Neural Network Approach for Camera-Based Indoor Positioning

Analysis and solution design

Martina Kuchtová

Im1, April 2024

Abstract

Indoor positioning remains a challenging task which has not been fully solved yet. The aim of this paper is to look at possible remedies to this problem, analyse them and discuss the relevant work conducted on this matter. Given the nature of camera-based positioning, several solutions are proposed to address the issue of protecting the identity of individuals who may be captured on camera indoors. For this purpose, methods from the field of computer vision and machine learning techniques will be used in the first place.

Keywords: neural network, indoor positioning, computer vision, anonymization

Introduction

Indoor positioning has grown significantly in recent years and is beneficial in a number of areas or sectors. In particular, it can be useful in large-scale buildings such as warehouses, airports or shopping centres, where with the right optimisation it can prove very effective.

The ability to precisely locate objects or people indoors has the potential to improve navigation. However, traditional methods such as GPS triangulation or Bluetooth beacons often fail in complex indoor environments due to the complex infrastructure. Thus, in buildings, other methods are used and combined to achieve the best possible result. Various sensors, technologies or artificial intelligence are employed, such as the classification and positioning based on real-time camera images presented in this thesis.

Inside, especially in buildings where there is a lot of foot traffic, there is the potential for people to be captured, so the challenge is to make them anonymous, to make sure that the people in the images are not clearly identifiable.

2 State of the Art

The need for accurate indoor positioning has led to the exploration of various techniques and approaches, each with particular strengths and capabilities. Traditional methods include WiFi-based trilateration, which uses WiFi signals and signal strength measurements to estimate the user's location based on the distances from multiple access points; Placement of low-energy Bluetooth transmitters in indoor environments to enable proximity-based positioning, often in conjunction with smartphone applications. A feature-based localisation that identifies distinctive visual features in the environment and matches them to a pre-existing map to estimate the camera's position and deep learning-based approaches that utilise neural networks to extract spatial information from camera images, enabling precise localisation without explicit feature extraction or map matching.

The demand for accurate indoor positioning has spurred interest in camera-based systems, offering high-precision localization without additional hardware. Researchers have explored various combinations of camera-based systems with other techniques, such as WiFi-based trilateration, Bluetooth Low Energy (BLE) beacons, and inertial navigation systems (INS), to enhance accuracy and robustness. These hybrid approaches leverage the strengths of different technologies to overcome individual limitations and provide more reliable indoor positioning solutions, which opens up new possibilities for Augmented Reality (AR) applications that incorporate both the physical environment and digital overlays.

Naturally, it is crucial to address the concerns related to privacy and security when using cameras and images for indoor positioning. In addition to improving accuracy, researchers must also prioritise the protection of individuals' anonymity and data security.

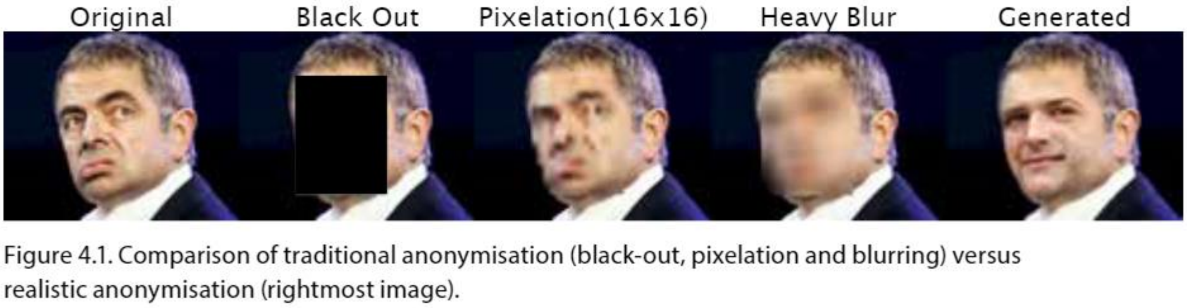


Figure 4.1. Comparison of traditional anonymisation (black-out, pixelation and blurring) versus realistic anonymisation (rightmost image).

Fig. 1 Different face anonymisation approaches [1]

While GPS-based navigation technologies have revolutionised outdoor navigation, their effectiveness indoors is often limited, necessitating the exploration of alternative solutions. One such innovative approach, as demonstrated by Reni Hena Helan et al, is the integration of augmented reality (AR) technology with built-in sensors available in mobile devices to create user-friendly and cost-effective indoor navigation systems. Their research highlights the shortcomings of traditional GPS-based approaches in indoor environments and advocates the adoption of AR-based solutions to overcome these challenges. Utilising the AR Core technology, their proposed system will provide users with an immersive navigation experience by using estimated routes to display AR guidance in indoor environments. This approach not only improves accuracy and usability, but also eliminates the need for additional hardware, reducing deployment costs and improving accessibility. Furthermore, the methodology emphasises the importance of user feedback and validation through surveys and emphasises the iterative nature of the development and optimisation of the system. Through their indoor navigation smartphone app and extensive testing, they demonstrate the feasibility and effectiveness of their approach in real-world scenarios. [2]

