2024 IEEE International Joint Conference on Biometrics (IJCB) | 979-8-3503-6413-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/IJCB62174.2024.10744466

Identity-Preserving GAN for Cross Spectral Iris Recognition

Hannah Anderson, Moktari Mostofa[1], Nasser Nasrabadi, Jeremy Dawson[2]
West Virginia University
Morgantown WV, USA
{hba0002, mm0251}@mix.wvu.edu, {Nasser.Nasrabadi, Jeremy.Dawson}@mail.wvu.edu

Abstract

Cross spectral iris recognition has been shown to cause a degradation in iris matching scenarios due to the inherent differences between the NIR and visible spectra. This led us to explore methods of iris domain translation, allowing us to generate images between the NIR and visible domains using generative adversarial networks (GANs). We train a GAN network with an additional classifier component to act as an identity-preserving module allowing the generator to produce not only high quality, but identity-specific images. We apply this method on three cross-spectral iris datasets, namely, the Cross-eyed-cross-spectral iris database, the PolyU bi-spectral database and the WVU Multispectral database collected from our lab. We implement image enhancement techniques on the cropped iris images and unrolled, normalized iris images, allowing for the generator to learn the iris texture with minimal noise surrounding the iris and to show the performance of the generated images in different matching scenarios. We show the performance of our model by matching the generated iris images against the true iris images in their translated domain. We show that applying this image translation technique as a preprocessing step increases the matching performance when applied to iris matching software, such as Neurotechnology's commercial iris recognition software, VeriEye and an open-source iris recognition software, OSIRIS. Lastly, we perform an ablation study for each set of experiments by removing the classifier component and comparing the results with our model, showing that the competition between the generator and classifier has an important role in learning identity-specific features.

1. Introduction

Iris recognition is a widely used biometric modality due to the unique variations of the iris tissue, offering high accuracy whether it be for identification or verification [3]. However, some applications of iris recognition, such as cross spectral iris recognition that could be used to match opportunistic iris images extracted from face images to NIR iris galleries, are still experiencing challenges to implementation. This is due to the inherent differences between the two typical near infrared (NIR; 800-900 nm) iris images and those captured in the visible spectrum (400-700 nm).

Iris images are generally captured in the NIR spectrum which offers the ability to image the iris texture clearly regardless of eye color. Visible iris images, as shown from [1], also suffer from a lack of clarity in texture when the individual has darkly pigmented irises, resulting in a poor-quality image with less discriminate features than NIR images of the same individual. This brings a new complication when trying to match images from two disparate spectral ranges, as individuals with darker pigmented irises could see a decrease in match reliability where individuals with lighter-colored irises would not. However, there are several benefits to visible iris imaging that outweigh these shortcomings, including the ability to capture high quality 50+MP images, potential long-distance imaging, and the lower cost and higher availability of visible iris sensors (i.e., visible cameras). Visible images are also highly desirable for lighter eyes since more texture can be captured than in an NIR image. Existing data from surveillance videos and facial images can also be extracted, allowing for an expansion of the field.

Because of the limitations of both spectra, it would be significant for researchers to develop methods to translate between the near infra-red (NIR) domain, where a majority of large scale iris galleries are collected, and the visible (VIS) domain, where a larger population will have access to recognition systems.

Since a majority of iris galleries are obtained in the NIR spectrum, it would be highly strenuous to ask the participants to re-enroll using visible wavelength systems, as some

[1]Moktari Mostofa is now an ORISE Fellow with the US Food and Drug Administration, Silver Spring, MD 20993 (email: moktari.mostofa@fda.hhs.gov). This work is conducted inspired by her previous work during her graduate student status at the Department of Computer Science and Electrical Engineering, Morgantown, WV 26506, USA.

[2]The results presented here were obtained for a project supported by the Center for Identification Technology Research and the National Science Foundation under Grant No. 1650474.

large-scale datasets can have millions of samples, creating a discrepancy with newly captured iris probes and previously enrolled iris galleries. This is where a domain-translation architecture can be significant, allowing for previously enrolled images or newly enrolled images to be translated between domains to replicate intra-spectral matching. If a module has the ability to learn the spectral differences between iris images captured in the NIR vs VIS wavelengths, while keeping identity-specific features in tact, this would allow for the enrollment of images between domains with only one capture.

The goal of this paper is to analyze the performance of deep generative models on both unrolled, normalized iris images and segmented iris images to generalize the translation performance of NIR to VIS and VIS to NIR. A module dedicated for identity preservation is added to a Generative Adversarial Network (GAN) to keep the translated image as true to the original identity as possible. The trained model is then evaluated visually and empirically to show the performance enhancement with identity preservation as a key component to the success of the generator. The primary contributions of this work are:

- The proposal of an identity preserving module intended to guide the generator to class-specific features.
- An evaluation of the architecture on cropped and unrolled iris images, allowing for a generalization of the model and removing the need of the preprocessing step, especially when conducting matching analysis on closed source software such as VeriEye, which doesn't accept the input of an unrolled iris
- An evaluation of three cross-spectral datasets, highlighting the benefits and drawbacks to each in different processing scenarios.

