ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 004.85:519.172:519.816 **doi: 10.32620/reks.2025.3.12**

Viacheslav MOSKALENKO, Alona MOSKALENKO, Yuriy MOSKALENKO

Sumy State University, Ukraine

A METHOD FOR IMPROVING THE ROBUSTNESS OF NEURAL NETWORK MODEL FOR AERIAL IMAGE MATCHING

Neural network–based image matching techniques are increasingly employed in aerial image analysis—particularly for UAV navigation, localization, and mapping. However, their sensitivity to structured visual distortions (e.g., shadows, illumination changes, and terrain variability) limits robustness under real-world conditions. Addressing this challenge, we propose a training methodology that enhances both robustness and cross-domain generalization of feature-matching models by integrating adversarial procedural noise with activation-function modification. During training, structured noise patterns (Perlin, Gabor, and Worley) are synthesized and applied in an adversarial manner, while the standard ReLU activation is replaced by a hybrid LeakyReLU6 to mitigate sensitivity to local perturbations. We evaluate our approach on both detector-based (SuperPoint + SuperGlue) and detector-free (LoFTR) architectures using the Aerial Image Matching Benchmark Dataset and further assess cross-domain performance on the HPatches dataset. Experimental results show that our method yields over 4 % absolute improvements in matching precision and recall on noisy test data for both classes of models. Ablation studies confirm that these gains are attributable to the synergistic effect of procedural noise and LeakyReLU6. Moreover, models trained with our procedure exhibit significantly smaller performance drops when transferred to HPatches, demonstrating enhanced generalization relative to conventionally trained counterparts. To our knowledge, this is the first work to combine adversarial procedural-noise training with activation-function constraints for aerial image matching. Beyond improved noise resistance, our method advances cross-domain applicability and is readily extendable to diverse neural-network architectures.

Keywords: *image matching; robustness; adversarial training; procedural noise; aerial image.*

1. Introduction

1.1 Motivation of research

In recent years, aerial image matching has emerged as a crucial task in various computer vision applications, including drone navigation, autonomous inspection, cartography, and geospatial analysis. For a long time, traditional feature detection and matching algorithms such as SIFT, SURF, ORB, and FAST were widely used in these domains [1]. In classical approaches, feature detection, description, and matching are treated as independent processes and are not jointly optimized. As a result, they often fail to adequately account for the contextual information of the scene and individual keypoints, which limits their accuracy and robustness compared to neural network-based methods.

Among neural network-based approaches for keypoint detection and description, models such as SuperPoint, R2D2, and XFeat have gained widespread adoption due to their computational efficiency and high reliability [2]. For matching feature descriptors, SuperGlue and LightGlue are commonly employed. Furthermore, detector-free dense feature matching methods, such as LoFTR, COTR, and DGC-Net, have also demonstrated promising results [3].

However, the performance of these models tends to degrade significantly in real-world conditions, where surface texture changes caused by uneven lighting, diurnal or seasonal illumination variations, partial occlusions, and digital artifacts are common [4]. Natural variations in lighting and texture, sensor-induced distortions, and scene clutter can be interpreted as a form of adversarial noise. This presents considerable challenges for achieving robust image matching, particularly in aerial imagery, where variable viewpoints, terrain structures, and sensor-specific noise are prevalent. Structured noise models such as Gabor, Perlin, or Worley noise can be leveraged during training to simulate such adversarial conditions and improve the resilience of image matching models.

Despite recent advancements, there remains a research gap in developing methods that explicitly enhance the robustness of neural network-based image matching models under structured adversarial noise. Addressing this challenge is a promising area of research, particularly for aerial applications, where consistent and reliable feature correspondence is essential for navigation, mapping, and analysis.

1.2. Objectives and Contributions

The main objective of this research is to improve the robustness of neural network models for aerial image matching by developing a novel training approach that enhances performance under structured visual distortions, measured by Matching Accuracy (MA) and Matching Recall (MR). Robustness is defined as the ability to maintain high MA and MR on noisy test data, with improved accuracy as a secondary outcome**.** The proposed method aims to increase the resilience of learned feature descriptors and matching mechanisms to real-world variations in scene appearance, sensor artifacts, and environmental noise through adversarial procedural noise and activation-function modification.

The key issues are as follows:

– analysis of existing methods for enhancing neural network robustness, with a particular focus on their application to image matching tasks;

– development of a training approach that improves robustness by incorporating structured synthetic noise such as Gabor, Perlin, and Worley noise during the learning process;

– comparative evaluation of models trained with the proposed method versus conventional training procedures, using MA and MR to quantify robustness and accuracy improvements.

Structurally, the paper is organized as follows. Section 2 provides an overview and analysis of related work in the field of robust image matching. Section 3 presents the proposed training method using structured noise to simulate adversarial conditions. Section 4 details the experimental setup, dataset description, and evaluation results, including comparisons with baseline models and robustness testing under various distortions. Section 5 discusses the method's particularities, benefits, and limitations. Finally, Section 6 concludes the paper and outlines potential directions for future research.

2. The State-of-the-Art

2.1. Applications of Image Matching in Unmanned Aerial Systems

Image matching is a foundational component in Unmanned Aerial Systems (UAS), enabling crucial functionalities such as autonomous navigation, visual localization, mapping, and environmental monitoring [5]. By aligning current UAV-captured imagery with reference datasets—such as orthophotos, satellite images, or previous flight captures—UAS platforms can perform robust geolocation and track environmental changes over time. However, the reliability of these operations heavily depends on the robustness of the underlying image matching algorithms, particularly under challenging visual conditions.