3 Data

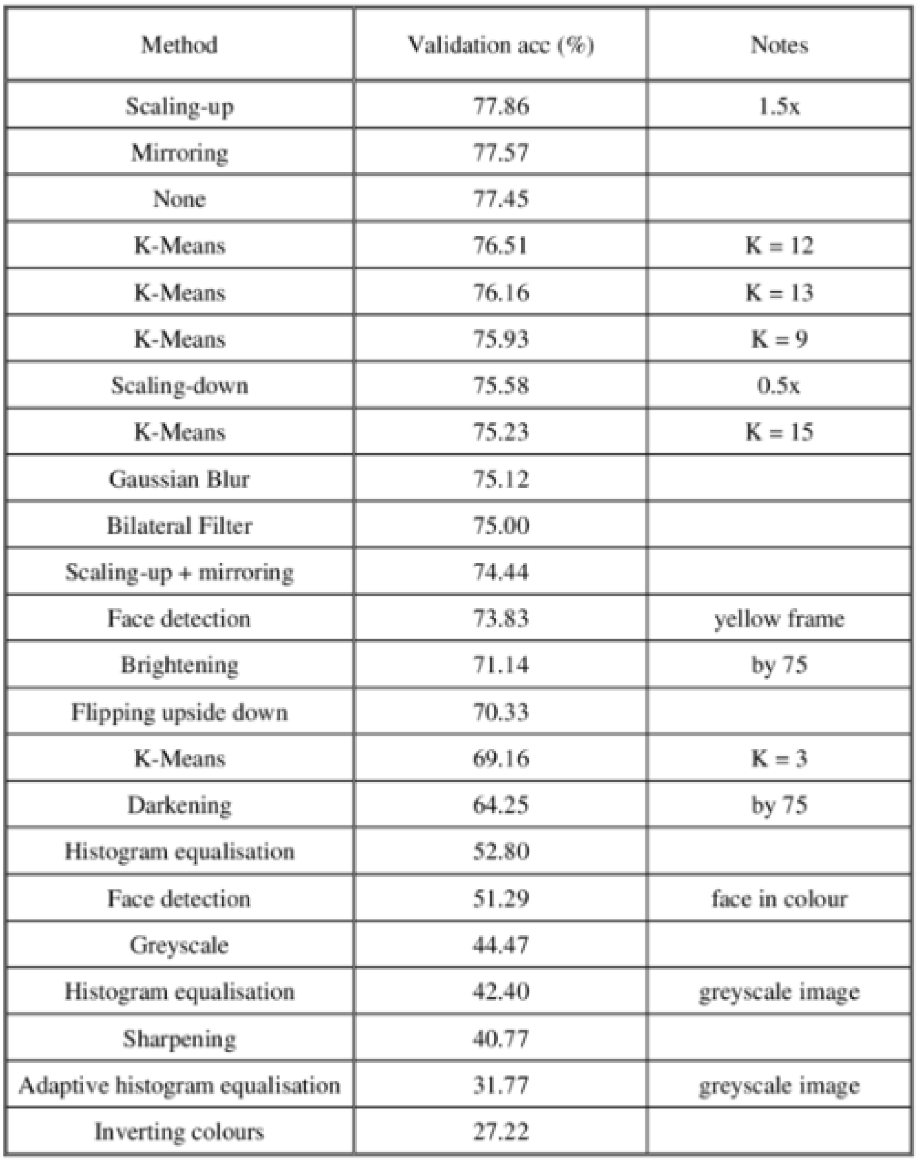
Based on the knowledge in the field of image analysis, where the achievement of ideal results depends on the pre-processing of the inputs (e.g. for the Canny detector), we decided to analyse the impact of a similar approach in the case of neural network image classification. The focus was on the validation dataset, and the main focus of the research was whether there was a method that would improve the accuracy of the classification itself, ideally for all the chosen classification tasks.

In the past, the very problem of modifying images for the purpose of classification has been the subject of a number of research studies that have produced results, for example in the area of medical images, or in improving classification accuracy through data augmentation, where a training set is artificially expanded so that a neural network can learn on multiple different instances.

In our case, we are and will be dealing with the calibration of images for indoor positioning purposes. A similar approach was used in our bachelor's thesis, where we analysed the content of the selected datasets and the proposed methods, which were expected to produce results in the form of an increase in the quality of the stored information, which could play a significant role in the classification and thus lead to its improvement, or alternatively, the impact of a given method was otherwise of interest to us. The proposed methods that were implemented operated on the values of individual pixels, resulting in changes in colour, brightness or clarity of detail. However, we also used more complex algorithms, such as those used for face recognition or 'whitening' images by means of decorrelation. We first evaluated the impact of each method and the relevance of their results for the purpose of analysis on a small sample of data, and then applied them to the entire validation set.

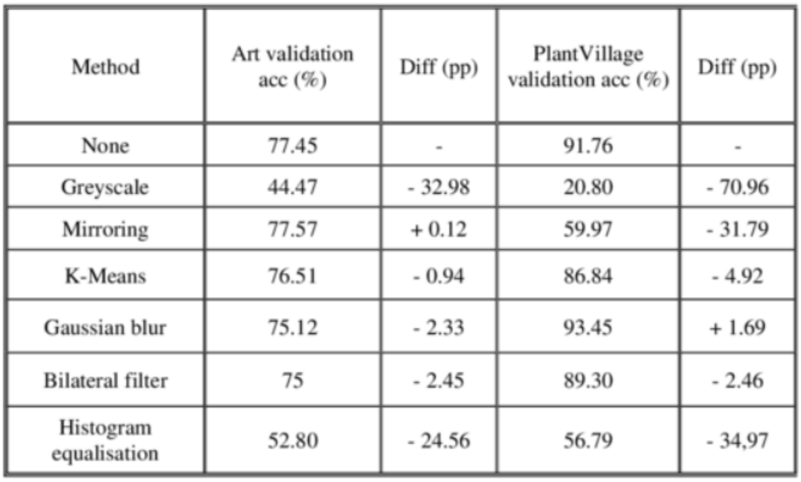
The results for each method use on Art image dataset are shown in Table 1 below and somparision of few methods that were used to two different datasets and their impact on classification accuracy is being compared in the Table 2.