2. Related Work

Cross-spectral iris recognition has received the attention of many researchers, and there are several approaches to try to mitigate the effects of sensor differences. Nalla and Kumar [16] suggests using Markov random fields and domain adaption to synthesize features between domains. Vyas and Kanumuri [27] extracted features of the different spectrum's using template partitioning. Recently, there have been multiple research papers that have focused on increasing the match performance of cross-spectral iris recognition using deep learning based approaches. Wei et al. [29] extracts features by learning a device-specific band that then translates between the two domains, training an adversarial model on learning these distributions. Wang and Kumar [28] implement deep hashing to extract sparse features and reduce the size of iris templates for matching.

Since the goal of this network is to generate iris images through adversarial learning, it is important to note the work in the area of iris and periocular synthesis. Mostofa et al. [14] produced a model to learn the common embedding of visible and NIR normalized iris images through a latent vector in the generator model. They also created a translation network from a 16-layer ResNet model to translate between the NIR and VIS spectrum's [15]. Both of these methods have only experimented on unrolled iris images. Poster et el. [22] and Hernandez Diaz [7] used a similar technique for the synthesis of periocular images using the Cross-Spectral and POLYU Bispectral datasets. These methods both show that Coupled GAN (cpGAN) and conditional GAN (cGAN) architectures have a great generalizability and can learn key differences between domains well.

Though these approaches achieve low EER and high performance for their task, they did not directly have a component for preserving the identity of the individual during the image transformation. They also focused on iris codes, which require preprocessing and cannot be used in every iris recognition software (VeriEye [19], for example only accepts segmented iris images) or the periocular region, which although is an important biometric trait in it's own right, the goal of this method is to improve the generators ability to replicate iris texture, and the added noise of the surrounding periocular region would force the generator to learn features not related to the iris. This is why we have applied an iris crop method as proposed in [13] to eliminate as much noise surrounding the iris in order to force the generator to be more selective in it's synthesized features.

Though the scope of this paper is through the lens of iris recognition, there are also implementations of generative models through facial recognition that can be transferred between domains. Shen et al. [25] implement a three-player GAN with an additional classifier component to preserve identity features across poses and shows promising results for generative models. Other researchers are applying identity-preserving features for face super-resolution [2; 11]. Since these networks are gaining popularity for facial recognition, it poses a question of whether these methods can be translated to other modalities such as iris recognition.

3. Datasets

Three cross-spectral datasets were used for the training and testing of these experiments. The Cross-Eyed Iris/Periocular Dataset [24] consists of 120 individuals and 8 images per eye, resulting in 1,920 images per spectrum at 400x300 resolution. The images in this dataset were taken simultaneously with a custom dual-lensed sensor, allowing for the NIR and VIS images of each individual to be paired.

The second dataset used in this work is the Hong Kong Polytechnic University Cross-Spectral Iris Images Database

(PolyU) [17], consisting of 209 individuals and 15 images per eye, resulting in 6,270 images per spectrum at 640x480 resolution. The images in this dataset were taken on a cross-spectral camera, allowing for the NIR and VIS images of each individual to be paired.

The final dataset is a dataset collected from our research lab at West Virginia University, which we will refer to as the WVU Multispectral Iris Dataset (this dataset is available upon request) which consists of 1,248 individuals with 2 images for the left iris and 2 images for the right iris, resulting in 4,992 images per spectrum at 640x480 resolution. This dataset does not take images synchronously similar to the other two datasets. This gives us the ability to experiment with the real-world scenario where the gallery is composed of images taken at a different time as the probes.

3.1. Preprocessing

In order to evaluate the overall performance of a generative model on iris texture, the datasets were processed in two ways. The iris boundary is found using the method from [30], applying median filtering to the image and enhancing the features to extract the iris region and the iris is then localized and normalized according to the rubber sheet format introduced by John Daugman [4] by converting the polar coordinates of the iris to a 64x512 sheet. The resulting normalized image was then enhanced using the methods from Nagam et al. [20], where the background-subtracted image is enhanced using contrast-limited adaptive histogram equalization (CLAHE) to obtain the final enhanced normalized iris image. We test the normalized iris images using the OSIRIS software.

The second method for processing the datasets was direct enhancement of the original cropped iris image. This method was conducted to translate directly between the captured iris image without the need to be normalized and allows for researchers to transform the images as needed without being limited to the iris code, which cannot be directly translated back to an iris image. The iris detection and segmentation technique from the method described above was used. The mask boundaries were then converted to coordinates to crop the original iris image and eliminate the periocular region around the iris. This acquires an image similar to Figure 1, with some of the inner periocular region still visible. For this reason, we adopted the preprocessing step from [13], removing the region outside of the iris to eliminate as much noise as possible from the training images. The last step is feature enhancement, where histogram equalization was applied to each image, which is intended to distribute the image histogram evenly across the entire intensity of the image. Figure 1 shows that it greatly increases the details of the images. It is also worth noting that we chose to use the gray-scale version for the visible images, which was chosen so the generator focuses on the texture more than the color. We test the cropped images using the VeriEye software.

