

Vein Pattern Visualisation using Conditional Generative Adversarial Networks

Ali Keivanmarz*, Hamid Sharifzadeh*, Rachel Fleming†

* School of Computing, Unitec Institute of Technology, Auckland, New Zealand

E-mails: keivaa01@myunitec.ac.nz, hsharifzadeh@unitec.ac.nz

† Forensic R&D, Institute of Environmental Science and Research (ESR), Auckland, New Zealand

E-mail: Rachel.Fleming@esr.cri.nz

Abstract—Utilising vein pattern as a biometric attribute for forensic identification in crime investigation has been challenging because vein patterns are almost undetectable in standard RGB images. Significant research efforts for uncovering vein patterns have been recently made based on various computational techniques such as artificial neural networks, optical vein disclosure, auto-encoders, etc. While some promising results have been achieved using these methods, comparing with the NIR reference images, these computational techniques are still struggling to provide reliable outcomes.

In this paper, we propose a new method that performs vein pattern visualisation based on a conditional Generative Adversarial Network (GAN). GANs have shown promising results on image translation tasks in other areas. Therefore, for the first time, a specialised conditional GAN is proposed for translating colour RGB images to NIR images in this paper. The performance evaluation conducted on a small dataset shows the efficiency of our proposed method by uncovering over 80% of vein pixels in forearms of eleven subjects.

Keywords: Vein pattern, GAN, Forensics, Image translation, CSAM

I. INTRODUCTION

In fighting with child sexual abuse material (CSAM), which is regarded as one of the rapidly growing crimes [1], identification of criminal and victim is a vital factor. Child pornography is associated with a series of other criminal offences, including human trafficking and sexual abuse. Thus it is of paramount importance in investigative operations of the law enforcement agencies and police across the world. Identifying a person based on typical colour images is possible through the face characteristics, so there has been considerable research effort to develop biometric feature extraction via face recognition [2], [3], [4], [5], [6].

While typical biometric identification methods are effectively used in conventional evidence images, in the CSAM cases, these methods are not useful due to the typically covered face of the offenders. To avoid recognition, child abusers are often cautious not to reveal their faces in the footage.

The internet has resulted in an explosion in the illicit trade of CSAM and due to the lack of effective investigative tools, crimes associated with the CSAM are growing significantly [7]. Therefore, developing reliable methods for identifying child abusers based on partial skin images can be considered as the most efficient way of fighting with the CSAM and associated crimes.

In the CSAM cases, tattoos and skin traces may provide some clues; however, tattoos are not always unique, or some people have no tattoos, while others might have similar tattoos. Additionally, the skin exposed in evidence images may not have enough visible skin marks for the person identification [8]. Therefore, alternative biometric traits and identification methods are required to be developed for criminal identification in these cases.

Recently, some computational methods have been introduced that aim to uncover the vein patterns of subjects - as unique to an individual as their fingerprint or DNA - in normal digital images [9], [10], [11], [12]. To visualise vein patterns in the body limbs, we usually require images to be captured by a Near-Infrared (NIR) camera (as special equipment) while in the CSAM cases, the evidence images have been captured by a normal digital camera (i.e. not a NIR image). Thus, through visualising vein patterns from a small portion of non-facial skin image captured by a normal digital camera, the ultimate aim of these computational methods are to provide an image with similar characteristics to a NIR image. While some promising results have been achieved, these approaches still struggle in providing a reliable outcome with the necessary details of vein patterns similar to what a NIR image can provide.

In this paper, within an image translation framework, for the first time, we propose GANs in the conditional setting to map RGB images to NIR images. GANs have shown very promising results on image translation tasks in other areas [13], [14], [15], [16]; therefore, a specialised conditional GAN is proposed for translating colour images to NIR counterparts in this paper. While GANs based methods use a generator model as part of the architecture, in our proposed model, a conditional generator model is used [13], [14]. This would be a suitable approach for the RGB-to-NIR image translation because we can apply conditions to the generative model in the conversion of the input to output images in corresponding to the target images.

