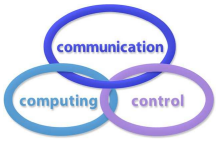
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 2, Month: April, Year: 2025
Article Number: 7031, https://doi.org/10.15837/ijccc.2025.2.7031



CCC Publications

Fingerprint image scale estimation for forensic identification systems

T. Oblak, J. Videnović, H. Kupinić, V. Štruc, P. Peer, Ž. Emeršič

Tim Oblak*
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
113 Večna pot, Ljubljana, Slovenia
tim.oblak@fri.uni-lj.si

Jovana Videnović*
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
113 Večna pot, Ljubljana, Slovenia
jv8043@student.uni-lj.si

Haris Kupinić*
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
113 Večna pot, Ljubljana, Slovenia
hk8302@student.uni-lj.si

Vitomir Štruc
Faculty of Electrical Engineering
University of Ljubljana, Slovenia
Tržaška 25, Ljubljana, Slovenia
vitomir.struc@fe.uni-lj.si

Peter Peer
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
113 Večna pot, Ljubljana, Slovenia
peter.peer@fri.uni-lj.si

Žiga Emeršič
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
113 Večna pot, Ljubljana, Slovenia
ziga.emersic@fri.uni-lj.si

* These authors contributed equally to this work.

Abstract

The large majority of modern software solutions intended for fingermark processing in a forensic context is heavily dependant on the correct image scaling. Fingermark images captured with digital cameras at a crime scene require the use of physical rulers or labels. While the resolution of a fingermark image can be calibrated manually by a forensic examiner in a lab, we propose an automated approach, which could be integrated directly into existing identification systems and would eliminate the need for human intervention. Our approach consists of a CNN regressor, which directly predicts the PPI of stochastically-sampled local patches based on the friction ridge information contained within. In a range of PPI between 500 and 1500, our method achieves a mean average error of around 24 PPI for fingerprint and fingermark images.

Keywords: fingerprint, fingermark, image scale, ppi prediction, forensics

1 Introduction

Fingermarks (latent fingerprints) are special type of fingerprints, found in uncontrolled environments, such as crime scenes, and used to identify subjects of interest during a forensic investigation. Due to the uncontrolled deposition process, as well as different development and capture techniques,

fingermarks contain a wide variety of imperfections, which reduces the quality of identifiable features and increases the difficulty of successful identification. Modern forensic practices heavily rely on Automated Identification Fingerprint Systems (AFISs), which greatly reduce the amount of manual labour and increase identification rates. However, with digitization of fingermark images new constraints were introduced into the forensic process.

When a fingermark image is enrolled into an AFIS, a biometric template is generated, usually using a combination of automated feature extraction and, if required, manual intervention by dactyloscopic experts. In order to detect relevant features, automated feature extractors often leverage domain knowledge and known friction ridge properties to calculate type and location of feature in images. Many of these are detected based on several assumptions, including the scale at which the fingerprint image was captured. Specifying the wrong scale parameters to the AFIS has detrimental effects to the matching process. It is therefore crucial for the AFIS to have accurate information about the scale of an image. In practice, the scale translates into pixel density and is most commonly given in terms of pixels per inch (PPI). Best practice guides for forensic fingerprint identification specify 250 PPI as the minimum pixel density to derive first level features (ridge orientation, singular points), 500 PPI as the minimum for second level features (minutiae points), and 1000 PPI for third level details (such as individual pores, incipient ridges, etc.) [5].

To capture information about scale, forensic technicians place rulers or measuring labels adjacent to the fingermark being captured. The PPI is then calculated by measuring the number of pixels within a measure of distance. However, fingermark images are often subjected to preprocessing, which includes cropping, modifications of colorspace, enhancement, etc. In some cases, these operations might remove the original information about scale. This problem also extends to storage of fingermark data, for example, evidence retention of cold cases, or for purposes of large-scale biometric IT systems. The PPI also directly relates to the quality of a biometric sample [7, 8, 13, 14, 15].

To address this problem, we propose an approach to detect the scale of a fingermark image (in terms of PPI). To achieve this, we leverage known frequency of friction ridges in adult population. Furthermore, we present a new deep learning model, which does not require domain knowledge to derive the PPI, which makes it more robust to differences between male and female friction ridge impressions. Our main contributions in this paper are:

- A new light-weight patch-based deep learning approach to detect the scale of a fingermark images without the help of rulers or measuring labels in the original image.
- We compare our approach to Fourier transform-based solutions, commonly used to predict friction ridge frequency in related literature.
- Thorough evaluation of the proposed approach on a modern fingermark dataset (NIST SD 302), including some qualitative results.

2 Related work

In practice, the scale of a fingerprint image relative to physical distance measurements is already determined at the time of capture, either from the sensor itself or by manual annotation of the metadata. However, the information about the scale of an impression is also encoded indirectly in an friction ridge image and can be retrieved by correlating image pixel density with known characteristics of finger impressions. While forensic software for fingerprint analysis typically include features to manually determine the resolution based on ridge spacing, our goal is to estimate this in an automated manner. Based on the reviewed literature, we define two potential approaches to solve the problem of fingerprint image scale estimation.

2.1 Relative scale estimation for sensor interoperability

Several publications focus on fingerprint scaling from the perspective of sensor interoperability. The goal of such methods is to estimate the scaling factor between pairs of images, captured with different sensors at different resolutions. Ren *et al.* [10, 11] use spectral analysis to calculate average

inter-ridge distance in a fingerprint image, which is then compared to inter-ridge distance of another fingerprint to estimate the relative scaling factor s. In a similar fashion, Kunsuk *et al.* [6] use 2 different methods to determine the scaling factor for contactless finger photos. The first method uses Fast Fourier Transform (FFT) to find the dominant frequency in the entire image and the second uses Short-time Fourier Transform (STFT) to analyze narrow frequency bands in local areas of the fingerprint. All referenced methods report an improvement in identification performance when images are scaled with the estimated scaling factor. However, these methods require two fingerprint samples to establish the said scaling factor. In contrast, our aim is to estimate an absolute measure of scale for forensic fingermark images.

