IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 February 2025, accepted 12 April 2025, date of publication 25 April 2025, date of current version 13 May 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3564496*

RESEARCH ARTICLE

# Improving Image Embeddings With Colour Features in Indoor Scene Geolocation

**OPEYEMI BAMIGBADE[1], MARK SCANLON[2], (Senior Member, IEEE), AND JOHN SHEPPARD[1], (Member, IEEE)**

[1]Department of Computing and Mathematics, South East Technological University, Waterford, X91 HE36 Ireland
[2]School of Computer Science, University College Dublin, Dublin, D04 V1W8 Ireland

Corresponding author: Mark Scanlon (mark.scanlon@ucd.ie)

This work was supported by the SETU Ph.D. Scholarship Programme.

**ABSTRACT** Embeddings remain the best way to represent image features, but do not always capture all latent information. This is still a problem in representation learning, and computer vision descriptors struggle with precision and accuracy. Improving image embedding with other features is necessary for tasks like image geolocation, especially for indoor scenes where descriptive cues can have less distinctive characteristics. This work proposes a model architecture that integrates image N-dominant colours and colour histogram vectors in different colour spaces with image embedding from deep metric learning and classification perspectives. The results indicate that the integration of colour features improves image embedding, surpassing the performance of using embedding alone. In addition, the classification approach yields higher accuracy compared to deep metric learning methods. Interestingly, different saturation points were observed for image colour-improved embedding features in models and colour spaces. These findings have implications for the design of more robust image geolocation systems, particularly in indoor environments.

**INDEX TERMS** Classification, color descriptor, deep metric learning, embeddings, image geolocation, image retrieval, indoor scenes.

## I. INTRODUCTION

Image geolocation is the act of determining the exact location or narrowing down the possible search space to the location where an image was taken. This task is crucial for various applications, such as urban planning, tourism, autonomous driving, and investigation of criminal acts such as human trafficking and child sexual exploitation [1], [2], [3], [4]. Although this task may seem straightforward with the presence of geolocation cues and metadata, the difficulty increases as images are often stripped of this location information intentionally for illegal activities. In other instances, the information is lost through instant messaging/social media platforms [5]. In addition, enhancement techniques aimed at improving image quality sometimes inadvertently remove or distort geolocation data embedded in the image.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

At a broad level, the scene of the image can be indoors or outdoor. Indoor scenes are often contained within the boundaries of a building, such as rooms and corridors, while outdoor scenes include open areas, roads, and landscapes. With the presence of distinct cues and public references in outdoor scenes, there is a better chance of geolocation compared to indoor scenes, where GPS signals and location cues are often not reliable or readily available [6], [7]. This constraint to the low level and latent information in indoor scenes and images necessitates the need for fine-grained, descriptive cue generation and enhanced feature representation for effective geolocation [8].

One of the best ways to effectively capture both low-level and high-level visual features of a scene is the use of embedding techniques. Embedding is a powerful computer vision approach that allows high-dimensional data, such as images, to be transformed into compact and meaningful representations in a lower-dimensional space, easing tasks such as similarity matching, clustering, and classification

that can be used to facilitate more accurate geolocation [6]. With recent and continuous advances in deep learning, embedding techniques have been extensively applied to image geolocation modelling.

Embedding can capture a great amount of latent features, including the texture and colour information of an image. However, the proportion of information learnt by an image embedding model is more tilted towards texture [9]. Indoor environments, such as hotel rooms, often lack the distinctive structural landmarks present in outdoor scenes, making visual geolocation inherently more difficult. These environments tend to share similar layouts, furniture, and lighting conditions in different locations, which can limit the discriminative power of standard image embeddings. However, colour schemes such as wall tones, bedding, and decor often vary significantly between hotels and serve as subtle yet powerful cues for differentiation. Enhancing image embeddings with colour features helps capture these subtle variations, thereby improving the model's ability to distinguish between visually similar indoor scenes. For example, in hotel room identification, leveraging dominant colour patterns can significantly boost the accuracy of matching a query image to its correct location within a large database of rooms. This work proposes a model architecture that integrates image $N$-dominant colour and colour histogram vectors with image embedding techniques using deep metric learning (DML) and classification methods to narrow down the search space in indoor scene geolocation. It shows experimentally using two major colour spaces that this approach achieves competitive performance on Hotels-50K; an indoor scene dataset provided to combat human trafficking.

In summary, the main contributions are the following:

- A proposed model architecture that generates and integrates the colour-improved image embedding features to enhance the geolocation of indoor scenes. The code is available open source at https://github.com/OBA-Research/colourNembedding.
- Through experiments, this work observed and reported on the saturation point of image embedding and colour feature fusion.
- The work observed and reported on the model sensitivity to the colour-improved features through accuracy convergence rate and loss decay trends for classification and deep-metric learning, respectively.
- For the task of combating human trafficking, this work shows that the proposed method reduces the geolocation search space with better chances of retrieving hotels of interest.

## II. RELATED WORKS

### A. IMAGE GEOLOCATION

This has been a subject of extensive research, encompassing both outdoor and indoor scene geolocation. Various methodologies and techniques have been explored to accurately determine the geographic location of images, some leveraging the power of computer vision algorithms and spatial analysis [10], [11]. [12] introduced `PlaNet model`, which employed a Convolutional Neural Network (CNN) to predict the location of photos based on visual content by subdividing the Earth's surface into geographic cells. Reference [13] approached the problem by matching a ground view query image to a reference database of aerial/satellite images. Although these cover a large geographical area, their application to a specific domain or enclosed regions is highly limited. The indoor scene problem is more complicated as a result of increased variations in light, shape, layout, and severe occlusions [8]. Due to these, only a few research works focus on indoor image and scene geolocation [2], [14] with an increasing trend in hotel room identification to combat human trafficking [15], [16], [17], [18]. Recently, [18] proposed an object-centric approach to hotel recognition, which involves ensembling object features extracted from the image. [2] incorporated contextual information at different spatial resolutions, as well as more specific features, in the CNN learning process. These approaches suggest that breakthroughs in indoor scene geolocation using computer vision techniques are highly dependent on image feature engineering and representation.

In computer vision, the problem of image geolocation is often seen as a classification or image retrieval problem. State-of-the-art methods treat the task of geolocation as a classification problem [2], [12], [19], [20] such as [12] and [21] subdividing the world map into a number of classes for deep classification. With the classification approach, the model is often trained using softmax loss, and the embeddings produced by the penultimate layer of the model can be used to facilitate geolocation through similar image retrieval [15]. Although higher accuracy, fine-grained geolocation, efficient processing, and interoperability are often seen with this approach, it can be rigid in its adaptability to unseen classes [22], [23].

Image geolocation has also been addressed using DML in large-scale datasets via embedding losses such as contrastive loss, triplet loss, and ArcFace loss. A model is trained to explicitly learn the embedding of data in a latent space, where the similarity is maximised for semantically similar data points and minimised otherwise [13], [15], [24], [25], [26]. The main advantages of this model are flexibility in querying, scalability, and adaptability, but it can be computationally expensive [27], [28], [29].

