

# Face Recognition in Cross-spectral Environment using Deep Learning

Sana Khan  
CSE,

All Saints College of Technology, Bhopal, India

Zuber Farooqui  
CSE,

All Saints College of Technology, Bhopal, India

## ABSTRACT

Face recognition is very much popular in the present era, and many researchers are working on face recognition and produce the most promising results in terms of recognition and human identification. It has many applications for authentication and verification. Along with these advancements, face recognition is still challenging the heterogamous environment, such as near-infrared and visible spectrum. Matching of face images capture in the near-infrared spectrum (NIR) to face images of the visible spectrum (VIS) is a very challenging task.

In this research work, we have proposed a deep learning-based model for cross-spectral face matching (face recognition). The 26-layered deep residual network is extracted discriminative features from the face images and learn the common feature of the subject in the cross-spectral for the matching. To trained the proposed model, we have applied both VIS and NIR images with corresponding labels. For the performance evaluation of the proposed cross-spectral matching algorithm, experiments are performed on publicly available CASIA 2.0 NIR-VIS face datasets. The proposed method produced significant improvement in GAR. Our method gives recognition accuracy of 98.55 %.

## Keywords

Biometric system, Face Recognition, Near-infrared spectrum image, Visible spectrum image, Deep learning, Residual Network.

## 1. INTRODUCTION

The face recognition system which is very popular nowadays. It has very useful applications such as forensic, person identification, bank card identification [1], access control [2] and surveillance [3 4]. Face recognition has many challenges due to illumination variations, large dimensionality, uncontrolled environments, aging and pose variations, expression, motion blur, and occlusion, etc. One of the challenging problems in face recognition is face recognition in the cross-spectral environment. Cross spectrum matching means face image of one spectrum (Visible spectrum) is matched to another spectrum image (night vision or near-infrared spectrum) [5]. Face recognition in such scenarios is a very challenging and complicated task due to the visible and near-infrared spectrum has a deferent wavelength, the visual spectrum has a wavelength from  $.4\mu\text{m}$  to  $.7\mu\text{m}$  and NIR spectrum  $.7\mu\text{m}$  to  $1.4\mu\text{m}$ , and NIR spectrum is illumination invariant. Figure1 shows the sample images from the visible spectral and near-infrared (NIR).



Figure 1: Example of NIR images (top row) and visual spectrum (bottom row) from the CASIA database 2.0 [6]

In this research work, we have proposed a deep learning-based approach for face recognition in the cross-spectral environment. A cross-spectral environment such as near-infrared images, is matched with visual spectral images. The main objective of our research work is to develop an efficient convolutional neural network (CNN) model which efficiently and accurately recognizes the face in such a difficult scenario. In the proposed approach, we develop a deep residual neural network for face recognition. The developed method has been trained on the cross-spectral dataset from scratch. We have also used the data augmentation technique for improving network performance.

## 2. LITERATURE REVIEW

Face Recognition in the mismatched condition is recently very challenging problem and many researchers have been working on this problem. Data non-uniformity in face recognition are often addressed either by finding mismatched condition invariant algorithm or by modeling the non-uniformity.

Tan and Triggset al. [7] reduce the difference between NIR and VIS images by preprocessing based on Gamma correction, Difference-of-Gaussian (DoG) filtering, klare et al. [8] combine the histogram gradients (HOG) features with LBP to describe the face images. Light Source Invariant Features (LSIFs) is proposed to reduce the gap between VIS and NIR face image [9]. Goswami et al. introduced an efficient preprocessing chain to cut back the difference between VIS and NIR facial pictures supported Gamma correction, Difference-of-Gaussian (DoG) filtering and distinction deed [10]. Liao et al. [11] advised encryption both VIS and NIR face pictures victimization Multi-block LBP (MBLBP) followed by DoG filtering. Light AdaBoost and R-LDA were conducted for more feature choice. Following this work, Binary Laplacian of Gaussian (LoG) was also

investigated. Recently, Liu et al. projected light-weight Source Invariant options (LSIFs) to fill the gap between VIS and NIR face pictures [12-13].