2.2 Friction ridge density

The density of papillary ridges in a finger impression is known to be relatively consistent across the adult population with slight variation based on gender and geographical location [5, 12]. We can correlate the average inter-ridge distance with ridge distances, calculated from a friction ridge image. Most commonly, methods counting the number of periods in a sine-like waveform use Fourier transform. For example, the NIST Fingerprint Image Quality (NFIQ 2) [14] first orients a local window based on image derivatives so that the friction ridges are aligned vertically. Then, the image is averaged across all columns to retrieve the one-dimensional sine wave of the ridge pattern. Finally, FFT is used to calculate the dominant frequency in the sine wave. Some methods calculate spatial frequency directly from 2D signals [13], either from a local window or from the entire image. The approach is relatively slower, but does not require rotating the local window initially. Several publications related to friction ridge processing also use the Short-time Fourier Transform (STFT) to calculate the frequency spectrum more locally [6, 15], which enables them to isolate specific frequency sub-bands. The methods, outlined here work well for good quality fingerprints with a high contrast impression. However, forensic fingermarks usually contain complex background patterns or distortions within the impression. This often results in dominant frequencies, which are not a part of the friction ridge impression.

3 Fingermark scale prediction

In this section, we present our approach for determining the scale of fingermark images (in terms of pixels-per-inch, PPI). In order to retain image scale and aspect ration, we propose a patch-based approach to determine the PPI. The approach consists of three main components: (a) Sampling of local patches from the input image, (b) a CNN model, which then predicts PPI of individual patches, and (c) an aggregator function, which joins the individual predictions into the final PPI value. Our approach is intentionally designed to be simple and efficient. Traditional preprocessing techniques, such as segmentation, often rely on knowing the image resolution (PPI) in advance. To avoid this dependency, our method operates without any preprocessing and offers an assumption-free solution.

3.1 Patch sampling strategy

The first step in our approach involves extracting local patches from the input fingermark image. Since our goal is to estimate the image scale (PPI) with minimal prior assumptions, we design a flexible sampling strategy that adapts to different potential fingermark placements. The only assumption we make is regarding the approximate position of the fingermark within the image.

We define two distinct spatial sampling strategies: Gaussian sampling and uniform sampling. The choice between these strategies depends on whether the fingermark is centrally located with surrounding background or spans the entire image.

Gaussian sampling. If the impression is centered in the image with background regions around it, we employ a Gaussian distribution to sample patches. The probability of selecting a patch decreases with distance from the image center, ensuring that the majority of patches capture friction ridge details rather than background noise. This strategy is particularly effective when images contain substantial

non-fingermark regions, as it prioritizes high-information areas. In Fig. 1, we illustrate the size of local patches on images with different scale as well as the effect of the σ parameter.

Uniform sampling. If the impression extends across the entire image, a uniform sampling strategy is more appropriate. In this case, patches are sampled randomly across the entire image without any spatial bias. This ensures a more balanced representation of different regions, especially when there is no clear separation between the fingermark and the background.

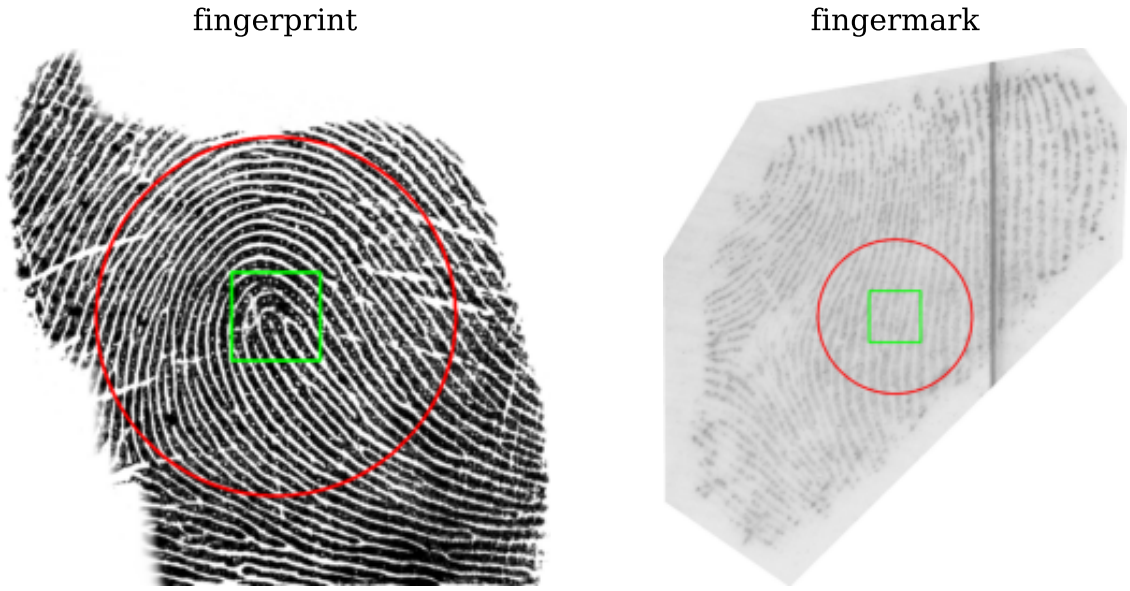


Figure 1: Example of local patch extraction from an image. The green square represent the size of a local patch. The perceived difference in size comes from the difference between PPI values of these images. The red circle indicates the size of the σ parameter of the Gaussian distribution, which cam be changed based on the characteristics of the input data.