### B. IMAGE FEATURE REPRESENTATION

Texture, colour distribution, and shape properties in an image can be used as distinguishing features for image classification, matching, and retrieval, allowing the correlation of similar images with known geolocations [6], [30], [31]. These image properties have been individually explored for effective geolocation, with good results in different scenarios and problem domains. To accommodate the complexity of scene types, both global and local spatial information needs

to be explored holistically [8]. Early works adopted handmade image feature representation before advances in deep learning [32], [33], [34], [35], recent works leverage CNNs to learn image embeddings which continue to demonstrate a good descriptive representation of image features in computer vision [12], [21]. Reference [10] worked on geolocating panoramic images on a 2-D cartographic map based on learning a low-dimensional embedded space. Reference [36] proposed a model for learning image representations that integrate context-aware feature reweighting to focus on regions that positively contribute to geolocation effectively. Although these works achieve good results, it is believed that the conglomeration of their approaches with other features, such as image colour, will yield better performances.

### C. IMAGE COLOUR DESCRIPTOR

Considerable work has been done in designing and applying efficient colour descriptors, among which are colour histograms and dominant colour. A colour histogram is one of the most commonly used colour descriptors that characterise the colour distribution in an image, while the dominant colour descriptor gives the distribution of the salient colours in the image using algorithms such as clustering and median cut quantization [30], [37], [38], [39]. Reference [37] presented colour difference histograms, which count the perceptually uniform colour difference between two points under different backgrounds with regard to colours and edge orientations for image retrieval. Reference [39] proposed an algorithm that used the peaks in 3D histograms of colours to segment a colour image. Reference [40] extended this algorithm and used the colour histogram for multiband image segmentation.

Dominant colour extraction and analysis is another commonly adopted method in colour-based image retrieval systems. Although this has some shortcomings, especially for object-based image retrieval, this has been widely used in well-defined problems [41], [42]. Reference [43] extracted image dominant colour features based on region growth, while [44] extracted dynamic image dominant colour using the centroid of partitions in the image, and [45] extracted prominent colours from small image regions as dominant colours. All of these works achieved good results, but emphasised integration with other image features that are learnt on a global scale, since dominant colours and colour histograms are local to the image.

In summary, the major research gap in related work lies within the advancement in feature engineering to enrich image embeddings with other features, such as colour patterns, to capture more latent information that could improve the accuracy of geolocation models.

## III. METHODS

### A. PROBLEM DEFINITION

The objective of this work is to improve the representation of image features for the augmentation of discriminative attributes, thus improving performance in classification and information retrieval models, particularly within the domain of indoor scene geolocation. Given a finite number of images $n$, each having latent and inherent features, the objective is to enhance the embedding of the image with the integration of colour features.

Let $\mathbb{X}_i$ be input image, $i = 1,2,\ldots,n$, where $n$ is the total number of images.

Let $\mathbb{E}_i$ represent $\mathbb{X}_i$ embedding of size 128 for the $i^{th}$ image,

Let $\mathbb{C}_i^S$ represent the colour feature vector for $\mathbb{X}_i$ in a given colour space $S$, such that $S \in \{RGB, HSV\}$.

The combined feature representation $\mathbb{F}_i$ can be defined as the concatenation of the image embedding $\mathbb{E}_i$ with the colour feature vector $\mathbb{C}_i{}^S$:

$$\mathbb{F}_i = [\mathbb{E}_i || \mathbb{C}_i^S] \tag{1}$$

*where || denotes concatenation*

Therefore, geolocation model $\mathbb{M}$ can use classification with deep learning or information retrieval with deep metric learning model as a function of $\mathbb{E}_i$ or $\mathbb{F}_i$:

$$\mathbb{M}(\mathbb{E}) or \mathbb{M}(\mathbb{F}) \tag{2}$$

### B. HYPOTHESES

Having the embedding $\mathbb{E}$ with a constant size of 128 and a colour feature vector size varying with $N$ for both colour feature extraction approaches in defined colour spaces, the following hypotheses can be:

- Hypothesis 1: The model performance on colour features improved embedding surpasses the performance on using embedding alone across both classification and deep metric learning.

$$\mathbb{M}(\mathbb{F}) \geq \mathbb{M}(\mathbb{E}) \tag{3}$$

- Hypothesis 2: The proportions of colour features in the fused features influence the convergence rate of models during training, impacting their ability to learn and generalise effectively.
- Hypothesis 3: A saturation point $k$ exists wherein additional colour features, beyond $k$ do not notably improve model performance, indicating an optimal balance between feature enhancement and model effectiveness.

$$\mathbb{M}(\mathbb{F}_{ik}) \approx \mathbb{M}(\mathbb{F}_{ik+1})$$
$$\textit{for all } i \textit{ with } k \textit{ being the saturation point} \tag{4}$$

### C. EMBEDDING

With a pretrained CNN, specifically `EfficientNet_B0`, as the backbone model of the proposed architecture, the semantic and textural features of the image were learnt and the embedding layer (a fully connected layer) was used to learn the embedding vector $\mathbb{E}$ with a constant size of 128, serving as a condensed representation of the features of the image.

`EfficientNet_B0` is the baseline model in the EfficientNet architecture family, characterised by its balance

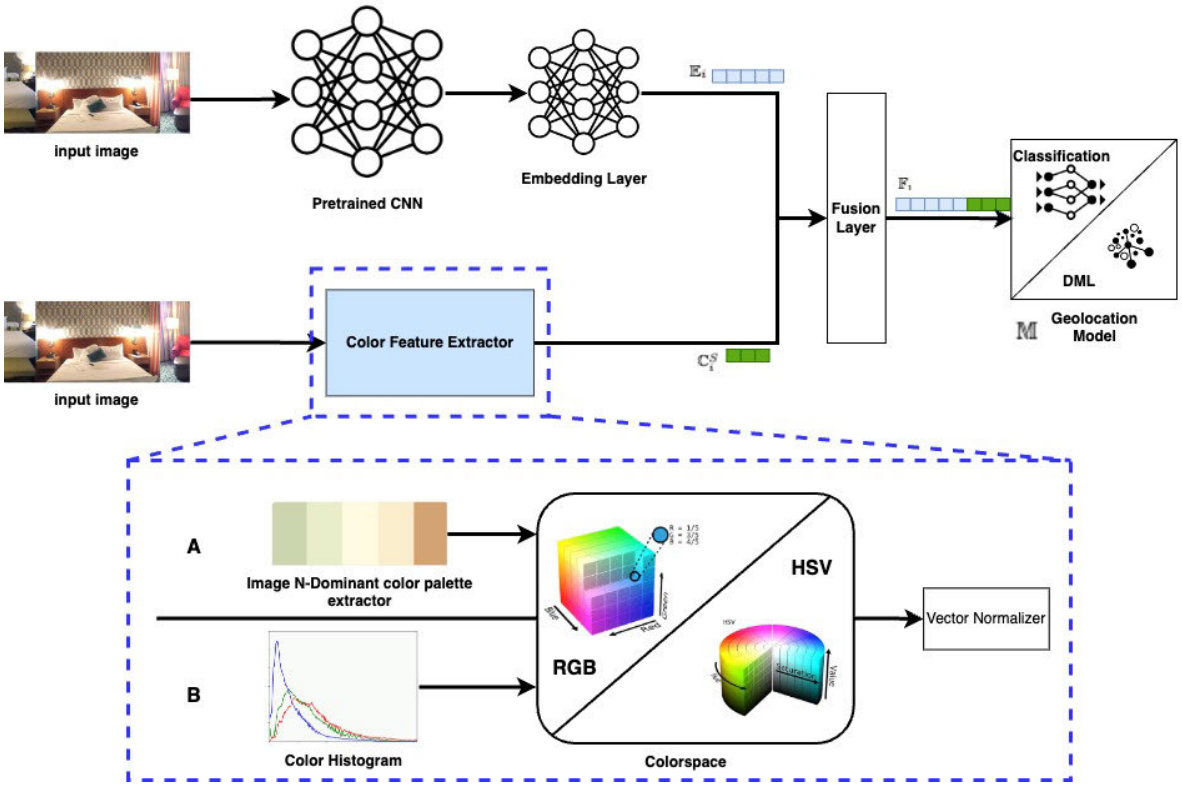


**FIGURE 1.** Illustration of the proposed model architecture to improve image embedding with colour features. The colour feature extractor uses colour palettes and colour histograms to compute a normalised feature vector that is concatenated with the image embedding extracted using Efficient Net as the backbone model.