Example images from each dataset are shown in the top row of Figure 1, as well as the processed images showing the enhancement processes of the normalized iris images and cropped iris images.

4. Model Architecture

The structure of our model was based on Mostofa et al. [14; 15] with additional components to improve the robustness and identity-preserving aspect of the model, inspired by Shen et al. [25]. A generative adversarial network was trained on both the normalized iris images and cropped iris images to compare the performance of both preprocessing techniques.

Figure 2 shows the basic architecture of the adversarial network and the classifier module. Similar to the work in [14; 15], a UNet [23] architecture with a Resnet-18 [6] backbone was implemented for both generators. We implemented a PatchGAN architecture introduced by Isola et al. [8] for the discriminators, to improve the quality of the images. We implemented a Resnet-18 network as the identity-preserving component, acting as a classifier which predicts the class the generated images belong to. The learning rate for each model was fixed at 1e-4 and we used Adam optimizer [12] for each model.

5. Objective Functions

In a conventional GAN architecture [5], the generator (G) and Discriminator (D) are competing in a min-max game to improve the quality of generated images based on a constant competition between the two objective functions. When the generator model produces a synthetic (or fake) image, the image will run through multiple convolutional blocks in the discriminator, resulting in a final value between 0 and 1, where 0 is classified as a fake sample, and 1 is classified as a real sample, and finally passing the information back to the generator to update it's weights accordingly. In our approach, two generative adversarial networks with players G_{VIS}, G_{NIR}, D_{VIS} and D_{NIR} are trained with an additional player, Classifier C. G_{NIR} and G_{VIS} will receive an input image in the VIS or NIR spectrums, respectively. They then produce an image $x_{VIS \rightarrow NIR}$ or $x_{NIR \rightarrow VIS}$. The goal of G_{VIS} and G_{NIR} in a standard GAN model is to produce realistic images within the target distribution that D_{VIS} and D_{NIR} cannot distinguish between a real image, and also minimize any auxiliary losses, commonly including structural and feature loss. The sections below go in detail about the loss functions that guide our model to the optimal output. To simplify the next section, we will be assigning x as the feature space we are translating to, either NIR or VIS, and each translation model

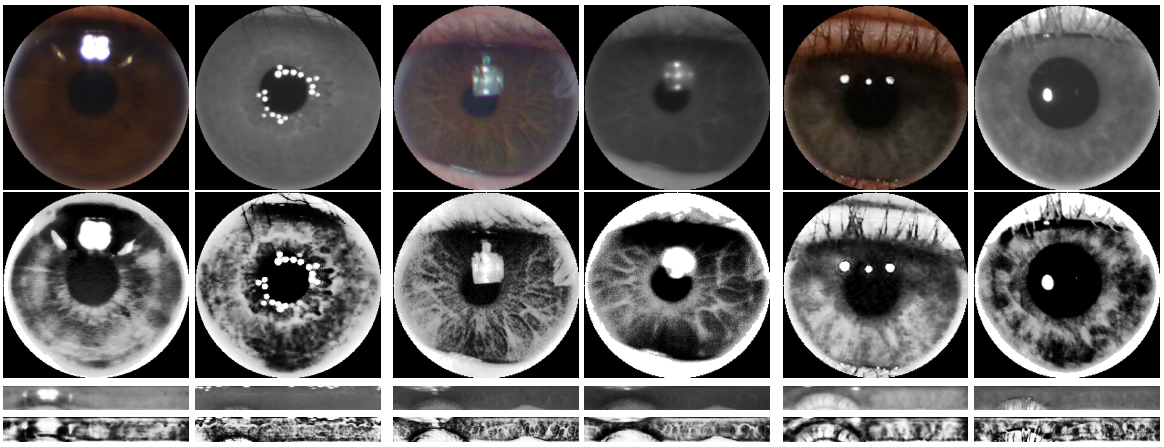

Figure 1. Examples of visible and NIR images from each dataset, as well as the processing step for each image. The first row is the original images cropped to just show iris area, with the order from left to right being: POLYU VIS, POLYU NIR, Cross-Eyed VIS, Cross-Eyed NIR, WVU Multispectral VIS, WVU Multispectral NIR. The second row shows the enhanced circle-cropped images for each eye, the third row shows the normalized iris codes for each eye, and the fourth row shows the enhanced normalized iris codes for each eye. The second and fourth rows are used for training the GANs.

