Università degli Studi di Napoli Federico II
Corso di Laurea Magistrale in Informatica

# CNN-based transcoding from visible to near-infrared images for improved iris segmentation and recognition

Relatore:
Prof. Daniel Riccio

Correlatore:
Prof. Antonio Origlia

Candidato:
Salvatore Davide Amodio
N97000369

# Introduction

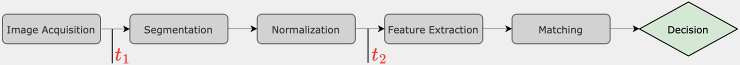
Researchers have shown that the near-infrared (NIR) spectrum is more effective than the visible (VIS) spectrum for iris recognition.

However, NIR-based iris recognition systems have limitations [1]:

- availability of near-infrared cameras
- in close-range acquisition scenarios

A potential solution could involve converting VIS iris images to NIR to enhance the performance of iris recognition systems under visible light conditions.

Critical steps in the pipeline of an iris recognition system are acquisition, segmentation, and feature extraction.



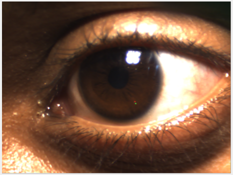

This study leverages the inherent correlation between visible (VIS) and near-infrared (NIR) spectra, proposing two transcoding models to:

- to mitigate some noise factors in iris images to improve iris segmentation
- to refine normalized (segmented) iris images to result in more discriminating features

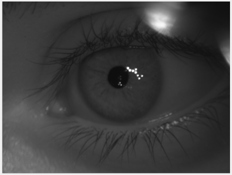
# Image quality enhancement for iris segmentation

In the NIR domain, iris images exhibit:

- reduction of the impact of visible light sources
- higher contrast between the iris and sclera
- higher distinction between the iris and pupil



VIS acquisition



NIR acquisition

**Figure:** Iris images extracted from the PolyU Database.

The UNB transcoding model consists of a U-shaped encoder-decoder with skip connections and residual blocks.

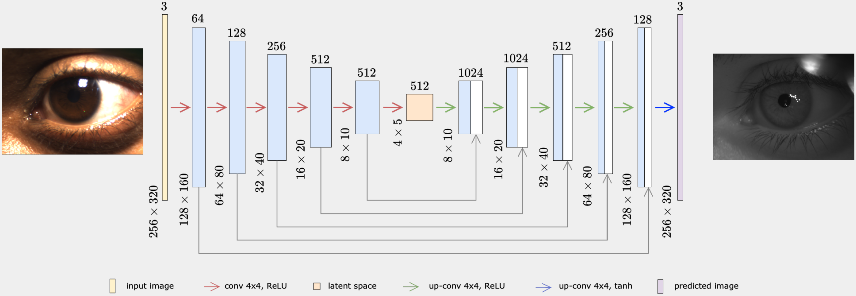



**Figure:** U-Net-Based Architecture

The figure illustrates two CNNs combined: the generator (PPB transcoding model) and the discriminator. The way of generating synthetic images is unchanged.

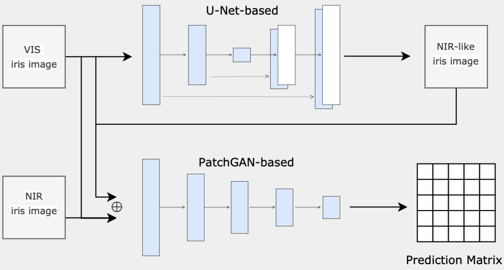



**Figure:** Pix2pix-Based Architecture

The discriminator can identify an iris image from a real distribution or generator network output by paying attention to local features combined with global structure.

