Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Automated Biometric Identification using Dorsal Hand Images and Convolutional Neural Networks

To cite this article: Mahsa Mohaghegh and Ash Payne 2021 *J. Phys.: Conf. Ser.* **1880** 012014

Automated Biometric Identification using Dorsal Hand Images and Convolutional Neural Networks

Mahsa Mohaghegh[1, a, *] and Ash Payne[1, b]

[1]Auckland University of Technology, Auckland, New Zealand

[a, *] mahsa.mohaghegh@aut.ac.nz, [b]ash@diveunlimited.com.au

Abstract. The identification of perpetrators, present in Child Sexual Abuse Imagery (CSAI), is a significant challenge due to the use of anonymisation techniques that mask their identities. Consequently, researchers have investigated the use of uncommon biometric identifiers such as knuckle patterns, palmprints and the dorsal side of the hand. This research proposes a Convolutional Neural Network (CNN) based, fully automated approach to biometric identification using dorsal hand images. The identification performance of three different CNN architectures, AlexNet, ResNet50 and ResNet152, is experimentally determined against two similar datasets, the 11k Hands and IITD dorsal hand databases. A transfer learning approach is used and the final output layers of the CNNs are modified to match the number of classes present in the datasets. The results showed that ResNet CNNs achieved identification accuracies greater than 99.9% on both datasets, whereas the AlexNet CNN achieved between 80.1% and 93.7%. These results demonstrate that it is feasible to use deep, off-the-shelf CNNs, such as ResNets, for automated biometric identification using dorsal hand images. This highlights the potential of using dorsal hand images to identify perpetrators of child sexual abuse from CSAI.

1 Introduction

The proliferation of mobile devices capable of capturing high quality images and video has, in recent years, led to the exponential growth of Child Sexual Abuse Imagery (CSAI) reports. A United States Department of Justice report published in 2017 stated that between 2004 and 2013, 72% of all federal prosecutions for child exploitation were for the possession or distribution of child pornography [1]. Further to this, Bursztein et al. [2] note that of the 23.4 million reports of CSAI lodged between 1998 and 2017, 9.6 million of those reports occurred in 2017 alone. Since 2013 staff at the National Center for Missing & Exploited Children (NCMEC) have been categorising CSAI content into one of three file formats, images, videos and other [2]. By 2018, the NCMEC had identified 20,568,240 images, 2,335,536 videos and 337,575 other files [2]. The scale of the problem creates a need for law enforcement agencies and associated organisations to find and prosecute those that create this material and perpetrate the abuse. However, existing processes for identifying both perpetrators and victims present in CSAI rely on manually reviewing the content. According to Matkowski, Chan, and Kong [3], since 2002 the NCMEC has identified more than 10,900 child abuse victims by manually reviewing more than 164 million images and videos. Law enforcement agencies and other organisations face significant challenges when it comes to identifying the abusers featured in the images and videos, due to the use of anonymisation methods. Perpetrators often conceal their identities by hiding or obscuring their faces and other distinguishing features [3]. Since distinguishing features such as faces or tattoos are often concealed, researchers have been investigating the use of other biometric modalities for forensic identification. These include finger knuckle patterns [4], palmprints [5], and the dorsal side of the hand [6].

Researchers have long been interested in the dorsal region of the hand for forensic image comparison. However, this has traditionally involved a manual process carried out by trained experts. One such process has been used in the UK criminal justice system to secure 28 life sentences against perpetrators of child sexual abuse [7]. This system used the unique vein patterns, skin folds and blemishes present on the hands and forearms to identify child abusers. However, this system has limitations as it is resource intensive and can only be employed by trained professionals. Therefore, as can be seen from the volume of material that requires review, an automated approach is required to investigate CSAI in a timely manner.