With the dynamic selection of the sampling strategy we aim to avoid strong assumptions about the input image. The sampling strategy can be selected based on the general characteristics of expected input data, the type of sensor device or technique, used to capture the image. In Fig. 2, we show several patches, which were samples from images with different PPI values. There is also a clear difference in the quality of patches extracted from fingerprint and fingermark images. Fingerprint images have better contrast and a well-defined structure of ridges with minimal noise and no background interference. In contrast, fingermark images, which are often captured on various underlying surfaces, tend to be more distorted, low contrast and have ambiguous structural characteristics.

3.2 CNN-based PPI predictor

The second component of our approach is a lightweight convolutional neural network (CNN) designed to estimate the pixels-per-inch (PPI) of individual image patches. The model we have chosen is the compact MobileNetV3 [3], which offers a similar performance to its contemporaries at the benefit of a smaller model size and computational complexity.

Design considerations. The choice of MobileNetV3 is driven by the need for a lightweight model that can operate efficiently in real-world forensic scenarios. Our PPI prediction pipeline is intended to be integrated into existing automated fingerprint identification system (AFIS), or into resource-constrained capture devices, which can be used directly in the field. To support these use-cases, minimizing computational overhead is crucial. The CNN operates at the patch level, meaning that it receives individual local patches extracted from the input fingermark image and predicts a local PPI value for each patch. Compared to traditional analytical approaches that rely on Fourier Transform, a CNN-based predictor offers a much more robust way to extract frequency information, even from noisy, distorted patches, or regions with highly curved ridge patterns.

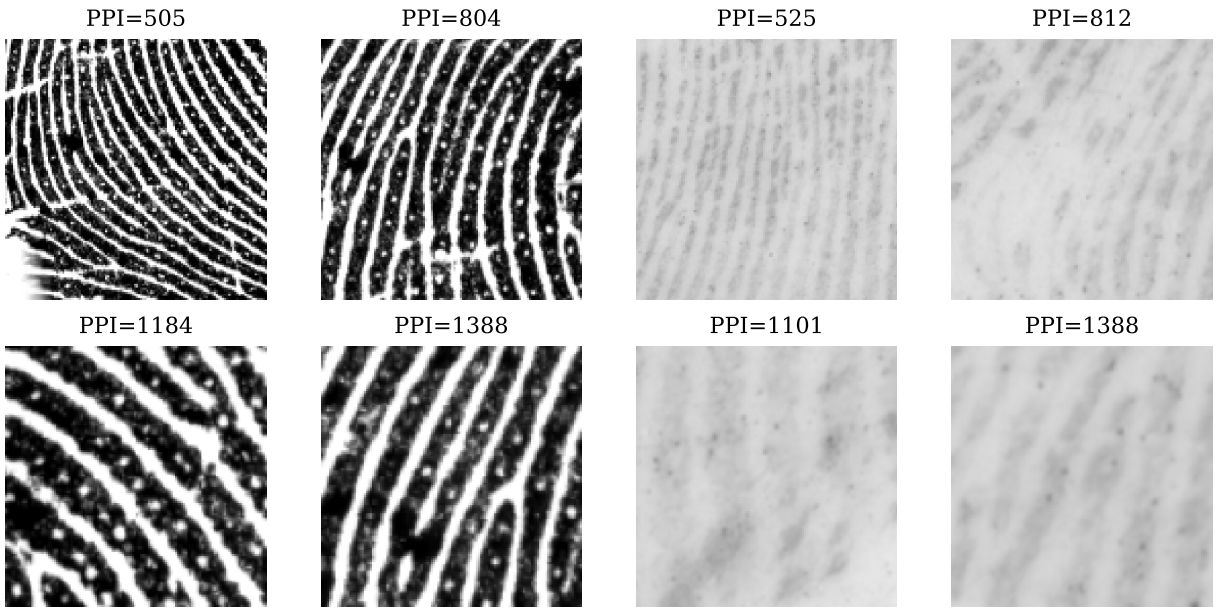


Figure 2: Illustration of local patches extracted from a rolled fingerprint image (left) and a fingermark (right). Each patch effectively captures the frequency of ridges in a local area. These patches also show the added complexity when processing fingermark images.

Learning objective. Due to the stochastic sampling process, some extracted patches might contain little to no visible ridge pattern. The CNN predictor is thus trained to estimate the PPI of local patches by minimizing the Mean Absolute Error (MAE), also known as L1 loss. The loss function is specifically chosen for its robustness to outliers and the interpretable error measurements in regression tasks.

Given a training dataset of N image patches, where each patch x_i has a corresponding ground-truth PPI label y_i, the network learns to predict the estimated PPI value $\hat{y}_i$. The L1 loss function is defined as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{1}$$

Model training for involves two phases. Initially, we train the model using high-quality fingerprint images. In the second phase, we continue with fine-tuning the model specifically for fingermark PPI detection. Opting to train the entire model is motivated by the importance of the initial layers, which concentrate on low-level features like edges, aligning well with our problem domain.

3.2.1 Score aggregation and inference

The local PPI estimates produced by the CNN are then passed to the aggregation function, which combines these predictions to determine the final PPI value for the entire image. The choice of aggregation function is flexible and depends on factors such as the distribution of PPI values across patches and the number of samples extracted. As an initial approach, we propose using either the mean or median of the predicted values. The mean value is better-suited for images where the ridge pattern remains relatively consistent across patches. In contrast, the median offers more robustness to outliers and skewed distributions, which may appear due to noise, partial fingermarks, or varying image quality.