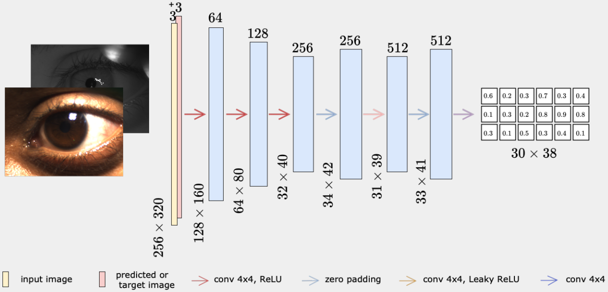



**Figure:** PatchGAN-based discriminator

The adversarial loss measures the gap between the distribution of generated data in a GAN and the distribution of real data [2].

$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) &= \mathcal{L}_{real\_dist}(D) + \mathcal{L}_{fake\_dist}(G, D) \\ &= \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]\end{aligned} \quad (1)$$

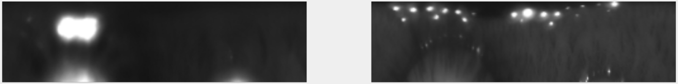
$G$ can only affect the distribution of fake data:

$$\mathcal{L}_{gen}(G, D) = \mathcal{L}_{fake\_dist}(G, D) + \lambda_P \mathcal{L}_{pw}(G) \quad (2)$$

where $\mathcal{L}_{pw}(G)$ is a traditional per-pixel loss and $\lambda_P$ weights its contribution within the overall loss function.

# Image quality enhancement for feature extraction

NIR light enables the device to acquire the complex structure of the iris region rather than its pigmentation. Being less prone to noise, NIR cameras obtain superb acquisitions even on dark-colored irises.



normalized VIS iris image

normalized NIR iris image

**Figure:** Sampling of a normalized VIS-NIR pair of iris images from the PolyU database.

It's crucial for synthetic images to faithfully replicate intricate iris texture details in NIR representations for accurate recognition, rather than just being realistic.

Daugman, in his study of iris as biometric, used a feature extractor to capture the distinctive pattern of the iris.

A transcoding loss function, including a module $\mathcal{L}_{fb}(U, F)$, assesses the distance between generated and NIR target images by evaluating Daugman's features for iris-specific characteristics.

- Daugman leverages Gabor filters to effectively capture and highlight the distinctive information inherent to the iris.
- To encode the information into an image, the information extracted from the filters is reorganized into an image:

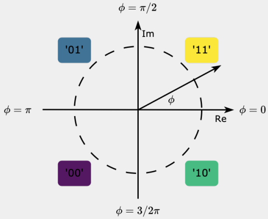
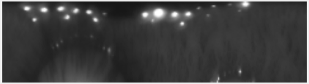
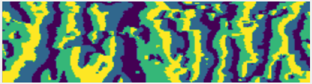



**Figure:** Phase-quadrant demodulation



normalized NIR iris image



NIR feature image

LogGaborNet receives a normalized near-infrared (NIR) iris image as input and emulates the complex response of the Daugman-based algorithm. LogGaborNet constitutes the module $\mathcal{L}_{fb}(U, F)$.

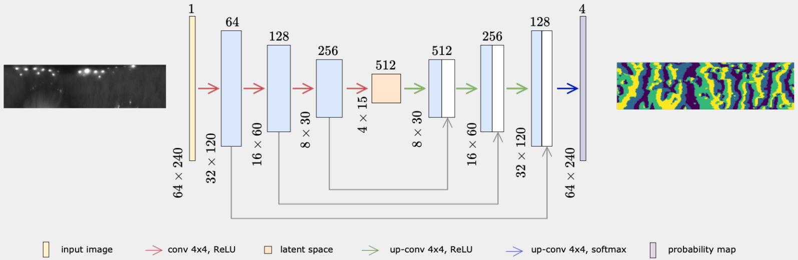



**Figure:** LogGaborNet architecture

The UNB transcoding model is fed by normalized VIS iris images of size $64 \times 240 \times 1$ and produces normalized NIR-like iris images of the same size.

