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

UDC 004.8

doi: 10.32620/reks.2025.3.11

Volodymyr VOZNIAK[1], Oleksander BARMAK[1], Iurii KRAK[2,3]

[1] ***Khmelnytskyi National University, Khmelnytskyi, Ukraine***
[2] ***Taras Shevchenko National University of Kyiv, Kyiv, Ukraine***
[3] ***Glushkov Cybernetics Institute, Kyiv, Ukraine***

METHOD FOR MATCHING SATELLITE AND UAV IMAGES FOR VISUAL PLACE RECOGNITION WITH CROSS-VIEW COLOR NORMALIZATION

*The **subject of this** article is visual place recognition (VPR), specifically matching satellite images with images captured by unmanned aerial vehicles (UAVs). VPR is critical for autonomous UAV navigation, particularly in GPS-denied environments such as urban canyons or* areas with significant infrastructure coverage *where GNSS signals are unreliable. Despite its practical importance, accurately matching UAV images to satellite imagery remains challenging due to significant viewpoint, scale, illumination, and texture discrepancies. Traditional approaches that rely on handcrafted descriptors or classical local features often fail under such cross-view conditions. This study **aims** to design a robust visual place recognition method for matching UAV and satellite imagery, employing deep learning-based embeddings and advanced color normalization to improve reliability across cross-view scenarios. The **tasks** addressed in this article are: firstly, designing a YOLO-based method is designed for extracting global image embeddings, which utilizes YOLO's multi-scale feature extraction capabilities to encode semantically significant landmarks in the scene. Second, a novel preprocessing technique based on aligning statistical color distributions between UAV and satellite images was developed and implemented to enhance their visual congruence. Finally, these components are integrated into a complete VPR system and evaluated for effectiveness using the challenging VPAIR dataset, emphasizing urban settings. The **methods** employed include deep learning techniques, particularly fine-tuning a YOLO11 neural network on a dataset specifically annotated for building segmentation. Statistical alignment techniques based on cumulative distribution functions (CDF) were used to standardize image appearances between the two distinct image domains. **Conclusions.** The experiments demonstrate significant improvements in UAV-to-satellite image matching performance using the proposed method. Fine-tuning YOLO11 specifically for building segmentation resulted in a robust embedding generation method that achieved high segmentation accuracy (F1-score of 0.722). The color preprocessing technique further improved the recognition performance, with Recall@1 reaching 19.5% for urban terrain within a localization radius of 3, substantially outperforming the traditional methods. This study provides an effective solution for UAV localization tasks, particularly in complex urban environments, highlighting the importance of integrated embedding extraction and domain-specific image preprocessing in cross-view visual place recognition.*

Keywords: *visual place recognition; UAV; YOLO; image preprocessing; deep learning; image segmentation.*

1. Introduction

1.1. Motivation

Unmanned Aerial Vehicles (UAVs) are increasingly employed for tasks ranging from environmental monitoring and disaster response to smart agriculture and urban planning [1]. A core requirement in these applications is the accurate self-localization of the UAV. While Global Navigation Satellite Systems (GNSS) like GPS are the de facto solution, they often fail or degrade in performance under signal blockages or interference (e.g., urban canyons, areas with significant infrastructure coverage). Vision-based localization offers an attractive alternative in GPS-denied scenarios, providing low-cost, rich-information positioning that does not suffer cumulative drift [1]. One such approach is Visual Place Recognition (VPR) is one such approach, wherein the UAV's onboard camera is used to recognize its location by matching the current view against a database of geo-referenced images [2]. Matching UAV-captured images to satellite imagery has emerged as a practical solution for global localization, since satellites provide broad coverage and readily available maps. This cross-view image matching problem, identifying the same place from drastically different viewpoints (ground oblique vs. overhead), is the focus of extensive research due to its importance for autonomous UAV navigation in GNSS-denied environments [3].

However, UAV-satellite image matching poses significant challenges [4]. The two image domains differ in viewpoint (oblique/side perspective vs. top-down), scale, and resolution and often exhibit stark appearance discrepancies in color, illumination, and texture [5]. Seasonal changes and weather conditions can alter the view

of the UAV, while satellite images may be captured at different times or spectral bands, thereby exacerbate visual mismatches. Moreover, aerial scenes often contain repetitive patterns (e.g., rooftops and fields) with few distinct landmarks, making correspondence ambiguous. Traditional feature-based methods struggle in this context. For instance, the direct matching of keypoints between UAV and satellite images is unreliable due to extreme viewpoint differences and loss of 3D information [3]. As a result, the early approaches to UAV localization that relied on classical local features (SIFT, SURF) or handcrafted descriptors had limited success in cross-view settings.

These challenges motivate the use of learned image embeddings to bridge the domain gap between UAV and satellite imagery [6]. The rapid progress of deep learning in computer vision has led to powerful convolutional neural network (CNN) features and architectures that significantly improve the robustness of place recognition [3]. Deep models can learn viewpoint-invariant and appearance-invariant representations by training on large datasets, succeeding where handcrafted descriptors fail. Visual localization systems have begun to leverage such deep embeddings to achieve high recall despite perspective changes, effectively overcoming many limitations of traditional methods.

Appropriate image preprocessing techniques play a crucial role in enhancing the effectiveness of these deep learning-based approaches. Preprocessing steps, such as geometric transformations, color normalization, illumination correction, and scale adjustment, significantly reduce domain discrepancies between UAV and satellite imagery [7]. For example, geometric rectification and alignment methods can normalize perspectives, reduce viewpoint variability, and simplify cross-domain matching. Illumination and color normalization techniques mitigate the effects of lighting variations and atmospheric conditions, thereby stabilizing visual features across diverse environmental conditions. Additionally, preprocessing can emphasize relevant semantic features while suppressing irrelevant or ambiguous details, thereby enabling a more consistent feature extraction by the subsequent deep learning models.

Given the constrained compute resources on UAV platforms, methods that are not only accurate but also efficient are needed. In this regard, the YOLO family [8] of object detection networks stands out for its speed and accuracy balance, even on edge devices [1]. YOLO-based models process images in real-time on limited hardware, making them promising candidates for UAV place recognition. Moreover, YOLO's architecture provides multiscale feature extraction and focuses on salient objects, enriching place descriptors with semantically meaningful cues. These considerations underpin the proposed approach, which exploits the strengths of YOLO to create robust image embeddings for cross-view VPR while simultaneously addressing the appearance gap through meticulous image preprocessing.

1.2. State of the art

Modern visual place recognition is typically formulated as an image retrieval problem: a query image (e.g., a UAV snapshot) is compared against a large database of geo-tagged reference images (e.g., satellite map tiles), and the most similar match is returned as the hypothesized location [3]. The key to this process is a reliable image descriptor or embedding that makes matching accurate and efficient. Early VPR systems (around the 2000s) used handcrafted global descriptors built on local features, such as bag-of-visual-words representations [9]. Notably, FAB-MAP [10] introduced an appearance-based place recognition method using a bag-of-words algorithm over SURF features, and DBoW2 [11] improved speed with binary feature quantization. Although effective for moderate viewpoint changes, these classical methods degrade severely under the wide baseline differences in UAV vs. satellite imagery.

The advent of deep learning caused a step-change in descriptor quality: CNN-based embeddings proved far more robust to illumination and viewpoint variation than engineered features. A seminal example is NetVLAD [12], which combined a CNN backbone with a VLAD aggregation layer to produce compact global descriptors, dramatically outperforming prior approaches on place recognition benchmarks. Subsequent research has refined global embeddings through various means, e.g., multiscale feature fusion and local feature integration in Patch-NetVLAD [13], or transformer-based context aggregation in recent methods – all with the aim of capturing distinctive scene signatures that remain stable despite viewpoint changes.

A typical VPR pipeline consists of (1) feature extraction, where each image is converted to a descriptor, and (2) feature matching/retrieval, where the descriptor of a query is compared to those in the reference database (often via nearest-neighbor search in the embedding space). To handle cross-view scenarios, such as UAV-to-satellite matching, specialized architectures are used. A common approach is a dual-branch network (Siamese or triplet network) that learns to map UAV and satellite images into a common embedding space, usually by training with metric learning objectives (contrastive or triplet loss) so that true match pairs come together in that space. This training paradigm, known as cross-view metric learning, has been widely adopted in recent studies.