Fig. 3 illustrates the final pipeline. During the inference phase, we sample N patches from each input image. Each sampled patch is passed through the MobileNetV3 model, which outputs an individual PPI estimate. When utilizing a GPU, this step can be further optimized by processing all extracted patches as a mini-batch, allowing them to be processed by the CNN in parallel. The set of predicted PPI values is then combined using a suitable aggregation function to yield the final PPI

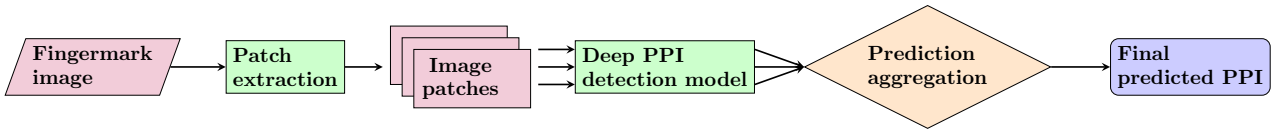


Figure 3: Illustration of the predictive pipeline. We use a pretrained MobileNet v3 model. We fine-tune it sequentially on fingerprint data and then on fingermark data. The model is designed to process patches extracted from both fingerprint and fingermark images, and generates a PPI score per patch.

estimate for the entire image.

4 Experiments

In this section, we present our experiment evaluation. First, we outline our implementation and hardware details, then introduce the dataset and the evaluation metrics. We evaluate out approach only on patches as well as on entire fingermark and fingerprint images. We also compare out CNN-based solution to two alternative methods that utilize Fourier Transform and present some qualitative results.

4.1 Experimental setup

The predictive pipeline is implemented in Python, with the MobileNetV3 model sourced from the torchvision library [3, 9]. For optimization, we employ a gradually decreasing learning rate using *ReduceLROnPlateau* from PyTorch, starting with an initial learning rate of 10^{-4}. To prevent overfitting, we apply early stopping, setting the patience parameter to 10. All deep learning models were trained and evaluated on an NVIDIA GeForce RTX 3080 GPU with 10GB of memory.

4.2 Datasets and metrics

We develop and evaluate our approach using the NIST SD302 dataset [1], which includes both fingerprint and fingermark images. The dataset comprises approximately 10,000 fingermarks, along with an even larger number of rolled and plain fingerprints. The fingermarks were developed and captured by trained forensic examiners in a controlled, simulated environment at scales ranging from 1042 to 1433 PPI, closely resembling real-world operational conditions. In contrast, the fingerprints were acquired using various sensor devices at fixed resolutions of 500 or 1000 PPI.

In total, we use 10.920 fingerprint images and 2525 fingermark images. Since the SD302 contains a substantial number of fingermark images without visible ridge structure, we only selected fingermarks from a subset of fingermarks, for which NIST also provides manual minutiae annotations, meaning these contain at least some visible ridge structure. We reserve 20% of all images as a **holdout test set** $\mathcal{M}$, which we use to evaluate the final pipeline. The rest are then used for the development of models and interim evaluations on the image patches. Specifically, we use these images to create a set of patches, which are then used to train the patch-level CNN. Similarly, we also create a **test set of individual patches** $\mathcal{P}$. To artificially inflate the set of patches, we use data augmentation techniques, which are described in detail the following section.

To evaluate the performance of the trained models, we employ Mean Absolute Error (MAE), which gives us a more explicit idea of how large the error is within the expected output range.

4.3 Data Preparation

Patch extraction differs between the training and inference phases to best suit their respective objectives. During training, the primary goal is to ensure that the extracted patches best capture the full range of PPI values while focusing on regions where the ridge pattern is clearly visible. The range of PPI values we aim to predict falls between 500 and 1500. However, fingerprint images are typically available only at fixed PPI values. Since our goal is to estimate any PPI value within this range, we

employ data augmentation to create a more continuous and nearly uniform distribution of PPI values. We sampled the desired PPI for each patch from a uniform distribution spanning 500 to 1500. The area of the local patch around the randomly sampled location is determined based on the source and target PPI values:

$$s = \left\lfloor \frac{PPI_{orig}}{PPI_{gen}} \times p \right\rfloor \tag{2}$$

where s is the width and height of the patch, PPI_{orig} is the original PPI of the image, PPI_{gen} is the newly generated PPI, and p is fixed patch size. We generated 25 patches, where each patch is assigned one randomly generated PPI value from the target range. In contrast, during inference, patch extraction follows a general approach for all types of data to avoid making assumptions about the input image. To find the ideal inference parameters, we experiment with different patch sizes, patch extraction techniques, and number of patches per image.

We also apply various image augmentations, such as contrast and brightness, to better simulate the appearance variations commonly observed in fingermark images. We also experimented with augmenting patch orientation, but have observed no added benefits in the final evaluation.

4.4 Baseline methods

To evaluate the effectiveness of our approach, we compare it against two traditional approaches based on Fast Fourier Transform (FFT), which are commonly used for estimating ridge frequencies in forensic and biometric applications. These baseline methods, referred to as FFT1 and FFT2, replace the CNN predictor in our pipeline while keeping the overall framework unchanged.

- *FFT1* follows a ridge alignment and 1D frequency estimation approach. First, the orientation of a local image patch is determined based on image gradients and the orientation of the principal variation axis. Once aligned, the patch is averaged across all columns, reducing the 2D ridge pattern to a 1D sine wave representation. A 1D Fast Fourier Transform (FFT) is then applied to extract the dominant frequency, corresponding to the ridge spacing in the original image. While FFT1 is computationally efficient, it relies heavily on the accuracy of the ridge alignment step, which can be unreliable in the presence of noise, poor contrast, or irregular ridge patterns.