After the evaluation, it was hypothesised that the colour of the images can have a significant impact on the accuracy and classification results. This could also be reflected and taken into account in this work.If the neuron is trained on images and recordings of empty corridors in a building, the presence of a person (even depending on how they are dressed) can have a significant effect on the results. The same principle could apply if people are present in the training sample but a particular type/fabric of clothing (colour palette) dominates - the classification may not be accurate if, for example, we have people dressed in neutral colours and garish or neon colours appear in the test.



Method	Validation acc (%)	Notes
Scaling-up	77.86	1.5x
Mirroring	77.57	
None	77.45	
K-Means	76.51	K = 12
K-Means	76.16	K = 13
K-Means	75.93	K = 9
Scaling-down	75.58	0.5x
K-Means	75.23	K = 15
Gaussian Blur	75.12	
Bilateral Filter	75.00	
Scaling-up + mirroring	74.44	
Face detection	73.83	yellow frame
Brightening	71.14	by 75
Flipping upside down	70.33	
K-Means	69.16	K = 3
Darkening	64.25	by 75
Histogram equalisation	52.80	
Face detection	51.29	face in colour
Greyscale	44.47	
Histogram equalisation	42.40	greyscale image
Sharpening	40.77	
Adaptive histogram equalisation	31.77	greyscale image
Inverting colours	27.22	

Tab. 1 Art image classification performance [3]



Method	Art validation acc (%)	Diff (pp)	PlantVillage validation acc (%)	Diff (pp)
None	77.45	-	91.76	-
Greyscale	44.47	- 32.98	20.80	- 70.96
Mirroring	77.57	+ 0.12	59.97	- 31.79
K-Means	76.51	- 0.94	86.84	- 4.92
Gaussian blur	75.12	- 2.33	93.45	+ 1.69
Bilateral filter	75	- 2.45	89.30	- 2.46
Histogram equalisation	52.80	- 24.56	56.79	- 34,97

Tab. 2 Comparison of methods used on 2 different classification tasks [3]

3.1 Imclust

Imclust incorporates advanced image clustering techniques that employ NN models to extract meaningful features from images resized to 224x224 pixels. This feature extraction is accomplished using state-of-the-art Convolutional Neural Networks (CNNs) pre-trained on the ImageNet dataset. Principal Component Analysis (PCA), a statistical technique used in machine learning and scientific research to simplify data and identify patterns, is subsequently used to reduce these high-dimensional features.

Various image clustering methods are used to categorise images into different classes in the Imclust, namely:

- *Imclust Agglomerative* - the method employs hierarchical clustering, starting from individual images as separate clusters and progressively merging pairs based on their similarity or distance. It is particularly suitable for smaller datasets or when the data exhibits hierarchical structures, such as photographs taken under different lighting conditions or varying weather.
- *Imclust CNN* - features extracted using CNNs are dimensionally reduced via PCA. Clustering is then performed using the CommonNNClustering algorithm, ideal for large datasets and capturing deeper similarities beyond mere texture or color, such as the presence of specific objects within the images.
- *Imclust DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* - it does not require specifying the number of clusters beforehand. It identifies clusters based on the density of data points, marking outlier points that deviate significantly from the norm as noise. This method relies on two parameters: epsilon (the maximum distance between neighboring points) and min_samples (the minimum number of neighbors a point must have to be considered a core point).
- *Imclust K-Means* - this classic clustering approach iteratively assigns points to the nearest cluster center, recalculating the center after each iteration. It requires the number of clusters to be specified and is sensitive to the initial placement of these cluster centers.

- *Imclust K-Medians* - similar to K-Means, this method minimizes the median of the distances between the points in a cluster and the cluster center, making it more robust against outliers within the dataset.
- *Imclust K-Medoids* - using actual data points as centers (medoids), this method is resistant to noise and outliers. It is suitable for smaller datasets where the distance between any two points suffices, avoiding the need for iterative center updates as in K-Means.
- *Imclust K-Modes* - designed for categorical data, this method updates cluster centers based on the mode (most frequent category) rather than the mean, accommodating non-numeric data types effectively.
- *Imclust Spectral Clustering* - utilizes eigenvalues and eigenvectors of a similarity matrix constructed from data relationships. This method is often enhanced with techniques like Singular Value Decomposition (SVD) and employs a Gaussian similarity function, where the parameter sigma controls the width of the function, affecting similarity sensitivity with respect to distance.

The effectiveness of clustering is evaluated using several metrics that can help us better understand how well the clustering method had performed on the data:

- *Mean Silhouette Coefficient (MSC)* - measures how similar an object is to its cluster compared to other clusters,
- *Calinski-Harabasz Score (CHS)* - quantifies the ratio of variance between clusters to the variance within clusters,
- *Davies-Bouldin Score (DBS)* - reflects the average similarity between each cluster, with lower values indicating better-separated clusters,
- *Cluster Purity Index (COP)* and *SDbw Index* - assess cluster compactness and separation comprehensively.