For example, [14] pioneered ground-to-aerial localization by learning CNN features to match street-level images with satellite images. [15] further demonstrated deep regression of geo-coordinates from ground images

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

using aerial reference data. Building on these, the introduction of the University-1652 dataset by [16] brought UAV drone imagery into the mix, enabling learning-based geo-localization where drones capture building facades and are matched to overhead maps. University-1652 and its variants framed UAV VPR as an image retrieval task and spurred the development of numerous deep models. For instance, the dataset was used to train cross-view networks with classification and triplet-loss schemes, significantly improving retrieval accuracy for drone views.

In recent years, new benchmarks have continued to push the state of the art: VIGOR [17] introduced a validation beyond one-to-one matching by allowing multiple correct matches and negative mining, and SUES-200 [18] provided a large-scale cross-view dataset with multi-altitude drone images and diverse scenes to evaluate robustness across different flight heights. Even specialized datasets such as ALTO [19] focus on large-scale outdoor environments for UAV place recognition, or the very recent ComplexUAV [20] emphasizes complex multi-terrain scenarios, reflecting the community's drive toward more realistic evaluations.

The VPAIR dataset [21] complements these existing benchmarks by addressing challenges specific to medium to high altitude aerial scenarios. The dataset includes downward-facing camera images paired with high-resolution, rendered reference imagery, dense depth maps, and precise 6-DoF reference poses captured from a light aircraft at altitudes between 300 and 400 m. The 107-kilometer-long trajectory spans diverse landscapes, including urban, agricultural, and forested regions, highlighting challenges such as large viewpoint differences and significant in-plane rotations. Experiments on VPAIR underscore the need for rotation-robust and computationally efficient image descriptors tailored explicitly for aerial-to-aerial VPR tasks.

Several techniques have been proposed to generate robust image embeddings for VPR. Researchers have explored local feature-based recognition beyond the CNN+VLAD approaches, with the hypothesis being that explicit keypoint matching could complement global descriptors. For example, a 2023 study asked "Are local features all you need for cross-domain VPR?" and found that incorporating learned local features can boost performance in challenging domain shifts. Hybrid methods now often combine global and local cues: a global retrieval to shortlist candidates, followed by local feature matching to re-rank or verify (a strategy employed by Patch-NetVLAD [13] and others). Meanwhile, the rise of Vision Transformers has influenced embedding design; self-attention can encode long-range context useful for place recognition, as seen in transformer-based global descriptors and reranking models.

Numerous recent studies have investigated the use of lightweight and multimodal feature representations for place recognition. For instance, MinkLoc [22] uses sparse 3D convolutions to enhance recognition performance at scale, especially when applied to LiDAR or depth map inputs. Nevertheless, integrating such additional sensors can be unfeasible for UAVs due to strict power and weight constraints. CosPlace [23] applied a classification-oriented training strategy on the San Francisco XL dataset, which includes 40 million GPS-annotated directional images, to address challenges in image-based localization. Following that, MixVPR [24] proposed a feature mixing approach based on MLPs, trained using the GSV-Cities dataset [25], which consists of 530,000 images spanning 62,000 different locations worldwide. These cases highlight the growing reliance on large-scale, highly curated datasets in contemporary VPR research.

Another notable trend is leveraging foundation models and self-supervised learning. [13] and [26] advocated general, pre-trained features for VPR. Recently, AnyLoc [27] demonstrated that using a pre-trained vision transformer (without any task-specific fine-tuning) can achieve state-of-the-art results on multiple VPR benchmarks. This suggests that rich semantic embeddings from models like OpenAI's CLIP [28] or Meta's DINOv2 [29] are sufficiently discriminative for place recognition across modalities. Indeed, self-supervised models trained on massive image data have shown remarkable robustness to viewpoint and appearance changes, making them appealing for cross-view tasks. On the other hand, domain-specific training is still crucial to squeeze out maximum performance; methods such as EigenPlaces [30] explicitly train on multi-view data to achieve viewpoint invariance. In summary, the state-of-the-art in image embeddings for VPR encompasses a spectrum from task-specific deep models to generic pretrained features, often with an ensemble of global and local representation techniques.

An important aspect of cross-view VPR is the handling of the photometric disparities between UAV and satellite images. Color preprocessing techniques attempt to normalize or reduce these differences before feature extraction, thereby making the recognition model's task easier. Applying histogram matching is a straightforward approach: adjust the color histogram of the UAV image to mimic that of the satellite image (or vice versa). While histogram matching can align global color distributions, it often leads to unnatural color distortions.

Recent research has proposed more nuanced color transfer methods. For example, [2] performed color alignment in a decorrelate color space by matching the mean and variance of UAV image channels to those of the satellite image, thereby avoiding some artifacts

caused by direct histogram warping. This kind of preprocessing (sometimes called color constancy or style transfer) essentially standardizes the UAV image appearance to be closer to the satellite domain before feeding it to the network. The benefit is a reduced domain gap, the feature extractor does not have to learn to ignore irrelevant color differences. Indeed, [2] reported that applying such color normalization to UAV images before training improved geo-localization accuracy in their experiments.

Another popular strategy is data augmentation via style variation: instead of directly altering the images at test time, the model is trained on various appearance conditions so that it becomes invariant. For instance, [5] used neural style transfer to create augmented training samples with different weather, lighting, and seasonal conditions, thereby improving the model's robustness to environmental changes. This approach has the model effectively learn to "see through" color and illumination differences. However, appearance normalization has limitations. As noted by [2], additional color correction may yield diminishing returns if images already contain rich color content, and over-normalizing can even wash out discriminative details. Ultimately, geometric discrepancies (viewpoint and object arrangement) pose a greater challenge than color; thus, color preprocessing is a helpful but partial solution. It is most beneficial when the UAV and satellite images have systematic color biases (e.g., one captured at dusk and the other at noon), in which case normalizing brightness and tone can significantly aid matching. In current state-of-the-art pipelines, a combination of techniques is often used: basic normalization (such as per-channel mean subtraction and scaling as done in ImageNet preprocessing) is almost always applied, and in specialized cases, histogram equalization or learned style translation may be added to further reduce appearance gaps.

The [31] method complements existing preprocessing techniques by introducing a style alignment strategy to reduce intraclass visual discrepancies between UAV and satellite images caused by differing illumination conditions and camera parameters. They use a cumulative distribution-based mapping function derived from the RGB channels of satellite images to unify drone imagery to a consistent satellite-like visual style. This alignment significantly mitigates differences in lighting and color, thus improving feature consistency and matching accuracy across disparate views. This strategy particularly benefits scenarios with strong variations in drone imagery due to sunlight or camera-induced chromatic aberrations.

1.3. Objectives and tasks

he objective of this work is to build an enhanced visual place recognition system that aligns UAV images with satellite views, integrating deep embedding models and color preprocessing techniques to achieve greater robustness and precision, especially in urban cross-view conditions. The target system is designed to reliably recognize places from a low-altitude UAV perspective by matching against satellite imagery, even under significant viewpoint and appearance changes. Based on the insights from the state-of-the-art, our approach integrates a YOLO-based feature extractor with color normalization techniques into the VPR pipeline. The motivation for using YOLO lies in its demonstrated efficiency and ability to capture multi-scale features and semantic objects in the scene [1], which is hypothesized to provide a rich descriptor for place recognition [32], [33]. Moreover, the use of YOLO for visual place recognition represents a novel contribution because these architectures have not been previously applied in this domain. YOLO currently ranks among the most accurate models for image detection and segmentation, and its exceptionally high inference speed ensures effective real-time performance compared to transformer-based approaches. Meanwhile, color preprocessing is employed to mitigate the domain gap between UAV camera and satellite images, unifying their visual characteristics and thus easing the learning task.

The tasks addressed in this work are as follows:

1. **YOLO Embedding Generation.** A method is designed and implemented to extract global image descriptors from a YOLO network. Instead of using YOLO solely for object detection, its internal feature maps are repurposed to serve as image embeddings for the entire scene. By doing so, the embedding inherently encodes information about prominent objects and structures that could act as stable landmarks for place recognition (learned through YOLO's detection training).

2. **Cross-Domain Color Preprocessing.** The preprocessing steps are investigated to unify the color distributions of UAV and satellite images, with experiments using methods such as histogram matching and learned color transfer to standardize images across both domains. The goal is to produce a consistent look between UAV and satellite imagery, for example, by adjusting the UAV images to approximate the spectral response of satellite sensors or vice versa. The impact of converting images to different color spaces and applying global normalization will be evaluated.