In the absence of GNSS signals, such as in urban canyons, forests, or indoor environments, UAVs rely on image matching for visual localization. Although numerous methods have been proposed for this task, including matching based on handcrafted features and deep-learning-based matching, many exhibit limited robustness when exposed to real-world visual perturbations—such as lighting variations, seasonal changes, partial occlusions, or motion blur [6].

For instance, Kim et al. (2024) proposed a method for UAV localization by matching onboard images to aerial maps, achieving good performance in ideal conditions but facing reduced accuracy under shadowed and low-contrast scenes [7]. Similarly, Liu et al. (2024) benchmarked multiple image matching models for UAV positioning and emphasized the sensitivity of feature matchers to viewpoint and illumination variations [8].

In environmental monitoring tasks, image matching is used to detect subtle changes in terrain or vegetation over time. However, these applications often suffer from inconsistent image quality due to varying weather, light, and sensor conditions. Śledziowski et al. (2022) utilized UAV imagery to monitor underwater coastal structures but noted that image matching reliability degraded significantly in the presence of surface reflections and turbidity [9].

Advanced techniques such as deep local descriptors and dense feature matchers have been introduced to address limitations of classical algorithms. For example, Zhang et al. (2021) introduced a learning-based method for matching infrared UAV imagery to satellite images, yet the system remained sensitive to structural distortions and non-rigid changes in appearance [10]. Koch et al. (2016) proposed a feature detector tailored to UAV imagery, but their evaluation highlighted instability under complex textures and occlusions [11].

In summary, while image matching has enabled numerous UAS capabilities, current methods often fall short in operational scenarios characterized by dynamic lighting, seasonal appearance changes, sensor-specific artifacts, and occlusions. These challenges underline the need for more robust and adaptive matching approaches capable of withstanding the structured visual distortions typical of real aerial missions.

2.2. Neural Networks for Image Matching

The task of image matching has undergone significant transformation with the advent of neural networks, which have enabled data-driven learning of keypoint representations, dense correspondence, and matching confidence. Traditional pipelines relied on separate stages of detection (e.g., Harris, FAST), description (e.g., SIFT, ORB), and matching (e.g., nearest neighbor search), often lacking robustness to changes in scale, illumination, and viewpoint. Modern learning-based methods offer

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

more integrated and robust alternatives, especially important for aerial imagery where appearance variations and geometric distortions are common.

A major direction in neural image matching is based on learned keypoint detectors and descriptors. SuperPoint [12] is a widely adopted model that jointly learns keypoint detection and description using self-supervised and supervised stages. R2D2 [13] improves on this by adding reliability and repeatability scoring, allowing the network to emphasize semantically stable and geometrically consistent features. XFeat [14] introduces a modular and lightweight architecture that separates detection, description, and matching, making it suitable for real-time and low-power applications such as UAV systems.

Another important advancement is the development of context-aware matching networks that refine correspondences between keypoints. SuperGlue [15] uses graph neural networks with self- and cross-attention to model interactions between keypoint sets, significantly improving matching precision under challenging conditions such as occlusion or repetitive textures. LightGlue [16] provides a more efficient alternative optimized for speed and memory, while preserving the core structure of attention-based reasoning.

In contrast to detector-based pipelines, detector-free methods such as LoFTR [17] propose a radically different approach by eliminating explicit keypoint detection altogether. LoFTR performs dense matching at the feature level by combining coarse-to-fine Transformer-based descriptors, producing high-quality correspondences across textureless or low-detail regions. This makes it particularly attractive for aerial imagery, where traditional keypoint detectors often fail due to repetitive patterns, vegetation, or low contrast.

Each direction offers complementary strengths. While sparse pipelines benefit from interpretability and efficiency, dense methods provide higher coverage and are more resilient to local feature loss. However, all of these models still face significant challenges under structured visual perturbations, such as shadows, seasonal variation, and natural noise — common in real-world aerial applications. Enhancing their robustness to such factors remains a critical research goal.

2.3. Approaches to Ensuring Robustness of Neural Networks for Image Analysis and Matching

Neural networks have significantly advanced the field of image analysis and matching, underpinning applications such as object recognition, scene understanding, and image alignment [1]. However, their deployment in real-world scenarios is often hindered by a pronounced sensitivity to complex visual conditions, including variations in lighting, occlusions, noise, and other perturbations.

One commonly explored direction is image preprocessing using denoising autoencoders [18]. While this approach can partially mitigate input noise, it introduces additional computational overhead, which is problematic in resource-constrained environments such as onboard UAV systems. Moreover, denoising autoencoders typically fail to handle perturbations that simultaneously affect multiple domains — spatial, temporal, and latent feature space — limiting their practical robustness.

Various gradient-masking techniques have also been proposed to counter adversarial perturbations. These include non-differentiable input transformations (e.g., JPEG compression, random padding and resizing), defensive distillation, dropout-based ensemble strategies, generative reconstruction, and discrete atomic compression methods [19]. While these methods can obscure gradient-based attacks to some extent, they lack comprehensive resilience to naturally occurring image variations and structural noise, which frequently affect aerial imagery.

Standard data augmentation — involving rotations, scaling, color jittering, and other visual transformations — has also been used to increase network generalization [20]. Although effective in controlled scenarios, such augmentations often fail to capture the full complexity of real-world disturbances. Models trained with these techniques still suffer significant performance degradation when exposed to naturally occurring occlusions, motion blur, or textured clutter, underscoring the gap in robustness.

Adversarial training offers another approach, where networks are trained on intentionally perturbed samples [21]. While this can enhance resistance to certain attack types, it often suffers from poor generalization to unseen perturbations. Moreover, adversarial training increases the overall computational load and can cause overfitting to the specific distortions present in the training set.