To briefly provide an idea of the proposed method, the block diagram in Fig. 1 outlines the technical steps of our method. As shown in Fig. 1, first the RGB and NIR images are globally enhanced by contrast adjustments. After initially enhancing both synchronised RGB and NIR images, conditional GANs are used to generate plausible NIR images. Finally, for vali-

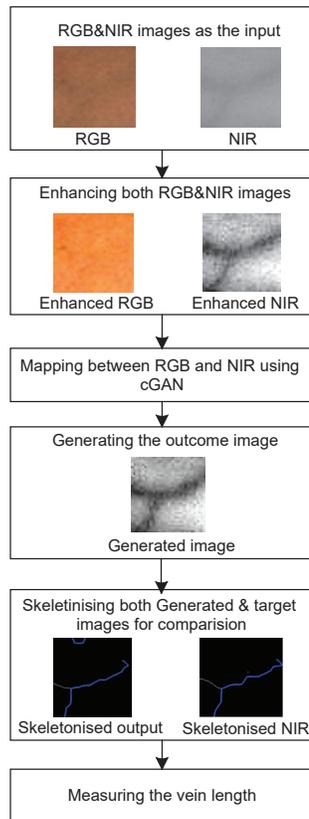


Fig. 1. The technical procedure for the proposed visualisation method. RGB and NIR images are first enhanced before being trained by the proposed cGAN. Both the outcome of the generative model and enhanced target NIR image are then skeletonised for the purpose of measuring the vein lengths.

dating the results of the proposed cGAN, both reference NIR and generated NIR images are skeletonised and the length of vein pixels is measured. Section 3 discusses further technical details of the proposed method.

The rest of this paper is organised as follows: Section II briefly evaluates the related literature. Section III explains the cGAN based image transformation approach as the core of our proposed method, and Section IV discusses skeletonisation and vein length measurement. Section V reports the performance evaluation and experimental findings. Finally, Section VI offers conclusive remarks.

II. RELATED LITERATURE

For the purpose of criminal and victim identification in the scenes where the faces are masked, uncovering vein pattern from colour digital images can be considered as a new biometric tool. As mentioned in Section I, a variety of computational models have been developed for this purpose. As a very early effort for extracting epidermis characteristics in standard images, Claridge et al. [17] developed a method based on Pigmented Skin Lesions technique that was used

to reveal melanin intensity, dermal depth, and haemoglobin, using textured skin lesions. Their method was mainly developed for clinical applications, i.e. vein patterns or forensic investigations were not the focus of their work.

In 2011, Tang et al. [18] presented a model using optics and skin biophysics called OBVU. Their proposed model had a different approach from that of Claridge et al. and was focused on using vein structures for forensic research. The OBVU used the Kubelka-Munk theory [19], [20] to model the formation of skin colours. They indicated that the method was sensitive to green channel noise, and various lighting conditions can influence the performance of vein pattern extraction. Later, to address some of these shortcomings, Zhang et al. [8] suggested an uncovering vein pattern algorithm moderately based on OBVU and introduced a new technique for colour-optimisation and adaptive image intensity adjustment. The drawback with Zhang’s method was poor outcomes on diverse skin properties and also on various image illuminations [9], [21].

To overcome the issues raising from skin diversity, Sharifzadeh et al. [9] proposed an uncovering method based on a basic artificial neural network but with banks of mapping models. More recently, Varastehpour et al. [21] developed a deep neural network using sparse-auto-encoder to enhance the image before the vein pattern extraction phase. In their method, three parameters were adjusted: Sparsity-Proportion, L2-Weight-Regularisation and Sparsity Regularisation.

By and large, all these techniques of vein extraction are still subject to certain limitations. In addition to the limitations in efficiently addressing skin diversity, or lighting conditions, or image quality issues, one of the main drawbacks in these methods might be the commonly used framework of mapping RGB images to NIR images. These methods assume there is an optimum pixel to pixel mapping that can efficiently convert three channels of RGB into one channel of grey-scale to generate a NIR image; however, given consideration to the outcomes of these methods, this assumption might not be the best approach. Many problems in computer vision are modelled as mapping and converting an input image into a corresponding output image. When using deep learning in the image to image translation, networks learn to reduce an error function which is an objective that rates the value of output image. Therefore, a great deal of manual adjustments needs to be performed to define an effective loss function despite automatic learning stage. Additionally, the current methods are often framed as pixel-by-pixel classification and regression in which each resulted pixel is isolated from other pixels.