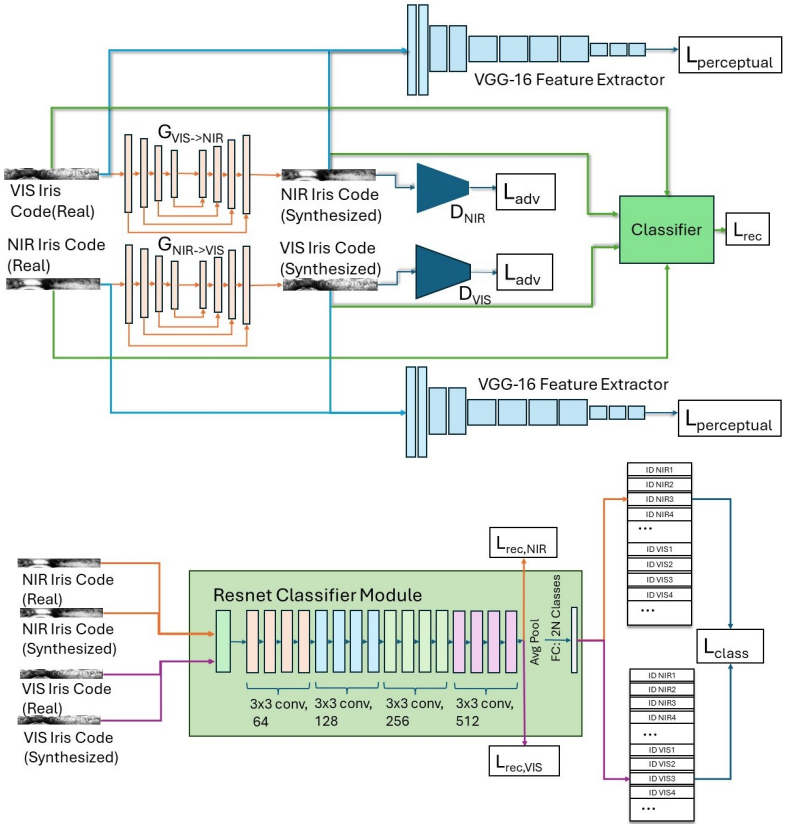

Figure 2. The proposed method, (Top) consisting of a UNet generator and PatchGAN discriminator. Each image is passed through it's respective generator and translated to the opposite modality. From there, the fake image is passed through the discriminator, which will evaluate the quality of the image. A perceptual module is used, where the feature map extracted from the 11th layer of a VGG16 network is used to preserve the high-level features of the fake images. (Bottom) Shows the process of the classifier network. The classifier is a ResNet-18 architecture with an auxiliary component before the fully connected layer to help guide the generator. The classifier is then optimized using a cross entropy loss where it learns to classify real and fake images into their original class before transformation.

will inherit the same loss functions. The two translation methods were also trained simultaneously but disjoint, so G_{VIS} has no knowledge of the progress of G_{NIR} and vice versa.

5.1. Adversarial Loss

We made use of the relativistic discriminator, first introduced by Jolicoeur-Martineau [10], which is used to stabilize training by making it's predictions based on a relationship between the real data and fake data. Here, the discriminator is trying to determine if the real data is *more* realistic than the fake data. Using this method for our network and defining these equations for each GAN, D(x) is defined as:

$$D(x) = \sigma(L(x_R) - L(x_F)), \tag{1}$$

where L(*) is the non-transformed layer of the discriminator, $\sigma(*)$ is the activation function and x_R is the real sample in the x domain and x_F is the fake sample in the x domain, will give an adversarial loss of:

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x_R}[\log D(x)] \\ +\mathbb{E}_{x_F}[\log(1 - D(G(x)))]\end{aligned} \tag{2}$$

, where G(x) is the generated data in x domain, and D(*) is the discriminator prediction. $\mathbb{E}_{x_R}[*]$ is the discriminator output on real images, which the discriminator attempts to maximize. $\mathbb{E}_{x_F}[*]$ is the discriminator output on fake images, which the generator attempts to minimize and the discriminator attempts to maximize.

5.2. Classifier Loss

The addition of the classifier is for the purpose of guiding the generator to class-specific synthesis. The competition between the generator and classifier will stabilize the generator to not only produce high quality images, but images that keep the individual features in tact. With this in mind, the classifier has the job of classifying real and synthesized images in their *original* class, where the generator is attempting to pull classes of real and fake images together. For example, if we are translating between NIR to VIS, the classifier is punished if it classifies the fake VIS image in the visible spectrum, and is encouraged to classify the fake VIS as it's corresponding NIR class.

For each generator, we will define real samples as x_R and fake samples as x_F as we did earlier. Since the classifier has the job of classifying both spectrum's of iris images, the fully connected layer consists of 2*N nodes, where N is the number of classes shared between the visible and NIR spectrum's. This brings the classifier loss to:

$$\begin{aligned}L_{C_R} &= -\sum_{c=1}^{2*N} (\vec{\ell}_{id}^{R})_j \log(C(x_R)_j) \\ L_{C_F} &= -\lambda_{C_F} \sum_{c=1}^{2*N} (\vec{\ell}_{id}^{F})_j \log(C(x_F)_j)\end{aligned} \tag{3}$$

where $\vec{\ell}_{id}^{R}$ is the label vector for the real sample and $\vec{\ell}_{id}^{F}$ is the label vector for the fake sample, and C(*) is the output of the classifier's fully-connected layer. Using the same logic as [25], the real images should carry more weight during classification. This is why a weight term, λ_{C_F}, is added into the final classifier loss, which we set to 0.4.