- *FFT2* eliminates the need for ridge alignment by directly analyzing the spatial frequency content of the local patch using a 2D Fast Fourier Transform (2D FFT). In this approach, the patch is extracted without orientation adjustments, and the 2D FFT is computed over the entire region, transforming the spatial domain into the frequency domain. The power spectrum of the transformed image is then analyzed to identify the dominant frequency. Although FFT2 is more robust to variations in ridge orientation, it is computationally more expensive due to the higher-dimensional FFT operation.

Both methods output the period of the sine signal, detected within the given image window, or the inverse of the frequency of ridges. To convert the period to PPI value, we apply following formula:

$$PPI = win\_size \times \frac{C}{f}, \tag{3}$$

where C is a constant that represents the average number of ridges per inch in the adult population, *win_size* represents size of the window, and f represents the sine wave period based on respective FFT method. The parameter C represents the mean value from several statistical studies conducted on a larger population size and is fixed to 58.9 ridges per inch [5].

4.5 Quantitative analysis

In this section, we evaluate the performance of our proposed approach and baseline methods using the Mean Absolute Error (MAE) regression metric. We first assess the accuracy of PPI predictions on individual extracted patches. Then, we extend the evaluation to full fingerprint and fingermark images, where we use different sampling strategies and aggregation techniques.

Patch-level performance. Since we have fingerprint and fingermark images, we train the model in different ways. We experimented with three different training strategies:

- *Fingermark only.* The model is trained exclusively on fingermark patches, learning directly from the target domain without any influence from fingerprint images.
- *Combined Dataset.* The model is trained on a merged dataset consisting of both fingermark and fingerprint patches, allowing it to generalize across both image types from the start.
- *Sequential Fine-Tuning.* The model is first trained on fingerprint patches, then fine-tuned on fingermark patches. In this case, them model first learns on the structured ridge patterns of fingerprints before adapting to the noisier fingermark domain.

In addition to the training strategy, we apply data augmentation to either one, both, or neither of the training subsets. It is important to note that augmentation is applied exclusively during training and is not used for validation or testing. All models are trained, validated, and tested on patches of fixed size. The final evaluation is conducted using the holdout test set of patches $\mathcal{P}$, with results summarized in Table 1.

First, we observe that larger patch sizes lead to better results. Patch size of 160×160 pixels performs best, but after that we observed diminishing results. The better performance of larger patches may be due the added relevant information. Particularly due to the noisy nature of fingermark images, the likelihood of capturing high-quality fingermark regions is increased with larger patch size. When evaluating the training strategy, sequential fine-tuning performed similar to the combined dataset strategy, highlighting the importance of including fingerprint data at some stage in the development of models. In contrast, training only on fingermarks achieved worse results in any other experiment. Image augmentation was found to be beneficial, but only when applied to fingerprints. Augmenting fingermarks may degrade the already sparse information present in fingermark images. In all evaluation stages, the 95% confidence intervals were narrow, indicating that the results were consistent and the predictions are made with high confidence. The best result achieves a MAE value of 22.27 within the range of possible PPI values from 500 to 1500.

To further analyze the best-performing configuration, we present the absolute errors across the entire PPI range in Figure 4(a). The error statistics are calculated within discrete PPI sub-ranges (e.g., 500–599, 600–699, etc.), with each sub-range containing approximately 2,500 patches. The results indicate that the MAE within each sub-range remains relatively consistent with the overall mean absolute error. However, we observe a slight improvement in prediction accuracy for patches with PPI values above 1000. This could be attributed to the larger patch size used in this configuration. At lower PPI values, patches cover a larger relative area of the impression, resulting in a denser ridge structure with increased curvature, which makes accurate PPI estimation more challenging. In contrast, at higher PPI values, the same patch size corresponds to a more localized region of the impression, containing fewer ridges and less curvature, leading to more precise predictions.

Image-level performance. Based on the results of the patch-level evaluation, we continue our experiments using the best-performing configuration identified in the previous section. For this experiment, we assess the entire prediction pipeline on full fingerprint and fingermark images. The goal of this evaluation is to (a) select the optimal sampling strategy, (b) determine the ideal number of local patches per image, and (c) identify the most effective aggregation method for combining patch-level predictions into a final PPI estimate.

In Table 2, we present the results of the full-image evaluation. We first compare various sampling techniques. The Gaussian sampling strategy assumes that the fingermark is centered in the image. To achieve this, we set the standard deviation of the Gaussian distribution to $\sigma = \left\lfloor \frac{\min(h,w)}{6} \right\rfloor$, where h and w represent the image height and width, respectively. Our findings suggest that uniform sampling yields the best results in general, with 25 patches per image being sufficient for accurate predictions when using median aggregation. This is likely due to the relatively large patch size (160×160), which can lead to redundant patches when extracting a higher number. Comparing Gaussian and uniform sampling, we observe only minor differences in performance, as increasing the number of extracted patches causes the results to converge. This occurs because a larger number of patches increases the