Generative Adversarial Networks (GANs) as a recent development in deep learning algorithms can potentially address some of the issues in current methods by learning an error function by itself and classify the output image into positive class if it is what we are looking for while simultaneously training a generative model to minimise that error function [22], [23], [24]. Since GANs are able to learn an error function related to the synchronised pairs of input and target images, they are able to automate this stage of the process as well [14].

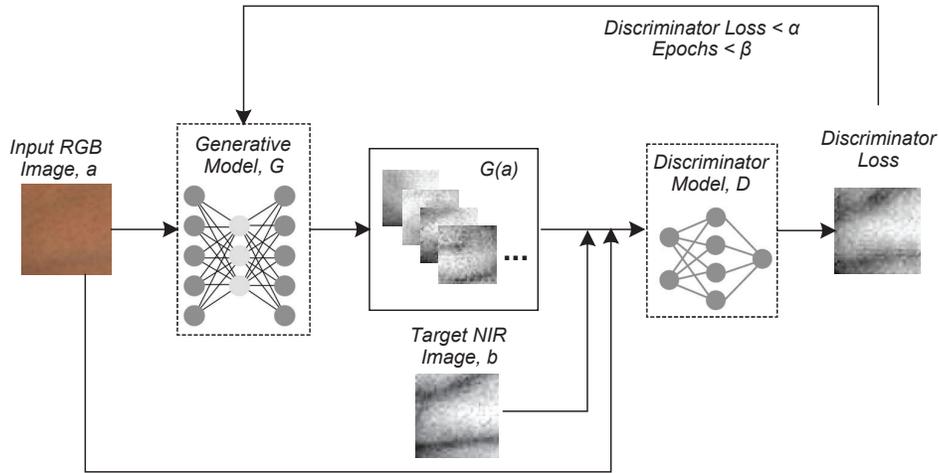


Fig. 2. Our proposed conditional GAN training procedure: G as the generative model learns to deceive D as the discriminative model which simultaneously is learning to distinguish between real (i.e. reference NIR image) and fake (i.e. the NIR images generated by G). This training process continues until D cannot distinguish the real target NIR from fake G generated image (i.e. the best NIR generated image similar to the reference NIR image). Additionally, both G and D are observing input image as the condition.

Thus, this paper proposes an entirely new approach based on translating RGB images to NIR counterparts. Unlike current methods that are mainly built upon RGB to NIR image conversion using feature extraction and fusion for mapping from colour to grayscale, in this paper an image to image translation method based on GANs is developed and evaluated. We also define a condition on an RGB input image and correspondingly create a NIR output image.

As shown in Fig. 1, this method consists of a two-stage algorithm based on synchronised RGB and NIR images: The first stage is the translation of the enhanced RGB images in three separate channels into the target NIR image using cGAN, and the second stage is the skeletonisation procedure that visualises an estimation of the topology of vein pixels; the outcome of this stage is also used for the objective measurements. In the following sections, we discuss the details of our proposed method and provide the results of the evaluative performance analysis.

III. PROPOSED CGAN FRAMEWORK

Transformation of a given input image to a reference image is a controlled conversion. Such a transformation requires special models and loss functions. The proposed cGAN architecture here is an adjusted approach specialised for the translation of RGB images into NIR images. For this purpose, a cGAN based technique is developed, which generates a NIR image, based on a provided input RGB image. The GAN algorithm within this setting adopts the error function to adjust the generated image to be close in content, and to look visually plausible to the target image (i.e. NIR image).

Fig. 2 shows the training process for our proposed RGB-NIR translation cGAN network outlining the mechanism in which the target NIR and input RGB images are fed into the model for the generation of the best possible NIR counterpart

from the corresponding RGB image. In the following subsections, first, a brief overview of cGAN architecture as an extension of GAN network is provided and then the technical components of our proposed cGAN model for the purpose of RGB to NIR translation, as shown in Fig. 2, are described.