The effectiveness of adversarial training can be improved by tailoring the perturbation types used during training to more accurately reflect the kinds of noise and distortions encountered in operational settings. For aerial image matching tasks, structured noise generators — such as Gabor, Perlin, or Worley noise — have shown promise in this regard [22, 23]. These types of noise simulate real-world conditions: Gabor noise mimics repetitive structures and cast shadows; Perlin noise models soft gradients due to illumination or vegetation; and Worley noise imitates irregular patterns like road textures or surface degradation. Studies have demonstrated that using such structured perturbations during training can improve robustness in classification and detection tasks by teaching networks to ignore unstable or non-repeatable features [23]. However, their use in aerial image matching remains underexplored.

Architectural improvements also play a critical role in enhancing robustness. For example, activation functions with bounded output ranges such as ReLU6 have been empirically shown to improve resistance to both adversarial perturbations and structured noise [24]. Additionally, transformer-like architectures with self-attention mechanisms can better integrate contextual information around local features, allowing the model to remain robust when portions of the image are degraded or obscured by noise or shadows [2, 17].

In summary, multiple approaches have been proposed to improve the robustness of neural networks for image analysis and matching. However, no single method has proven universally effective. Each technique presents specific advantages and limitations, and the optimal strategy often depends on the specific nature of the application domain. Consequently, the integration and adaptation of multiple robustness-enhancing techniques warrant further investigation, particularly for real-world tasks such as aerial image matching under complex visual conditions.

3. A Method for Enhancing the Robustness of Aerial Image Matching Model

To improve model robustness under structured adversarial conditions, we propose replacing all ReLU-like activations in the network with a hybrid LeakyReLU6 function. This modification combines the advantages of two well-studied activation design choices in adversarial robustness research.

Leaky ReLU introduces a non-zero gradient in the negative domain, which improves gradient flow during adversarial training and reduces the likelihood of inactive ("dead") neurons — a known limitation of standard ReLU under gradient-based attacks. This smooth gradient profile supports better convergence and more stable loss landscapes during adversarial fine-tuning [25].

ReLU6, a bounded variant of ReLU, clips the activation output to a fixed upper limit (typically 6), which has been shown to reduce the impact of activation spikes caused by input perturbations during inference. By constraining the dynamic range of activations, ReLU6 helps limit the amplification of structured noise in deeper layers, contributing to improved generalization in visually degraded conditions [2].

The proposed LeakyReLU6 formulation aims to benefit from both effects — preserving smoothness for robust optimization, while maintaining a bounded response for improved stability under perceptual perturbations such as Perlin, Gabor, and Worley noise.

Figure 1 illustrates the stages of the proposed method, where the procedural noise can be generated in the form of Gabor, Perlin, Simplex, Voronoi, Worley noise, or other variants.

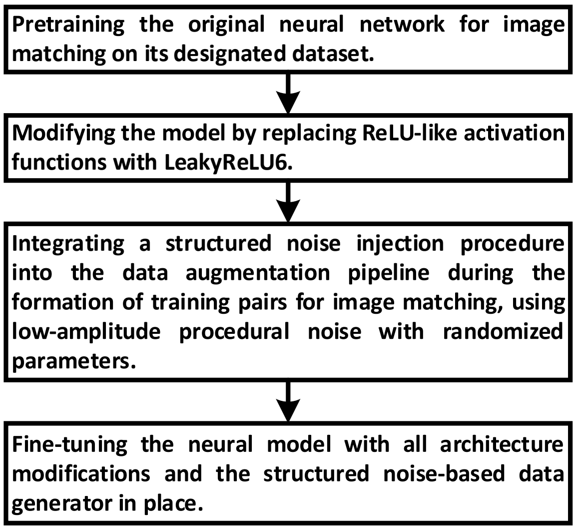


Fig. 1. Stages of the Proposed Method for Improving the Robustness of Neural Network-Based Image Matching Models

When adapting a pretrained model for robustness through activation replacement, we recommend substituting ReLU-like activations with LeakyReLU6 primarily in the deep convolutional layers and feature extraction blocks, where improved gradient flow and bounded activation outputs help suppress noise amplification and dead neurons. Do not modify activations in the final classification layers, where changes could disrupt output distributions, nor in Transformer-style attention blocks, which typically use GELU or similar smooth activations designed for stability. Additionally, avoid replacing activations in the initial layers directly following input, as these may act as implicit noise filters. When batch normalization is used, replacement is acceptable only when the activation follows normalization (i.e., BN → ReLU), to maintain expected activation distributions.

To simulate natural yet adversarial visual distortions that commonly degrade the performance of image matching systems, we propose the injection of structured procedural noise during training. Specifically, we utilize Gabor, Perlin, and Worley noise patterns to perturb the input data in a way that approximates real-world conditions such as shadows, surface textures, occlusions, and lighting inconsistencies.

Unlike unstructured white or Gaussian noise, these patterns introduce semantically plausible perturbations that mimic natural signal variability, thereby forcing the model to focus on stable, semantically meaningful features rather than texture-biased or fragile keypoints.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Perlin Noise – models organic variations like illumination gradients or natural surfaces. It is generated using interpolated gradients at grid points and typically implemented recursively to create fractal behavior. The Perlin noise is constructed as a fractal sum of smoothed co-frequencies in the frequency space [22]:

$$S_{per}(x, y) = \sum_{n=1}^{\Omega} p\left(x\frac{2^{n-1}}{\lambda_x},\ y\frac{2^{n-1}}{\lambda_y}\right),$$

λ_x , λ_y are the wavelength parameters along the respective axes;
Ω is the number of octaves (i.e., levels of granularity);
p $(\cdot)$ denotes the classical 2D Perlin noise function.