The generator is attempting to pull the real and fake classes closer together, trying to minimize the distance between the real and fake images of the same identity. From Figure 2, we add an auxiliary loss to the classifier, C, to extract the feature map before the fully connected layer. When the generator produces an image very similar to the original image, these feature maps will become closer together decreasing the distance. However, the classifier will suffer since it is attempting to separate the fake image into it's original class, hence allowing the generator and classifier to compete. If we define the loss obtained from the feature maps as the identity-preserving loss, and the output of the feature layer of the classifier as F(*), we take the cosine embedding loss to show the distance between the two extracted feature layers,

$$L_{IP,G} = 1 - \frac{F(x_R) \cdot F(x_F)}{|F(x_R)||F(x_F)|}, \tag{4}$$

allowing the generator to minimize this function.

5.3. Auxiliary Losses

We implement perceptual loss [9] in order to preserve the high-level features of the generated images. Feature maps extracted from the 11-th layer of a pretrained VGG-16 [26] are extracted from both the real and fake images and we calculate the mean squared error of those feature maps to feed back to the generator. We also implement a reconstruction loss by taking the Mean squared error of the real and fake images.

6. Experimental Results

A series of experiments were run on each dataset, and the trained generators were evaluated on how well their translated images performed against a gallery of the true images. We explored the results of the generated normalized irises by first matching the non-translated NIR and VIS iris codes through OSIRIS [21], an open source iris recognition software, and then applying our translated images to the

software as well. Our generated cropped images were run through a licensed software, VeriEye [19] and the matching performance of true images was compared with the matching performance of the generated images.

We show the results from matching each probe image to each gallery image, which we will call the All-to-All method, as well as applying score fusion to the results, which we will call the Score Fusion method. Score fusion is the process of combining scores to achieve an optimal result. In our case, we fused the scores of the same identity to output a single value. This is useful if there are multiple images in the gallery of each individual, so noisy or incorrectly segmented images don't hold as much weight during matching. In order to quantify the performance of the generated images, we report the EER for each test, which is the point at which the false acceptance rate and true acceptance rate are equal, and show the Detection Error Tradeoff (DET) curves. We also show the performance of our model with and without a classifier for the normalized iris method.

6.1. POLYU Bi-spectral Dataset

For the first set of experiments, we trained our model on the POLYU Bi-spectral dataset. We used the same train/test split as [7; 14; 18]. For each class, the first 10 instances are used for training and the remaining 5 instances are used for testing. We report the DET curves for the original non-translated NIR vs VIS, Fake NIR vs Real NIR, trained with and without a classifier, and Fake VIS vs Real VIS, trained with and without a classifier. It can be seen from Figure 3 that our method significantly improves the recognition performance. Applying translation without a classifier reduces the EER from 0.2331 to 0.1315 and 0.1401, respectively for the All-to-All method. Applying the classifier component decreases the values even more to 0.1203 and 0.1266, respectively. Applying score fusion significantly reduces the EER to 0.0550 and 0.0718, with the best performing model being the VIS vs Fake VIS with a classifier component. We did not include the cropped results for the POLYU dataset, as the model was not able to outperform the original images.

6.2. Cross-Eyed Dataset

For our next set of experiments we trained our model on the Cross-Eyed dataset. We followed the train/test split from [7; 14; 18] again by using the first 5 instances of each subject as the training set and the remaining 3 as the test set. Figure 4 (Top) shows that the best performing method for All-to-All matching is the VIS vs Fake VIS with a classifier component, decreasing the baseline EER from 0.0995 to 0.0648. When applying score fusion, the VIS vs Fake VIS with a classifier component decreases the baseline EER from 0.0542 to 0.0250, showing the best performance.

Applying the model to the cropped iris images, Figure 4 (Bottom) shows an increase in performance for the NIR vs Fake NIR for both the All-to-All and score fusion methods. For the All-to-All method, we experience the highest drop in EER from the NIR vs Fake NIR model, decreasing the EER from 0.2925 to 0.1909. Applying score fusion decreases the EER even more, from 0.1638 for the baseline to 0.0798 for NIR vs Fake NIR.

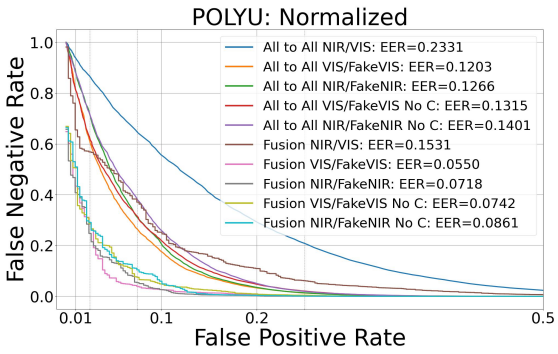

Figure 3. DET Curves from the POLYU dataset on the normalized iris images, showing the All-to-All matching method and the score fusion method.