A. cGAN Architecture

The generator models in GANs aim to learn a mapping function between random vector noise c and the resulted image b [25],

$$G : c \rightarrow b \tag{1}$$

The generator models in cGANs aim to figure the mapping from input image a and c to b [26],

$$G : a, c \rightarrow b \tag{2}$$

G , like the generator, creates images that D , as the discriminator model, cannot distinguish from the target image. At the same time, D is being trained and continuously improved to detect the undesirable or fake images in GANs context.

The conditional GAN's objective is stated as:

$$LcGAN(G, D) = E_{a, b}[\log D(a, b)] + E_{a, b}[\log(1 - D(a, G(a, c)))] \tag{3}$$

where $LcGAN$ is the loss of the cGAN model, E is the expectation, and G minimises the error while the adversarial D attempts to maximise it, i.e. $G^* = \arg \min_G \max_D LcGAN(G, D)$.

Previous approaches use a mixture of $L2$ distance with GAN objective [27]. While the discriminator's task remains the same, the generator model's job is to deceive the D in addition to generating an image close to the target image in

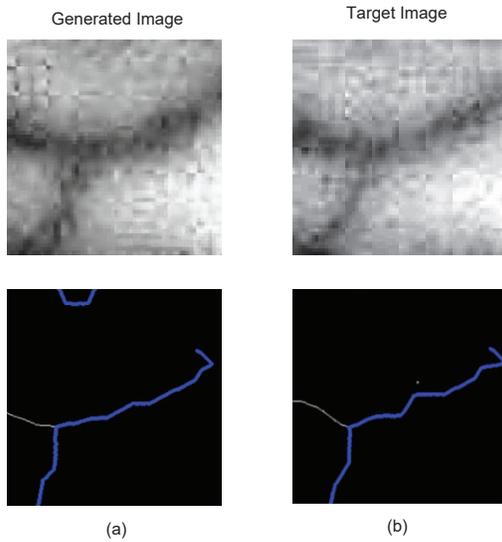


Fig. 3. Vein Skeletonisation: (a) Output of the proposed model, (b) Target NIR image.

an $L2$ sense. Instead of using $L2$, we noticed that using $L1$ results in less blurry outcomes in RGB to NIR generation:

$$LL1(G) = E_{a,b,c} |b - G(a,c)| \quad (4)$$

The concluded G is:

$$G^* = \lambda LL1(G) + \arg \min (G) \max (D) LcGAN(G, D) \quad (5)$$

where $LL1$ is the loss of $L1$ and λ is used to control the combination.

Although the network is able to learn the mapping from a to b without considering c , yet the results will be deterministic and do not coordinate with any distribution except the delta function [14]. Previous cGANs addressed the issue and introduced Gaussian noise c as an additional parameter to a when applied to the generator [28]. In our case, we did not find Gaussian noise c as an effective strategy and instead, we applied dropout as noise to the layers of our generator model during training and testing phases as was previously tested by Isola et al. [14].

Our specialised cGAN architecture includes an accurate specification of a discriminator model, a generative model, and the optimisation process. In the following, we describe each of these functioning blocks.

B. Generator Model

We design our generator architecture based on the consideration that the input (RGB) and output (generated NIR) are only different in surface look, but both are similar in the core structure. As a result, the structure in the input RGB image is approximately adjusted with the structure in the output image (i.e. generated NIR image).

Previous approaches [28], [27] to similar tasks of image translation utilised an encoder-decoder network as the generator [29]. In these models, all information move across all the generative layers, including the bottleneck [30]. In our case, there is a large amount of low-level information shared between the input and output images for the RGB-to-NIR image translations (such as the location of edges) and it would be beneficial to move such information straight through the network layers instead of going through the generative process. So as an alternative solution, a U-Net model [31], [32] can be used for the generative model which works very similar to auto-encoder-decoder and includes down-sampling and up-sampling to create output images with the addition of links or skip-connections between layers of the same size in the encoder and the decoder in order to avoid low-level processing information especially in the bottleneck of the network.

