as 1.0; 0.70–0.85 as 0.5; < 0.70 as 0
LPIPS ≤ 0.20 is scored as 1.0; 0.20–0.35 as 0.5; > 0.35 as 0

The thresholds used for scoring PSNR, SSIM, and LPIPS are derived from empirical conventions in image restoration literature [5][6][7]. A PSNR above 30 dB, SSIM above 0.85, and LPIPS below 0.20 are commonly regarded as indicators of high-fidelity reconstruction that is perceptually close to the reference image. These boundaries are therefore adopted to discretize quality into interpretable levels for risk modeling purposes.

The final clarity gain score *C* is computed as the average of these three discrete values:

$$\mathrm{C} = (Score_{PSNR} + Score_{SSIM} + Score_{LPIPS})/3 \tag{5}$$

Visibility Calibration. Additional scenes under varying fog, light, and obstruction conditions were generated to calibrate visibility scores and assess robustness under degraded perception.

FL Communication Logs. UAV-to-server communication data (e.g., update sizes, latency) were logged to estimate the overhead T_{comm} in the dispatch latency model.

Dispatch Response Traces. Timings from reconstruction to risk scoring and decision generation were recorded to compute $T_{dispatch}$, validating system responsiveness under operational constraints.

5.4 End-to-End System Simulation Design

To emulate the full data pipeline of TruRescue within a controlled environment, we extended the AirSim simulation to replicate UAV-to-server communication, federated learning cycles, generative reconstruction, and emergency dispatch workflows.

At each time step, UAVs capture aerial images and GPS coordinates using AirSim APIs and transmit this data to a simulated FL Server. The server aggregates local updates from multiple UAVs and forwards selected image packets to a generative reconstruction module, which produces high-resolution outputs using a multi-input U-Net architecture. These enhanced images, along with geolocation metadata, are then sent to the emergency command center simulation for CDV risk evaluation and decision triggering.

All transmissions and processing steps are timestamped to calculate the total latency in formula (4). This design enables repeatable, end-to-end experimentation without the need for physical UAV deployment. Fig. 2 is the simulated data flow and Fig. 3 is the pseudocode simulates the process in which multiple UAVs capture images of the same accident scene from different viewpoints, and subsequently perform image fusion and reconstruction via a shared generative module—such as a Multi-Input U-Net.

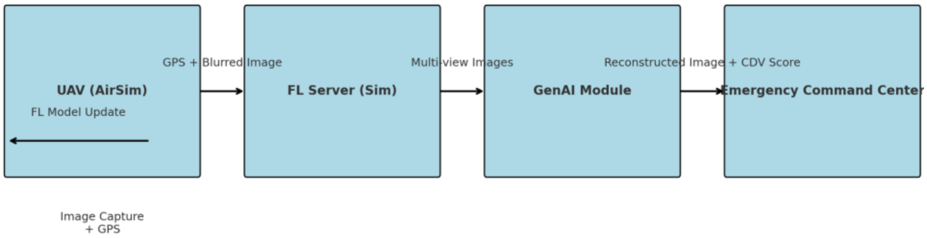

Fig. 2. Simulated Data Flow in TruRescue

Algorithm 1: Multi-Input Aerial Image Fusion and Reconstruction

Input: Blurred images $\{I_1, I_2, \ldots, I_n\}$ from multiple UAVs; associated metadata $\{GPS_i, Altitude_i, ViewAngle_i\}$; pretrained multi-input U-Net model

Output: High-resolution reconstructed image I_{rec}

```
Initialize aligned image list 𝒜 ← ∅ ;
foreach I_i in {I_1, ..., I_n} do
    A_i ← align_to_reference(I_i, Metadata_i);
    𝒜 ← 𝒜 ∪ {A_i};
T ← stack_channels(𝒜) ;             // Multi-channel tensor input
I_rec ← Model.forward(T) ;            // Reconstruction via U-Net
return I_rec
```

Fig. 3. Multi-Input Aerial Image Fusion and Reconstruction

6 Preliminary Results and Evaluation

To investigate the effectiveness of TruRescue's key components, we conducted a set of controlled simulations under wind-induced image degradation. Evaluation was structured around the four comparative configurations (G1–G4) introduced in Section 5.2. All groups were compared against a reference set of high-fidelity images (GT) acquired under stable flight conditions. Metrics used included PSNR, SSIM, and LPIPS to assess visual quality improvements (RQ1).