To enhance visual contrast, researchers often apply a sinusoidal transformation:

$$G_{per}(x, y) = \sin\left(2\pi \cdot \phi_{sine} \cdot S_{per}(x, y)\right),$$

where ϕ_{sine} is a sine frequency parameter.

Gabor Noise – mimics directional patterns such as cast shadows or repetitive structures (e.g., fences, rooftops). Gabor noise is defined as the convolution of sparse white noise with a Gabor kernel [23]:

$$g(x, y) = e^{-\pi\sigma^2(x^2+y^2)}\cos\left(\frac{2\pi}{\lambda}(x \cdot \cos(\omega) + y \cdot \sin(\omega))\right),$$

where σ is the Gaussian envelope width;
λ is the wavelength (harmonic period);
ω is the orientation.

The resulting Gabor noise is generated as a sum of convolutions:

$$S_{gab}(x, y) = \frac{1}{\xi}\sum_{n=1}^{\xi}\sum_{i}\ g(x - x_i,\ y - y_i; \sigma, \lambda, \omega + \frac{n\pi}{\xi}),$$

where ξ is the number of orientations (degree of isotropy);
$(x_i,\ y_i)$ are the randomly selected kernel placement points.

Worley (Cellular) Noise – simulates structural irregularities like stone patterns, cracked surfaces, or uneven soil. It is defined by computing the distance to the nearest point in a grid of pseudo-random feature points [22]:

$$W(x, y) = \min_{p_i \in F} \|(x, y) - p_i\|,$$

where F is the set of random control points (feature points), typically generated 1–2 per cell of a regular grid; (x, y) are the coordinates of a pixel or a point in space.

To ensure the adversarial effectiveness of these noise patterns, their parameters (orientation, frequency, phase, feature spacing, etc.) are randomized per training iteration or batch, while the amplitude is constrained to a small range (e.g., 3–8% of dynamic range) to maintain visual plausibility and avoid trivializing the task. The resulting image is formed through additive noise superposition:

$$I_{noisy} = I + \epsilon \cdot N(x, y)$$

where ϵ is the noise amplitude parameter, $\epsilon \in [0.03,\ 0.08]$.

This results in imperceptible yet semantically disruptive perturbations, making them ideal candidates for structured adversarial training. By exposing the model to such distortions, we promote resilience to the types of visual variability commonly encountered in real aerial imagery.

To evaluate the effectiveness of the proposed method, it was decided to consider both keypoint detection–based and detector-free models. The most popular models for keypoint detection and description are the neural network combinations of SuperPoint and SuperGlue [16]. One of the most well-known and effective detector-free models is the LoFTR neural network [17].

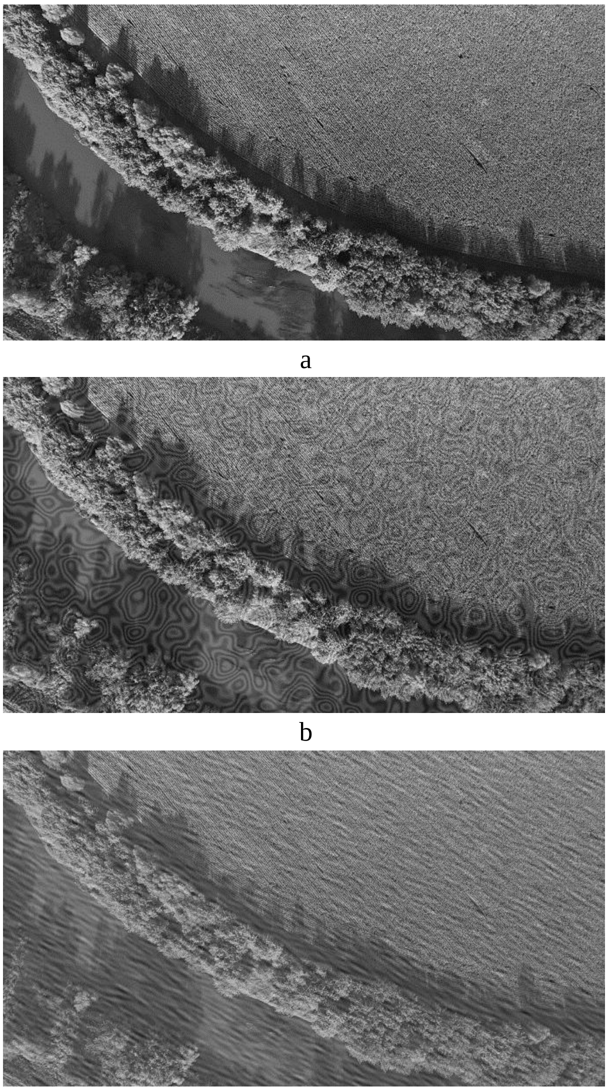
4. Experiments

4.1 Dataset and Image Processing Description

The experiments utilize the Aerial Image Matching Benchmark Dataset [26], comprising high-resolution aerial image pairs (1–2 m/pixel, urban and rural scenes) with ground-truth homographies, sourced from ISPRS-EuroSDR and IEEE DataPort. The dataset includes over 10,000 image pairs, split into training (70%), validation (15%), and test (15%) sets, ensuring robust evaluation across diverse scenes (e.g., buildings, forests, roads). The HPatches Dataset is used for cross-domain evaluation, containing 1,500+ planar patches with controlled viewpoint and illumination changes, also with ground-truth homographies. Sample images include urban rooftops, rural fields, and textured surfaces, processed with random homographies (e.g., rotation, scaling) and photometric distortions (e.g., brightness, contrast). Procedural noise (Gabor, Perlin, Worley) is applied during training to simulate real-world distortions like shadows, illumination gradients, and surface textures, with parameters randomized per batch (e.g., 3–8% amplitude). These datasets' diversity and scale ensure trustworthy evaluation of robustness, as validated by consistent results across splits.