The generative network is trained by applying both an adversarial loss for the discriminative network and the $L1$ or a mean absolute pixel difference between the generated image and the intended target image. The adversarial loss and the $L1$, which is used to update the model generator, are added together as a compound loss function. As previously mentioned in subsection III-A, the loss of $L2$ has also been considered for our model but resulted in blurred images, so we focused on the $L1$.

The adversarial loss has an impact on the ability of the generator model in the target domain to produce images that are realistic, while the $L1$ loss regularises the generator model to produce images that are a plausible conversion of the input image. As such, a new λ , which has been set to 10 or 100, is used to control the combination of the $L1$ loss in an adversary loss in our model:

$$GeneratorLoss = AdversarialLoss + \lambda \times L1Loss \quad (6)$$

C. Discriminator Model

The discriminator model uses the source domain RGB image and a target domain NIR image to calculate the probability of how realistic the generated image is.

The approach to the discriminatory model emphasises the need for an image dataset composed of combined source and target images during model training. Our proposed model uses a PatchGAN [14] as opposed to the standard GAN model which uses a deep CNN for image classification. Instead of classifying the entire dataset, PatchGAN is used to classify patches of data image sets. The proposed discriminator model classifies the generated images as real or fake images considering how plausible and similar they are compared to the target images (i.e. how similar/plausible a generated NIR image is to a target NIR image). The discriminator model output is a single characteristic map of real/fake estimates that can be averaged for a single result. The RGB-to-NIR translation task is determined to be effective with a patch size of 70×70 in our model.

The discriminatory model is trained independently of a standard GAN model, which minimises the negative log-like

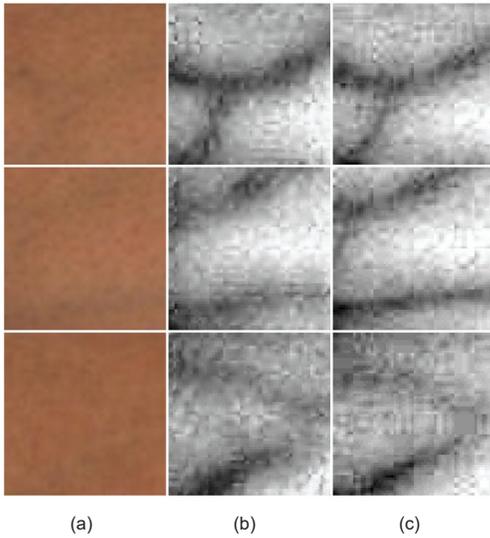


Fig. 4. Vein uncovering results: (a) Input RGB image, (b) cGAN generated image, (c) Target NIR image.

identification of real and fake images, even though it is conditioned on a generated image. The training of discrimination against the generator is too fast; thus, the loss of discrimination is halved to reduce the speed of the training process for the discriminator.

$$DiscriminatorLoss = 0.5 \times DiscriminatorLoss \quad (7)$$

D. Optimisation

To optimise the proposed model, we apply the approach obtained from [25]: We shift between a step down on D, as the discriminator model, then one step down on G as the generative model. We use the Adam solver and mini-batch SGD (Stochastic Gradient Descent) [33], which takes just one example at a time to take a single step. Also, the learning rate is set to 0.0002.

For the training phase, we define α as the threshold for the discriminator loss in order to restrict the number of training loops. We also define β as the threshold for the number of epochs. For the purpose of these experiments, we evaluated these thresholds and set α to 0.7 and β to 150 epochs for the training phase. During the testing phase, we run the generator model like the way we run the model on the training phase, which means we do not apply dropout during the test phase.

We also apply batch normalisation using the test batch statistics [33]. The batch size of 1 has been tested to be effective on the image translation tasks [34]. We also use batch size 1.

IV. SKELETONISATION

The outcome of the proposed cGAN network is passed through an adaptive thresholding algorithm for skeletonisation

[35]. The algorithm performs thresholding over local neighbourhoods, allowing for the extraction of structure over a broad dynamic range.