The clustering results are visualised in HTML format, providing an intuitive presentation of data and images grouped by their respective clusters.

While the implementation discussed herein uses code developed by R. Jaksa and further contributions made by A. Gajdoš, it is important to clarify that the underlying clustering methods are based on well-established algorithms in the field of image processing. The authors' contribution consists primarily of adapting these algorithms into practical code for specific applications, such as their project on human head scanning.

The ability to accurately group images into clusters using algorithms such as K-Means, DBSCAN and hierarchical clustering, among others, greatly enhances our understanding and categorisation of complex image data. For example, in applications such as indoor positioning, image clustering can provide key insights into the optimal number of segments into which a building should be divided to improve location accuracy. Similarly, in projects involving automated data labelling, effective image clustering can greatly assist in the preliminary organisation of image data, which in turn facilitates more accurate label assignment.

4 Model

Within the context of camera-based indoor positioning systems, neural networks offer powerful tools for improving accuracy and reliability. Two primary approaches within neural networks are classification and regression, both of which can be used effectively depending on the specific requirements of the indoor positioning system.

Classification by neural networks involves classifying input data into pre-defined classes. In an indoor positioning context, this could mean identifying specific rooms or areas (such as 'Lobby', 'Meeting Room', 'Canteen') based on camera images. This approach is particularly useful in environments where there are distinct, recognisable landmarks or features, and the goal is to determine which of these predefined spaces the camera is currently viewing. Neural network models such as Convolutional Neural Networks (CNNs) are particularly adept at extracting and learning from the visual features required for such classification tasks.

Regression, on the other hand, involves predicting a continuously varying value that can be applied directly to indoor positioning by estimating precise coordinates (such as x, y (and z) positions within a building). This method is useful when you need to determine a more precise location, rather than just identifying a type of area. For example, regression models can continuously track the movement of a device through space, providing not just the room or area, but the specific location within that space. Techniques such as CNNs combined with Recurrent Neural Networks (RNNs) or Long Short-Term Memory Networks (LSTMs) can analyse sequential image data to accurately predict these coordinates.

Both approaches can be integrated with other data sources and technologies - such as electromagnetic fields in this specific work - to improve the positioning accuracy and robustness of the system. Hybrid neural network models, which combine features of both classification and regression, can dynamically switch between determining general area type and precise location based on available data and specific user requirements.

Ultimately, the choice between classification and regression approaches in indoor neural networks depends on the level of granularity required by the application. For environments where knowing the general area is sufficient, classification may be the preferred approach. However, for applications that require precise positioning and navigation assistance, such as guiding a user to a specific product in a large retail store, regression models are a more appropriate solution. As indoor positioning technology evolves, the use of advanced neural network models and their integration with other technologies will continue to improve the ability to effectively navigate complex indoor spaces.

4 Anonymisation

In the field of camera-based indoor positioning systems, the need to anonymise personal data cannot be overstated, given the deep privacy implications of processing, and storing identifiable visual information. By exploring a variety of anonymisation methods integrated into neural network approaches, the aim of this work is not only to identify the most effective strategies for preserving anonymity, but also to evaluate their impact on the accuracy and efficiency of the positioning system.

4.1 Censorship and blur (privacy filters)

Techniques such as pixelation and blurring are key tools in anonymisation efforts, as they effectively anonymize individuals by obscuring identifiable features in images, such as faces. These methods allow the anonymity of subjects to be skilfully maintained without compromising the overall usefulness of the images for location purposes. In our opinion, it is better to use more subtle forms of censorship than black 'squares', which look too harsh and can distort the structure of the image. A similar approach is used e.g. in media or Street view by Google.

Complementing this, Haar features - a technique adept at identifying patterns, particularly human faces, within images - increase the precision of where pixelation or blurring should be applied. Together, they provide an approach to anonymisation that is both automated and targeted. By blending Haar features with pixelation or blurring, the system efficiently protects privacy by obscuring only necessary elements of the image, thereby preserving the indoor positioning accuracy.

However, in these cases of image modification, instead of focusing on the faces or figures of people, a simple blur can be used to affect the whole image and the effect on the ability to position could be observed.

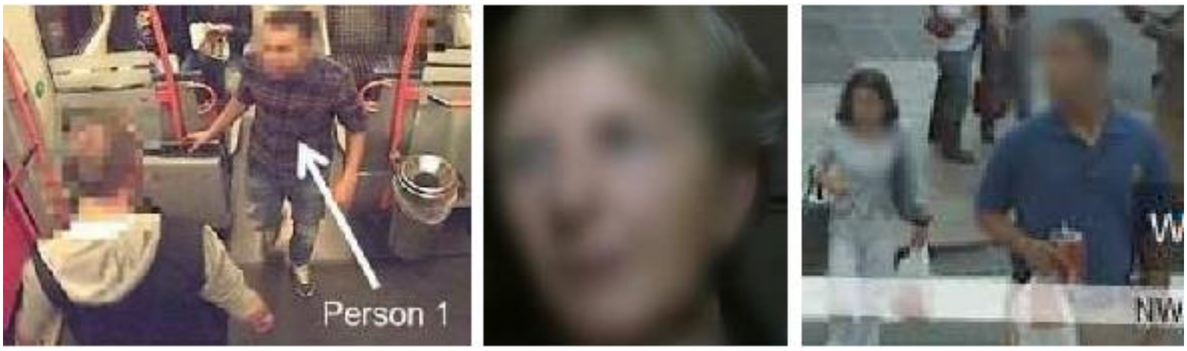


Fig. 2 Censorship in the media [4]