Figure 2 presents an example image from the Aerial Image Matching Benchmark Dataset [26] together with two versions perturbed by procedural noise. As seen in

Fig. 2b and Fig. 2c, procedural noise of limited amplitude does not distort the scene semantics; during training it can enhance the network's ability to extract local features and maintain correct correspondences along field boundaries, forest shelterbelts, and shadowed regions. This, in turn, improves matching robustness and reduces the number of false correspondences under challenging illumination conditions.



a b c

Fig. 2. Example aerial image and its versions perturbed by procedural noise: (a) original aerial image; (b) image with Perlin noise; (c) image with Gabor noise

4.2 Experimental Setup and Scheme

The goal of the experiments is to enhance model robustness under structured noise, evaluated via Matching Accuracy (MA) and Matching Recall (MR). The scheme involves: (1) pretraining SuperPoint+SuperGlue and LoFTR on the Aerial Image Matching Benchmark Dataset; (2) fine-tuning with procedural noise augmentation (Gabor, Perlin, Worley) to simulate adversarial conditions; (3) replacing ReLU activations with LeakyReLU6 in convolutional layers to reduce noise amplification. Initial conditions include pretrained models (public weights from [12, 17]), learning rate 1e-4, batch size 16, 100 epochs, and Adam optimizer. Image pairs are generated by applying random homographies and photometric distortions, with noise parameters randomized per batch. Evaluation uses ground-truth homographies to compute MA and MR, with a threshold of 3 pixels.

We generate image pairs by sampling random homographies and applying random photometric distortions to aerial images. The underlying images come from the set of aerial images in the Aerial Image Matching Benchmark Dataset [26], split into training, validation, and test sets.

Since the homography between the original image and its randomly transformed version is known, the correctness of the matching can always be verified using the following formula [27]

$$\|\mathrm{H} \cdot \mathrm{x} - \mathrm{x}'\| < \theta \Rightarrow \text{match is correct},$$

where H is the random homography applied to the original image to generate the image pair;
x, x' are the original image and its modified version after applying random homographies and random photometric distortions;
θ is the threshold value (typically 3 or 5 pixels).

Matching precision (P) and recall (R) are evaluation metrics and computed from the ground truth correspondences.

Matching Recall (or Matching Score) is the proportion of true correspondences that have been correctly identified by the matching model:

$$\mathrm{Recall}_{\theta} = \frac{\text{Number of correct matches (within } \theta)}{\text{Total number of ground truth matches}}.$$

Matching Accuracy (or Matching Precision) is the proportion of true correspondences among all those predicted by the matching model:

$$\mathrm{MA}_{\theta} = \frac{\text{Number of correct matches (within } \theta)}{\text{Total number of predicted matches}}.$$

4.3 Results

To improve the robustness of the models against structured distortions simulated using procedural noise (Gabor, Perlin, Worley), additional training was conducted with the use of a modified LeakyReLU6 activation function. The results obtained during testing of the trained models on test data with the same type of procedural noise applied are presented in Table 1.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 1
Results on the Aerial Image Matching Benchmark Dataset

Local features	Matcher	Matching Accuracy (θ = 3)	Matching Recall (θ = 3)
SuperPoint	SuperGlue	82.5	75.0
LoFTR	LoFTR	85.3	78.2
SupePoint+ LeakyReLU pretrained with procedural noise	SuperGlue+ LeakyReLU pretrained with procedural noise	88.1	81.5
LoFTR + LeakyReLU pretrained with procedural noise	LoFTR + LeakyReLU pretrained with procedural noise	90.1	84.3

Analysis of Table 1 shows that the proposed method improves the matching accuracy of noisy images for both feature-based models (by 5.5%) and detector-free models (by 5%). However, the individual contribution of procedural noise and activation function modification to the final result remains unclear. Therefore, we conduct an ablation study in which procedural noise is retained as part of the data pair augmentation during training, but the LeakyReLU activation function is not used (see Table 2).

Tables 1 and 2 show that the use of the modified LeakyReLU6 activation function in combination with procedural noise resulted in a 4.7% increase in matching accuracy and a 5.5% increase in matching recall for SuperPoint+SuperGlue. For LoFTR, the use of LeakyReLU6 led to a 4.1% improvement in matching accuracy and a 5.1% improvement in matching recall.

Table 2
Ablation study: matching performance without LeakyReLU6 activation

Local features	Matcher	Matching Accuracy (θ = 3)	Matching Recall (θ = 3)
SuperPoint pretrained with procedural noise	SuperGlue pretrained with procedural noise	83.4	76.0
LoFTR pretrained with procedural noise	LoFTR pretrained with procedural noise	86.0	79.2

The obtained results depend not only on the model and training method but also on the training and test datasets. Therefore, it is important to evaluate the effectiveness of the proposed method on other publicly available datasets as well. In this context, it is useful to compare the results achieved using the proposed method with those obtained using a conventional approach. Table 3 presents the results reported in [27] using SuperPoint + SuperGlue, alongside the results obtained with the proposed method on HPatches Dataset.

Table 3
Results on the HPatches Dataset for neural networks pre-trained on the HPatches Dataset

Local features	Matcher	Matching Accuracy (θ = 3)	Matching Recall (θ = 3)
SuperPoint	SuperGlue	91.5	95.5
SuperPoint+ LeakyReLU pretrained with procedural noise	SuperGlue+ LeakyReLU pretrained with procedural noise	92.9	97.1

Analysis of Table 3 shows that the well-known HPatches Dataset also demonstrates an improvement in Matching Accuracy and Matching Recall compared to the baseline reported in [27].