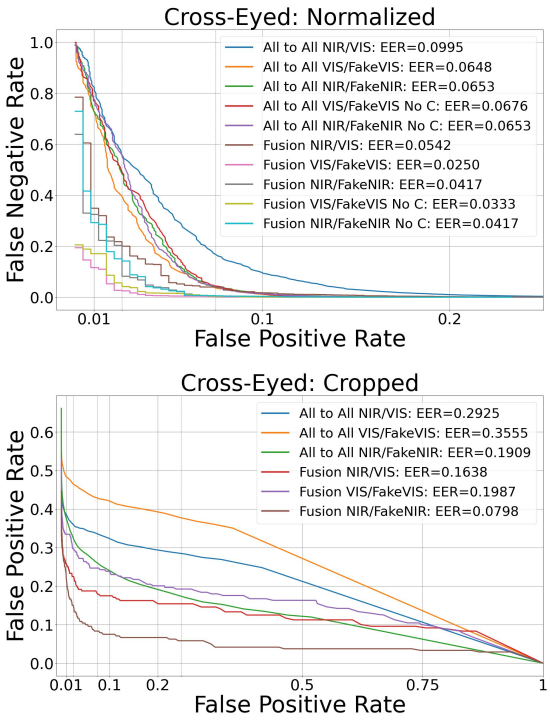

Figure 4. DET Curves from the Cross-Eyed dataset on the normalized iris images (top) and cropped iris images (bottom), showing the All-to-All matching method and the score fusion method.

6.3. WVU Multispectral Iris Dataset

For the next experiment we trained our model on the WVU Multispectral Iris dataset. Since each class only has 2 images and there are 2,384 classes, we chose to split the dataset differently than the first two. We chose this so the training set had more samples per class, and allowing it to better learn class-specific features instead of splitting the dataset evenly to only contain one sample per class. We adopted the method used in [14] where we use 750 samples as the testing set. In order to mimic a real-world situation, we trained on both samples from the first 1,634 classes and one sample from the final 750 classes. The test set contains the second image from the final 750 classes.

Figure 5 shows the DET curves for the normalized iris images as well as the cropped images for the WVU Multispectral Iris dataset. Our All-to-All method saw an increase in performance for the normalized iris images, with VIS vs Fake VIS with a classifier component slightly exceeding the VIS vs Fake VIS without a classifier component. The baseline EER is decreased from 0.2733 to 0.1587. We applied our method to the cropped iris images as well, however, none of the models were able to outperform the baseline, which obtained an EER of 0.2434.

The performance of this dataset is much lower than the POLYU and Cross-Eyed datasets, and two things can attribute to that. 1. The images in this dataset were taken disjointly, so there are multiple iris images that have different pupil dilation's, reflections, and rotated iris images. This affected the learning ability of the model. This is especially difficult for the cropped iris images since there is large variability in the images, requiring further research. 2. Because of the use of a classifier, more images per class (or iris) are ideal for training a classifier model. However, these limitations reflect real-world scenarios, since iris images won't likely be taken in both spectra simultaneously. The performance of this dataset should be further explored to reflect these situations.

This dataset obtained a lower overall EER for the normalized iris images, however the cropped iris images were able to obtain a much higher TAR at lower FAR than the normalized model. The normalized iris images obtained 0.661 TAR@FAR=0.1 and 0.1559 TAR@FAR=0.01, and 0.0366 TAR@FAR=0.001 where the cropped iris images obtained 0.6875 TAR@FAR=0.1, 0.597 TAR@FAR=0.01, and 0.5531 TAR@FAR=0.001. Allowing the threshold value to be set low allows for a very low false acceptance rate, while rejecting lower quality genuine matches, which is less detrimental for security purposes.

6.4. Generated Images

This section shows the visualization of the generated images, showing high-performing generated images as well as lower-performing images. Though the image enhancement technique greatly increases the iris features, there are limitations for noisy and dark images. We show three examples of the generators performance on the datasets in Figures 6 ,7 and 8, with different performances.