3. **System Integration and Evaluation.** The YOLO-based embedding and color preprocessing are integrated into a complete VPR pipeline for UAV localization. This includes building a reference database of satellite image descriptors, processing incoming UAV frames through color normalization and embedding extraction, and performing a fast nearest-neighbor search to retrieve candidate matches. Then, a thorough evaluation is performed on the benchmark dataset VPAIR [21],

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

comparing the proposed method against classical feature-based baselines and recent deep-learning methods. Key evaluation metrics include Recall@K [26] and localization accuracy under varying conditions, enabling the quantification of improvements. The contribution of each component (embedding network vs. color preprocessing) is analyzed through ablation studies to confirm the tangible benefits of each proposed element.

The expected outcome is a state-of-the-art image matching system that substantially improves UAV-satellite place recognition performance by accomplishing these tasks. In summary, the contributions of this paper are as follows:

1) a novel YOLO-derived image embedding tailored for cross-view place recognition;

2) a color normalization approach to bridge UAV and satellite image appearance;

3) an extensive evaluation demonstrating superior performance to existing methods on challenging cross-view datasets.

Collectively, these contributions advance the field of visual place recognition for aerial robots, moving closer to reliable GPS-independent UAV navigation in real-world environments.

The paper is structured as follows: Section 1 introduces the problem of visual place recognition (VPR) for UAVs, highlighting the challenges of cross-view image matching between UAV and satellite imagery, and provides a comprehensive review of related work, objectives, and motivation for the research. Section 2 describes the materials and methods used in the study, detailing the proposed approach that combines a YOLO-based embedding extraction with statistical color preprocessing, as well as the evaluation methodology and metrics for assessing system performance. Section 3 presents the results and discussion, including the YOLO model fine-tuning for building segmentation, the image preprocessing method's effectiveness, and experimental comparisons using the VPAIR dataset. Section 4 concludes the paper by summarizing the main findings, discussing the advantages and current limitations of the proposed method, and outlining potential directions for future research.

2. Materials and methods of research

2.1. Methods

Determining the location of Unmanned Aerial Vehicles (UAVs) is typically addressed through Visual Place Recognition (VPR). This study proposes a method that compares UAV-captured query images against a geotagged satellite imagery database. The system identifies the UAV's approximate global position by finding the most closely matched satellite image or group of images. Subsequently, a precise coordinate determination method is used to align the UAV's query image with the matched satellite imagery.

Thus, the primary task can be divided into the following subtasks:

1) global location estimation: identifying the satellite image (tile) most closely resembling the UAV's query image within an extensive database;

2) precise coordinate estimation: determining the exact coordinates (latitude and longitude) within the selected tile.

Figure 1 demonstrates the general workflow employed to determine the UAV coordinates.

This study specifically targets the determination of global UAV locations in urban areas using CNN-based methods. A detailed schematic overview of the proposed approach is presented in Figure 2.

The following notation is introduced. Let $\mathcal{Q} = \{\mathrm{q}\}$ denote a query image captured by the UAV at a particular location and orientation, and let $\mathcal{R} = \{\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_\mathrm{n}\}$ represent a set of predefined images (typically satellite imagery) with known coordinates.

Additionally, the following mapping is defined to obtain vector representations (feature vectors) from images:

$$\mathrm{F}: \mathrm{I} \to \mathrm{V}, \tag{1}$$

where I is the image, V is the feature vector.

Consequently, the overall objective of UAV global location determination to identify an image $\mathrm{r}_\mathrm{j} \in \mathcal{R}$ whose spatial position $\ell(\mathrm{r}_\mathrm{j})$ most closely aligns with the query image position $\ell(q)$. Formally, this is represented as follows:

$$\mathrm{k} = \underset{1 \le \mathrm{j} \le \mathrm{n}}{\operatorname{argmin}} \, \mathrm{d}\left(\mathrm{F}(\mathrm{q}), \mathrm{F}(\mathrm{r}_\mathrm{j})\right), \tag{2}$$

where $\mathrm{d}(\cdot,\cdot)$ is a metric (e.g., Euclidean distance) operating on images represented as feature vectors, and k is the identified satellite image.

A satellite imagery database covering predetermined flight routes must be accessible to facilitate UAV global location identification. Using deep learning architectures, vector representations or image features can be derived (1). In this study, the YOLO11 model is employed, specifically fine-tuned using segmented building images. Feature vectors are extracted from the final backbone layer, as this layer encapsulates the richest image information obtained through the convolutional structure of YOLO11. Crucially, the model is identical for obtaining vector features from both UAV and satellite images. Thus, the refined YOLO11 model generates image vectors for the satellite database, enabling accurate UAV global localization. YOLO11 is used to process real-time

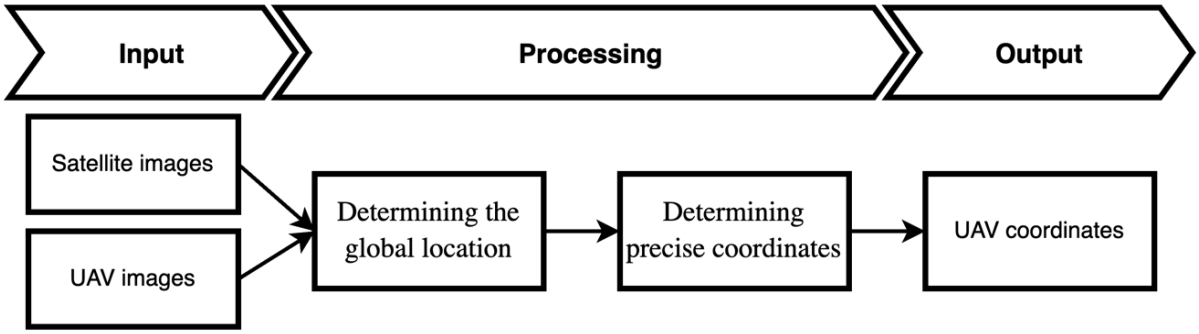

Fig. 1. General data processing scheme for determining UAV location coordinates

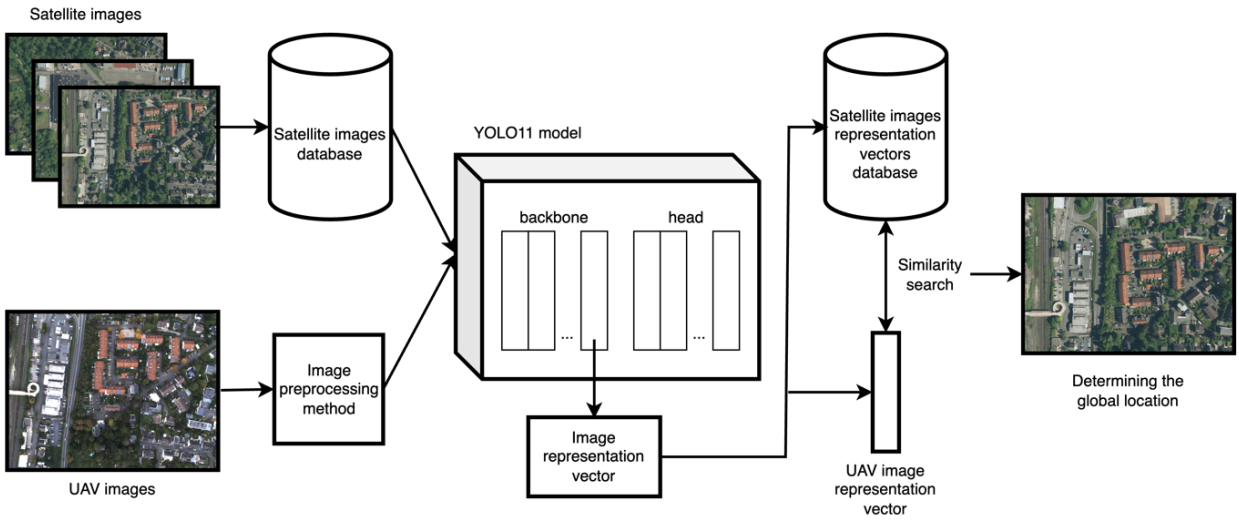

Fig. 2. Processing scheme for UAV global localization using an image preprocessing method and a YOLO11 model fine-tuned on a segmented building dataset

UAV images to obtain vector representations (1), after which the best-matching satellite image (2) is determined using a similarity metric (e.g., Euclidean distance). This matched satellite image represents the approximate global position of the UAV.