The proposed method shares many similarities with previous works on data augmentation and adversarial training aimed at improving the generalization ability and robustness of neural networks to observation variability [27]. Therefore, it is worth investigating the generalization capabilities of neural networks trained on the Aerial Image Matching Benchmark Dataset when tested on clean (non-noisy) data from the HPatches Dataset (see Table 4).

Table 4
Results on the HPatches Dataset for neural networks pre-trained on the Aerial Image Matching Benchmark Dataset

Local features	Matcher	Matching Accuracy (θ = 3)	Matching Recall (θ = 3)
SuperPoint	SuperGlue	88.1	90.3
SuperPoint+ LeakyReLU pretrained with procedural noise	SuperGlue+ LeakyReLU pretrained with procedural noise	91.9	96.2

Analysis of Table 4 shows that the SuperPoint+SuperGlue models trained on the Aerial Image Matching Benchmark Dataset exhibit a 3.4% decrease in Matching Accuracy and a 5% decrease in Matching Recall when tested on the HPatches Dataset. In contrast, the models trained using the proposed method show only a 1% decrease in Matching Accuracy and a 0.9% decrease in Matching Recall. This confirms the improved generalization capability of the models trained with the proposed approach.

Thus, the proposed method for enhancing the robustness of image matching neural networks is suitable for both feature-based and detector-free models. Moreover, improvements in accuracy metrics were achieved during both training and testing across different datasets.

5. Discussion

The proposed method enhances robustness by combining LeakyReLU6 activation and procedural noise training, addressing real-world challenges in aerial image matching. LeakyReLU6 reduces noise amplification by bounding activations (up to 6) and improving gradient flow, while procedural noise (Gabor, Perlin, Worley) mimics natural distortions like shadows and terrain textures, forcing models to prioritize stable features. Benefits include over 4% improvements in Matching Accuracy (MA) and Matching Recall (MR) on noisy data (Tables 1–2) and better cross-domain generalization on HPatches (Tables 3–4), with only a 1% MA drop compared to 3.4% for baselines. This makes the method suitable for UAV applications like navigation and mapping under adverse conditions (e.g., shadows, seasonal changes).

Limitations include the lack of optimization for noise parameters (e.g., frequency, amplitude), which may affect robustness, and evaluation restricted to two models (SuperPoint+SuperGlue, LoFTR), limiting generalizability to other architectures. The synergy of LeakyReLU6 and procedural noise is confirmed by ablation studies (Table 2), but further analysis of individual contributions (e.g., noise types) is needed. The method's platform-agnostic nature supports deployment on various UAVs, but real-time implementation requires additional optimization.

The gains achieved by the proposed training scheme are particularly valuable for visual navigation in GNSS-denied settings (e.g., forest edges, urban canyons, and dense built-up areas). By injecting procedural perturbations during training, we encourage the network to attend to stable, semantically meaningful boundaries—field/parcel borders, shorelines and water edges, forest margins and shelterbelts, as well as building footprints, roof edges, and other architectural contours—rather than transient appearance cues. Consequently, feature extraction and matching remain more consistent across shadows and seasonal illumination changes, which reduces drift in visual localization and improves alignment to aerial maps. Figure 2 illustrates that the proposed augmentations preserve the visibility of these semantically salient boundaries (e.g., agricultural parcels, tree lines, and building edges); when used during training, this emphasis biases the network toward such stable structures, ultimately improving in-flight map-matching reliability.

6. Conclusions

This study introduces a novel method to enhance the robustness of neural network models for aerial image matching by integrating LeakyReLU6 activation functions and adversarial procedural noise (Perlin, Gabor, Worley) during training. The contribution lies in combining these techniques to improve Matching Accuracy (MA) and Matching Recall (MR) under structured noise, achieving over 4% improvements for both feature-based (SuperPoint+SuperGlue) and detector-free (LoFTR) models on the Aerial Image Matching Benchmark Dataset (Tables 1–2). The novelty is the first application of this combined approach to aerial image matching, enhancing cross-domain generalization on HPatches (Tables 3–4, with 1% vs. 3.4% MA drop). Practically, the method improves UAV navigation and mapping reliability under adverse conditions like shadows and seasonal changes.

Limitations include untested noise parameter optimization and evaluation limited to two models. Future research will focus on: (1) optimizing noise parameters (e.g., frequency, amplitude) to balance robustness and generalization; (2) applying meta-learning to adapt models to diverse noise types dynamically; (3) evaluating additional architectures like R2D2 and XFeat; (4) exploring real-time implementation for UAV onboard systems to enhance practical deployment.

Future research will focus on the application of meta-learning, regularization, and architectural improvements to enhance the robustness and resilience of neural network-based aerial image matchers.

Contribution of authors: development of conceptual provisions and methodology of research, formulation of conclusions – **Viacheslav Moskalenko**; Review and analysis of references; development of software for modeling – **Alona Moskalenko**; development of mathematical models, analysis of research results – **Yuriy Moskalenko**.

Conflict of interest: the authors declare that they have no conflict of interest in relation to this research, whether financial, personal, author ship or otherwise, that could affect the research and its results presented in this paper.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Financing: the work presented in this paper is supported by research project No. 0124U000548 "Information Technology for Ensuring the Resilience of the Intelligent Onboard System of Small-Sized Aircraft" (2024–2026).