Within the skeletonisation process, first, a mask is constructed by applying morphological operators to remove unnecessary small regions. Then we reduce the image into skeleton form with one-pixel width using a Medial Axis Transform technique [36]. The skeletonised image shows the vein location and pattern.

To find the vein lengths, we remove the tiny branches on the image. Every remaining graphic entity in the resulting image is transformed into a graph. We use such graph to find the longest path through the skeleton to determine the length of the veins. The other branches that are away from the longest vein are removed.

In order to check the outcome of the proposed method and skeletonisation process, we use a pair of synchronised RGB and NIR images. As an example, Fig. 3 shows the outcome of the proposed image translation and vein skeletonisation on a small part of a forearm along with the corresponding reference NIR image.

V. PERFORMANCE EVALUATION & DISCUSSION

To evaluate our proposed cGAN model for the task of RGB-to-NIR image translation, we trained and tested our method on the images in the NTU Forensic Image Skin dataset [37]. As the early steps towards using cGAN for the NIR generation, eleven subjects from the NTU dataset were used to provide training and testing samples for the proposed method. Synchronised pairs of NIR and RGB images were obtained from 22 forearms of these eleven subjects (left and right hands) and then segmented into 277 images in the size of 32x32. Although small in size, they were able to offer a diversity of skin tones and variety of forearms to some extent.

To balance the proportion of vein and skin pixels on each segment of test and training images, we calculated the number of the vein and skin pixels based on the NIR-image and then chose the segments with more than 25% of vein pixels. 223 skin images were used for the training and 54 images were put aside for the testing. The samples for each subject were

TABLE I
COMPARISON OF UNCOVERED VEIN LENGTHS FOR EACH SUBJECT MEASURED IN NUMBER OF PIXELS (THE OVERALL ACCURACY SITS AT 81.02%).

Subject No.	Number of Test Forearms Images	Overall Target Length (Pixel)	Overall Output Length (Pixel)	Accuracy Percentage (%)
Subject 01	4	1529.80	1122.53	73.38
Subject 02	5	2161.28	1891.66	87.52
Subject 03	5	1705.27	1337.94	78.46
Subject 04	5	1655.25	1889.06	85.87
Subject 05	4	1936.50	2386.60	76.76
Subject 06	6	2277.87	2059.03	90.39
Subject 07	4	1869.58	2285.45	77.76
Subject 08	5	1453.72	1097.39	75.49
Subject 09	4	1666.48	1916.73	84.98
Subject 10	6	2708.99	2028.98	74.90
Subject 11	6	2763.51	3156.96	85.76
Total	54	21728.26	21172.33	81.02

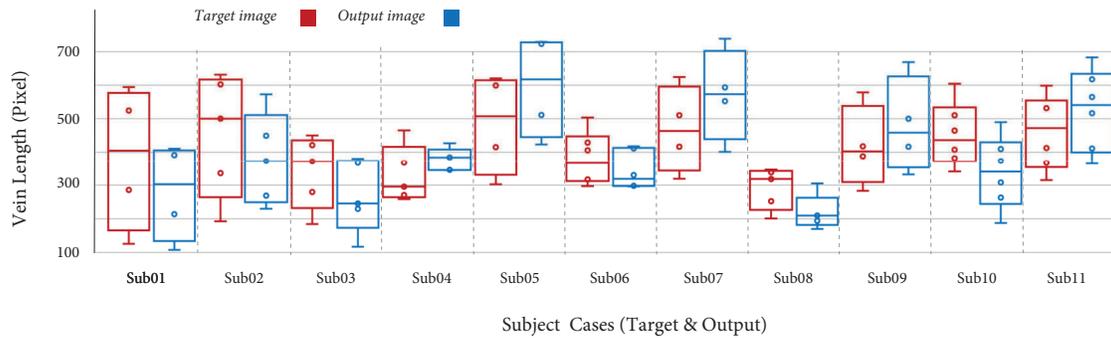


Fig. 5. Box-plot representation of vein lengths estimated by the proposed cGAN model for each subject.

randomly mixed and split into training and testing sets in the ratio of 1 to 4.