4.2 Detection & inpainting models

Taking advantage of advances in deep learning, we can explore methods beyond simple image pixelation for anonymisation purposes. Generative Adversarial Networks (GANs) are a promising approach. Here, one neural network (generator) generates realistic images without people, while another (discriminator) refines the generated image to appear authentic. Training these models yields anonymised images that closely resemble the originals, preserving scene context for accurate localisation while removing identifying information.

A number of models exist that are trained on huge amounts of data and are able to remove people from the image. Most often, it is the people in the images that are desired to be removed. Such models can be very handy for this kind of problem, since most of the information is retained, perhaps even the information that was originally obscured by the person in the image. However, the cost can be computationally intensive, as the model needs to be as fast and efficient as possible to perform this anonymisation in real time.

A good example and model to use is LaMa - Resolution-robust Large Mask Inpainting with Fourier Convolutions, which excels at removing objects, including people. It analyses the image and the surrounding area to understand the context of the scene. This allows LaMa to inpaint the region where the person was located, effectively "filling the gap" with a realistic background. This approach has several benefits: it maintains scene fidelity, which is critical for accurate positioning, and avoids the visually disruptive nature of pixelation. While LaMa may require more computing resources than simpler methods, its ability to preserve scene detail makes it a strong choice for anonymising images in your indoor positioning system.

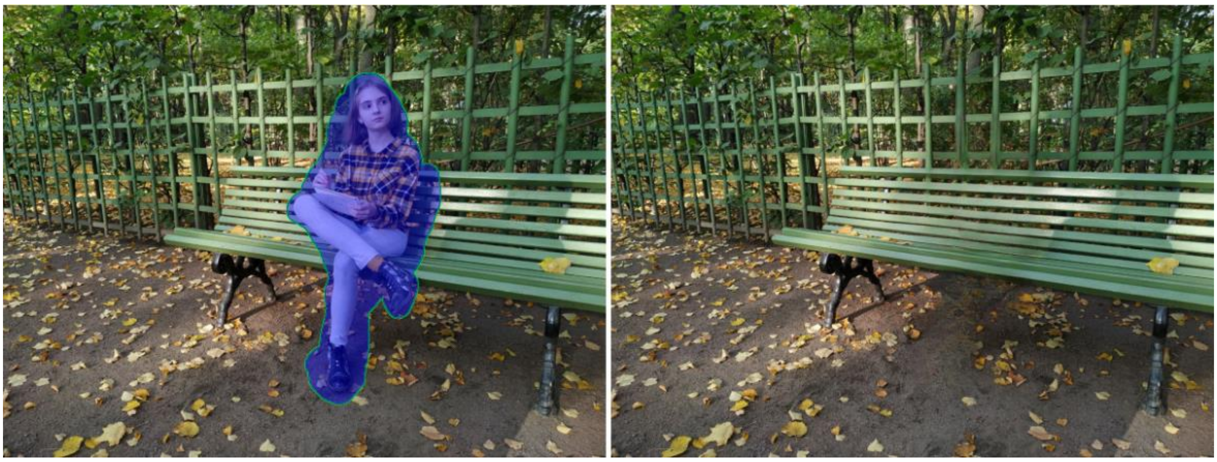


Fig. 3 LaMa inpainting result [5]

4.3 Face filters

While face filters have gained immense popularity on social media platforms, primarily for fun and playful manipulations, their underlying technology holds promise for anonymization in your indoor positioning system. It is well known that such filters can often change a person beyond recognition and make the picture far from reality when not resembling the actual individual.

Of particular interest is the ability to work with the image quickly and efficiently in real time, something that can be seen when using any online filter. This is why such a filter and its use can be beneficial to us.

One approach is to replace entire faces with natural elements such as materials or geometric shapes. This can be effective in obscuring identities while maintaining some scene context. Another option is to use image inpainting techniques, similar to LaMa, but specifically focused on replacing faces with realistic, anonymised representations. This could involve replacing plausible faces that lack identifiable features, or using stylistic elements to create a more uniform, avatar-like appearance. Such creative approaches offer a balance between privacy and scene realism that could potentially improve user acceptance.

Crafting a custom filter tailored for your indoor positioning neural network presents an interesting approach. By using deep learning frameworks such as TensorFlow, a filter could be developed specifically for the concrete environment. This filter could potentially replace faces with pre-defined objects, or even use image manipulation techniques to seamlessly integrate anonymised faces that resemble famous people (e.g. historical figures). It also allows you to modify the features and change the person to the point of unrecognizability, for example by adding a beard or changing the "weight".

There are a number of environments, platforms and SDKs that allow you to create your own face filter. Here is one of the options that might be interesting for us and involves AR.

The filter can be saved and used later in image preprocessing when it is applied to the image before feeding it to the neural network.