2 Related work

The study of computer vision, a subset of machine learning, provides a potential solution to the problem of biometrically identifying suspects of child abuse offences. Computer vision techniques use a combination of feature extraction and classification to achieve biometric identification. Feature extraction can be performed using manual, handcrafted methods, or automated, deep learning methods. Classification can then be realised using machine learning classifiers such as Support Vector Machines (SVMs) or neural network classification layers such as the Softmax classifier. Recent research into biometric identification using the dorsal side of the hand tends to focus on the unique vein patterns present. Examples from the literature that use these vein patterns show a combination of approaches that include using handcrafted features with neural network classifiers [8], neural network feature extraction with machine learning classifiers [9], or end-to-end neural network feature extraction and classification [9], [10], [11]. The studies that use neural networks for either feature extraction or end-to-end extraction and classification tend to favour Convolutional Neural Networks (CNNs). This is because CNNs are well suited to the problem of image classification and several architectures have previously been developed for this purpose. Another commonality among the studies that use dorsal hand vein patterns is the use of Near Infrared (NIR) light to increase the visibility of the veins. This is because the haemoglobin in the blood absorbs this wavelength of light making the veins appear black [12]. However, this approach does not suit the scenario of identifying suspects from video footage unless that footage has been captured using a night vision mode. Instead, the footage is more likely to be shot using natural or artificial white light. However, studies using images of the dorsal side of the hand taken in natural or artificial white light are lacking. Only one such study was found in the literature, in the research by Afifi [6] the author proposes the use of an AlexNet CNN to extract features to feed an ensemble of SVM classifiers.

3 Methods

The recent studies conducted using dorsal hand images for biometric identification use either a single CNN architecture, seek to compare the performance of several existing architectures, or propose novel architectures that achieve similar identification accuracies. The commonly used existing CNN architectures are AlexNet [13], VGG-16 [14], VGG-19 [14], and ResNets [15]. This research compares the biometric identification performance of three, pre-trained CNNs, AlexNet, ResNet50, and ResNet152. These CNNs have been chosen to investigate biometric identification using dorsal hand images since AlexNet approaches [6], [9] and ResNet approaches [10] have previously demonstrated high identification accuracies. For example, in [9] the authors used transfer learning with the AlexNet CNN to achieve an identification accuracy of 95.51%. In addition, in [10], the authors demonstrated that an identification accuracy of 99.51% could be reached using ResNet50.

3.1 Transfer learning approach

To produce CNN models suited to the problem of biometric identification using dorsal hand images, a transfer learning approach was employed. Transfer learning refers to the process of fine-tuning a CNN model that has been pre-trained on a dataset that relates to a different, but similar, classification task.

Examples from the literature demonstrate that transfer learning is an effective research technique for the evaluation of CNN-based biometric identification. For example, research by Minaee, Azimi, and Abdolrashidi [16] uses transfer learning to fine-tune the pre-trained ResNet50 model for biometric identification using fingerprints. Similarly, Minaee and Abdolrashidi [17] apply the same technique to the problem of biometric identification using iris images. The literature also includes examples of using transfer learning to repurpose several other pre-trained CNNs and apply them to the problem of biometric identification using dorsal hand images [6] [9] [10]. These papers cite their use of transfer learning as being due to the relatively small datasets in use which make it impractical to fully retrain a CNN for the new task. As this research also replies on a relatively small dataset, a similar approach was used.

3.2 Datasets
The 11k Hands dataset, previously collected by Afifi [6], is the primary dataset that will be used for both model training and evaluation. The dataset contains a total of 11,076, high resolution (1600x1200), natural light images collected from 190 subjects of varying ages (18-75 years old). The dataset contains images of both the dorsal and palmer sides of the left and rights hands of the subjects. The author of [6] states that the dataset contains an average of 58 images per subject with an average of 30 images of the dorsal side and an average of 28 images of the palmer side of the hands.

The Hong Kong Polytechnic University Contactless Hand Dorsal Images Database [4] (IITD dataset) is comprised of images taken from both male and female subjects. This database was largely acquired in the IIT Delhi Campus, in The Hong Kong Polytechnic University campus and in some villages in India during 2006-2015. Several devices were used to capture the images including mobile phones and handheld digital cameras. This database has 4650 dorsal hand images from the right hand of 502 different subjects. The images include the dorsal side of the palm and the backs of all four fingers that illustrate the knuckle patterns in each. The images all have a resolution of 1600x1200 pixels and are taken in either natural or artificial white light.