Fig. 4 shows three examples of uncovering results from the proposed method, comparing the input RGB image, cGAN generated image and the target NIR image. As it is evident in this figure, cGAN generated images are very similar to the reference NIR images in visualising vein patterns.

Table I summarises the vein lengths obtained from samples of eleven subjects’ forearms and compares them with their reference NIR images. As it is seen in Table I, the proposed cGAN can efficiently generate the vein pixels close to the target NIR image with an average accuracy of 81.02% across all samples. With close to 90% accuracy, subjects 2 & 6 show the highest accuracy. Comparing with other subjects, this might be because of less amount of hair on the skins of these subjects. Also, the existence of skin marks, amount of fat, and skin tones can be considered as the other contributing factors to the range of different accuracies from 73% to 90%.

We can also examine the length of veins obtained per-subject in the box-plot of Fig. 5. From the plot, we can see that there is quite a wide range of vein lengths for subject 1 in the reference NIR images while this is the case for subject 7 in the generated cGAN images; however, in both subjects, mean values and spread of vein lengths are almost similar between the NIR and cGAN generated images. On the other side, the short-range of vein lengths for some subjects such as 4 & 8 indicate that our vein length measurement method may well be quite suitable for some subjects but not for all. Subject 6 that has the highest accuracy in Table I, shows a small spread of vein lengths in testing images as well.

It should also be noted that the objective measurement of vein lengths for comparing the cGAN generated outcome with the target NIR images might be prone to misinterpretation due to reasons such as: a) simplified assumptions in the skeletonisation procedure, b) considering the length of veins as the objective value while the pattern and topology of veins are not evaluated.

In summary, with more than 80% overall accuracy, some promising evaluative results have been achieved using the proposed cGAN generated images. These early results indicate

the efficiency of this new approach for the challenging task of uncovering vein patterns from RGB images; however, we should be conservative in generalising these results as further expansions and experiments with more number of subjects should be conducted.

VI. CONCLUSION

In this paper, we proposed a new method for the visualisation of the vein patterns based on a conditional GAN. Within a GAN framework, we developed generator and discriminator models adjusted for the RGB to NIR translation. Our proposed cGAN takes an RGB image as the input and condition to the GAN model and generates a NIR image that is plausible to the reference NIR image to the extent that the discriminator of our model is not able to distinguish it from the reference NIR image.

To evaluate the efficiency of our method, we skeletonised the generated and reference NIR images and compared the length of uncovered veins with each other. With over 80% accuracy, the results tested on a small dataset of eleven subjects suggested that conditional GAN can be seen as a promising approach for the RGB-to-NIR translation tasks.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Institute of Environmental Science and Research (ESR), New Zealand for partially funding this research.

REFERENCES

- [1] Y. Akdeniz, *Internet Child Pornography and the Law: National and International Responses*. Routledge, May 2016.
- [2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210–227, Feb. 2009.
- [3] C. Ding, C. Xu, and D. Tao, “Multi-Task Pose-Invariant Face Recognition,” *IEEE Transactions on Image Processing*, vol. 24, pp. 980–993, Mar. 2015.
- [4] G. M. Zafaruddin and H. S. Fadewar, “Face recognition: A holistic approach review,” in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 175–178, Nov. 2014.
- [5] S. Yang, P. Luo, C. C. Loy, and X. Tang, “WIDER FACE: A Face Detection Benchmark,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5525–5533, June 2016.