between model size and performance [46]. It consists of multiple blocks with separable convolutions in-depth, along with efficient scaling techniques such as compound scaling and model architecture search (MNAS). This serves as a sufficient baseline for the proposed architecture to observe the effect of the colour features on the image-embedding model performance.

### D. COLOUR FEATURE EXTRACTION

This involves two key approaches: extracting $N$-dominant colour palettes and computing $N$-bin colour histograms. As seen in Table 1, the value $N$, representing the palette or bin size, is calculated based on the proportion *'P'* of colour features relative to the fused embedding colour features. Mathematically,

$$P = \frac{TotalColourFeatures}{TotalColourFeatures + Embedding} \times 100 \quad (5)$$

$$P = \frac{3N}{3N + \mathbb{E}} \times 100 \quad (6)$$

Therefore:

$$N = \frac{P \times \mathbb{E}}{3(100 - P)} \quad (7)$$

where:

- $N$: Palette or Bin size
- $P$: Proportion of colour features (as a percentage).
- $\mathbb{E}$: Embedding size
- '3': Number of colour channels in RGB or HSV.
- '100': percentage).

Equation 7, derived from the ratio of colour features to the total fused feature size, ensures that a specific proportion $P$ of colour features is incorporated in the experiments.

#### 1) $N$-DOMINANT COLOUR PALETTE

An $N$-dominant colour palette was extracted from each input image. This involved identifying a number of the most prominent colours present in the image, specifically achieved through median-cut colour quantisation. Each extracted colour in the palette was then converted to a specific colour space to obtain the colour values as a vector. The number of dominant colours i.e $N$ and the colour space $S$ were parameterized such that $N \in \{5, 11, 18, 28, 43, 64, 100\}$ and $S \in \{RGB, HSV\}$ to observe the effect of colour feature vector size on the constant embedding size of 128. Subsequently, the outputted colour feature vectors were normalised to ensure consistency and uniformity with image embedding fusion.

#### 2) $N$-BIN COLOUR HISTOGRAM

In parallel with $N$-dominant colour palette extraction, the colour histogram was computed for each input image within the chosen colour spaces. The colour histogram captures the frequency distribution of the colours in the image, divided

into a number of bins ($N$-bin colour histogram) such that $N \in \{5, 11, 18, 28, 43, 64, 100\}$. Following the computation of the colour histogram, the resulting colour histogram vector was normalised to maintain uniformity and ensure consistency in the fusion with the image embedding.

#### 3) COLOUR SPACE AND FEATURES FUSION

Experiments were performed in two commonly employed colour spaces: RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value). Each colour space offers unique advantages and insight into the colour characteristics of the input images. The RGB colour space represents colours as combinations of red, green and blue primary colours as shown in the upper row of Fig. 2. The HSV colour space represents colours based on their perceptual attributes: hue, saturation, and value shown in the lower row of Fig. 2. These two colour spaces were selected because of their complementary advantages: RGB aligns with how images are stored and processed in most computer vision pipelines, while HSV separates chromatic content from intensity, offering perceptual benefits that make it more robust to lighting variations, particularly useful in indoor scene analysis.

Having extracted the image colour features with different vector sizes depending on the value of $N$, the image embedding and the colour vector were fused as the final representation of the input image features. This fusion was performed continuously for every batch of data during model training and evaluation, as the model learns to better represent image embedding for every iteration with the colour features in consideration. Among several ways of achieving this feature fusion, concatenation was used because of its preservation of individual features, flexibility, simplicity, and compatibility.

### E. GEOLOCATION MODELS

As shown in Fig. 1, the final building block of the proposed architecture is the geolocation model. The application of two distinct computational approaches to this task is explored. Firstly, classification with deep learning and, secondly, deep metric learning. Both models are set up to process the fused features, which combine colour and image embedding, to enhance the predictive accuracy and robustness of the geolocation system.

#### 1) CLASSIFICATION LOSS AND OPTIMISATION

The cross-entropy loss 8 was used together with Adam (Adaptive Moment Estimation) optimiser for the classification approach. This loss function allowed measuring the disparity between the predicted probabilities and the actual geolocation labels, guiding the model toward more accurate predictions. To assign an image to one of several geolocation classes, cross-entropy loss is defined as:

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{C=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) \tag{8}$$

where:

- N is the total number of images in the dataset or batch.
- C represents the total number of classes. In this case, unique geolocation categories).
- $y_{i,c}$ is the true label for image i, encoded as a one-hot vector (1 if the image belongs to class $c$, otherwise 0).
- $\hat{y}_{i,c}$ denotes the predicted probability that the image $i$ belongs to class $c$, as produced by the softmax output of the model.

The logarithm penalises incorrect predictions by assigning higher loss values to lower predicted probabilities for the true class. Multiplying by the true label $y_{i,c}$ ensures that only the loss corresponding to the actual class is considered, effectively ignoring other classes in the summation. Averaging of overall samples ($\frac{I}{N}$) provides a mean loss value, facilitating consistent gradient updates during model training [47]

The Adam optimiser, introduced by [48], combines the advantages of two popular optimisers: AdaGrad and RMSProp. It is known for its efficiency in handling sparse gradients and acceleration of convergence. This was used to update the model parameters. Adam calculates adaptive learning rates for each parameter using estimates of the first moment (mean) and the second moment (uncentered variance) of the gradients. The update rule for a parameter $\theta_t$ is expressed as:

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

where:

- $\theta_t$: Parameter at time step $t$ being updated
- $\theta_{t-1}$: Previous parameter value.
- $\eta$: Learning rate
- $\hat{m}_t$: Bias-corrected first moment estimate
- $\hat{v}_t$: Bias-corrected second moment estimate
- $\epsilon$: Small constant to prevent division by zero, ensuring numerical stability.