3.3 Data Augmentation
In [6], the author states that the 11k Hands dataset had an average of 30 dorsal hand images per subject. However, analysis of the image distribution showed that the actual number of images per subject varied significantly. For example, one subject had a total of 101 dorsal hand images, whereas another subject had only 8 dorsal hand images. The IITD dataset also showed variation between subjects, although this variation was not as pronounced. Consequently, both datasets can be thought of as imbalanced, and according to Rochac, Zhang, Thompson and Oladunni [18], an imbalanced dataset can lead to a reduced identification accuracy in CNN classification scenarios. To address this imbalance, as well as the relatively small number of images for certain subjects, data augmentation (DA) techniques were applied. These included horizontal flipping, random rotation, and the addition of random noise. Following the application of the DA techniques, the number of samples per class was increased to 100. This resulted in two larger datasets for experimentation, one containing 37,800 images (based on the 11k Hands dataset) and another containing 50,200 images (based on the IITD dataset).

4 Experimental Results
To test the suitability of each CNN architecture for the problem of dorsal hand biometric identification, the three architectures, AlexNet, ResNet50, and ResNet152 were used as the base CNNs. The final output layer of each CNN was modified to create 6 new models, 3 for the 11k Hands dataset and 3 for the IITD dataset. The 3 models to be trained using the 11k Hands dataset were configured to have 378 outputs in the final classification layer, whereas the 3 models to be trained using the IITD dataset were configured to have 502 outputs. The number of outputs correspond with the number of classes in each dataset. When using the 11k Hands dataset, an increase in the average training and testing accuracies was observed after training for 75 epochs compared to 50 epochs. Consequently, when training the 3 models using the IITD dataset a training duration of 75 epochs was used. Since a duration of 75 epochs

was deemed optimal the results of the 50 epoch experiments have been omitted from the following sections. The purpose of training and testing the models on 2 different datasets was to validate the generalisation ability of the 3 CNN architectures to the dorsal hand biometric modality. The images contained in the 11k Hands and IITD datasets have different backgrounds, lighting conditions, and distances from the capturing device and therefore added credibility to the results.

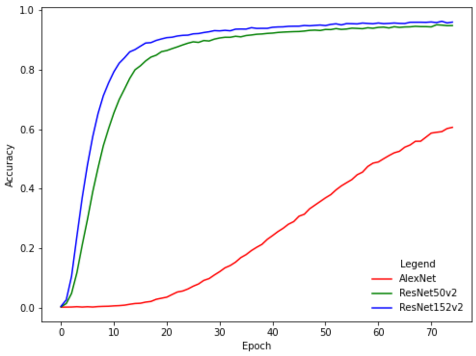

Figure 1. Training accuracies when using the 11k Hands dataset

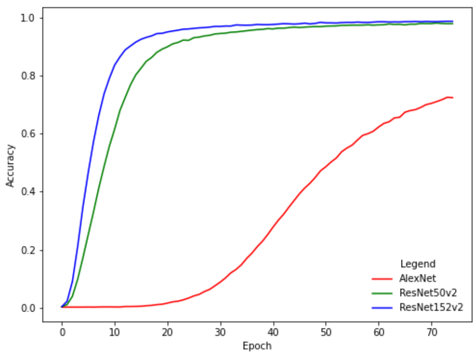

Figure 2. Training accuracies when using the IITD dataset

All experiments were implemented using the Python API of the MXNet deep learning framework using a PC with the following configuration: Windows 10 operating system, Intel Xeon Gold 6240 2.60GHz CPU, 192GB RAM, NVIDIA Quadro RTX 6000 GPU with 24GB dedicated memory.

4.1 Training

As can be seen in figure 1, training the AlexNet model with the 11k Hands dataset resulted in a shallow gradient for the first 10-20 epochs with minimal increase in training accuracy. After this point the accuracy increased more rapidly but not significantly enough to begin plateauing before the training cycle concluded. In contrast, the ResNet50 and ResNet152 models demonstrated a rapid increase in training accuracy during epochs 1 through 10. After this, the gradients begin to level out as the accuracies approached 1 and began to show signs of plateauing. The AlexNet model achieved a final average training accuracy 60.61%, whereas the ResNet50 model achieved 94.77% and the ResNet152 achieved 95.92%.