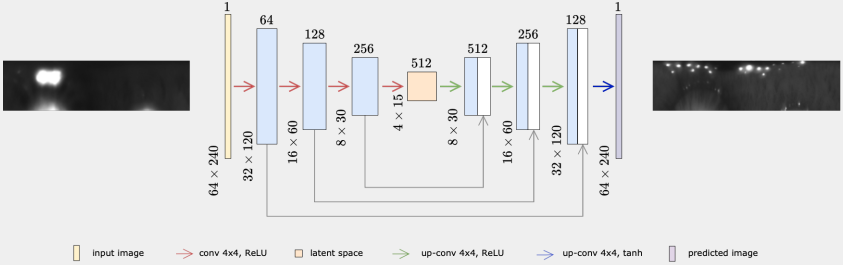



**Figure:** U-Net-Based architecture

A pixel-wise loss function $\mathcal{L}_{pw}(U)$ is combined with the feature-based loss function $\mathcal{L}_{fb}(U, F)$, resulting in:

$$\mathcal{L}_{ubm}(U, F) = \mathcal{L}_{pw}(U) + \underline{\lambda_F \mathcal{L}_{fb}(U, F)} \tag{3}$$

where $U$ is the current transcoding model, $F$ is the LogGaborNet model and $\lambda_F \in [0, 1]$ weights the contribution of $\mathcal{L}_{fb}(U, F)$ within the overall loss function.

- LogGaborNet is integrated within the Pix2pix-Based architecture.
- In the generator (the PPB transcoding model), the feature images act on the loss function.
- In the discriminator, the feature images join the network input.

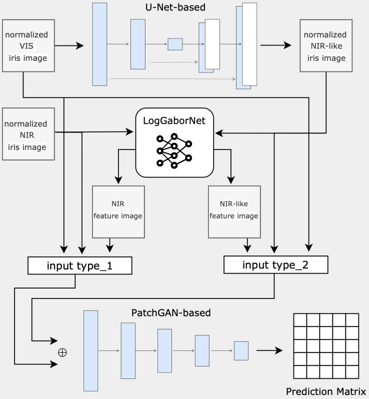



**Figure:** Pix2pix-Based architecture

The input is a concatenation of three images. In both cases, the input has a fixed size of $64 \times 240 \times 3$.

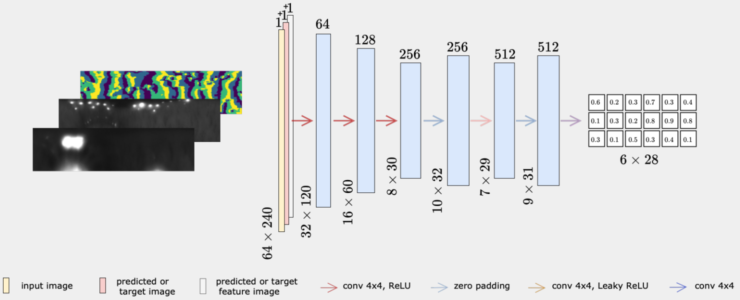



**Figure:** PatchGAN-based discriminator

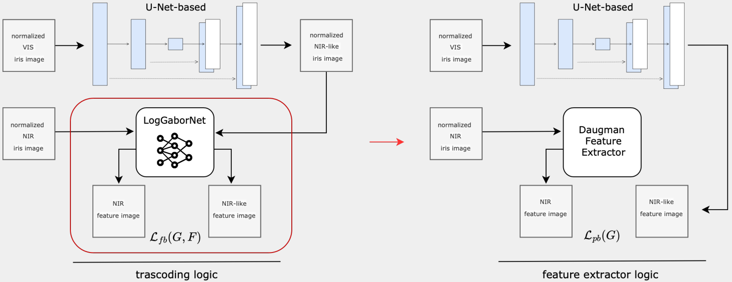
The adversarial loss measures the gap between the distribution of generated data in a GAN and the distribution of real data [2].