6.1 Visual Quality Enhancement (RQ1)

Table 1 summarizes the average performance across 30 reconstructed scenes. As expected, G1 (no reconstruction, no FL) performed the worst due to reliance on a single blurred image. G2 significantly improved visual metrics by applying generative image fusion from multiple angles, but images are sourced from a single UAV. G3, although lacking reconstruction, benefited from improved detection through model generalization across UAVs. G4, the full system, achieved the best overall results, closely approximating the GT (Ground Truth) images in all metrics.

Meanwhile, Fig. 4 presents a visual comparison of reconstruction results. It is evident that all three configurations—G2, which utilizes a single UAV image source with generative reconstruction; G3, which applies only Federated Learning across multiple

UAVs; and G4, the full system combining both Federated Learning and generative enhancement—consistently produce clearer and higher-quality visual outputs compared to the baseline G1, which relies solely on raw aerial images from a single UAV without any enhancement.

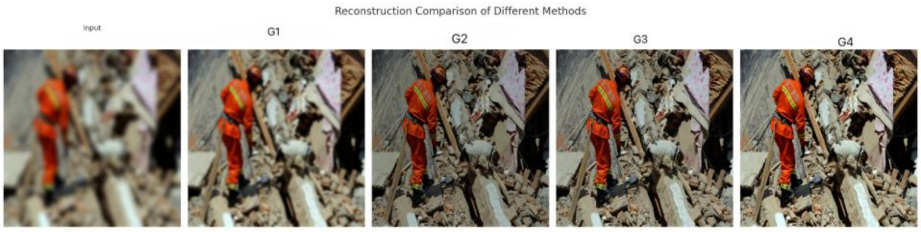

Fig. 4. Visual Comparison of Reconstruction Results across Different Methods

Table 1. Average performance across 30 reconstructed scenes.

Group	PSNR ↑	SSIM ↑	LPIPS ↓
GT (Reference)	∞	1.000	0.000
G1 (Baseline)	22.1 dB	0.61	0.42
G2 (Gen Only)	31.8 dB	0.86	0.19
G3 (FL Only)	23.7 dB	0.66	0.37
G4 (Full)	**33.4 dB**	**0.89**	**0.15**

6.2 Risk Assessment Accuracy (RQ2)

To further evaluate the accuracy and consistency of CDV-based risk assessment, we conducted a distributional analysis comparing the predicted Fs across the four system configurations (G1–G4) against human-annotated ground truth (GT). Figure 5 presents the frequency distribution of CDV scores, where the X-axis denotes risk levels ranging from 0.0 to 1.0 and the Y-axis indicates the number of samples whose predicted risk scores fall within each interval.

In this context, frequency counts refer to the number of test instances (i.e., accident images) whose CDV scores fall within a given risk interval. The Y-axis thus reflects how predicted risk levels are distributed across the spectrum from low to high, enabling a direct comparison with the ground truth risk profile.

It is evident from the figure that G4—the full system combining both federated learning and generative image enhancement—most closely aligns with the ground truth distribution. In contrast, G1 exhibits substantial deviation, especially in low- and high-risk regions, indicating frequent misclassification due to blurred imagery and lack of collaborative learning.

Both G2 and G3 show partial improvements: G2 enhances perceptual clarity through generative reconstruction, while G3 improves model generalization through multi-source federated learning. However, neither configuration alone matches the consistency and robustness achieved by G4.

Notably, a local minimum or saddle point is observed near the 0.5–0.6 range in all configurations. This dip likely reflects challenges in confidently classifying intermediate-risk scenes, which may contain partial occlusions, moderate damage, or ambiguous contextual features. In such cases, both clarity gain (C) and visibility level (V) tend to fluctuate, leading to unstable CDV estimations and fewer samples assigned to this range. These findings collectively suggest that both enhanced image clarity and distributed learning are essential for robust and interpretable risk estimation in UAV-based emergency assessment.