Figure 2 shows the average training accuracy curves when training the three models with the IITD dataset. As in figure 1, the initial gradient of the curve for the AlexNet model remains shallow for several epochs. However, between epochs 30 and 60 the gradient is steeper than the AlexNet curve in figure 1. As a result, the final average training accuracy for the AlexNet model is higher than in figure 1 at a value of 72.40%. As with figure 1, the accuracy curves in figure 2 for the ResNet models showed a much steeper gradient than for the AlexNet curves. However, the gradients for the ResNet models in figure 2 were steeper than their counterparts in figure 1. As a result, the final average training accuracies for the ResNet models were higher when trained with the IITD dataset. The ResNet50 model reached a value of 97.87% and the ResNet152 model achieved a value of 98.69%. A summary of the final average training accuracy values for each architecture, using each dataset, is given in table 1.

Table 1. The final accuracy and loss values obtained after training for 75 epochs

	11k Hands		**IITD**	
Model	**Accuracy**	**Loss**	**Accuracy**	**Loss**
AlexNet	60.61%	1.601	72.40%	0.952
ResNet50	94.77%	0.3136	97.87%	0.1399
ResNet152	95.92%	0.2383	98.69%	0.07902

Table 1 highlights the fact that the more complex ResNet models provided superior results compared to the AlexNet models. In addition, table 1 shows that the increase in accuracy between the AlexNet models and the ResNet models was significant (between 6% and 20%). Whereas the increases between the ResNet50 and ResNet152 models were much smaller (1-2%).

4.2 Testing
The values in table 2 show that during testing the AlexNet models had the lowest identification accuracies on both the 11k Hands dataset and the IITD dataset and achieved 93.7% and 80.1%, respectively. The ResNet50 models demonstrated superior results compared to the AlexNet models achieving identification accuracies of 99.96% and 99.85% on the 11k Hands and IITD datasets, respectively. This corresponds to an improvement of approximately 6% for the 11k Hands dataset and 20% for the IITD dataset when compared to the AlexNet results. In the case of the ResNet152 models, further improvements to identification accuracy were obtained. The gain in identification accuracy of the ResNet152 models over the AlexNet models was approximately equivalent to the gain seen with the ResNet50 models. That is, the ResNet152 performed around 6% better on the 11k Hands dataset and around 20% better on the IITD dataset than the AlexNet models. However, there was only a small improvement in identification accuracy when using ResNet152 versus ResNet50. Using the 11k Hands dataset, the ResNet152 model achieved an accuracy of 99.98%, an increase of 0.02% over the ResNet50 model. In the case of the IITD dataset the ResNet152 model demonstrated an identification accuracy of 99.92%, an increase of 0.07% compared to the ResNet50 model. As previously stated, table 2, also shows the accuracies obtained during training, for each of the three architectures (AlexNet, ResNet50, ResNet152). The average identification accuracies achieved during testing exceeded the final average accuracies obtained during training indicating that the models did not overfit the data.

Table 2. The training and testing accuracies for the three CNN architectures, for each dataset, after training for 75 epochs

	11k Hands		**IITD**	
Architecture	**Training**	**Testing**	**Training**	**Testing**
AlexNet	60.61%	93.7%	72.40%	80.1%
ResNet50	94.77%	99.96%	97.87%	99.85%
ResNet152	95.92%	99.98%	98.69%	99.92%

5 Analysis
The results from these experiments show that the ResNet models achieve greater biometric identification performance compared with the AlexNet models. This is likely due to the increased network depth of the ResNet models as this allows the deeper layers to learn increasing complex local features. In contrast, the shallow nature of the AlexNet architecture means that these models only learn more global features making them less suitable for this scenario. Prior research using images of the dorsal side of the hands, taken in normal lighting conditions, has focused on using CNNs solely for feature extraction purposes. However, the method detailed in [6] may not generalise well to other biometric modalities due to the inclusion of handcrafted features. Consequently, the technique may only prove optimal for dorsal hand-based biometric identification. Whilst these experiments show that the ResNet models are more suitable than the AlexNet models for this type of data, the performance of ResNet152 model is comparable to ResNet50 model. The average identification accuracy was only 0.02% greater during testing when using ResNet152 versus ResNet50, indicating there is no significant benefit to using the more complex ResNet152 architecture for dorsal hand biometric identification.