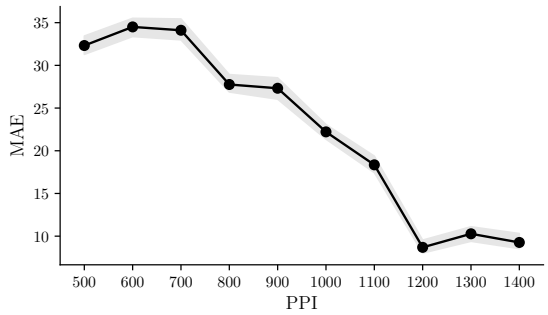
Patch size	Dev. strategy	Augmentation	MAE	95% confidence intervals
64x64	Fingermark Only	None	46.96	(46.41, 47.56)
64x64	Fingermark Only	Fingermark	56.26	(55.55, 56.92)
64x64	Combined Dataset	None	39.98	(39.37, 40.64)
64x64	Combined Dataset	Fingerprint	44.72	(44.17, 45.33)
64x64	Combined Dataset	Both	51.66	(51.01, 52.41)
64x64	Sequential Fine-tuning	None	38.67	(38.13, 39.20)
64x64	Sequential Fine-tuning	Fingerprint	44.45	(43.85, 45.14)
64x64	Sequential Fine-tuning	Both	50.38	(49.71, 50.99)
128x128	Fingermark Only	None	29.53	(29.14, 29.91)
128x128	Fingermark Only	Fingermark	40.43	(40.01, 40.89)
128x128	Combined Dataset	None	25.34	(24.97, 25.69)
128x128	Combined Dataset	Fingerprint	25.05	(24.71, 25.43)
128x128	Combined Dataset	Both	32.91	(32.44, 33.32)
128x128	Sequential Fine-tuning	None	28.12	(27.73, 28.53)
128x128	Sequential Fine-tuning	Fingerprint	27.67	(27.32, 28.02)
128x128	Sequential Fine-tuning	Both	32.54	(32.11, 33.02)
160x160	Fingermark Only	None	26.68	(26.27, 27.06)
160x160	Fingermark Only	Fingermark	36.16	(35.74, 36.53)
160x160	Combined Dataset	None	23.94	(23.62, 24.26)
160x160	Combined Dataset	Fingerprint	22.53	(22.21, 22.85)
160x160	Combined Dataset	Both	31.81	(31.36, 32.23)
160x160	Sequential Fine-tuning	None	24.56	(24.17, 24.94)
160x160	**Sequential Fine-tuning**	**Fingerprint**	**22.27**	**(21.92, 22.59)**
160x160	Sequential Fine-tuning	Both	25.89	(25.50, 26.23)

Table 1: Comparison of different model types for determining PPI on patch dataset $\mathcal{P}$. We present results with MAE metric and estimation of 95% confidence intervals with bootstrapping technique. The best result is in bold.

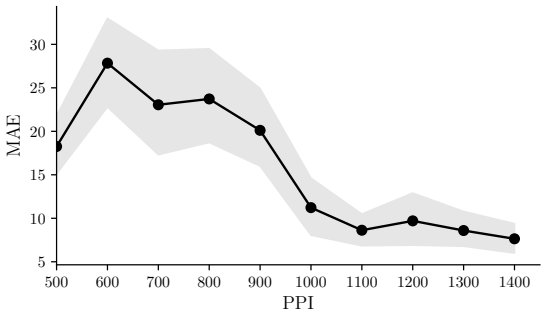
probability of obtaining similar patches. As expected, increasing the number of sampled patches from 2 to 10 reduces error uncertainty, but beyond this point, the performance stabilizes. The confidence intervals are quite narrow, indicating that the model performs reliably. An important observation is that image-level aggregation achieves a lower average error compared to patch-level predictions within individual sub-ranges. This leads to the conclusion that our initial decision to extract patches and subsequently aggregate PPI predictions was the correct approach for a more generalized PPI estimation.

In Fig. 4(b), we present the MAE values calculated on an full-image basis using the best sampling strategy (10 patches, median aggregation, uniform sampling distribution). The results show no significant deviations from the overall image-wise MAE. Similar to patch-level prediction performance in Fig. 4(a), the error decreases as PPI increases, stabilizing once PPI exceeds 1000. However, we observe a lower overall MAE across the entire range with slightly higher variation, suggesting that while full-image predictions improve accuracy, they introduce some uncertainty.

Finally, we compare our approach to the selected baseline method on the full-image dataset $\mathcal{M}$. The parameters of baseline methods are searched independently from the proposed pipeline to enable a fair comparison. In the end, optimal results for baseline methods were obtained with window size 64 and median aggregation function. The baseline performance is shown in Table 3. We observe that our approach outperformed both baseline methods. It is also worth mentioning that baseline



(a) MAE across PPI values for the $\mathcal{P}$ dataset, categorized into discrete sub-ranges of PPI from 500 to 1500. Bootstrap-estimated confidence intervals are in gray. Patch-wise MAE is ∼22.



(b) MAE across PPI values for the $\mathcal{M}$ dataset, categorized into discrete sub-ranges of PPI from 500 to 1500. Bootstrap-estimated confidence intervals are in gray. Image-wise MAE is ∼16.

Sampling distr.	**# patches**	**Aggregation**	**MAE (per image)**	**95% CIs (per image)**
Gaussian	2	Mean	21.86	(19.82, 23.72)
Gaussian	10	Mean	18.15	(16.34, 20.06)
Gaussian	25	Mean	17.61	(16.06, 19.03)
Gaussian	50	Mean	17.36	(15.76, 18.66)
Gaussian	10	Median	17.75	(16.12, 19.45)
Gaussian	25	Median	16.77	(15.14, 18.29)
Gaussian	50	Median	17.74	(15.97, 19.18)
Uniform	2	Mean	20.16	(18.21, 22.29)
Uniform	10	Mean	17.44	(16.03, 18.98)
Uniform	25	Mean	17.52	(16.10, 19.08)
Uniform	50	Mean	17.90	(16.36, 19.46)
Uniform	**10**	**Median**	**16.30**	**(14.86, 17.71)**
Uniform	25	Median	16.74	(15.16, 18.52)
Uniform	50	Median	16.49	(15.10, 18.10)

Table 2: Comparison of model's performance with different patches sampling techniques on fingermark dataset $\mathcal{M}$. We present results with MAE and estimation of 95% confidence intervals with bootstrapping technique. The best result is in bold.

methods require plenty of parameters' fine-tuning and still obtain suboptimal results compared to our end-to-end approach. Additionally, due to the nature of fingermarks and fact that certain patches of an image may not contain ridges, baseline methods may also completely fail in those scenarios.