To improve the visual congruence between UAV and satellite imagery, a method is proposed to align their statistical distributions. Unlike conventional methods that directly apply satellite cumulative distribution functions (CDF) to UAV images, which may lead to distortion from mismatched distributions [31], our method calculates transformations for each UAV frame individually, accommodating its specific distribution.

Our approach utilizes probability theory principles: given a random variable ξ with a known distribution function $F_\xi(x)$, applying its own distribution function generates a uniformly distributed variable $\gamma \sim U(0,1)$, $\gamma = F_\xi(\xi)$. Conversely, applying the inverse distribution function to a uniform variable $\gamma \sim U(0,1)$ recovers ξ with distribution $F_\xi(x)$. Satellite images consistently captured by similar sensors possess uniform color properties, whereas UAV images captured under varying conditions exhibit different statistical traits. Thus, pixel intensities in images behave like distinct random variables, producing random variable Y and X from UAV and satellite image intensities, respectively. By accurately computing $F_X(x)$ from satellite data and estimating $F_Y(y)$ for UAV images, the alignment of distributions becomes achievable via $F_X^{-1}(F_Y(y))$.

The proposed technique consists of two phases:

1) calculating the average cumulative distribution function from the satellite images;

2) applying this average distribution individually to UAV-captured images.

Algorithm 1 provides the pseudocode for computing the averaged cumulative distribution function from satellite imagery, and Algorithm 2 outlines the averaged function's application to individual UAV images.

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

Algorithm 1

Calculating the averaged cumulative distribution function from the satellite images

Input: $\mathcal{R} = \{r_1, r_2, \dots, r_n\}$ – n satellite images.

Initialization: $F^{sum} \leftarrow 0_{x \times c}, x \in [0; 255], c \in \{R, G, B\}$.

For i in 1..n

Step 1. Compute the normalized histogram (probability density function) for each color channel $c \in \{R, G, B\}$:

$$E_{c,i}(x) = \frac{H_{c,i}(x)}{\sum_{x=0}^{255} H_{c,i}(x)},$$

where $H_{c,i}(x)$ is the number of pixels with value x in channel c of the i-th satellite image.

Step 2. Compute the cumulative distribution function (CDF) for each channel:

$$F_{c,i}(x) = \sum_{k=0}^{x} E_{c,i}(k).$$

Step 3. Add the cumulative distribution function values to F^{sum} to later derive the average cumulative distribution function:

$$F_i(x) = [F_{R,i}(x), F_{G,i}(x), F_{B,i}(x)];$$
$$F^{sum}(x) = F^{sum}(x) + F_i(x).$$

End For

Step 4. The averaged cumulative distribution function (CDF) is computed as follows:

$$\hat{F}(x) = \frac{F^{sum}(x)}{n}.$$

Output: $\hat{F}(x)$ – averaged cumulative distribution function of satellite images, $x \in [0; 255]$.

Algorithm 2

Applying the averaged cumulative distribution function of satellite images to UAV images individually.

Input: $\mathcal{Q} = \{q_1, q_2, \dots, q_m\}$ – m UAV images.

For j in 1..m

Step 1. Compute the normalized histogram (probability density function) for each color channel $c \in \{R, G, B\}$:

$$D_{c,j}(y) = \frac{S_{c,j}(y)}{\sum_{y=0}^{255} S_{c,j}(y)},$$

where $S_{c,j}(y)$ is the number of pixels with value y in a channel c of the j-th UAV image.

Step 2. Compute the cumulative distribution function (CDF) for each channel:

$$G_{c,j}(y) = \sum_{n=0}^{y} E_{c,j}(n).$$

Step 3. Define the transformation function for each channel c by finding the value at which the cumulative distribution function (CDF) from the UAV images aligns with the averaged cumulative distribution function of satellite images:

$$M_{c,j}(y) = \hat{F}_{c,j}^{-1}\left(G_{c,j}(y)\right),$$

where $\hat{F}_c^{-1}$ is the inverse function of the averaged cumulative distribution function for satellite images in a channel c. Because $\hat{F}_c^{-1}$ might be non-analytical, interpolation is used as an approximation:

$$M_{c,j}(y) = \text{interp}\left(G_{c,j}(y), \hat{F}_c(z), z\right),$$

where interp is an interpolation function that finds the corresponding z for each $G_c(y)$, such that $\hat{F}_c(z) \approx G_c(z)$.

End For

Output: $M_{c,j}(y)$ – the resulting function that transforms the input pixels of the j-th UAV image for the given color channel c, $y \in [0; 255]$.

Histogram alignment ensures that UAV and satellite imagery share similar pixel intensity distributions, adjusting contrast, brightness, and tonal attributes, minimizing variations within classes and enhancing feature matching and recognition accuracy.

2.2. Evaluation

The standard evaluation metrics employed for image segmentation include mAP, Precision, Recall, and F1-score. Metrics such as Precision, Recall, and F1-score, are widely used across machine learning tasks, extending from binary classifications to segmentation. The absence of true-negative values in the confusion matrix is a unique aspect of segmentation tasks, although this does not affect the calculation of these metrics.

Additionally, mean Average Precision (mAP) warrants individual discussion. It is formally represented as follows:

$$AP_c = \sum_{n=1}^{N} (R_n - R_{n-1}) P_n, \quad (3)$$

$$mAP = \frac{1}{C} \sum_{c=1}^{C} AP_c, \quad (4)$$

where P_n and R_n are Precision and Recall at threshold n with $R_0 = 0$ and $R_N = 1$, C is the number of classes, AP_c – average precision for the class c.

AP_c can alternatively be viewed as the area under the Precision-Recall curve specific to class c. In the current study, which focuses exclusively on buildings, the mAP simplifies to AP_b.

Recall@N [26] is frequently used to assess localization techniques. According to this metric, if the corresponding database image appears within the top N search outcomes, search results are true-positive for a given query. This metric is widely recognized in the computer vision domain, particularly when additional processing can further filter incorrect matches:

$$\text{Recall@N} = \frac{M_{\mathcal{Q}}}{N_{\mathcal{Q}}}, \quad (5)$$

where N_Q is the total number of query images, and M_Q is the number of queries with at least one correct match within the top-N results.

An alternative metric version incorporates a localization radius, considering a match as true-positive if the spatial distance between the query and database images is within a specified range, defined in meters or frame count. This variant is particularly beneficial for scenarios involving overlapping reference imagery, allowing precise UAV localization even without exact image matches.

3. Results and Discussion

3.1. YOLO fine-tuning

Given that the standard YOLO model does not recognize buildings as a specific class, this study suggests fine-tuning the YOLO11 [34] model using a specialized dataset exclusively consisting of building segmentation. The dataset encompasses 9,665 images from cities including Tyrol (2999), Tripoli (1078), Kherson (1053), Donetsk (999), Mekelle (951), Mykolaiv (739), and Kharkiv (602). Figure 3 shows an example of building segmentation from the training dataset.

The YOLO11 model was fine-tuned using the building segmentation dataset [35] and the open-source Python library ultralytics [36], running on an Ubuntu operating system and an Nvidia MSI RTX 3060 GPU. The segmentation task targeted buildings exclusively, and the default YOLO11 architecture was implemented with 100 training epochs and 640-pixel image resolution.

The performance metrics for the refined YOLO model – mAP, Precision, Recall, and F1-score – were calculated. The F1-score was prioritized because of its balanced assessment of both false negatives (missed buildings) and false positives (incorrectly identified non-buildings). Metrics were averaged (Avg) and standard deviations (Std) were computed from seven randomized splits of the dataset into 80% training and 20% testing subsets to evaluate model reliability.

Figure 4 illustrates a sample of building segmentation results derived from an image within the test set used for the YOLO11 model's fine-tuning. Figure 5 depicts a similar building segmentation outcome using an image sourced from the VPAIR dataset [21].

Figure 6 shows the training and validation loss curves associated with the YOLO11 model, specifically fine-tuned using the optimal split of training and testing datasets. The YOLO model architecture incorporates multiple loss functions combined using a weighted summation approach. The following four distinct loss functions are employed for tasks related to image segmentation:

1) box_loss focuses primarily on the precise positioning of bounding boxes surrounding detected objects (assigned weight: 7.5).