Data availability: The **Aerial Image Matching Benchmark Dataset** used in this study can be obtained from two sources: the ISPRS-EuroSDR portal (https://ifpwww.ifp.uni-stuttgart.de/ISPRS-EuroSDR/ImageMatching/default.aspx) and IEEE DataPort (https://ieee-dataport.org/documents/aerial-image-matching-benchmark-dataset). Please note that both repositories require free user registration prior to download. The dataset comprises high-resolution aerial image pairs, ground-truth homographies, and official evaluation scripts. The **HPatches Dataset** [Balntas *et al.*, CVPR 2017] is available without restriction at https://hpatches.github.io/ (GitHub: https://github.com/uzh-rpg/hpatches-release). It includes planar image patches under controlled viewpoint and illumination changes, together with the corresponding ground-truth homography matrices. All raw images, homography files, and the code used to compute matching precision and recall in this paper can be accessed from these locations.

Use of artificial intelligence: the authors utilized artificial intelligence technologies to provide their verified results. The writing of the article's text was done without the use of artificial intelligence technologies.

Acknowledgments: the authors thank the Ministry of Education and Science of Ukraine for the support to the Laboratory of Intellectual Systems in the framework of research project No. 0124U000548 "Information Technology for Ensuring the Resilience of the Intelligent Onboard System of Small-Sized Aircraft" (2024–2026).

All authors have read and agreed with the published version of the manuscript.

References

1. ISIK, M. Comprehensive empirical evaluation of feature extractors in computer vision. *PeerJ Computer Science*, 2024, vol. 10, article no. e2415. DOI: 10.7717/peerj-cs.2415.
2. Verykokou, S., & Ioannidis, C. Image Matching: A Comprehensive Overview of Conventional and Learning-Based Methods. *Encyclopedia*, 2025, vol. 5, iss. 1, article no. 4. DOI: 10.3390/encyclopedia5010004.
3. Xu, S., Chen, S., Xu, R., Wang, C., Lu, P., & Guo, L. Local feature matching using deep learning: A survey. *Information Fusion*, 2024, vol. 107, article no. 102344. DOI: 10.1016/j.inffus.2024.102344.
4. Bonilla, S., Di Vece, C., Daher, R., Ju, X., Stoyanov, D., Vasconcelos, F., & Bano, S. Mismatched: Evaluating the Limits of Image Matching Approaches and Benchmarks (Version 2). *arXiv*, 2024. DOI: 10.48550/ARXIV.2408.16445.
5. Zhang, X., Xiao, F., Zheng, M., & Xie, Z. UAV image matching from handcrafted to deep local features. *European Journal of Remote Sensing*, 2024, vol. 57, iss. 1. DOI: 10.1080/22797254.2024.2307619.
6. Khurshid, M., Shahzad, M., Khattak, H. A., Malik, M. I., & Fraz, M. M. Vision-Based 3-D Localization of UAV Using Deep Image Matching. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, vol. 17, pp. 12020–12030. DOI: 10.1109/jstars.2024.3422310.
7. Kim, Y., Back, S., Song, D., & Lee, B.-Y. Aerial Map-Based Navigation by Ground Object Pattern Matching. *Drones*, 2024, vol. 8, iss. 8, article no. 375. DOI: 10.3390/drones8080375.
8. Śledziowski, J., Terefenko, P., Giza, A., Forczmański, P., Łysko, A., Maćków, W., Stępień, G., Tomczak, A., & Kurylczyk, A. Application of Unmanned Aerial Vehicles and Image Processing Techniques in Monitoring Underwater Coastal Protection Measures. *Remote Sensing*, 2022, vol. 14, iss. 3, article no. 458. DOI: 10.3390/rs14030458.
9. Zhang, X., He, Z., Ma, Z., Wang, Z., Wang, L. LLFE: A Novel Learning Local Features Extraction for UAV Navigation Based on Infrared Aerial Image and Satellite Reference Image Matching. *Remote Sensing*, 2021, vol. 13, iss. 22, article no. 4618. DOI: 10.3390/rs13224618.
10. Koch, T., Zhuo, X., Reinartz, P., & Fraundorfer, F. A new paradigm for matching UAV- and aerial images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, vol. III–3, pp. 83–90. DOI: 10.5194/isprs-annals-iii-3-83-2016.
11. Bian, X., He, Z., Gong, Z., & Ren, K. An optimized feature extraction algorithm based on SuperPoint. In C. Zuo (Ed.), *AOPC 2022: AI in Optics and Photonics*, 2023. 9 p. SPIE. DOI: 10.1117/12.2651943.
12. Petrakis, G. A SuperPoint neural network implementation for accurate feature extraction in unstructured environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023, vol. XLVIII-1/W2-2023, pp. 1215–1222. DOI: 10.5194/isprs-archives-xlviii-1-w2-2023-1215-2023.
13. Revaud, J., Weinzaepfel, P., De Souza, C., Pion, N., Csurka, G., Cabon, Y., & Humenberger, M. R2D2: Repeatable and reliable detector and descriptor (Version 2). *arXiv*, 2019. DOI: 10.48550/ARXIV.1906.06195.
14. Wang, T., Lin, X., Zhang, P. Enhancing astronomical image stitching with XFeat: A deep learning approach. *2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 2024, pp. 708–712. IEEE. DOI:

10.1109/icicml63543.2024.10958013.

15. Li, W. SuperGlue-based deep learning method for image matching from multiple viewpoints. *Proceedings of the 2023 8th International Conference on Mathematics and Artificial Intelligence*, 2023, pp. 53–58. ACM. DOI: 10.1145/3594300.3594310.

16. Lindenberger, P., Sarlin, P.-E., & Pollefeys, M. LightGlue: Local feature matching at light speed. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. IEEE. DOI: 10.1109/iccv51070.2023.01616.

17. Hoshi, S., Ito, K., & Aoki, T. Accurate and robust image correspondence for structure-from-motion and its application to multi-view stereo. *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 2626–2630. IEEE. DOI: 10.1109/icip46576.2022.9897304.