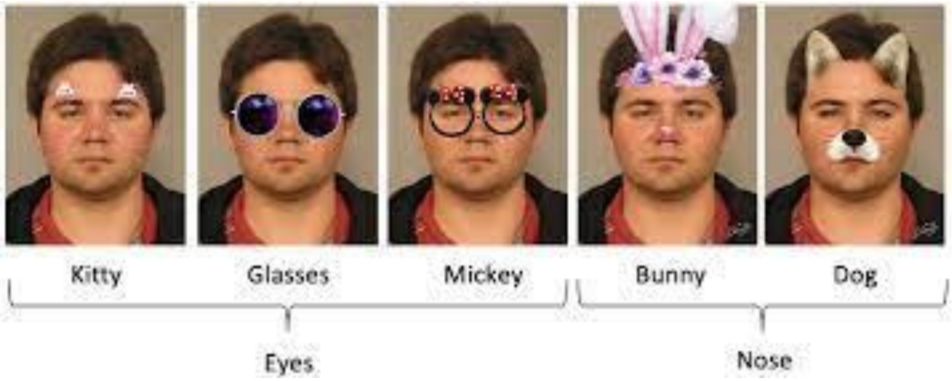


Fig. 4 AR face filters [6]

Fig. 5 Adding glasses to images [7]

4.4 “Slow shutter speed”

Using a slow shutter speed in combination with frame stacking. The technique captures multiple video frames at a slow shutter speed, effectively blurring moving objects such as people in each frame. By strategically stacking these blurred frames, a final image with a more consistent level of anonymisation for moving people can be achieved. While this approach offers several potential avenues for anonymisation, the trade-off is that the camera must remain stable to achieve better results - the building and surroundings will remain sharp, while the people and the trajectory in which they are moving will be blurred.

4.5 Floor Detection and Segmentation

The floor region is first identified in the captured images. Segmentation techniques are then employed to isolate features such as walls and furniture from the non-floor areas. By focusing on these features, the system can learn to accurately position people in space without directly analysing the people themselves. With this approach, it is possible to make sure that the areas in which a person might appear are "ignored", or else to blend them out, remove them and focus on other elements and features.

6 Current status & latest results

First of all, from the corridors of the Jesenná and Park Angelinum buildings, data and videos/recordings were collected.

A few segments were selected for which we collected "data", later we plan to use these segments in a smaller neural network model where we will test some basic operations and make some observations. As part of the data we collected, there were simulated situations where the hallway was empty, there were people in the hallway - stationary (as we walked past them), walking in front of us, walking towards us, and a selfie video as we walked down the hallway.

Several of the anonymisation techniques suggested above were applied - various forms of censorship such as black squares, blur, and mosaicking (pixelation). In these cases, haar cascade styles were used to detect the face and its features. The primary focus of our experiment was on face detection, though we also explored the potential of detecting upper bodies and full bodies in images. We conducted these experiments on photographs featuring large crowds to better understand the impact and effectiveness of our anonymization techniques. Despite the limited success in these initial trials, we plan to implement this technology on the data collected from cameras installed in various halls. To optimize face detection, we integrated three distinct cascade classifiers: one for detecting frontal faces, another for profiles, and a third for eyes. This combination proved particularly effective in identifying faces even when they were partially obscured or covered, enhancing the robustness of our detection system. This multi-faceted approach allowed us to refine our technique and improve the accuracy of detecting individuals in complex visual environments.

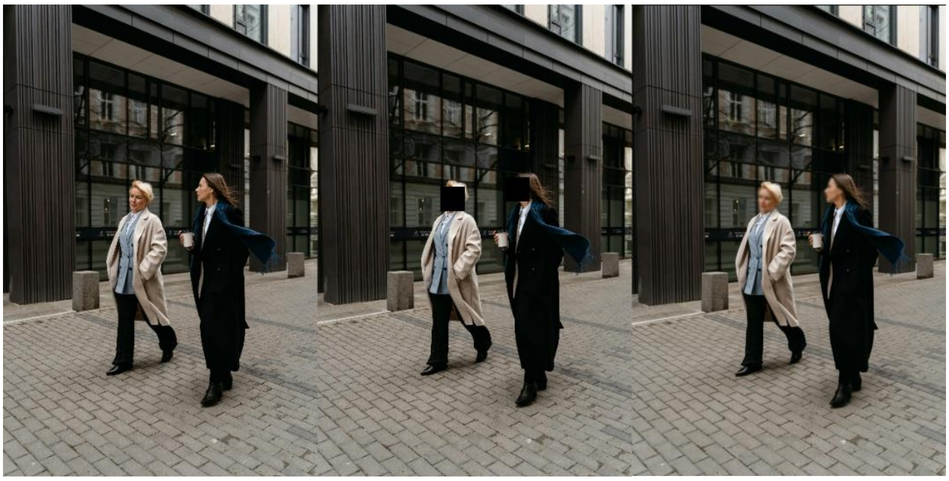


Fig. 6 Face anonymisation (censorship, blur)

Another technique is face-swapping, which uses a pre-trained model *shape_predictor_68_face_landmarks.dat* that is associated with open-source library dlib and helps identifying 68 specific points that correspond to key facial features such as eyebrows, eyes, nose, mouth, and jawline. Face-swapping uses this model to locate corresponding facial features on two different faces and by aligning them, seamlessly swap facial textures between the two images, creating the illusion that one face is being replaced by the other. We have attempted this type of anonymisation by using celebrity faces to overlay the faces in the videos, but we're also planning to use some AI-generated faces.