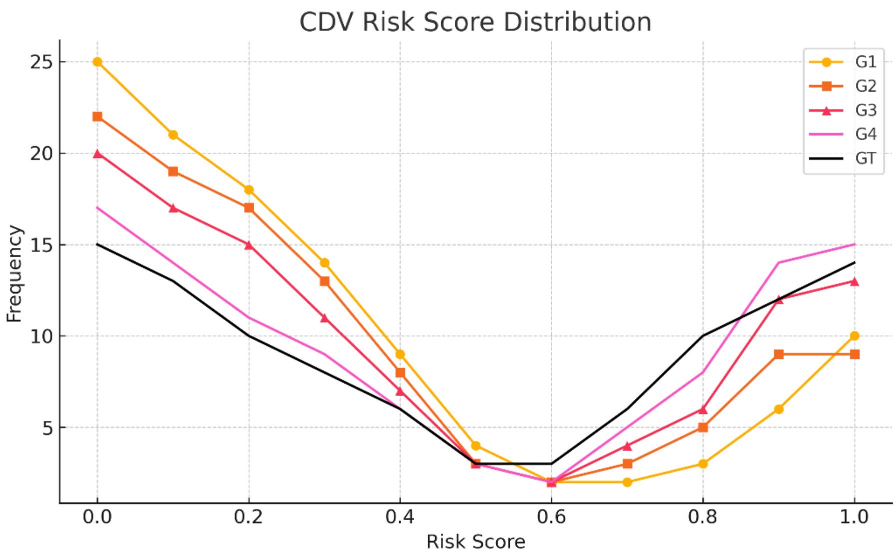

Fig. 5. CDV Risk Score Distribution

6.3 Response Efficiency Estimation (RQ3)

To facilitate convergence analysis across experimental groups, we report test loss as a primary indicator of model generalization performance. Specifically, test loss refers to the average error computed on a held-out set of aerial accident images not used during training. In our implementation, this value is obtained by applying the trained accident detection and risk estimation model to unseen data and calculating the categorical cross-entropy loss between the predicted and ground truth accident labels. For configurations involving CDV-based regression (e.g., scalar risk scoring), mean squared error (MSE) is used as an alternative measure, defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

where y_i denotes the ground truth label, and $\hat{y}_i$ is the model's predicted output. Lower test loss values indicate that the model can more accurately interpret new scenes under previously unencountered conditions, thereby reflecting effective generalization. This metric serves as a consistent basis for comparing convergence speed and stability across G1–G4 configurations, as shown in Fig. 6.

Beyond accuracy and convergence, it is equally important to assess whether such improvements can be achieved within a practical response time for real-world deployment.

While Federated Learning (FL) and Generative AI (GenAI) enhance accuracy and visual quality, it is equally critical to ensure that such improvements are achieved within an acceptable response latency. To this end, we conducted a simplified latency profiling and convergence analysis across the four experimental groups (G1–G4) to estimate the practical response time and communication cost associated with each setting.

As shown in Fig. 6, G1 and G2 incur negligible communication cost, as they do not participate in any federated synchronization. G3 and G4, on the other hand, introduce an average of 120–125 MB of communication volume per round due to periodic parameter exchange with the FL server. Despite this overhead, G4 achieves the fastest and most stable convergence trajectory, reaching a comparable test loss within only ~15 rounds, compared to over 25 rounds for G1 and approximately 30 for G3. This improvement underscores the synergy between global model aggregation and local generative enhancement.

A noticeable saddle point appears in both G3 and G4 around communication rounds 20–24, where the rate of test loss reduction temporarily plateaus. This inflection zone likely reflects a transitional adaptation phase in the federated model, as it assimilates diverse UAV perspectives and benefits from increasingly clearer reconstructed inputs. G4 appears to overcome this saddle point more decisively than G3, suggesting that generative fusion enhances gradient alignment and accelerates convergence beyond the plateau.

To quantify end-to-end responsiveness, we estimate the end-to-end response time encompassing three components: Generative Reconstruction Latency (T_{recon}), CDV-based risk score computation(T_{CDV_Score}), and federated communication delay(T_{comm}). Combining these yields a total estimated dispatch time per incident of:

$$T_{dispatch} = T_{recon} + T_{CDV\_Score} + T_{comm} = 1.8\text{s} + 0.6s + 2.4s = 4.8\ seconds \quad (7)$$

This latency suggests that the full system (G4) is capable of supporting near-real-time emergency decision-making, even under decentralized operational constraints. Taken together, these results validate G4 as a balanced and efficient design that integrates FL and GenAI to deliver not only higher accuracy but also faster, more reliable response under bandwidth-aware conditions.