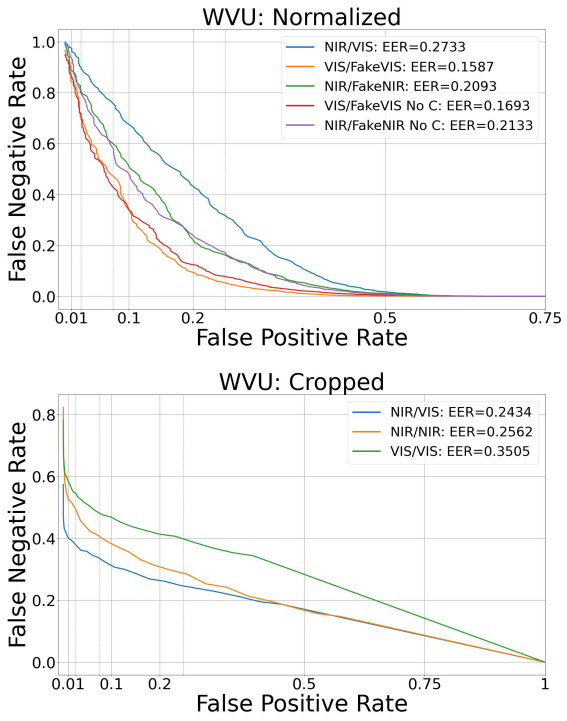

Figure 5. DET Curves from the WVU Multispectral Iris dataset, showing the All-to-All matching method for the normalized iris images (Top) and cropped iris images (Bottom).

It can be seen from the images in Figure 6 that the Cross-Eyed cropped models were able to learn class-specific features within the iris texture well, with visually appealing results. This translated well into the matching performance, with a significant increase in EER from the baseline.

Figure 7 shows an example from the WVU Multispectral dataset where the performance from one modality (NIR) outperforms the other (VIS). This is reflected in the DET curves, where the NIR vs Fake NIR model performed slightly worse than the baseline, however the VIS vs Fake VIS model performed significantly worse than both. We suspect that our model would perform better on this model with a larger sample of same-class images, and future work on this dataset could include data augmentation to synthesize same-class samples.

Lastly, Figure 8 shows the poor performance that our models showed on the cropped images of the POLYU dataset. Since many of the images had darkly pigmented irises, the texture was sparsely highlighted, unlike the Cross-Eyed and WVU Multispectral Iris datasets. This caused a degradation in the cropped images, while the nor-

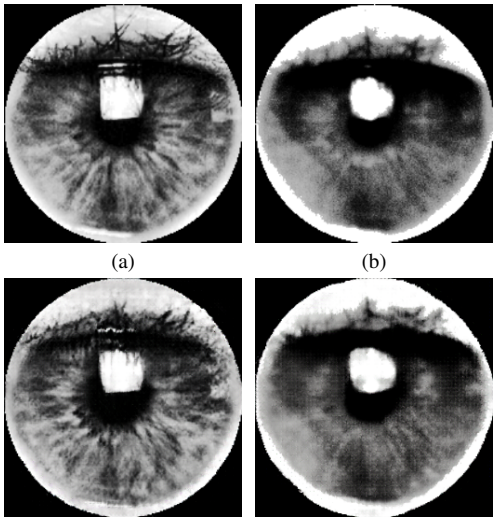

Figure 6. Samples of generated cropped images from the Cross-Eyed dataset, showing the translation between domains, with (a) real visible iris, (b) real NIR iris, (c) generated visible iris, and (d) generated NIR iris. This was a well-performing image on the matcher for both the NIR and VIS fake images.

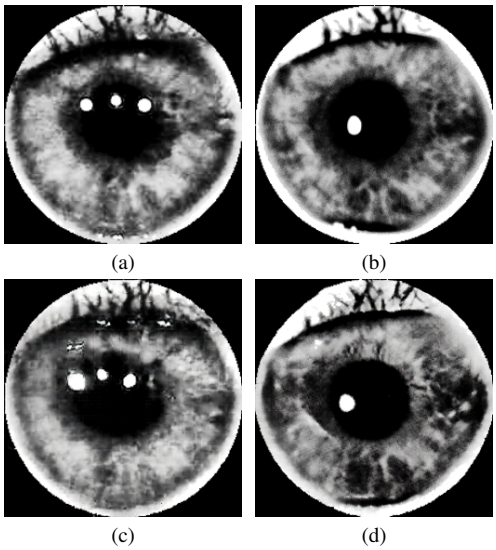

Figure 7. Samples of generated cropped images from the WVU Multispectral iris dataset, showing the translation between domains, with (a) real visible iris, (b) real NIR iris, (c) generated visible iris, and (d) generated NIR iris. This iris generated a higher match score than the baseline for the NIR vs Fake NIR but performed poorly on the VIS vs Fake VIS dataset

malized images were less affected. This dataset shows that normalizing iris images is still more robust with noisy and highly pigmented irises and more research is needed to highlight these features in cropped iris images consistently.

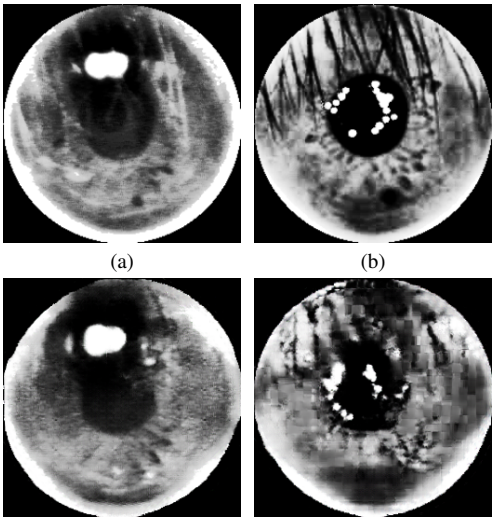

Figure 8. Samples of generated cropped images from the POLYU dataset learned by the generator, with (a) real visible iris, (b) real NIR iris, (c) generated visible iris, and (d) generated NIR iris. These images performed poorly in all macthing scenarios. This was the worst performing dataset for cropped iris images.