18. Kumbhar, M. F. Image denoising using autoencoders. *In Artificial Intelligence and Knowledge Processing*, 2023, pp. 137–144. CRC Press. DOI: 10.1201/9781003328414-13.

19. Moskalenko, V., Kharchenko, V., Moskalenko, A., & Kuzikov, B. Resilience and resilient systems of artificial intelligence: Taxonomy, models and methods. *Algorithms*, 2023, vol. 16, iss. 3, article no. 165. DOI: 10.3390/a16030165.

20. Pogorzelski, T. P. *Aerial Image Matching Benchmark Dataset [Dataset]*. IEEE DataPort, 2025. DOI: 10.21227/JD5S-AY89.

21. Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A. SuperGlue: Learning feature matching with graph neural networks (Version 2). *arXiv*, 2019. DOI: 10.48550/ARXIV.1911.11763.

22. Bai, T., Luo, J., Zhao, J., Wen, B., & Wang, Q. Recent Advances in Adversarial Training for Adversarial Robustness. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-21)*, 2021, pp. 4312–4321. DOI: 10.24963/ijcai.2021/591.

23. Athalye, A., Carlini, N., & Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. *arXiv*, 2018. DOI: 10.48550/ARXIV.1802.00420.

24. Qiu, P., Wang, Q., Wang, D., Lyu, Y., Lu, Z., & Qu, G. Mitigating Adversarial Attacks for Deep Neural Networks by Input Deformation and Augmentation. *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2020, pp. 157–162. DOI: 10.1109/ASP-DAC47756.2020.9045107.

25. Zhang, B., Tondi, B., Lv, X., Barni, M. Challenging the Adversarial Robustness of DNNs Based on Error-Correcting Output Codes. *Security and Communication Networks*, 2020, vol. 2020, pp. 1–11. DOI: 10.1155/2020/8882494.

26. Maurício, J., Domingues, I., & Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Applied Sciences*, 2023, vol. 13, iss. 9, article no. 5521. DOI: 10.3390/app13095521.

27. Assran, M., Caron, M., Misra, I., Bojanowski, P., Bordes, F., Vincent, P., Joulin, A., Rabbat, M., & Ballas, N. Masked Siamese Networks for Label-Efficient Learning. *Lecture Notes in Computer Science*, Springer Nature Switzerland, 2022, pp. 456–473. DOI: 10.1007/978-3-031-19821-2_26.

Received 28.04.2025, Accepted 25.08.2025

МЕТОД ПІДВИЩЕННЯ РОБАСТНОСТІ МОДЕЛІ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ СПІВСТАВЛЕННЯ АЕРОЗНІМКІВ

В. В. Москаленко, А. С. Москаленко, Ю. В. Москаленко

Нейронні мережі для задачі зіставлення зображень дедалі частіше використовуються в аналізі аерознімків, зокрема для навігації БПЛА, локалізації та картографування. Однак їхня чутливість до структурованих візуальних спотворень — таких як тіні, зміни освітлення та варіабельність рельєфу — обмежує стійкість у реальних умовах. Розроблення методів навчання, що підвищують стійкість зіставлення ознак за наявності змагальних шумів, є актуальним завданням. **Предметом** дослідження у статті є метод підвищення робастності моделей нейронних мереж для зіставлення аерознімків за наявності структурованого візуального шуму. **Метою** дослідження є розроблення методу навчання, який підвищує робастність та здатність до узагальнення моделей зіставлення зображень. **Методами** дослідження є: методи змагального навчання, методи генерації процедурного шуму, а також способи модифікації архітектури нейромереж. При цьому як процедурний шум розглядаються шум Габора, Перліна та Ворлі. Як модифікація архітектури нейромережі пропонується заміна ReLU-подібної функції активації на LeakyReLU6. Отримано такі **результати.** Запропонований метод покращує як точність (precision) зіставлення, так і повноту (recall) зіставлення більш ніж на 4% як для моделей, основаних на детектуванні ознак (SuperPoint+SuperGlue), так і для бездетекторних моделей (LoFTR) на зашумлених тестових даних. Дослідження абляції підтверджують, що підвищення надійності відбувається завдяки поєднанню процедурного шуму та LeakyReLU6. Крім того, моделі, навчені на тестовому наборі даних Aerial

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Image Matching Benchmark Dataset і оцінені на міждоменному наборі даних HPatches, демонструють покращене узагальнення, тобто вищі точністні характеристики порівняно з результатом традиційного методу навчання. **Висновки.** Вперше запропоновано метод підвищення робастності моделей зіставлення зображень на основі поєднання змагального навчання на даних з накладеним процедурним шумом, та модифікації функції активації для її обмеження під час прямого поширення. Запропонований метод окрім стійкості до шуму покращує міждоменне узагальнення, а також може бути застосований до різних архітектур нейронних мереж.

Ключові слова: зіставлення зображень; робастність; змагальне навчання; процедурний шум; аерознімки.

Москаленко В'ячеслав Васильович – канд. техн. наук, доц., доц. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Москаленко Альона Сергіївна – канд. техн. наук, доц. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Москаленко Юрій Васильович – асп. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Viacheslav Moskalenko – PhD, Associate Professor of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: v.moskalenko@cs.sumdu.edu.ua, ORCID: 0000-0001-6275-9803, Scopus Author ID: 57189099775.

Alona Moskalenko – PhD, Associate Professor of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: a.moskalenko@cs.sumdu.edu.ua, ORCID: 0000-0003-3239-1977, Scopus Author ID: 57148522500.

Yuriy Moskalenko – PhD Student of Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: yuriy.mosk@gmail.com, ORCID: 0009-0002-3635-3337.