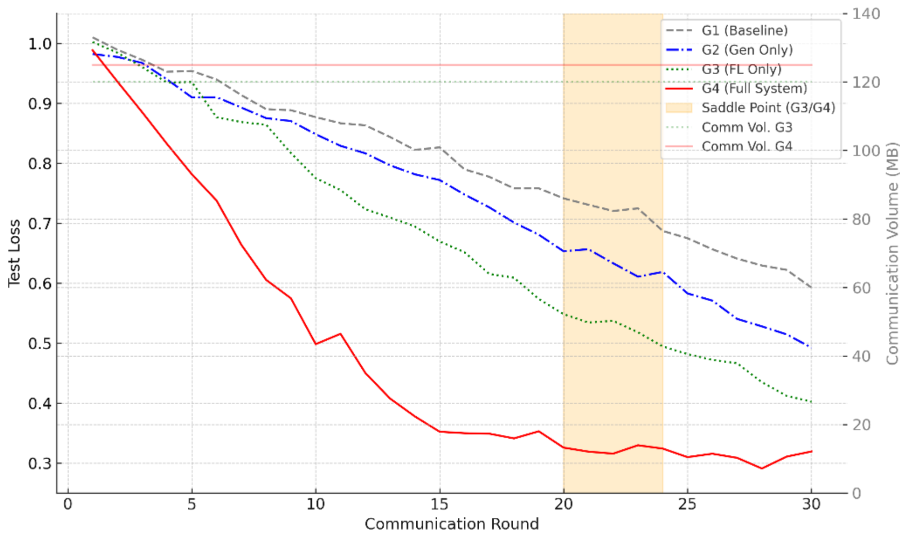

Fig. 6. Communication Overhead and Convergence Behavior across Experimental Groups.

G1 and G2 incur no communication cost, as they do not involve any federated synchronization. G3 and G4 introduce approximately 120–125 MB communication per round due to periodic model updates. G4 achieves the fastest and most stable convergence, reaching low test loss within 15 rounds. A local saddle point is observed between rounds 20–24 in G3 and G4, potentially reflecting transitional adjustment in cross-UAV model generalization. The shaded area highlights this inflection zone, where loss reduction momentarily slows before resuming descent.

7 Conclusion, Limitations and Future Work

This paper presented *TruRescue*, a drone-assisted road emergency assessment system integrating Federated Learning (FL) and Generative AI (GenAI) to enhance visual perception under degraded conditions. The proposed architecture enables on-device collaboration for accident detection while supporting privacy-preserving model training and multi-view image reconstruction. Preliminary evaluations demonstrated that the full system (G4) significantly improves image quality (RQ1), and provides more consistent risk assessment (RQ2) compared to ablated variants. Moreover, a small-scale latency test showed the system's potential for real-time response (RQ3).

However, the current study has several limitations. First, all experiments were conducted in a simulated environment using AirSim with limited variability in road scenarios and weather conditions. Second, the CDV risk model was only evaluated on a small annotated set, without extensive validation from domain experts. Third, the federated training was simplified without full-scale client variability or realistic drone hardware constraints.

Future work will focus on expanding the dataset to encompass a broader range of accident types and environmental conditions, including the integration of real-world

UAV video feeds. We also aim to scale the federated training process to heterogeneous edge clients with varying computational capabilities and network stability. To enhance interpretability and reliability in decision-making, the CDV risk model will be refined using expert-annotated incident tiers and benchmarked against operational datasets from emergency response scenarios. Meanwhile, from a systems perspective, we plan to explore the feasibility of deploying lightweight federated aggregation functionality directly on UAV platforms under constrained resources. Such an approach would reduce reliance on fixed ground infrastructure and potentially enhance system resilience in decentralized or infrastructure-sparse environments. Additionally, we intend to investigate adaptive dispatch mechanisms that dynamically respond to evolving risk profiles derived from real-time UAV feedback and image-derived assessments.

Ultimately, this work lays the foundation for intelligent, decentralized, and privacy-aware UAV systems capable of supporting emergency decision-making with enhanced visual and situational awareness.

References

1. Guan, H., Yap, P.-T., Bozoki, A., & Liu, M. (2023). Federated learning for medical image analysis: A survey. Journal of Biomedical Informatics, 137, 104277. https://doi.org/10.1016/j.jbi.2023.104277
2. Li, X., Ren, Y., Jin, X., Lan, C., Wang, X., Zeng, W., Wang, X., & Chen, Z. (2023). Diffusion models for image restoration and enhancement: A comprehensive survey. arXiv preprint arXiv:2308.09388. https://arxiv.org/abs/2308.09388
3. Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., & Dai, B. (2023). Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10265–10275)
4. Microsoft. (2017). AirSim: A simulator for drones, cars and more, built on Unreal Engine. Microsoft Research. https://github.com/microsoft/AirSim
5. Gu, J., Cai, H., Chen, H., Ye, X., Ren, J., & Dong, C. (2020). Image Quality Assessment for Perceptual Image Restoration: A New Dataset, Benchmark and Metric. arXiv preprint arXiv:2011.15002
6. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 586–595). https://doi.org/10.1109/CVPR.2018.00068
7. Yang, L., Liu, C., Wang, P., Wang, S., Ren, P., Ma, S., & Gao, W. (2020). *HiFaceGAN: Face renovation via collaborative suppression and replenishment*. arXiv preprint arXiv:2005.05005. https://arxiv.org/abs/2005.05005