7. Conclusion

The performance of our models show that the addition of the classifier component introduces a consistent increase in performance, guiding the generator with valuable class-specific information. We were able to show the benefits of applying a classifier to work as an identity-preserving module as well as introduce a new method of enhancement that hasn't been deeply explored in iris generation. This work also shows the performance of generation models on cropped iris images, which is a much more difficult task due to the noise introduced from the pupil and reflections, but can be valuable and more robust with a smaller acceptable FAR, which is valuable for security, especially biometrics.

In future work, we hope to create better feature-enhancement techniques for highly variable datasets such as the POLYU dataset, as well as improving the performance of the model with disjoint images, which becomes a challenge when the images aren't normalized. However, one benefit to cropped iris images is the lower FAR, as shown in the VeriEye results for the WVU Multispectral Iris dataset. Though the EER is higher for the cropped images, they are able to achieve a much higher TAR with a low FAR. This allows a higher threshold value to be applied for score-based matchers such as VeriEye, with a low FAR, which is more detrimental than a false reject. Future research would be valuable to move from normalized to cropped iris images, as the overall performance isn't caught up yet, however, the results are promising and should be explored further.

References

[1] M. A. M. Abdullah, J. A. Chambers, W. L. Woo, and S. S. Dlay. Iris biometrie: Is the near-infrared spectrum always the best? In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 816–819, 2015.

[2] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Towards open-set identity preserving face synthesis, 2018.

[3] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2):281–307, 2008.

[4] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[7] K. Hernandez-Diaz, F. Alonso-Fernandez, and J. Bigün. Cross-spectral periocular recognition with conditional adversarial networks. *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9, 2020.

[8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[9] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.

[10] A. Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard GAN. *CoRR*, abs/1807.00734, 2018.

[11] J. Kim, G. Li, I. Yun, C. Jung, and J. Kim. Edge and identity preserving network for face super-resolution. *Neurocomputing*, 446:11–22, July 2021.

[12] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[13] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore. Synthetic iris presentation attack using idcgan. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 674–680, 2017.

[14] M. Mostofa, S. Mohamadi, J. Dawson, and N. M. Nasrabadi. Deep gan-based cross-spectral cross-resolution iris recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):443–463, 2021.

[15] M. Mostofa, F. Taherkhani, J. Dawson, and N. M. Nasrabadi. Cross-spectral iris matching using conditional coupled gan. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9, 2020.

[16] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image Processing*, 26(1):208–221, 2017.

[17] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image Processing*, 26(1):208–221, 2017.

[18] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image Processing*, 26(1):208–221, 2017.

[19] Neurotechnology. Verieye sdk, https://www.neurotechnology.com/verieye.html.

[20] A. Nigam, Lovish, A. Bendale, and P. Gupta. Efficient iris recognition system using relational measures. In *Computational Forensics: 5th International Workshop, IWCF 2012, Tsukuba, Japan, November 11, 2012 and 6th International Workshop, IWCF 2014, Stockholm, Sweden, August 24, 2014, Revised Selected Papers*, page 55–66, Berlin, Heidelberg, 2015. Springer-Verlag.

[21] N. Othman, B. Dorizzi, and S. Garcia-Salicetti. Osiris: An open source iris recognition software. *Pattern Recognition Letters*, 82:124–131, 2016. An insight on eye biometrics.

[22] D. Poster and N. M. Nasrabadi. Synthesis-guided feature learning for cross-spectral periocular recognition. *CoRR*, abs/2111.08738, 2021.

[23] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[24] A. Sequeira, L. Chen, P. Wild, J. Ferryman, F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, and J. Bigun. Cross-eyed - cross-spectral iris/periocular recognition database and competition. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2016.

[25] Y. Shen, P. Luo, J. Yan, X. Wang, and X. Tang. Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 821–830, 2018.

[26] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[27] R. Vyas, T. Kanumuri, and G. Sheoran. Cross spectral iris recognition for surveillance based applications. *Multimedia Tools and Applications*, 78, 03 2019.

[28] K. Wang and A. Kumar. Cross-spectral iris recognition using cnn and supervised discrete hashing. *Pattern Recognition*, 86:85–98, 2019.

[29] J. Wei, Y. Wang, Y. Li, R. He, and Z. Sun. Cross-spectral iris recognition by learning device-specific band. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6):3810–3824, 2022.

[30] Z. Zhao and A. Kumar. An accurate iris segmenta-

tion framework under relaxed imaging constraints using total variation model. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3828–3836, 2015.