$G$ can only affect the distribution of fake data:

$$\mathcal{L}_{gen}(G, D, F) = \mathcal{L}_{fake\_dist}(G, D, F) + \lambda_P(\mathcal{L}_{pw}(G) + \underline{\lambda_F \mathcal{L}_{fb}(G, F)}) \quad (4)$$

where $\lambda_P$ weights the contribution of the generator reconstruction error made by $\mathcal{L}_{pw}(G)$, a per-pixel loss function, and $\mathcal{L}_{fb}(G, F)$ the feature-based loss function, weighted by $\lambda_F$.

- The architecture can be streamlined bypassing the generation of iris representation and proceeding to feature generation.
- This constraints iris recognition systems to incorporate this specific feature extractor in their architecture.



The UNB feature extractor is fed by normalized VIS iris images of size $64 \times 240 \times 1$ and produces a probability map of size $64 \times 240 \times 4$.

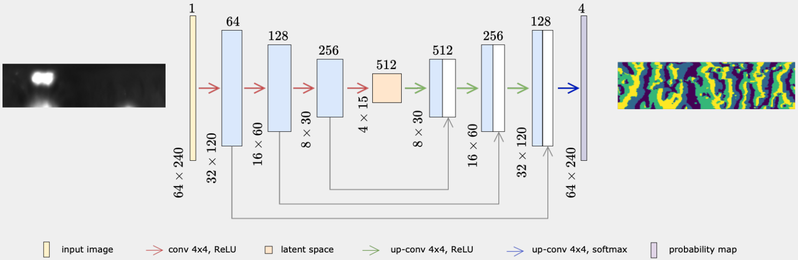



**Figure:** U-Net-Based architecture

The figure illustrates two CNNs combined: the generator (PPB feature extractor) and the discriminator. The way of generating NIR feature images is unchanged.

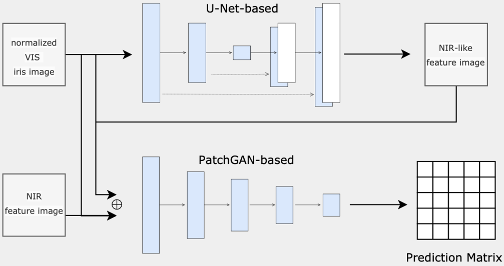



**Figure:** U-Net-Based architecture

The input is a concatenation of two images. In both cases, the input has a fixed size of $64 \times 240 \times 2$.

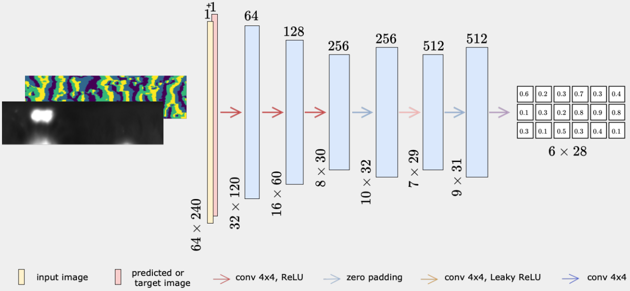



**Figure:** PatchGAN discriminator

The adversarial loss measures the gap between the distribution of generated data in a GAN and the distribution of real data [2]:

$G$ can only affect the distribution of fake data:

$$\mathcal{L}_{gen}(G, D) = \mathcal{L}_{fake\_dist}(G, D) + \lambda_P \mathcal{L}_{pw}(G) \tag{5}$$

where $\lambda_P$ weights the contribution of the generator error made by $\mathcal{L}_{pw}(G)$.

# Evaluation

- The PolyU database comprises 15 cross-spectral acquisitions, each consisting of left and right eye images, with the size of $640 \times 480 \times 3$.
- These acquisitions are available for 209 subjects.
- The database includes normalized (segmented) iris images of size $64 \times 512 \times 1$ corresponding to the same set of subjects [3].