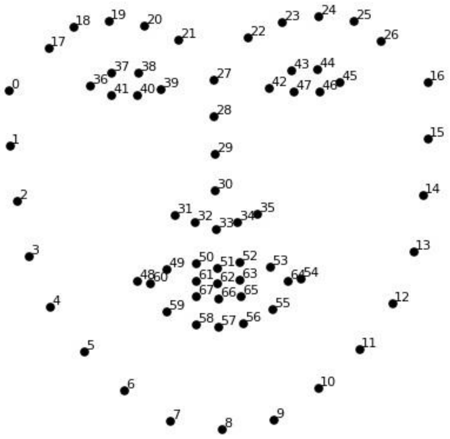


Fig. 9 Relevant points (68) for Shape prediction [6]

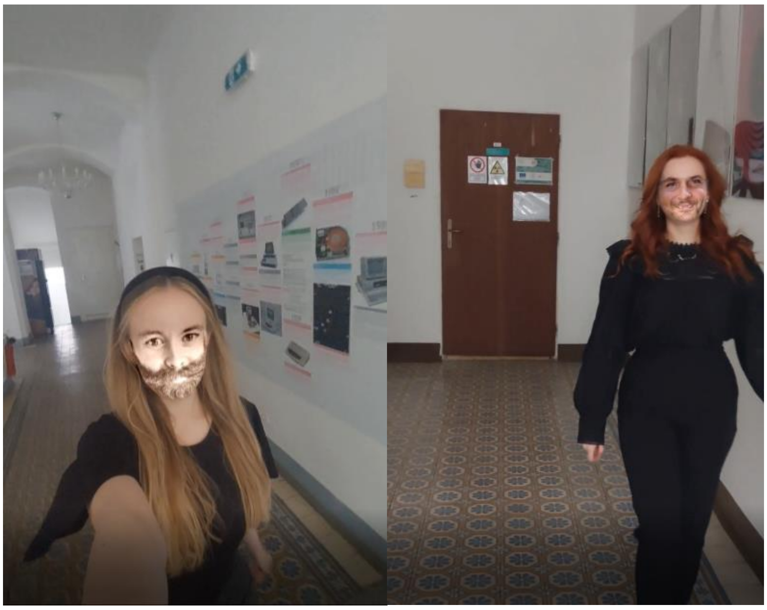


Fig. 7 Face swapping

The data, which consisted of videos, was meticulously processed by breaking them down into individual frames; specifically, we extracted every fourth frame for analysis. To categorize these frames, we utilized the Imclust tool, which enabled us to classify all the available frames/images into distinct clusters or classes. This clustering technique is anticipated to be particularly useful for future applications, such as automatic labelling. For instance, it could be employed to evenly distribute data across clusters, ensuring that there is a balanced number of examples for each class in both training and validation sets. We tested various clustering methods offered by Imclust, including Hierarchical and K-Means clustering, which have shown to yield the best results for our specific tasks. These methods were evaluated across different neural network models to assess their effectiveness and adaptability to our clustering needs. This comprehensive testing framework allowed us to optimize our approach for classifying and managing large sets of image data efficiently.