#### 2) DEEP METRIC LEARNING LOSS AND OPTIMISATION

The Triplet Margin Loss function 10 was used to learn a feature space where distances directly correspond to the similarity of geolocation. This loss function facilitated the optimisation process by penalising the model for incorrect pairwise distance relationships between anchor, positive, and negative samples within triplets. Additionally, to enhance the effectiveness of triplet selection, a combination of Triplet Margin Miner and cosine similarity was used to mine informative triplets during training, creating semihard triplets. This approach ensured that the model learned from the most relevant and informative sample triplets, improving the performance of the capture of subtle distinctions in geolocation similarities.

$$\mathcal{L} = \sum_{i=1}^{N} \max\left(0,\ d(f(x_i^a), f(x_i^p)) - d(f(x_i^a), f(x_i^n)) + \alpha\right) \tag{10}$$

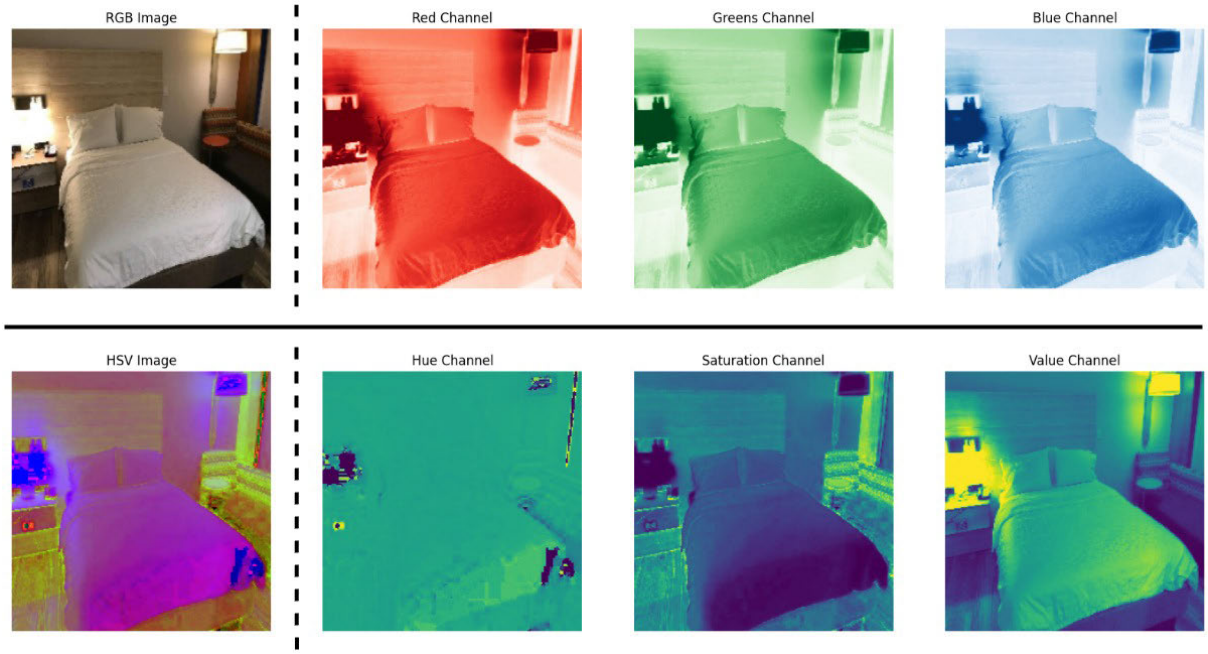


**FIGURE 2.** Example of indoor images in the selected colour spaces. On the left are the images in RGB and HSV colour spaces, on the right are the images in each channel.

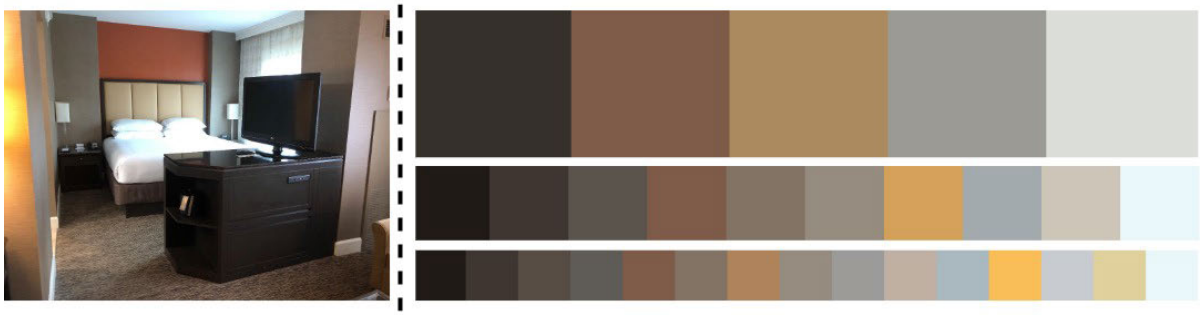

**FIGURE 3.** Example of extracted colour palette with varying palette sizes from Hotels-50K dataset.

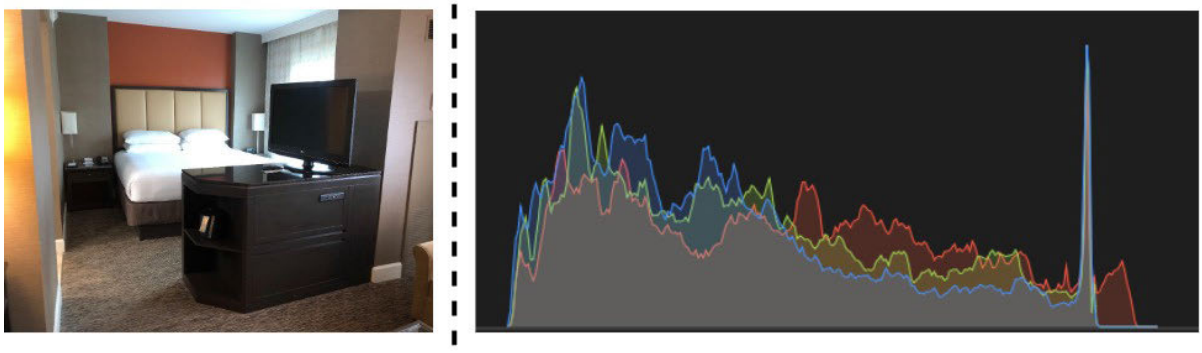

**FIGURE 4.** Example of the computed colour histogram from the Hotels-50K dataset.

where:

- $\mathcal{L}$: The triplet loss.
- $N$: The number of triplets in the batch.
- $x_i^a, x_i^p, x_i^n$: Anchor, positive, and negative samples, respectively.
- $f(x)$: The embedding function.

**TABLE 1.** *N*-Dominant and colour Histogram features extraction and concatenation with embedding. Where *N* is the palette and bin sizes, $\mathbb{C}$ is the resulting colour feature vector size, $\mathbb{E}$ is the embedding size, and $\mathbb{F}$ is the fused features size.

| Colour Space | *N* | $\mathbb{C}$ size | $\mathbb{E}$ size | $\mathbb{F}$ size | Proportion P(%) |
|---|---|---|---|---|---|
| RGB,HSV | 5 | 15 | 128 | 143 | 10 |
| RGB,HSV | 11 | 33 | 128 | 161 | 20 |
| RGB,HSV | 18 | 54 | 128 | 182 | 30 |
| RGB,HSV | 28 | 84 | 128 | 212 | 40 |
| RGB,HSV | 48 | 144 | 128 | 272 | 50 |
| RGB,HSV | 64 | 192 | 128 | 320 | 60 |
| RGB,HSV | 100 | 300 | 128 | 428 | 70 |

- $d(u, v)$: The distance metric (cosine similarity).
- $\alpha$: The margin parameter, which ensures that positive pairs are closer than negative pairs by at least $\alpha$.