2) seg_loss – targets the accurate delineation of segmentation masks around the identified objects (assigned weight: 7.5);

3) cls_loss – emphasizes the correct categorization of detected objects (assigned weight: 0.5);

4) dfl_loss – emphasizes differentiation between visually similar or challenging-to-distinguish objects by highlighting unique characteristics (assigned weight: 1.5).

The presented graphs confirm the capacity of the model to effectively learn and generalize underlying features from the dataset, as evidenced by a steady decline in the training loss and its stabilization when evaluated on the validation dataset.

Fig. 3. An example of segmented buildings from one image of the training dataset

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

Fig. 4. An example of building segmentation on an image from the test dataset used for fine-tuning YOLO11

Fig. 5. An example of building segmentation on an image from the VPAIR dataset [21]

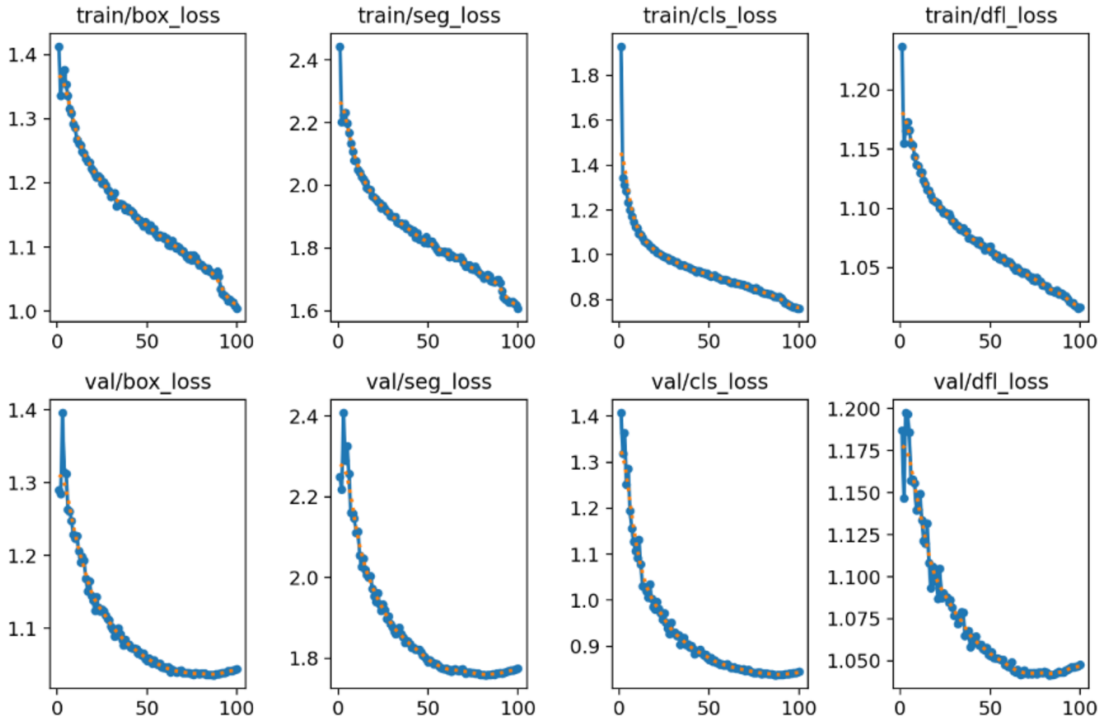

Fig. 6. Loss function plots for the training (train) and validation (val) sets

Figure 7 provides the confusion matrix, excluding true-negative metrics since defining such metrics is ambiguous within segmentation contexts.

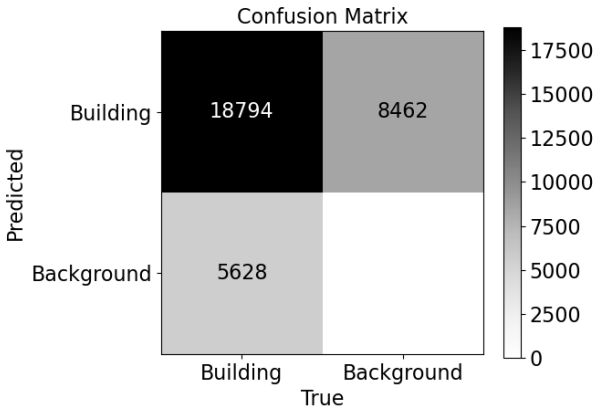

Fig. 7. The confusion matrix for the YOLO11 fine-tuned model on the dataset of segmented buildings

Table 1 summarizes the evaluation metrics for the YOLO11 model, specifically fine-tuned on the building segmentation dataset. This summary includes metrics, such as mean Average Precision (mAP), Precision, Recall, and F1-score, computed over seven unique splits of training and testing data subsets. Additionally, average (Avg) values and standard deviations (Std) are provided to provide insights into the performance stability of the model.

Figure 8 presents the Precision-Recall curve for the top-performing split of training and testing data, achieving an area under the curve (AUC [37]) of 0.76. Such a curve is particularly significant for segmentation tasks because it underscores the model's ability to accurately identify positive instances (buildings) of notable class imbalance (buildings versus background).

For building segmentation tasks, especially those involving partially obscured structures or buildings with intricate outlines, achieving an F1-score of 0.722 on the test dataset reflects a favorable result, notably considering the YOLO models' inherent advantage of real-time performance. This suggests that the fine-tuned model reliably recognizes buildings, a critical capability for producing vector data necessary for UAV-based global positioning. Additionally, the standard deviation values for all evaluated metrics are consistently under 0.5% across both the training and testing subsets, highlighting the model's robustness and uniformity. Future research directions might include modifications to the structure of the YOLO11 neural network and fine-tuning hyperparameters to further enhance the F1-score, particularly for building segmentation scenarios.

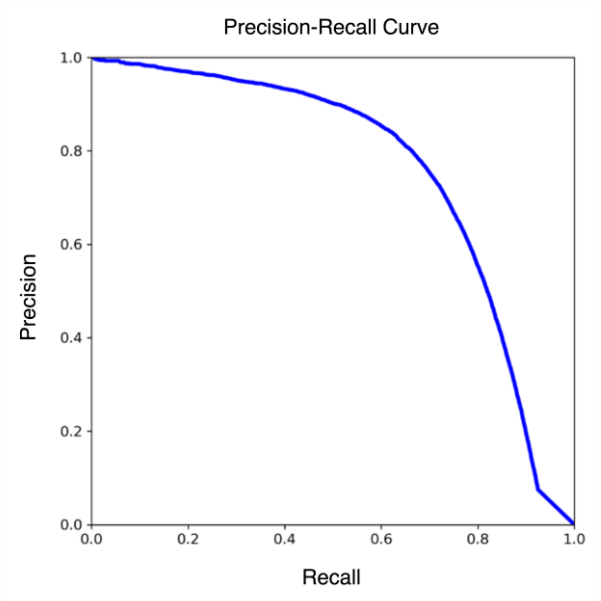

Fig. 8. Precision-Recall curve for YOLO11 fine-tuned model on the dataset of segmented buildings

Table 1

Evaluation metrics obtained from the fine-tuned YOLO11 model on the segmented building dataset

Metric		Random splitting							Avg	Std
		1	2	3	4	5	6	7		
mAP	Train	0.813	0.821	0.820	0.823	0.813	0.817	0.814	0.817	0.0043
	Test	0.748	0.760	0.757	0.755	0.757	0.753	0.749	**0.754**	0.0043
Recall	Train	0.727	0.737	0.736	0.738	0.728	0.730	0.728	0.732	0.0047
	Test	0.677	0.685	0.681	0.679	0.681	0.680	0.673	**0.680**	0.0037
Precision	Train	0.809	0.819	0.815	0.820	0.809	0.813	0.813	0.814	0.0042
	Test	0.764	0.773	0.775	0.768	0.778	0.770	0.769	**0.771**	0.0047
F1	Train	0.766	0.775	0.773	0.777	0.766	0.769	0.768	0.771	0.0044
	Test	0.718	0.727	0.725	0.721	0.726	0.722	0.718	**0.722**	0.0037

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

3.2. Visual place recognition with image preprocessing method

The VPAIR dataset [21] was utilized to validate the method designed to align statistical distributions between UAV and satellite images. Specifically created for UAV localization tasks under challenging scenarios, this dataset was compiled during a 107-kilometer flight stretching from Bonn into the mountainous Eifel area in Germany. The data collection, performed on October 13, 2020, covered various terrains, such as urban zones, agricultural fields, and forested regions. Images were captured using a single-lens color camera with a resolution of 1600x1200 pixels, subsequently downsized to 800x600 pixels for dataset inclusion, alongside highly accurate GNSS/INS positioning (rotational error: 0.05°, positional accuracy: <1 meter). The complete dataset contains 2706 query images from the flight, an equal number of matching satellite images, and an additional 10,000 distractor images from a separate region near Düsseldorf.