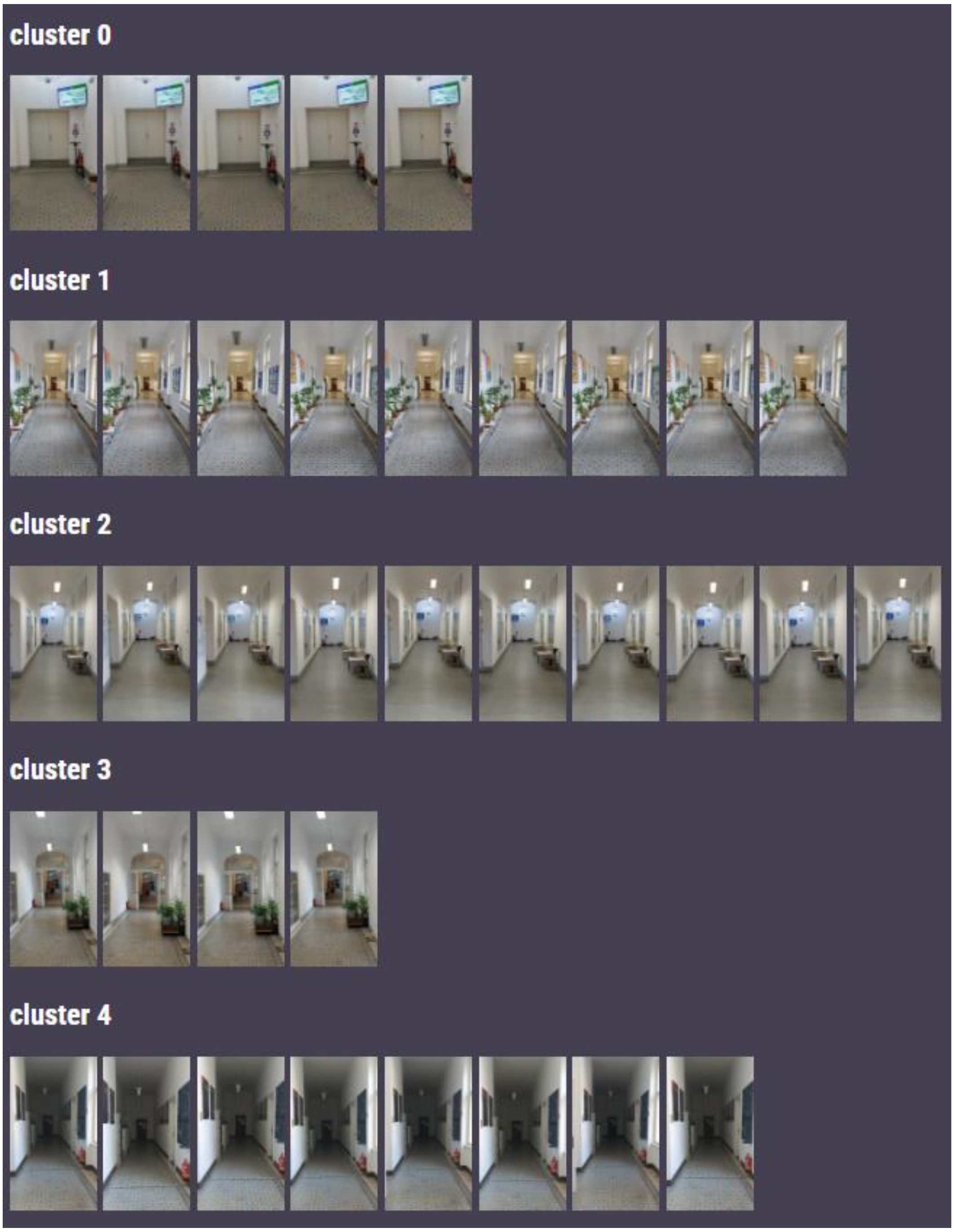


Fig. 8 Imclust Agglomerative output snapshot result (hall classification)

Conclusion

Our focus has been on the use of neural networks to enhance camera-based indoor positioning systems. Through detailed examination and analysis, we have recognised that the indoor positioning challenge involves complex interactions of various factors, including the physical layout of the environment, the presence and movement of people, and the inherent limitations of different technologies. We have also explored the strengths and weaknesses of using classification and regression approaches within neural networks to address these challenges. As the demand for more accurate and reliable indoor positioning systems grows, it becomes critical to select the most appropriate technological and methodological approach.

In addition, we have identified several key issues and problems that require further investigation. These include optimising the computational efficiency of neural network models to enable real-time processing, enhancing privacy and security measures to protect individuals captured in camera feeds, and improving the adaptability of systems to dynamically changing indoor environments. Each of these areas presents its own set of challenges and opportunities for deepening the understanding and application of neural network technologies in indoor positioning systems.

Moving forward, the focus of our work will be to select and refine the best approaches to address these issues. We will carry out extensive experimentation and analysis to determine the optimal balance between accuracy, efficiency, and data sensitivity.

In addition, there are plans to develop a smaller experimental neural network that will operate with only a few segments. This streamlined approach will allow us to gain deeper insights into the problem at hand and further explore the results and methods discussed in this article. By focusing on a more compact model, we aim to refine our techniques and improve our understanding of the underlying patterns in the data. This experimental network will serve as a testing ground where we can apply different strategies and analyse their effectiveness in a controlled environment, allowing us to incrementally improve our main project based on the results.

References

[1] Hukkelås, H., & Lindseth, F. (2024). Realistic Face Anonymisation. In B. Sloot & S. Schendel (Eds.), The Boundaries of Data (pp. 53-64). Amsterdam: Amsterdam University Press. https://doi.org/10.1515/9789048557998-004.

[2] Helan, R. R. H., Vivekanandan, S. J., Deepak, A., Imran, S., & Hemanth, T. (2023). Indoor navigation using augmented reality. Dhanalakshmi College of Engineering.

[3] Kuchtová, M.. (2023). Method of image modification for Neural Network Classification Univerzita P.J. Šafárika v Košiciach PF UPJŠ ÚINF (Master's thesis).

[4] Ruchaud, N., & Dugelay, J. L. (2016). Automatic Face Anonymization in Visual Data: Are we really well protected? In International Conference on Image Processing, Applications and Systems. Available at: https://api.semanticscholar.org/CorpusID:27373146

[5] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2021). Resolution-robust Large Mask Inpainting with Fourier Convolutions. arXiv preprint arXiv:2109.07161.

[6] Botezatu, C., Ibsen, M., Rathgeb, C., & Busch, C. (2022). Fun Selfie Filters in Face Recognition: Impact Assessment and Removal. arXiv preprint arXiv:2202.06022v1 [cs.CV].

[7] Hedman, P., Skepetzis, V., Hernandez-Diaz, K., Bigun, J., & Alonso-Fernandez, F. (2022). On the effect of selfie beautification filters on face detection and recognition. Pattern Recognition Letters, 163, 104-111. https://doi.org/10.1016/j.patrec.2022.09.018.

[8] PySource. "Face Landmarks Detection OpenCV with Python." PySource. https://pysource.com/2019/03/12/face-landmarks-detection-opencv-with-python/. Accessed: April 22, 2024

Might use later:

Hassan, M. U., Stava, M., & Hameed, I. A. (2023, July). Deep privacy based face anonymization for smart cities. Paper presented at the 2023 International Conference on Smart Applications, Communications and Networking (SmartNets). https://doi.org/10.1109/SmartNets58706.2023.10215996