- For the transcoding task involved in iris segmentation, named $t1$, the dataset comprises the first 4 acquisitions per eye of all the 209 subjects.
- To facilitate the learning, all right irises are horizontally reflected to make the dataset more uniform as if it presented only left irises.
- The training set has $1,200$ samples where each sample is a pair of ' VIS, NIR' iris images.
- The pre-processing step involved a Gaussian noise injection referred to as random jitter, a random mirroring, and normalization into $[-1, 1]$.

U-Net-based (UNB) and Pix2pix-based (PPB) architectures are illustrated to address the problem.

The evaluation of proposed models comes across three metrics:

- The Mean Absolute Error (MAE)
- The Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM)

The best configuration is obtained by setting a UNB model with $\mathcal{L}_{pw}(G) =$ VVG loss function with $\lambda_P = 1000$.

| | MAE | SSIM | PSNR |
|---|---|---|---|
| $UNB\_model_{t1}$ | **0.0787** | **0.87 ± 0.01** | **26.34 ± 13.05** |
| $PPB\_model_{t1}$ | 0.1123 | 0.81 ± 0.01 | 23.29 ± 11.06 |

**Figure:** Comparison of U-Net-based and Pix2pix-based models.

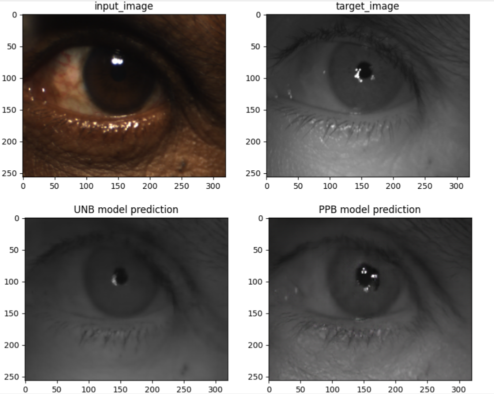



**Figure:** Predictions of $UNB\_model_{t1}$ and $PPB\_model_{t1}$ on subject 167, left eye, forth acquisition.

- For the transcoding task involved in feature extraction, named $t2$, the dataset comprises all the 15 normalized (segmented) iris images of all the 209 subjects provided by PolyU.
- The LogGaborNet, considers only the normalized NIR iris images as a dataset and their NIR feature images extracted through the Daugman-based algorithm.
- The image transcoding architecture incorporating LogGaborNet considers $4,500$ samples, where each sample is a pair of normalized 'VIS, NIR' iris images.

The evaluation for LogGaborNet models comes across four metrics:

- Accuracy and F1-score
- Dice coefficient and Jaccard index.

The best configuration is obtained by setting the model with a Dice loss rather than IoU loss or categorical cross-entropy loss.

| | Dice | IoU | Accuracy | F1-score |
|---|---|---|---|---|
| *fe_model* | 0.93 | 0.87 | 0.96 | 0.93 |

**Figure:** Evaluation of the selected model.

The best LogGaborNet model is incorporated into the general UNB and PPB architectures for image transcoding.

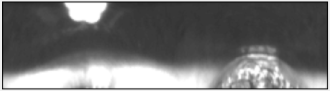
Due to a limited number of samples, the data augmentation step is necessary and involves in:

- horizontal and vertical flipping
- circular shifting

The evaluation of proposed models comes across two metrics:

- Equal Error Rate (EER) and Decidability (Dec)

| | Dec | EER |
|---|---|---|
| *normalized VIS iris images* | 1.2880 | 0.3399 |
| | | |
| | | |
| | | |

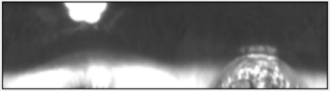


normalized VIS iris image

**Figure:** Iris recognition performance.

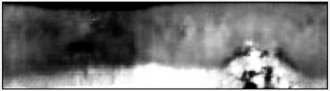
| | Dec | EER |
|---|---|---|
| *normalized VIS iris images* | 1.2880 | 0.3399 |
| *UNB_LF=Dice_$\lambda_F$=0.5_model$_{t2}$* | 1.2410 | 0.3772 |
| ***PPB_LF=Dice_$\lambda_F$=1_model$_{t2}$*** | **1.1671** | **0.1203** |
| | | |