### F. EVALUATION METRICS

The primary metric used to assess the performance of the models was the mean average precision at 5 (mAP@5). This metric calculates the average precision of the top five predictions, providing a robust measure of the model's ability to rank true locations highly among the top results. mAP@5 is defined as:

$$mAP@5 = \frac{1}{U}\sum_{u=1}^{U}\sum_{k=1}^{min(t,5)} P(k) \times rel(k) \tag{11}$$

where $U$ is the number of images, $P(k)$ is the precision at cutoff $k$, $t$ is the number of predictions per image, and rel$(k)$ is an indicator function equalling 1 if the item at rank $k$ is a relevant correct label, zero otherwise [49].

Furthermore, Precision at 1 was used, which is the closest equivalent of classification accuracy, to evaluate the immediate relevance of the top prediction. Precision at 1 directly measures the accuracy of the model in identifying the most likely geolocation as the top result, offering insights into the precision of the model with the most confident prediction.

## IV. EXPERIMENTS

### A. DATASET

The experiments used the 2022 Hotels-50K dataset [49] created to combat human trafficking. A validation set containing images of hotels with more than one image was created. To ensure consistency in image dimensions and reduce computational complexity, each image was resized to 256 by 256 pixels. Furthermore, occlusion files associated with each hotel class were excluded to maintain data integrity and minimise noise in colour features. This precautionary measure was designed to prevent noise due to mask colour and ensure the reliability of the colour feature extraction.

### B. EXPERIMENTAL SETUP

In the experimental setup, two key configurations were compared: classification and deep metric learning. These models used embedding alone as a baseline, and the models leveraged concatenated features for enhanced performance. This setup served as a reference point to evaluate the effectiveness of incorporating additional features. In the experimental variation, both the classification and the deep metric learning models used concatenated features, combining the embedding of images with the features of the colours. This configuration aimed to explore the impact of integrating colour information on model performance. To benchmark the work, the approach and configuration used in the original Hotels-50K paper [16] for deep metric learning were reproduced. A neural network was trained using triplet loss with batch hard mining, a miner that samples the hardest positive and hardest negative triplets for each anchor [15]. The classification approach was benchmarked against the pretrained ''EfficientNet B4''. Both models and configurations were trained and evaluated with the Hotels-50K dataset.

#### 1) PARAMETERS AND CONFIGURATION

For each of the colour feature extraction methods with $N \in \{5, 11, 18, 28, 43, 64, 100\}$, colour features were extracted in the selected colour spaces, as shown in Table 1. $\mathbb{C}$ and $\mathbb{E}$ sizes are the colour vector and embedding sizes, respectively, with $\mathbb{F}$ being the concatenated features size. The proportion is a measure of the colour vector size compared to the embedding size in the concatenated features. With a batch size of 32, models were trained using 20 epochs with an early stopping set to 5 and a learning rate of 0.001.

### C. IMPLEMENTATION DETAILS

Pylette [50], a colour palette extractor written in Python, was used with the median cut colour quantisation algorithm to extract $N$-Dominants colours in all the images in order of luminance, with an example shown in Figure 3. Also, with OpenCV [51], the colour histogram was calculated for each image as shown in Figure 4. All colour features were extracted before the modelling stage for smooth integration via late fusion during training and validation to reduce the overall computational time. Geolocation models were implemented using PyTorch [52] for classification and PyTorch Metric Learning [53] for the deep metric learning approach. All experiments involving computational graphs were conducted using the Metal Performance Shaders (MPS) on an Apple M2 device as the back-end for PyTorch.

### D. EVALUATION AND RESULTS

#### 1) ACCURACY CONVERGENCE AND LOSS DECAY

Classification with deep learning primarily emphasises accuracy during training to minimise classification errors, while deep metric learning models focus on optimising embedding loss to learn feature representations that facilitate similarity-based tasks. Convergence in both model accuracy

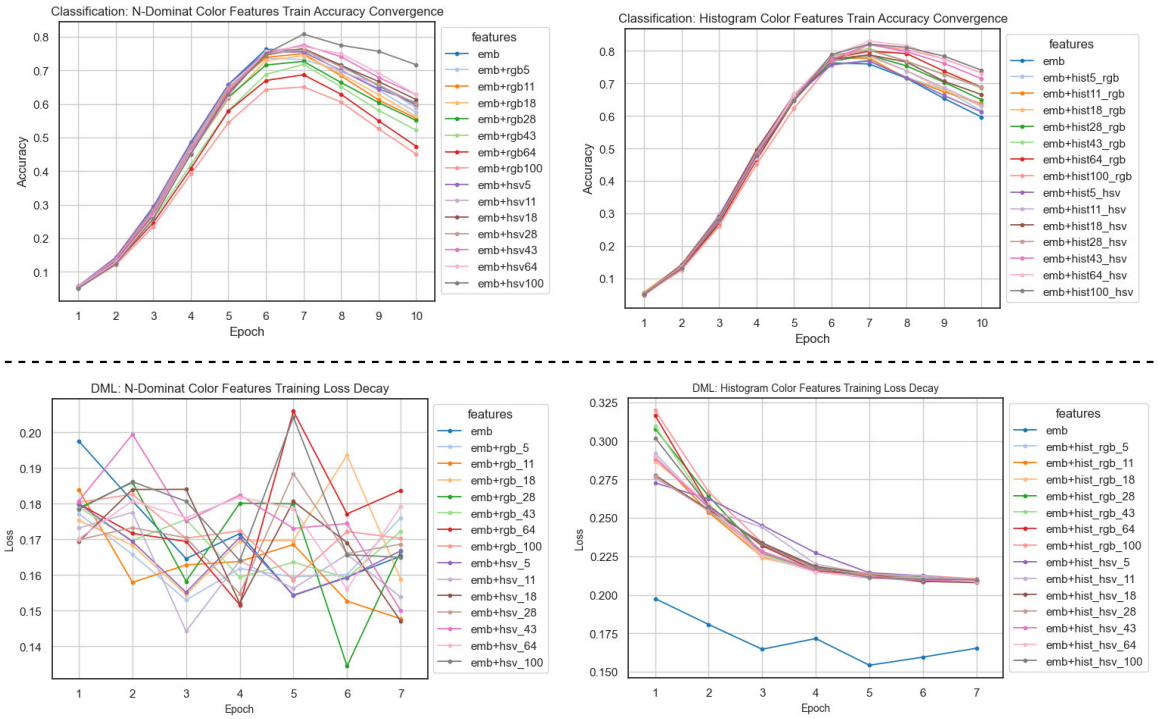


**FIGURE 5.** Analysis of model accuracy convergence and loss decay for the early epochs during training on all features. At the top is classification model accuracy convergence, and below is DML model loss decay.