Because the terrain-type annotations from the original authors were not publicly available, a new classification was developed into four distinct terrain categories: urban (primarily buildings and roads), field, forest, and water.

The statistical distribution alignment method between UAV and satellite images for UAV localization was validated using the VPAIR dataset [21], employing the Recall@1 metric with a localization radius of 3. Metrics were assessed across varied terrain types, including urban, field, forest, and water. This study prioritized urban terrain, reflecting the specialized fine-tuning of the YOLO11 model on segmented buildings for generating feature vectors.

Three comparative experiments were executed between the proposed YOLO11-based method and established UAV localization approaches (CosPlace [23]): one using original, unprocessed VPAIR dataset [21] images, another utilizing grayscale image conversions, and a third applying the proposed preprocessing method.

Figure 9 demonstrates the visual outcomes obtained using the introduced averaged cumulative distribution function (CDF) approach on UAV imagery, matching their statistical properties with corresponding satellite images.

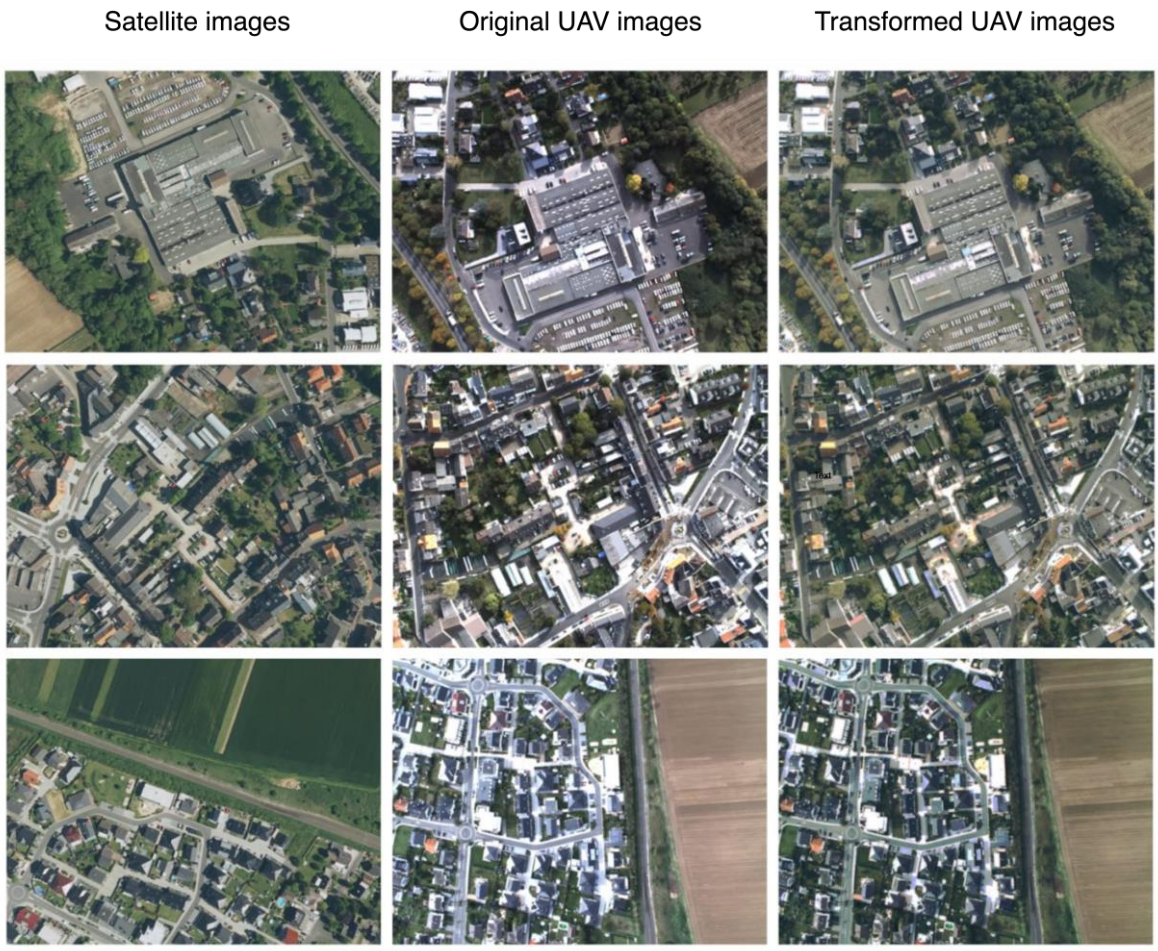

Fig. 9. Visual results of the proposed averaged cumulative distribution function (CDF) method applied to UAV images, aligning their statistical distributions with those of the satellite images. From left to right: satellite images – original UAV images – transformed UAV images

Table 2 provides a comprehensive snapshot of how each stage of the evaluation pipeline influences the final localization accuracy. For every UAV query image in the VPAIR dataset [21], the following sequence was executed three times, once per color-handling variant, before recording Recall@1 with a localization radius of 3. A step-by-step methodology to achieve these results is provided below.

1. **Color handling** – (a) leave the RGB image unchanged, (b) convert it to single-channel grayscale, or (c) apply the proposed averaged cumulative distribution-function (CDF) transfer that aligns UAV pixel statistics to the satellite domain.

2. **Embedding extraction** – feed the pre-processed image to either CosPlace [23] or to our fine-tuned YOLO11 model.

3. **Nearest-neighbour retrieval** – perform an L2 search against the 2706-image satellite reference set; the closest match is taken as the predicted place.

4. **Localization test** – declare a hit if the matched satellite tile lies within three reference frames of the ground-truth tile; otherwise record a miss.

5. **Metric aggregation** – accumulate hits over all 2706 queries separated into four terrain labels (Urban, Field, Forest, Water) and report the proportion of hits as Recall@1 for that terrain.

The findings reveal enhanced performance for CNN-based localization techniques (CosPlace [23], YOLO) when the proposed preprocessing strategy is employed, validating its effectiveness in improving global localization accuracy for UAVs. While CosPlace [23] shows superior performance for terrains such as fields, forests, and water, the proposed YOLO-based technique excels particularly in urban areas and targeted terrains. This performance gain arises primarily from the YOLO11 model being fine-tuned using segmented building data.

Given the significant challenges typically encountered in UAV global localization tasks, where sustaining high accuracy at top ranks remains difficult, achieving a Recall@1 score of 0.195 (19.5%) within a localization radius of 3 for urban settings is highly encouraging. This underscores the YOLO11 approach's competitiveness and robustness, even when faced with input data variability. Additionally, the performance of this approach surpasses established methods, such as CosPlace [23], highlighting its current effectiveness within CNN-based localization research.

However, the proposed UAV global localization method has several limitations, including its limited applicability to urban regions during daytime and favorable weather conditions and its dependence on a predetermined collection of satellite images along the anticipated UAV flight paths.

4. Conclusions

The main goal of this study was effectively accomplished: the development of a resilient UAV–satellite image matching method that leverages deep embeddings and color normalization to strengthen precision and robustness under challenging cross-view and urban conditions. Fine-tuning the YOLO11 model with a dataset containing segmented building data generated vector representations that considerably boosted the accuracy of matching UAV images to satellite images.

The proposed preprocessing technique, which focuses on synchronizing the statistical distributions between satellite imagery and images captured by UAVs, demonstrated notable benefits. It elevated visual consistency vital for accurate localization, surpassing established approaches like CosPlace [23], especially within urban and specific terrain contexts. The quantitative results include achieving an F1-score of 0.722, indicating robust and reliable performance in building segmentation, which is critical for generating effective vector representations. Moreover, the obtained Recall@1 metric of 19.5% within a localization radius of 3 significantly exceeds existing urban terrain benchmarks, underscoring the suggested approach's enhanced robustness and competitive edge.