**Figure:** Iris recognition performance.



normalized VIS iris image

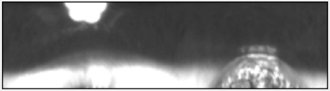
UNB transcoding model



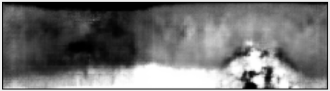
PPB transcoding model

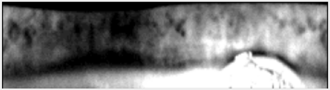
| | Dec | EER |
|---|---|---|
| *normalized VIS iris images* | 1.2880 | 0.3399 |
| *UNB_LF=Dice_$\lambda_F$=0.5_model$_{t2}$* | 1.2410 | 0.3772 |
| ***PPB_LF=Dice_$\lambda_F$=1_model$_{t2}$*** | **1.1671** | **0.1203** |
| *normalized NIR iris images* | 1.4681 | 0.0394 |

**Figure:** Iris recognition performance.


normalized VIS iris image


UNB transcoding model


PPB transcoding model


normalized NIR iris image

- For the feature extraction task, named $t3$, the dataset comprises all the 15 normalized (segmented) iris images of all the 209 subjects provided by PolyU.
- The pre-processing step involves into a data standardization with a final clipping to range in $[-1, 1]$.

The evaluation of proposed models comes across two metrics:

- Equal Error Rate (EER) and Decidability (Dec)

U-Net-based (UNB) and Pix2pix-based (PPB) architectures illustrated to address the problem achieve similar performance:

| | Dec | EER |
|---|---|---|
| *normalized VIS iris image* | 1.2880 | 0.3399 |
| *UNB_LF=Dice_model$_{t3}$* | 1.3746 | 0.1123 |
| ***PPB_LF=Dice_model$_{t3}$*** | **1.3638** | **0.1009** |
| *normalized NIR iris image* | 1.4681 | 0.0394 |

**Figure:** Evaluation iris recognition performance using Daugman algorithm on normalized VIS and NIR iris images, and UNB/PPB models on normalized VIS iris images.

Transcoding models for VIS-to-NIR domain conversion and a CNN-based feature extractor are integrated into the iris recognition system pipeline.

The system employs:

- $IS_{IS}$ method for iris segmentation [4]
- Rubber-sheet method for the iris normalization
- Daugman-based feature extraction method
- Pairwise Euclidean distances for matching

Transcoding models are tested together to evaluate their overall effectiveness regarding iris recognition.

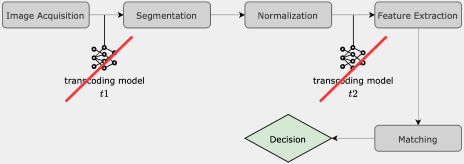



**Figure:** Transcoding models inside the iris recognition system.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| | | |
| | | |

**Figure:** IRS performance.

Transcoding models are tested together to evaluate their overall effectiveness regarding iris recognition.

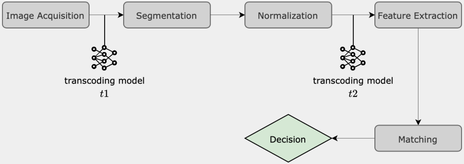



**Figure:** Transcoding models inside the iris recognition system.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| ***UNB_model***$_{t1}$ + ***PPB_LF=Dice_***$\lambda_F$***=1_model***$_{t2}$ | **0.8728** | **0.1462** |
| | | |

**Figure:** IRS performance.

Transcoding models are tested together to evaluate their overall effectiveness regarding iris recognition.