- [6] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, June 2015.
- [7] M. Chopra, M. V. Martin, L. Rueda, and P. C. Hung, "Toward New Paradigms to Combating Internet Child Pornography," in *2006 Canadian Conference on Electrical and Computer Engineering*, pp. 1012–1015, May 2006.
- [8] H. Zhang, C. Tang, A. W.-K. Kong, and N. Craft, "Matching vein patterns from color images for forensic investigation," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 77–84, Sept. 2012.
- [9] H. R. Sharifzadeh, H. Zhang, and A. W.-K. Kong, "Vein Pattern Visualization through Multiple Mapping Models and Local Parameter Estimation for Forensic Investigation," in *2014 22nd International Conference on Pattern Recognition*, pp. 160–165, Aug. 2014.
- [10] C. Tang, H. Zhang, and A. Wai-Kin Kong, "Using multiple models to uncover blood vessel patterns in color images for forensic analysis," *Information Fusion*, vol. 32, pp. 26–39, Nov. 2016.
- [11] S. Varastehpour, H. Sharifzadeh, I. Ardekani, and X. Francis, "Vein Pattern Visualisation and Feature Extraction using Sparse Auto-Encoder for Forensic Purposes," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–8, Sept. 2019.
- [12] R. Zhang, D. Huang, Y. Wang, and Y. Wang, "Improving feature based dorsal hand vein recognition through Random Keypoint Generation and fine-grained matching," in *2015 International Conference on Biometrics (ICB)*, pp. 326–333, May 2015.
- [13] J. Lin, Y. Xia, T. Qin, Z. Chen, and T.-Y. Liu, "Conditional Image-to-Image Translation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5524–5532, June 2018.
- [14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, July 2017.
- [15] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2868–2876, Oct. 2017.
- [16] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward Multimodal Image-to-Image Translation," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 465–476, Curran Associates, Inc., 2017.
- [17] E. Claridge, S. Cotton, P. Hall, and M. Moncrieff, "From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions," *Medical Image Analysis*, vol. 7, pp. 489–502, Dec. 2003.
- [18] C. Tang, A. W. K. Kong, and N. Craft, "Uncovering vein patterns from color skin images for forensic analysis," in *CVPR 2011*, pp. 665–672, June 2011.
- [19] L. Yang and B. Kruse, "Revised Kubelka–Munk theory. I. Theory and application," *JOSA A*, vol. 21, pp. 1933–1941, Oct. 2004.
- [20] M. Doi and S. Tominaga, "Spectral estimation of human skin color using the Kubelka-Munk theory," in *Color Imaging VIII: Processing, Hardcopy, and Applications*, vol. 5008, pp. 221–228, International Society for Optics and Photonics, Jan. 2003.
- [21] S. Varastehpour, H. Sharifzadeh, I. Ardekani, X. Francis, and N. Baghaei, "An Adaptive Method for Vein Recognition Enhancement Using Deep Learning," in *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–6, Dec. 2019.
- [22] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 640–651, Apr. 2017.
- [23] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning Representations for Automatic Colorization," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), Lecture Notes in Computer Science, (Cham), pp. 577–593, Springer International Publishing, 2016.
- [24] S. Xie and Z. Tu, "Holistically-Nested Edge Detection," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1395–1403, Dec. 2015.
- [25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.
- [26] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv:1411.1784 [cs, stat]*, Nov. 2014.
- [27] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, June 2016.
- [28] X. Wang and A. Gupta, "Generative Image Modeling Using Style and Structure Adversarial Networks," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), Lecture Notes in Computer Science, (Cham), pp. 318–335, Springer International Publishing, 2016.
- [29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, pp. 504–507, July 2006.
- [30] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational Autoencoder for Deep Learning of Images, Labels and Captions," in *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds.), pp. 2352–2360, Curran Associates, Inc., 2016.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), Lecture Notes in Computer Science, (Cham), pp. 234–241, Springer International Publishing, 2015.
- [32] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 749–753, May 2018.
- [33] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167 [cs]*, Mar. 2015.
- [34] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," *arXiv:1607.08022 [cs]*, Nov. 2017.
- [35] K. Rezaee, J. Haddadnia, and A. Tashk, "Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization," *Applied Soft Computing*, vol. 52, pp. 937–951, Mar. 2017.
- [36] A. D. Ward and G. Hamarneh, "The Groupwise Medial Axis Transform for Fuzzy Skeletonization and Pruning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1084–1096, June 2010.
- [37] "NTU Forensic Image Database." <https://www.ntu.edu.sg/home/adamskong/Resources.html>, Oct. 2018.