4.6 Qualitative analysis

In this section, we analyze our method's performance on both high-quality and low-quality fingermarks. We begin by examining patches with high absolute error in Fig. 4. These patches tend to share common characteristics, such as poorly defined ridge structures, very low contrast, or the presence of mixed structural elements, which make accurate PPI estimation challenging. Additionally, the model struggles with dark and low contrast patches, where ridge details are less distinguishable. Notably, many of the patches with high prediction errors correspond to background regions, which have been excluded from the visualization. In contrast, Figure 5 shows patches with near-zero absolute error. These patches feature higher contrast and well-defined ridge structures, making it easier to distinguish ridges and valleys. Importantly, we observe that the model does not require fully continuous ridge structures to make accurate predictions—partial but clear ridge patterns are sufficient for reliable PPI

Method	**MAE (per image)**	**95% CIs (per image)**
FFT1	248.58	(241.70, 256.29)
FFT2	268.53	(260.73, 276.35)
Ours	**16.30**	**(14.86, 17.71)**

Table 3: Comparison of ours and baseline's performance on fingermark dataset $\mathcal{M}$. We present MEA values and estimation of 95% confidence intervals with bootstrapping technique. The best result is in bold. The parameters for all methods are the same.

estimation.

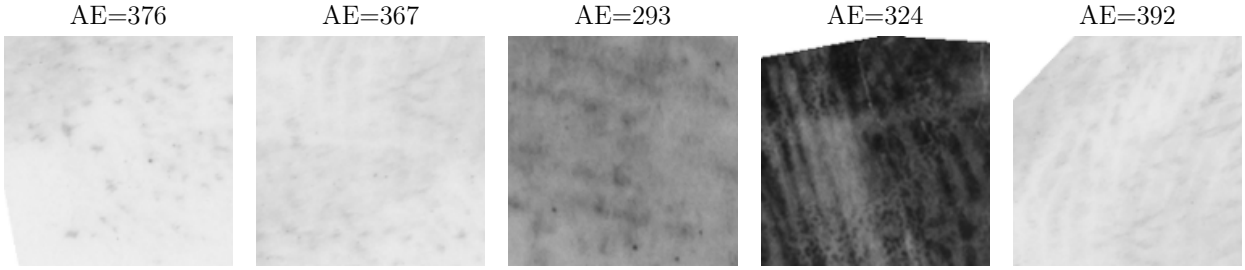


Figure 4: Examples of fingermark images with high absolute error (AE) values. Corresponding error values are indicated above patches. MAE (computed per patch) on this dataset is 25.

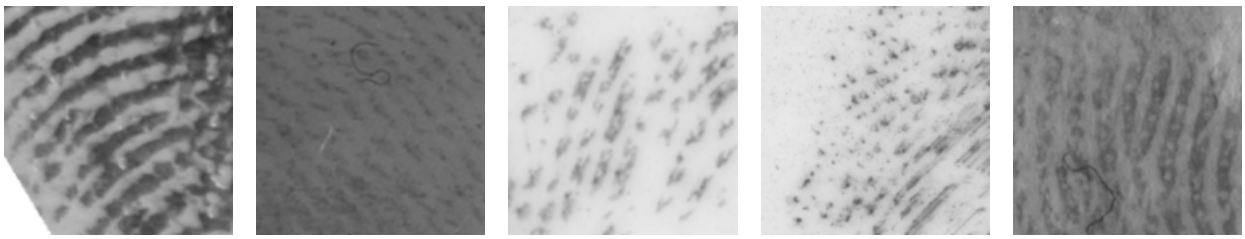
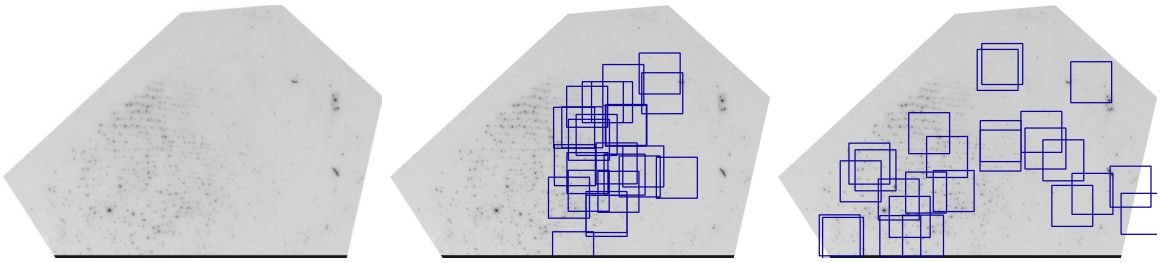


Figure 5: Examples of fingermark images with near-zero absolute error values. MAE (computed per patch) on this dataset is 25.

Additional filtering of patches could significantly reduce the number of irrelevant background regions, included in the aggregated PPI value. Gaussian sampling assumes that the fingermark is centered in the image, but this is not always the case. When this assumption fails, the model may extract patches from the background instead of the fingermark, leading to erroneous PPI estimates. On the other hand, uniform sampling makes no assumptions about fingermark placement, which works best for fingerprints, where the impression area is clearly defined. A comparison of these two sampling strategies is illustrated in Figure 6, where we highlight cases where Gaussian sampling resulted in background patches. As observed in the SD302 dataset, fingermark images often exhibit irregular and non-centered shapes, making it difficult for the Gaussian assumption to hold. In many cases, the fingermark structure is spread unevenly across the image, further complicating the effectiveness of fixed sampling strategies.