Table 2

Recall@1 metric results with a localization radius of 3 across different terrain types from the VPAIR dataset [21]

Method	Urban	Field	Forest	Water
(a) Without color preprocessing				
CosPlace [23]	0.140	**0.097**	**0.122**	**0.364**
YOLO (Our)	**0.181**	0.049	0.055	0.345
(b) Using grayscale preprocessing				
CosPlace [23]	0.132	**0.093**	**0.114**	**0.366**
YOLO (Our)	**0.175**	0.030	0.038	0.255
(c) Using proposed averaged cumulative distribution function				
CosPlace [23]	0.145	**0.108**	**0.134**	0.374
YOLO (Our)	**0.195**	0.068	0.070	**0.545**

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

The key advantages of this research include the capability for swift and precise UAV localization in unreliable GPS signals, thus addressing significant weaknesses of existing localization frameworks. Combining YOLO11's inherent real-time processing efficiency with advanced vector-matching strategies provides a viable and efficient solution, which is particularly beneficial for practical applications involving urgent operations, such as emergency rescue, urban surveillance, and infrastructure evaluation.

However, despite these advancements, the proposed method's current application scope remains predominantly effective in urban settings under ideal conditions, namely, during daytime and favorable weather conditions. Its performance also depends heavily on the availability and quality of an existing database of satellite imagery aligned with possible UAV operational routes. Furthermore, to minimize the effects of seasonal variations on visual place recognition, the reference imagery should be captured in the summer, late spring, or early autumn, when the environmental appearance is most stable and consistent. Additionally, the imagery used should closely resemble those from the VPAIR dataset [21], with a minimum spatial resolution of 640×640 pixels, to ensure sufficient detail and comparability for robust feature extraction and matching. Moreover, the difference in resolution between UAV and satellite images is irrelevant because all images are converted to a resolution of 640×640 pixels when fed into YOLO.

Future research should aim to enhance the versatility and precision of the YOLO11 model for more extensive segmentation and localization tasks. Extensive image augmentation strategies, architectural improvements, and meticulous hyperparameter tuning can achieve this. Additionally, extending the localization capabilities to cover varied landscapes such as forests, agricultural fields, and aquatic regions would substantially increase the adaptability and practical utility of the proposed method, thereby expanding its applicability across a wider range of UAV mission scenarios.

Contributions of authors: conceptualization, methodology – **Volodymyr Vozniak, Oleksander Barmak, Iurii Krak;** formulation of tasks, analysis – **Volodymyr Vozniak, Oleksander Barmak, Iurii Krak;** development of model, software, verification – **Volodymyr Vozniak;** analysis of results, visualization – **Volodymyr Vozniak;** writing – original draft preparation, writing – review and editing – **Volodymyr Vozniak, Oleksander Barmak.**

Conflict of Interest

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

Financing

This study was conducted without any financial support.

Data Availability

Data will be made available upon reasonable request.

Use of Artificial Intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, as described in the research methodology section.

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

References

1. Wang, Y., Feng, X., Li, F., Xian, Q., Jia, Z.-H., Du, Z., & Liu, C. Lightweight visual localization algorithm for UAVs. *Scientific Reports*, 2025, vol. 15, no. 1, article no. 6069. DOI: 10.1038/s41598-025-88089-y.

2. Cui, Z., Zhou, P., Wang, X., Zhang, Z., Li, Y., Li, H., & Zhang. Y. A Novel Geo-Localization Method for UAV and Satellite Images Using Cross-View Consistent Attention. *Remote Sensing*, 2023, vol. 15, no. 19, article no. 4667. DOI: 10.3390/rs15194667.

3. Yao, Y., Sun, C., Wang, T., Yang, J., & Zheng, E. UAV Geo-Localization Dataset and Method Based on Cross-View Matching. *Sensors*, 2024, vol. 24, no. 21, article no. 6905. DOI: 10.3390/s24216905.

4. Fesenko, H. V., & Kharchenko, V. S. Vyznachennya optymal′noho marshrutu obl′otu zadanykh tochok terytoriyi potentsiyno nebezpechnoho ob″yektu flotom BPLA [Determination of an optimal route for flight over of specified points of a potentially dangerous object territory by UAV fleet]. *Radioelectronic and Computer Systems*, 2019, no. 3, pp. 63-72. DOI: 10.32620/reks.2019.3.07. (In Ukrainian).

5. Fan, J., Zheng, E., He, Y., & Yang, J. A Cross-View Geo-Localization Algorithm Using UAV Image and Satellite Image. *Sensors*, 2024, vol. 24, no. 12, article no. 3719. DOI: 10.3390/s24123719.

6. Tsekhmystro, R., Rubel, O., Prysiazhniuk, O., & Lukin, V. Impact of distortions in UAV images on quality and accuracy of object localization. *Radioelectronic and Computer Systems*, 2024, vol. 2024, no. 4, pp. 59-67. DOI: 10.32620/reks.2024.4.05.

7. Karapet, B., Savitskyi, R., & Vakaliuk, T. Method of comparing and transforming images obtained

using UAV. *Radioelectronic and Computer Systems*, 2024, vol. 2024, no. 1, pp. 99-115. DOI: 10.32620/reks.2024.1.09.

8. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

9. Lowry, S., Sunderhauf, N., Newman, P., Leonard, J. J., Cox, D., Corke, P., & Milford, M. Visual Place Recognition: A Survey. *IEEE Transactions on Robotics*, 2016, vol. 32, no. 1, pp. 1–19. DOI: 10.1109/TRO.2015.2496823.

10. Cummins, M., & Newman, P. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 2008, vol. 27, no. 6, pp. 647–665. DOI: 10.1177/0278364908090961.

11. Galvez-López, D., & Tardos, J. D. Bags of Binary Words for Fast Place Recognition in Image Sequences. *IEEE Transactions on Robotics*, 2012, vol. 28, no. 5, pp. 1188–1197. DOI: 10.1109/TRO.2012.2197158.

12. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., & Sivic, J. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, vol. 40, iss. 6, pp. 1437-1451 DOI: 10.1109/TPAMI.2017.2711011.

13. Hausler, S., Garg, S., Xu, M., Milford, M., & Fischer, T. Patch-NetVLAD: Multi-Scale Fusion of Locally-Global Descriptors for Place Recognition. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 14136-14147. DOI: 10.1109/CVPR46437.2021.01392.

14. Workman, S., Souvenir, R., & Jacobs, N. Wide-Area Image Geolocalization With Aerial Reference Imagery. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 3961-3969. DOI: 10.1109/ICCV.2015.451.

15. Lin, T.-Y., Cui, Y., Belongie, S., & Hays, J. Learning Deep Representations for Ground-to-Aerial Geolocalization. *015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 5007-5015. DOI: 10.1109/CVPR.2015.7299135.

16. Zheng, Z., Wei, Y., & Yang, Y. University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization. *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1395–1403. DOI: 10.1145/3394171.3413896.

17. Zhu, S., Yang, T., & Chen, C. VIGOR: Cross-View Image Geo-Localization Beyond One-to-One Retrieval. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 5316-5325. DOI: 10.1109/CVPR46437.2021.00364.

18. Zhu, R., Yin, L., Yang, M., Wu, F., Yang, Y., & Hu, W. SUES-200: A Multi-Height Multi-Scene Cross-View Image Benchmark Across Drone and Satellite. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, vol. 33, no. 9, pp. 4825–4839. DOI: 10.1109/TCSVT.2023.3249204.

19. Cisneros, I., Yin, P., Zhang, J., Choset, H., & Scherer, S. ALTO: A Large-Scale Dataset for UAV Visual Place Recognition and Localization. *arXiv:2207.12317*. DOI: 10.48550/arXiv.2207.12317.

20. Chen, J., Wen, G., Jian, H., & Fan, X. A Visual Localization Benchmark for UAVs in Complex Multi-Terrain Environments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, pp. 1–15. DOI: 10.1109/JSTARS.2025.3526695.

21. Schleiss, M., Rouatbi, F., & Cremers, D. VPAIR -- Aerial Visual Place Recognition and Localization in Large-scale Outdoor Environments. *arXiv.2205.11567*. DOI: 10.48550/arXiv.2205.11567.