**TABLE 2.** N-dominant colours with embedding results on testing set.

| | | | Classification | | DML | |
|---|---|---|---|---|---|---|
| Colour Space | $\mathbb{F}$ size | Proportion(%) | Accuracy | mAP@5 | P@1 | mAP@5 |
| - | 128 | 0 | 0.252 | 0.488 | 0.069 | **0.258** |
| RGB | 143 | 10 | 0.243 | 0.505 | 0.069 | 0.237 |
| RGB | 161 | 20 | 0.244 | 0.482 | 0.056 | 0.210 |
| RGB | 182 | 30 | 0.229 | 0.478 | 0.055 | 0.193 |
| RGB | 212 | 40 | 0.222 | 0.480 | 0.040 | 0.165 |
| RGB | 272 | 50 | 0.245 | 0.473 | 0.031 | 0.138 |
| RGB | 320 | 60 | 0.236 | 0.454 | 0.038 | 0.142 |
| RGB | 428 | 70 | 0.220 | 0.439 | 0.030 | 0.121 |
| HSV | 143 | 10 | 0.239 | 0.479 | **0.063** | **0.247** |
| HSV | 161 | 20 | 0.239 | 0.482 | 0.071 | 0.226 |
| HSV | 182 | 30 | 0.230 | 0.496 | 0.063 | 0.222 |
| HSV | 212 | 40 | **0.277** | **0.508** | 0.058 | 0.206 |
| HSV | 272 | 50 | 0.249 | 0.492 | 0.047 | 0.172 |
| HSV | 320 | 60 | 0.256 | 0.493 | 0.038 | 0.154 |
| HSV | 428 | 70 | 0.247 | 0.498 | 0.044 | 0.150 |

**TABLE 3.** Colour histogram with embedding results on testing set.

| | | | Classification | | DML | |
|---|---|---|---|---|---|---|
| Colour Space | $\mathbb{F}$ size | Proportion(%) | Accuracy | mAP@5 | P@1 | mAP@5 |
| - | 128 | 0 | 0.252 | 0.488 | **0.069** | 0.258 |
| RGB | 143 | 10 | 0.264 | 0.495 | 0.053 | 0.232 |
| RGB | 161 | 20 | 0.254 | 0.489 | 0.038 | 0.200 |
| RGB | 182 | 30 | 0.252 | 0.502 | 0.036 | 0.200 |
| RGB | 212 | 40 | 0.270 | 0.496 | 0.037 | 0.201 |
| RGB | 272 | 50 | 0.255 | 0.515 | 0.045 | 0.220 |
| RGB | 320 | 60 | 0.264 | 0.512 | 0.034 | 0.190 |
| RGB | 428 | 70 | 0.255 | 0.525 | 0.034 | 0.191 |
| HSV | 143 | 10 | 0.233 | 0.482 | 0.052 | 0.248 |
| HSV | 161 | 20 | 0.256 | 0.477 | **0.060** | **0.268** |
| HSV | 182 | 30 | 0.254 | 0.499 | 0.047 | 0.236 |
| HSV | 212 | 40 | 0.254 | 0.505 | 0.054 | 0.244 |
| HSV | 272 | 50 | **0.282** | **0.531** | 0.041 | 0.214 |
| HSV | 320 | 60 | 0.278 | 0.521 | 0.047 | 0.230 |
| HSV | 428 | 70 | 0.278 | 0.521 | 0.046 | 0.226 |

and loss decay was observed during training epochs, as shown in Figure 5. Across all experiments, there were minimal discrepancies in the classification accuracy during the early epochs. However, as training progressed, noticeable differences in convergence behaviours emerged, demonstrating the effect of the proportion of colour features on the embedding representation.

Significant fluctuations were observed in the decay of deep metric learning loss, especially with $N$-dominant colours. i.e., There is no discrete pattern in the DML training loss on the $N$-dominant colour-improved embedding compared with the colour histogram features. These fluctuations indicate that the embedding space holds more latent and sensitive information, and the palette sizes are more sensitive to the DML model than the bin sizes. This makes it difficult for the loss to follow a smoother path. It can be seen that the addition of colour histogram features to the image embedding increases the loss decay duration during DML model training, as the loss function computations are element-wise in the embedding space.

### 2) N-DOMINANT COLOUR PALETTE

- **Classification**: As shown in Table 2, classification accuracy and mAP@5 on the N-dominant colour palette improved embedding (specifically in the HSV colourspace) and were higher than those achieved with

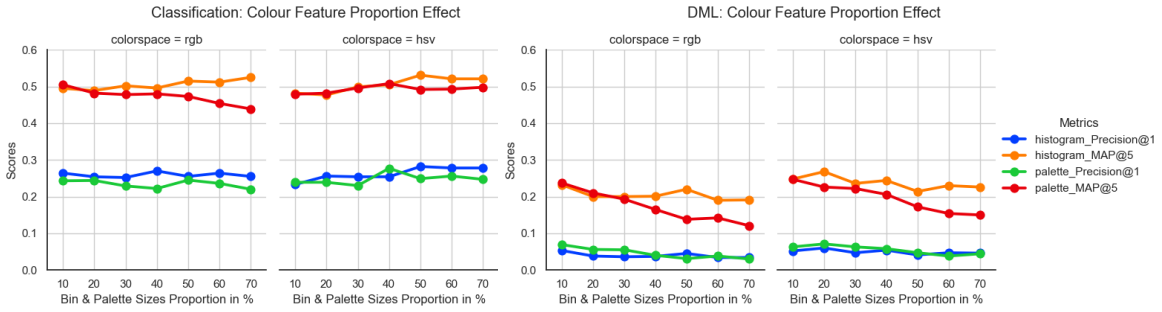


**FIGURE 6.** Effect of colour feature proportion on model scores in RGB and HSV colour spaces. On the left side is the classification model performance, and on the right side is the DML model performance.

embedding alone. This result validates hypothesis 1, indicating that the fusion of N-dominant colour features with embedding in the right proportion improves the performance of the model.

- **DML:** Using DML with N-dominant colour fused features also demonstrated superior performance with P@1 and mAP@5 in the HSV colourspace compared to using embedding alone. This also indicates that the incorporation of colour features in image embedding representation enhances the model's ability to capture descriptive features, thereby improving the retrieved geolocation ranking.
- **N-Dominant Colour Palette Saturation Point *K*:** In the classification approach, both the P@1 and mAP@5 show unique characteristics at points of inflection (40%) in RGB and HSV colour spaces, as shown in Figure 6. In the HSV colourspace, a minimal at 10% and a maximal at 40% were observed for the mAP@5. In the RGB space, 40% shows local minimal for P@1 and local maximal for mAP@5.
  With the DML approach, there is a direct correlation between the P@1 scores and the proportion between both colour spaces, but this is not consistent with mAP@5. In both colour spaces, the saturation point is found to be between 10% and 20% after which a downward trend in model performance was observed. This indicates that while these colour features can improve the performance of the model, only a small proportion is needed to avoid noise features in the embedding space.

### 3) N-BIN COLOUR HISTOGRAM

- **Classification:** The inclusion of colour histogram features specifically in the HSV colourspace significantly improved the embedding performance in both evaluation metrics, surpassing the use of embedding alone, as shown in Table 3. Models in the RGB colour space were found to outperform the base embedding model, which shows that these colour features contribute to the model's ability to identify the right class of the input image.