4.7 Ablation study

In this section, we evaluate whether zero-shot transfer from fingerprint to fingermark data is feasible by removing any fine-tuning on fingermark patches. The model is trained exclusively on the



(a) Fingermark image (b) Gaussian sampling (c) Uniform sampling

Figure 6: Example of a fingermark image where patch extraction resulted in poor patches, leading to the PPI being predicted from the background.

fingerprint subset of $\mathcal{P}$. We refer to this configuration as *Fingerprint Only*. The results are presented in Table 4. We observe a significant decline in performance compared to previous pipeline configurations. This demonstrates the limitations of directly applying fingerprint-trained models to fingermarks and underscores the importance of domain-specific fine-tuning.

Dev. strategy	**Augmentation**	**MAE**	**95% confidence intervals**
Fingerprint Only	None	183.56	(181.61, 185.20)
Fingerprint Only	Fingerprint	250.05	(247.66, 252.51)

Table 4: Comparison of models trained on the fingerprint subset of $\mathcal{P}$ and evaluated on fingermark images. We present MAE and estimation of 95% confidence intervals with bootstrapping technique.

4.8 Speed analysis

On the specified hardware, running the PPI detector on a single patch takes approximately 5 ms on GPU and 7 ms on CPU. When processing a full image with 10 patches, the total runtime averages 70 ms on GPU and 80 ms on CPU. The GPU does not provide a significant speedup in this case due to the overhead of data transfer. In contrast, batch processing on the GPU significantly accelerates the execution. By setting the batch size equal to the number of patches per image, the CNN predicts the PPI of all patches in parallel, resulting in a substantial speedup. In this optimized configuration, the entire pipeline runs in approximately 15 ms, regardless of the number of patches used in the process.

5 Conclusion

In this work, we introduced a CNN-based approach for estimating the PPI of fingermark images, addressing a critical step in forensic fingerprint processing. Unlike traditional frequency-based methods, our approach learns directly from local patches, which results in robust PPI predictions across a wide variety of ridge impressions.

We evaluated our model at both patch and full-image levels, optimizing sampling strategy, patch count, and aggregation methods. The results show the effectiveness of a larger patch size, uniform sampling and median aggregation when dealing with a generalized approach to PPI estimation. Finally, we demonstrated that our method outperforms FFT-based baselines, particularly in cases where fingermarks exhibit noise, distortion, or varying ridge structures. Our findings suggest that future research should focus on more advanced domain adaptation techniques as well as additional filtering of patches, which would minimize the impact of background regions on PPI estimation.

Acknowledgements

Supported by ARIS P2-0250, P2-0214.

References

[1] Fiumara, G.; Flanagan, P.; Grantham, J.; Ko, K.; Marshall, K.; Schwarz, M.; Tabassi, E.; Woodgate, B.; Boehnen, C., "NIST Special Database 302: Nail to nail fingerprint challenge", *National Institute of Standards and Technology*, 2019.

[2] He, K., Zhang, X., Ren, S., Sun, J., "Deep Residual Learning for Image Recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[3] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., Le, Q., "Searching for MobileNetV3", *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, 2019.

[4] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., "Densely Connected Convolutional Networks", *arXiv*, 1608.06993, 2018.

[5] Jain, A., Ross, A., Nandakumar, K., Swearingen, T., "Fingerprint recognition", *Introduction To Biometrics*, pp. 75–117, 2011.

[6] Kunsuk, S., Areekul, V., "Finger Photo Rescaling for Interoperability of Touchless and Touch-based Fingerprint Verification", *International Conference on Signal-Image Technology & Internet-Based Systems*, pp. 168–175, 2023.

[7] Oblak, T., Haraksim, R., Peer, P., Beslay, L., "Fingermark quality assessment framework with classic and deep learning ensemble models", *Knowledge-Based Systems*, 250, 2022.

[8] Oblak, T., Haraksim, R., Beslay, L., Peer, P., "Probabilistic Fingermark Quality Assessment with Quality Region Localisation", *Sensors*, 23, 2023.

[9] PyTorch [Online]. Available: https://github.com/pytorch/vision (Accesed on 29 January 2025).

[10] Ren, C.-X., Yin, Y.-L., Ma, J., Li, H., "Fingerprint scaling", *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues: International Conference on Intelligent Computing*, pp. 474–481, 2008.

[11] Ren, C., Guo, J., Qiu, D., Chang, G., Wu, Y., "A framework of fingerprint scaling", *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 11(3), pp. 1547–1559, 2013.

[12] Sharma, S., Shrestha, R., Krishan, K., Kanchan, T., "Sex estimation from fingerprint ridge density. A review of literature", *Acta Bio Medica: Atenei Parmensis*, 92(5), 2021.

[13] Swofford, H., Champod, C., Koertner, A., Eldridge, H., Salyards, M., "A method for measuring the quality of friction skin impression evidence: Method development and validation", *Forensic Science International*, 320, 2021.

[14] Tabassi, E., Olsen, M., Bausinger, O., Busch, C., Figlarz, A., Fiumara, G., Henniger, O., Merkle, J., Ruhland, T., Schiel, C., Schwaiger, M., "NIST Fingerprint Image Quality 2", *National Institute of Standards and Technology*, 2021.

[15] Yoon, S., Cao, K., Liu, E., Jain, A.K., "LFIQ: Latent fingerprint image quality", *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1–8, 2013.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Oblak, T.; Videnović, J.; Kupinić, H.; Štruc, V.; Peer, P.; Emeršič, Ž. (2025). Fingerprint image scale estimation for forensic identification systems, *International Journal of Computers Communications & Control*, 20(2), 7031, 2025.
https://doi.org/10.15837/ijccc.2025.2.7031