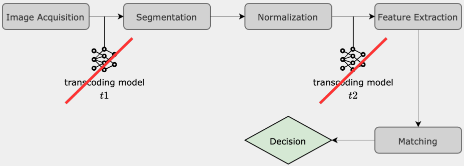



**Figure:** Transcoding models inside the iris recognition system.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| ***UNB_model$_{t1}$ + PPB_LF=Dice_$\lambda_F$=1_model$_{t2}$*** | **0.8728** | **0.1462** |
| *NIR iris images* | 1.5371 | 0.0993 |

**Figure:** IRS performance.

The transcoding model t1 and the CNN-based feature extractor are jointly tested to assess their overall effectiveness in iris recognition.

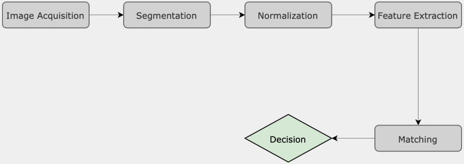



**Figure:** The iris recognition system without the transcoding model and the CNN-based feature extractor.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| | | |
| | | |

**Figure:** IRS performance.

The transcoding model t1 and the CNN-based feature extractor are jointly tested to assess their overall effectiveness in iris recognition.

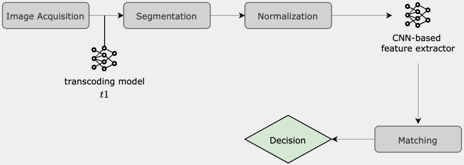



**Figure:** The iris recognition system constrained to the use of CNN-based feature extraction model.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| ***UNB_model***$_{t1}$ + ***PPB_LF=Dice__model***$_{t3}$ | **1.0151** | **0.1004** |
| | | |

**Figure:** IRS performance.

The transcoding model t1 and the CNN-based feature extractor are jointly tested to assess their overall effectiveness in iris recognition.

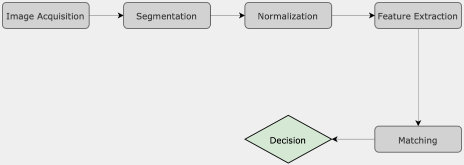



**Figure:** The iris recognition system without the transcoding model and the CNN-based feature extractor.

| | Dec | EER |
|---|---|---|
| *VIS iris images* | 0.9752 | 0.5246 |
| ***UNB_model$_{t1}$* + *PPB_LF=Dice__model$_{t3}$*** | **1.0151** | **0.1004** |
| *NIR iris images* | 1.5371 | 0.0993 |

**Figure:** IRS performance.

# Conclusions and future developments

In conclusion, this work demonstrates the feasibility of leveraging the relationship between the visible and near-infrared spectra to:

- develop a CNN-based transcoding model to enhance image quality and, consequently, iris segmentation.
- develop a CNN-based transcoding model to improve iris texture details and, thus, feature extraction.
- design a CNN-based feature extractor that reconstructs NIR-like features using the VIS spectrum.

In future improvements to this work, various aspects may be considered to refine the proposed approaches:

- For the transcoding model involved in iris segmentation:
  - expanded dataset for a more robust model, accommodating factors like mirror reflections, off-axis irises, blurring, and artifacts.
- For the model transcoding involved into feature extraction:
  - scaling up the dataset
  - changing the definition of the loss function that exploits different ways of comparing iris features.
- For the NIR-like feature extractor:
  - scaling up the dataset
  - changing the CNN architecture.

# Thank You
# for your attention!

[1] Mostofa M., Mohamadi S., Dawson J., and NasrabadiN . **“Deep gan-based cross-spectral cross-resolution iris recognition”.** In: *Transactions on Biometrics, Behavior, and Identity Science* 3.4 (2021), pp. 443–463.

[2] Isola P., Zhu J., Zhou T., and Efros A. **“Image-to-image translation with conditional adversarial networks”.** In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134.

[3] ***The Hong Kong Polytechnic University Cross-Spectral Iris Image Database.*** https://www4.comp.polyu.edu.hk/~csajaykr/polyuiris.htm. 2015.

[4] De Marsico M., Nappi M., and Riccio D. **“Is_is: Iris segmentation for identification systems”.** In: *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, pp. 2857–2860.