**TABLE 4.** Comparison of the proposed method using 40% and 20% colour features fusion proportion for Classification and DML respectively with Hotels-50K experiments configuration that used Efficient_B4.

| | | Classification | | DML | |
|---|---|---|---|---|---|
| Approach | Colour Space | Accuracy | mAP@5 | P@1 | mAP@5 |
| Original Hotels-50K (Efficient_B4) | - | 0.208 | 0.419 | 0.015 | 0.088 |
| Dominant colour+Efficient_B4 | RGB | 0.231 | 0.433 | **0.020** | **0.125** |
| | HSV | **0.252** | **0.442** | 0.019 | 0.124 |
| Colour Histogram+Efficient_B4 | RGB | 0.248 | 0.436 | 0.023 | 0.142 |
| | HSV | **0.260** | **0.451** | **0.024** | **0.145** |

- **DML:** With the DML, the model performance on colour histogram fused embedding is more pronounced with the P@1 than the mAP@5. This indicates that while colour histogram features enhance the model's ability to capture fine-grained geolocation similarities in some cases, their impact may vary depending on the evaluation metric used.
- **N-Bin colour Histogram Saturation Point K:** For classification, the histogram bin size proportion has a slight direct correlation with P@1 and mAP@5 up to 50% where the global maximal is observed as shown in Figure 6. Between 10% and 20%, the DML model with a colour histogram is seen to be saturated, just as in the case of the N-dominant colour features. This further validates that the embedding space can handle only minimal alteration before resulting in noisy features.
  With the saturation point found for both modelling approaches and hypotheses being true, comparison models were built using 40% and 20% colour features fusion proportion for classification and DML respectively. The results were compared with the Hotels-50K experiments configuration, which used Efficient_B4, as shown in Table 4. The proposed approach showed improved performance in both RGB and HSV colour spaces. To increase the chances of geolocating hotels of interest, an additional experiment was performed with k=20 (most relevant results for a given query image from a search space), further increasing the classification accuracy by 17% and a better chance of retrieving the hotel of interest in the dataset.

## V. CONCLUSION

This research demonstrates that integrating colour features into image embeddings significantly enhances model performance, surpassing the effectiveness of embeddings alone. Specifically, using colour information markedly improves the descriptive power of image features, validating one of the hypotheses. The experiments reveal that models leveraging the HSV colour space consistently outperformed those using the RGB space across both the $N$-dominant colour palette and colour histogram extraction methods.

The optimal saturation points were identified for colour feature integration: 40 to 50 percent for classification models and 10 to 20 percent for deep metric learning approaches. Balancing these proportions is crucial for maximising model performance. Although the integration of colour features did not significantly impact the convergence rate of classification accuracy, it introduced fluctuations in the loss decay for deep metric learning models. This reflects the complex learning process involved with this approach and the sensitivity of the embedding space to the colour feature vectors.

Although the integration of colour-based features introduces additional computational overhead, the resulting performance gains, particularly in indoor scene geolocation, justify this trade-off. The enhanced discriminative power provided by colour information, especially in visually similar environments, leads to improved model accuracy and retrieval effectiveness. Importantly, the extraction of colour features, such as dominant palettes and histograms, remains relatively lightweight compared to deep neural operations, making the approach feasible for deployment. This balance between performance and computational efficiency highlights the practical value of colour-enhanced embeddings in real-world geolocation systems.

The limitations of this work set the foundation for future research directions, such as exploring more sophisticated ways to combine colour features with image embeddings. This could include attention mechanisms or other fusion techniques that can further improve model performance. Furthermore, analyses that extend the model training to other colour spaces, assessing the trade-off of each, and determining if performance is better than RGB and HSV for geolocation and other use cases can be explored. Another avenue of research includes developing algorithms that can dynamically balance the contribution of colour features and embedding features during training, in line with real-time performance metrics. This work could also be extended to other datasets to evaluate the effectiveness of the proposed architecture in solving tasks beyond indoor geolocation. Furthermore, the approach has potential for real-world applications such as video-based indoor navigation, augmented reality, and mobile robotics, where accurate and efficient scene recognition is critical. By adapting colour-enhanced embeddings for real-time or sequential input, this method could contribute to improved localisation and environmental understanding in a variety of dynamic indoor settings.

## REFERENCES


[1] C. McManus, W. Churchill, W. Maddern, A. D. Stewart, and P. Newman, "Shady dealings: Robust, long-term visual localisation using illumination invariance," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 901–906.

[2] E. Müller-Budack, K. Pustu-Iren, and R. Ewerth, "Geolocation estimation of photos using a hierarchical model and scene classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 575–592.

[3] F. Breitinger, J.-N. Hilgert, C. Hargreaves, J. Sheppard, R. Overdorf, and M. Scanlon, "DFRWS EU 10-year review and future directions in digital forensic research," *Forensic Sci. Int., Digit. Invest.*, vol. 48, Mar. 2024, Art. no. 301685.

[4] X. Du, C. Hargreaves, J. Sheppard, F. Anda, A. Sayakkara, N.-A. Le-Khac, and M. Scanlon, "SoK: Exploring the state of the art and the future potential of artificial intelligence in digital forensic investigation," in *Proc. 15th Int. Conf. Availability, Rel. Secur.*, New York, NY, USA, Aug. 2020, pp. 1–10.

[5] B. Liu, Q. Yuan, G. Cong, and D. Xu, "Where your photo is taken: Geolocation prediction for social images," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 6, pp. 1232–1243, Jun. 2014.

[6] O. Bamigbade, J. Sheppard, and M. Scanlon, "Computer vision for multimedia geolocation in human trafficking investigation: A systematic literature review," 2024, *arXiv:2402.15448*.

[7] M. Werner, C. Hahn, and L. Schauer, "DeepMoVIPS: Visual indoor positioning using transfer learning," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–7.

[8] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 413–420.

[9] S. Bianco, C. Cusano, P. Napoletano, and R. Schettini, "Improving CNN-based texture classification by color balancing," *J. Imag.*, vol. 3, no. 3, p. 33, Jul. 2017.

[10] N. Samano, M. Zhou, and A. Calway, "You are here: Geolocation by embedding maps and images," in *Proc. ECCV*, Glasgow, U.K. Cham, Switzerland: Springer, Jan. 2020, pp. 502–518.

[11] J. Hays and A. A. Efros, "IM2GPS: Estimating geographic information from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[12] T. Weyand, I. Kostrikov, and J. Philbin, "PlaNet–photo geolocation with convolutional neural networks," in *Proc. ECCV*, Cham, Switzerland, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Springer, Jan. 2016, pp. 37–55.

[13] S. Cai, Y. Guo, S. Khan, J. Hu, and G. Wen, "Ground-to-Aerial image geo-localization with a hard exemplar reweighting triplet loss," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8390–8399.

[14] J. Choi, C.-W. Wong, A. Hajj-Ahmad, M. Wu, and Y. Ren, "Invisible geolocation signature extraction from a single image," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2598–2613, 2022.

[15] B. Tseytlin and I. Makarov, "Hotel recognition via latent image embeddings," in *Proc. Int. Work-Conf. Artif. Neural Networks*, Cham, Switzerland. Springer, Jun. 2021, pp. 293–305.

[16] A. Stylianou, H. Xuan, M. Shende, J. Brandt, R. Souvenir, and R. Pless, "Hotels-50K: A global hotel recognition dataset," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 726–733.

[17] R. Kamath, G. Rolwes, S. Black, and A. Stylianou, "The 2021 hotel-ID to combat human trafficking competition dataset," 2021, *arXiv:2106.05746*.

[18] S. S. Bhavanasi and A. Stylianou, "Hotel recognition using object ensembles," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Sep. 2023, pp. 1–8.

[19] F. Murgese, G. Alcaina, M. O. Mülâyim, J. Cerquides, and J. L. Fernandez-Marquez, "Automatic outdoor image geolocation with focal modulation networks," in *Proc. 24th Int. Conf. Catalan Assoc. Artif. Intell. Artif. Intell. Res. Develop.*, vol. 356, Oct. 2022, p. 279.

[20] G. Berton, C. Masone, and B. Caputo, "Rethinking visual geo-localization for large-scale applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4868–4878.