6 Conclusion
This research sought to demonstrate that a transfer learning approach can be applied to the problem of biometric identification using dorsal hand images. By using transfer learning and multiple datasets, this research demonstrates that existing, unmodified CNN architectures can be used to effectively perform

biometric identification using dorsal hand images. In addition, these experiments show that achieving high identification performance does not require enhancing the visibility of hand vein patterns using NIR light, as in other recent studies [9] [10] [11]. This paper also demonstrates that suitable identification accuracies, greater than 99.9%, can be achieved for automated biometric identification using the dorsal hand region as the biometric identifier. Further to this, the results show that such systems do not need to rely on NIR enhanced hand vein patterns, pre-processing, or modified CNN architectures to achieve these results. Whilst further work is required to fully evaluate the proposed approach for the purpose of identifying perpetrators from CSAI, the initial results show the potential of using off-the-shelf CNNs for this task. The experimental results suggest that the deep, ResNet CNNs can learn both the global and local features required to accurate classify individuals based on the dorsal hand region. The results also demonstrate that it is feasible to use a fully automated approach on relatively small datasets whilst maintaining highly accurate outcomes. To date, this study is the first to use deep learning methodologies for the end-to-end biometric identification of dorsal hand images taken in normal lighting conditions.

References

[1] Gottfried E D, Shier E K, Mulay A L. Child Pornography and Online Sexual Solicitation. Current Psychiatry Reports. 2020;22(3):10.

[2] Bursztein E, Clarke E, DeLaune M, Elifff D M, Hsu N, Olson L, et al. Rethinking the detection of child sexual abuse imagery on the Internet. The World Wide Web Conference. 2019:2601-7.

[3] Matkowski W M, Chan F K S, Kong A W K. A study on wrist identification for forensic investigation. Image and Vision Computing. 2019;88:96-112.

[4] Kumar A, Xu Z. Personal Identification Using Minor Knuckle Patterns From Palm Dorsal Surface. IEEE Transactions on Information Forensics and Security. 2016;11(10):2338-48.

[5] Fei L, Lu G, Jia W, Teng S, Zhang D. Feature Extraction Methods for Palmprint Recognition: A Survey and Evaluation. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2019;49(2):346-63.

[6] Afifi M. 11K Hands: Gender recognition and biometric identification using a large dataset of hand images. Multimedia Tools and Applications. 2019;78(15):20835-54.

[7] Pugh R. Dame Sue Black: ‘One girl’s case led me to pursue child sexual abusers’. The Guardian. 2019.

[8] Belean B, Streza M, Crisan S, Emerich S. Dorsal hand vein pattern analysis and neural networks for biometric authentication. Studies in Informatics and Control. 2017;26(3):305-14.

[9] Al-johania N, Elrefaei L A. Dorsal hand vein recognition by convolutional neural networks: Feature learning and transfer learning approaches. International Journal of Intelligent Engineering and Systems. 2019;12(3):178-91.

[10] Lefkovits S, Lefkovits L, Szilágyi L. CNN approaches for dorsal hand vein based identification. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. 2019;27:51-60.

[11] Wang J, Wang G. Hand-dorsa vein recognition with structure growing guided CNN. Optik. 2017;149:469-77.

[12] Das R, Piciucco E, Maiorana E, Campisi P. Convolutional Neural Network for Finger-Vein-Based Biometric Identification. IEEE Transactions on Information Forensics and Security. 2019;14(2):360-73.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25. 2012:1097-105.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.

[15] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:770-8.

[16] Minaee S, Azimi E, Abdolrashidi A. Fingernet: Pushing the limits of fingerprint recognition using convolutional neural network. arXiv preprint arXiv:190712956. 2019.
[17] Minaee S, Abdolrashidi A. Deepiris: Iris recognition using a deep learning approach. arXiv preprint arXiv:190709380. 2019.
[18] Rochac J F R, Zhang N, Thompson L, Oladunni T. A Data Augmentation-Assisted Deep Learning Model for High Dimensional and Highly Imbalanced Hyperspectral Imaging Data. 2019 9th International Conference on Information Science and Technology (ICIST). 2019:362-7.