22. Komorowski, J., Wysoczańska, M., & Trzcinski, T. MinkLoc++: Lidar and Monocular Image Fusion for Place Recognition. *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 2021, pp. 1-8. DOI: 10.1109/IJCNN52387.2021.9533373.

23. Berton, G., Masone, C., & Caputo, B. Re-thinking Visual Geo-Localization for Large-Scale Applications. *arXiv.2204.02287*, pp. 1-15. DOI: 10.48550/arXiv.2204.02287.

24. Ali-Bey, A., Chaib-Draa, B., & Giguère, P. MixVPR: Feature Mixing for Visual Place Recognition. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2023, pp. 2997-3006. DOI: 10.1109/WACV56688.2023.00301.

25. Ali-bey, A., Chaib-draa, B., & Giguère, P. GSV-Cities: Toward appropriate supervised visual place recognition. *Neurocomputing*, 2022, vol. 513, pp. 194–203. DOI: 10.1016/j.neucom.2022.09.127.

26. Zaffar, M., Ehsan, S., Momeni, L., & et al. VPR-Bench: An Open-Source Visual Place Recognition Evaluation Framework with Quantifiable Viewpoint and Appearance Change. *International Journal of Computer Vision*, 2021, vol. 129, pp. 2136–2174. DOI: 10.1007/s11263-021-01469-5.

27. Keetha, N., Mishra, A., Karhade, J., & et al. AnyLoc: Towards Universal Visual Place Recognition. *IEEE Robotics and Automation Letters*, 2024, vol. 9, no. 2, pp. 1286-1293. DOI: 10.1109/LRA.2023.3343602.

28. Radford, A., Kim, J. W., Hallacy, C., & et al. Learning Transferable Visual Models From Natural Language Supervision. *arXiv.2103.00020*, 2021, pp. 1-48. DOI: 10.48550/arXiv.2103.00020.

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

29. Oquab, M., Darcet, T., Moutakanni, T., & et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv.2304.07193*, 2024. DOI: 10.48550/arXiv.2304.07193.

30. Berton, G., Trivigno, G., Caputo, B., & Masone, C. EigenPlaces: Training Viewpoint Robust Models for Visual Place Recognition. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 2023, pp. 11046-11056. DOI: 10.1109/ICCV51070.2023.01017.

31. Shao, J., & Jiang, L. Style Alignment-Based Dynamic Observation Method for UAV-View Geo-Localization. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, pp. 1–14, article no. 3000914. DOI: 10.1109/TGRS.2023.3337383.

32. Gallo, I., Rehman, A. U., Dehkordi, R. H., Landro, N., La Grassa, R., & Boschetti, M. Deep Object Detection of Crop Weeds: Performance of YOLOv7 on a Real Case Dataset from UAV Images. *Remote Sensing*, 2023, vol. 15, no. 2, article no. 539. DOI: 10.3390/rs15020539.

33. Wu, T., & Dong, Y. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. *Applied Sciences*, 2023, vol. 13, no. 24, article no. 12977. DOI: 10.3390/app132412977.

34. Rainio, O., Teuho, J., & Klén, R. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 2024, vol. 14, article no. 6086. DOI: 10.1038/s41598-024-56706-x.

35. *Ultralytics. YOLO11*. Available at: https://docs.ultralytics.com/models/yolo11 (accessed 04.05.2025).

36. *Buildings Instance Segmentation - v1 raw-images*. Roboflow. Available at: https://universe.roboflow.com/roboflow-universe-projects/buildings-instance-segmentation/dataset/1 (accessed 04.05.2025).

37. Jocher, G., Qiu, J., & Chaurasia, A. *Ultralytics YOLO*. Python, Jan. 2023. Available at: https://github.com/ultralytics/ultralytics (accessed 04.05.2025).

38. Çorbacıoğlu, Ş. K., & Aksel, G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 2023, vol. 23, no. 4, pp. 195-198. DOI: 10.4103/tjem.tjem_182_23.

Received 16.05.2025, Accepted 25.08.2025

МЕТОД ДЛЯ СПІВСТАВЛЕННЯ СУПУТНИКОВИХ І БПЛА-ЗОБРАЖЕНЬ ДЛЯ ВІЗУАЛЬНОГО РОЗПІЗНАВАННЯ МІСЦЕВОСТІ ІЗ ВИКОРИСТАННЯМ МІЖРАКУРСНОЇ КОЛЬОРОВОЇ НОРМАЛІЗАЦІЇ

В. З. Возняк, О. В. Бармак, Ю. В. Крак

Предметом дослідження є візуальне розпізнавання місцевості (Visual Place Recognition, VPR), а саме зіставлення супутникових зображень із зображеннями, отриманими з безпілотних літальних апаратів (БПЛА). VPR є важливим для автономної навігації БПЛА, особливо в умовах, коли сигнали глобальних навігаційних супутникових систем (GNSS) ненадійні, наприклад, в умовах міської забудови чи територій із щільною інфраструктурною забудовою. Незважаючи на практичну значимість, точне зіставлення зображень БПЛА із супутниковими залишається складною задачею через значні відмінності в ракурсі, масштабі, освітленні та текстурі. Традиційні підходи на основі ручних дескрипторів або класичних локальних ознак часто неефективні в умовах таких міжвидових відмінностей. **Метою** цього дослідження є розробка надійного методу візуального розпізнавання місць для співставлення зображень, отриманих БПЛА та супутниками, із використанням вбудовувань на основі глибинного навчання та розширених методів нормалізації кольорів для підвищення надійності в умовах міжракурсних відмінностей. **Завдання**: по-перше, розробити метод генерації глобальних векторних представлень на базі YOLO, який використовує можливості багатомасштабної екстракції ознак для кодування семантично значущих орієнтирів місцевості; по-друге, створити та впровадити нову техніку передобробки зображень на основі вирівнювання статистичних розподілів кольору між зображеннями БПЛА та супутниковими зображеннями; по-третє, інтегрувати ці компоненти у завершену систему VPR та оцінити її ефективність на складному наборі даних VPAIR з акцентом на міські території. **Методи**, що використовуються, включають глибоке навчання, зокрема налаштування нейромережі YOLO11 на наборі даних, спеціально анотованому для сегментації будівель. Додатково застосовуються статистичні методи вирівнювання на основі кумулятивних функцій розподілу (CDF) для стандартизації вигляду зображень двох різних доменів. **Висновки.** Проведені експерименти показали значне покращення ефективності зіставлення зображень БПЛА із супутниковими завдяки запропонованому методу. Налаштування YOLO11 спеціально для сегментації будівель забезпечило створення надійних векторних представлень, що дозволило досягти високої точності сег-

ментації (F1-score – 0,722). Метод кольорової передобробки додатково покращив точність розпізнавання: значення Recall@1 досягло 19,5% для міських територій при радіусі локалізації 3, значно перевищуючи показники традиційних підходів. Дане дослідження пропонує ефективне рішення для задач локалізації БПЛА, зокрема в складних міських умовах, підкреслюючи важливість комплексного підходу до генерації векторних представлень та передобробки зображень.

Ключові слова: візуальне розпізнавання місцевості; БПЛА; YOLO; передобробка зображень; глибоке навчання; сегментація зображень.

Возняк Володимир Зіновійович – асп. каф. комп'ютерних наук, Хмельницький національний університет, Хмельницький, Україна.

Бармак Олександр Володимирович – д-р техн. наук, проф., зав. каф. комп'ютерних наук, Хмельницький національний університет, Хмельницький, Україна.

Крак Юрій Васильович – д-р фіз.-мат. наук, проф., зав. каф. теоретичної кібернетики, Київський національний університет імені Тараса Шевченка, Інститут кібернетики імені В. М. Глушкова, Київ, Україна.

Volodymyr Vozniak – PhD Student of the Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine,
e-mail: vozniakvz@khmnu.edu.ua, ORCID: 0009-0008-3055-5257.

Oleksander Barmak – Doctor of Technical Sciences, Professor, Head of the Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine,
e-mail: barmako@khmnu.edu.ua, ORCID: 0000-0003-0739-9678, Scopus Author ID: 57217176350.

Iurii Krak – Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Theoretical Cybernetics, Taras Shevchenko National University of Kyiv, Glushkov Cybernetics Institute, Kyiv, Ukraine,
e-mail: iurii.krak@knu.ua, ORCID: 0000-0002-8043-0785, Scopus Author ID: 6602577533.