[21] J. Theiner, E. Müller-Budack, and R. Ewerth, "Interpretable semantic photo geolocation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1474–1484.

[22] A. Stylianou, J. Schreier, R. Souvenir, and R. Pless, "TraffickCam: Crowdsourced and computer vision based approaches to fighting sex trafficking," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2017, pp. 1–8.

[23] G. Friedland, O. Vinyals, and T. Darrell, "Multimodal location estimation," in *Proc. 18th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2010, pp. 1245–1252.
[24] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5297–5307.
[25] J. Hays and A. A. Efros, "Large-scale image geolocalization," in *Multimodal Location Estimation of Videos and Images*, J. Choi and G. Friedland, Eds., Cham, Switzerland: Springer, Oct. 2014, pp. 41–62.
[26] H. J. Kim, E. Dunn, and J.-M. Frahm, "Predicting good features for image geo-localization using per-bundle VLAD," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1170–1178.
[27] M. Glistrup, S. Rudinac, and B. Jónsson, "Urban image geo-localization using open data on public spaces," in *Proc. Int. Conf. Content-Based Multimedia Indexing*, Sep. 2022, pp. 50–56.
[28] N. Vo, N. Jacobs, and J. Hays, "Revisiting IM2GPS in the deep learning era," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2640–2649.
[29] A. Stylianou, A. Norling-Ruggles, R. Souvenir, and R. Pless, "Indexing open imagery to create tools to fight sex trafficking," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2015, pp. 1–6.
[30] B. S. Manjunath, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[31] J. Herrmann, O. Bamigbade, J. Sheppard, and M. Scanlon, "Perceptual colour-based geolocation of human trafficking images for digital forensic investigation," in *Proc. Cyber Res. Conf.-Ireland (Cyber-RCI)*, Nov. 2024, pp. 1–8.
[32] M. Bansal, K. Daniilidis, and H. Sawhney, "Ultrawide baseline facade matching for geo-localization," in *Large-Scale Visual Geo-Localization*, A. R. Zamir, A. Hakeem, L. Van Gool, M. Shah, and R. Szeliski, Eds., Cham, Switzerland: Springer, 2016, pp. 77–98.
[33] M. Bansal, H. S. Sawhney, H. Cheng, and K. Daniilidis, "Geo-localization of street views with aerial image databases," in *Proc. 19th ACM Int. Conf. Multimedia*, New York, NY, USA, Nov. 2011, pp. 1125–1128.
[34] T. Senlet and A. Elgammal, "A framework for global vehicle localization using stereo images and satellite and road maps," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2034–2041.
[35] T.-Y. Lin, S. Belongie, and J. Hays, "Cross-view image geolocalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 891–898.
[36] H. J. Kim, E. Dunn, and J.-M. Frahm, "Learned contextual feature reweighting for image geo-localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3251–3260.
[37] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognit.*, vol. 46, no. 1, pp. 188–198, Jan. 2013.
[38] S. M. Lee, J. H. Xin, and S. Westland, "Evaluation of image similarity by histogram intersection," *Color Res. Appl.*, vol. 30, no. 4, pp. 265–274, Aug. 2005.
[39] G. Ramella and G. S. D. Baja, "Color histogram-based image segmentation," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Jan. 2011, pp. 76–83.
[40] F. Kurugollu, B. Sankur, and A. E. Harmanci, "Color image segmentation using histogram multithresholding and fusion," *Image Vis. Comput.*, vol. 19, no. 13, pp. 915–928, Nov. 2001.
[41] A. Talib, M. Mahmuddin, H. Husni, and L. E. George, "A weighted dominant color descriptor for content-based image retrieval," *J. Vis. Commun. Image Represent.*, vol. 24, no. 3, pp. 345–360, Apr. 2013.
[42] G. Xie, B. Guo, Z. Huang, Y. Zheng, and Y. Yan, "Combination of dominant color descriptor and hu moments in consistent zone for content based image retrieval," *IEEE Access*, vol. 8, pp. 146284–146299, 2020.
[43] A. Li and X. Bao, "Extracting image dominant color features based on region growing," in *Proc. Int. Conf. Web Inf. Syst. Mining*, vol. 2, Oct. 2010, pp. 120–123.
[44] M. B. Rao, B. P. Rao, and A. Govardhan, "Content based image retrieval using dominant color, texture and shape," *Int. J. Eng. Sci. Technol.*, vol. 3, no. 4, pp. 2887–2896, Jan. 2011.
[45] Y. Chang and N. Mukai, "Color feature based dominant color extraction," *IEEE Access*, vol. 10, pp. 93055–93061, 2022.
[46] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, Jan. 2019, pp. 6105–6114.
[47] A. Demirkaya, J. Chen, and S. Oymak, "Exploring the role of loss functions in multiclass classification," in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2020, pp. 1–5.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, Dec. 2014, pp. 1–23.
[49] A. Stylianou and S. Dane. (2022). *Hotel-ID to Combat Human Trafficking 2022-FGVC9*. Kaggle. [Online]. Available: https://kaggle.com/competitions/hotel-id-to-combat-human-trafficking-2022-fgvc9
[50] *GitHub-qTipTip/Pylette: A Python Library for Extracting Color Palettes From Supplied Images*. Accessed: Nov. 5, 2024. [Online]. Available: https://github.com/qTipTip/Pylette
[51] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, Jan. 2000.
[52] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Z. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Oct. 2017, pp. 1–4.
[53] K. Musgrave, S. Belongie, and S.-N. Lim, "PyTorch metric learning," 2020, *arXiv:2008.09164*.

**OPEYEMI BAMIGBADE** received the National Diploma degree in computer engineering from the Yaba College of Technology, Lagos, Nigeria, and the B.Sc. degree in systems engineering from the University of Lagos, Nigeria. He is currently pursuing the Ph.D. degree with the School of Science and Computing, South East Technological University, Waterford, Ireland. He also has experience as a Machine Learning Engineer, with a focus on the research, development, and deployment of solutions in artificial intelligence. His research interests include computer vision techniques for digital image processing and video understanding.

**MARK SCANLON** (Senior Member, IEEE) is currently an Associate Professor with the School of Computer Science, University College Dublin (UCD), and the Founding Director of the UCD Forensics and Security Research Group. He is a Fulbright Scholar of cybersecurity and cybercrime investigation. His research interests include evidence acquisition, evidence whitelisting and data deduplication, data encryption, file synchronization service forensics, network forensics, and digital forensics education. He is a Senior Editor of *Forensic Science International: Digital Investigation* (Elsevier) and is a keen reviewer and a conference organizer in the field. He is a member of the Board of Directors of Digital Forensics Research Workshop Inc. (DFRWS), a US 501(c) nonprofit organization responsible for defining and advancing the field of digital forensic science.

**JOHN SHEPPARD** (Member, IEEE) received the Ph.D. degree from University College Dublin (UCD). He is currently a Lecturer and a Researcher of AI techniques for cybersecurity and digital forensics with the Department of Computing and Mathematics, South East Technological University (SETU), Ireland. He is a Fulbright Cybersecurity TechImpact Scholar with Boston College. His research interests include digital forensics and incident response and the use of data mining/machine learning for intrusion detection, network forensic analysis, and the IoT and small device forensics. He is a reviewer of a number of international journals and conferences and an Organizing Committee Member of the Digital Forensics Research Workshop EU (DFRWS EU).