For NIR-VIS face recognition, this article introduces a heterogeneous facE recognition neTwork (TRIVET) deep Transfer NIR-VIS. First, we use the deep convolutional neural network (CNN) with ordinary measures to learn discriminative models to use big numbers of unpaired VIS face pictures. The ordinal activation function (Max-Feature Map) is used to select discriminative characteristics and to render the designs robust and lighter. Second, by fine-tuning with two kinds of NIRVIS triplet loss, we move these models to the NIR-VIS domain.

In [14] Authors extend the initial JB by using two distinct Gaussian distributions to model the gallery and test pictures to suggest a heterogeneous joint Bayesian (HJB) formula for face recognition of cross-modality. The suggested HJB explicitly models the distinction in the modality of the picture pairs and is thus prepared to discriminate more appropriately the same/different face pairs In[15] writers present an efficient Disentangled Spectrum Variations Networks (DSVNs) for VIS-NIR HFR. DSVNs are implemented with two main approaches to disconnect spectrum variations between two domains: Spectrum-adversarial Discriminative Feature Learning (SaDFL) and Step-wise Spectrum Orthogonal Decomposition (SSOD).

The SaDFL is made up of the sub-network Identity-Discriminative (IDNet) and the subnetwork Auxiliary Spectrum Adversarial (ASANet). On the one hand, the IDNet consists of a GH generator and a DU discriminator to extract identity-discriminatory characteristics. On the other hand, under the guidance of the

discriminator DM, ASANet is built by a generator GH and a discriminator DM to eliminate modality-varying spectrum information [16]

In [17] proposed re-ranking of high-dimensional profound local representation to match near-infrared (NIR) and visual (VIS) facial pictures, i.e., face recognition of NIR-VIS. The first way to construct a high-dimensional profound local representation is to extract and concatenate profound characteristics on local facial patches through a convolutional neural network (CNN). By comparing the compressed profound characteristics, the original NIR-VIS recognition ranking findings can be achieved. We then suggest a novel and effective LLRe-Rank method to refine the original ranking outcomes, which can discover useful data from the original ranking outcome.

### 3. METHODOLOGY

In this section, we describe the proposed deep learning method. Deep learning techniques are a very popular and powerful method to solve the pattern recognition problem in computer vision. The proposed method is based on the residual learning-based convolutional neural network named FaceRes15 which is consists of 15 layers, one residual block. Detail descriptions proposed approach is given in the following sections.

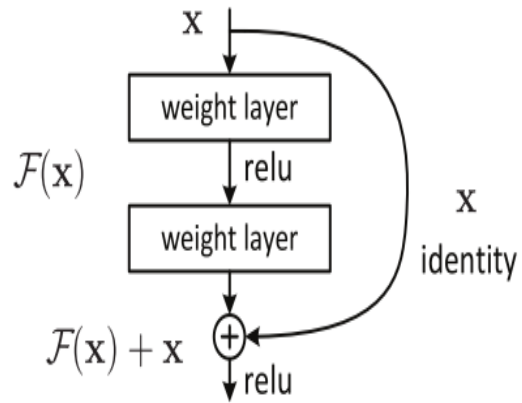


Figure 2: Residual Block

The name FaceRes15 is a short form of Residual Neural Network. It is a 15 layered convolutional neural network with a residual block. The main idea of ResNet is based upon “identity shortcut connection” that involves skipping one or more layers from the network. The residual block is represented in figure 2. The architecture of FaceRes15 is shown in figure 3.

### 3.1 Proposed Cross-spectral Face Recognition Method algorithm

We have used the concept of residual learning to the reorganization of human face images in the mismatched condition. Proposed model matched face images of near-infrared to the visible spectrum images. In the proposed approach, a ResNet15 model has trained on a training set of the face NIR-VIS images dataset. Test set images are given as the input to the trained model, which identifies the face images. Proposed approach algorithm is given as follows:

---

**Algorithm 1:** Reorganization of face images from the Cross-spectral environment.

**Input:** NIR query images and database images

**Output:** Recognized face VIS images corresponding to NIR query image.

---

**Begin**

- Step 1.** Read all images from the database.
- Step 2.** Divide dataset into train set, test set and validation set.
- Step 3.** Trained ResNet15 model by pass train set and validation set in the input.
- Step 4.** Now read test images and apply to the trained model.
- Step 5.** Test images are given as input to the trained model and Recognized face VIS images corresponding to NIR query image.

**End**

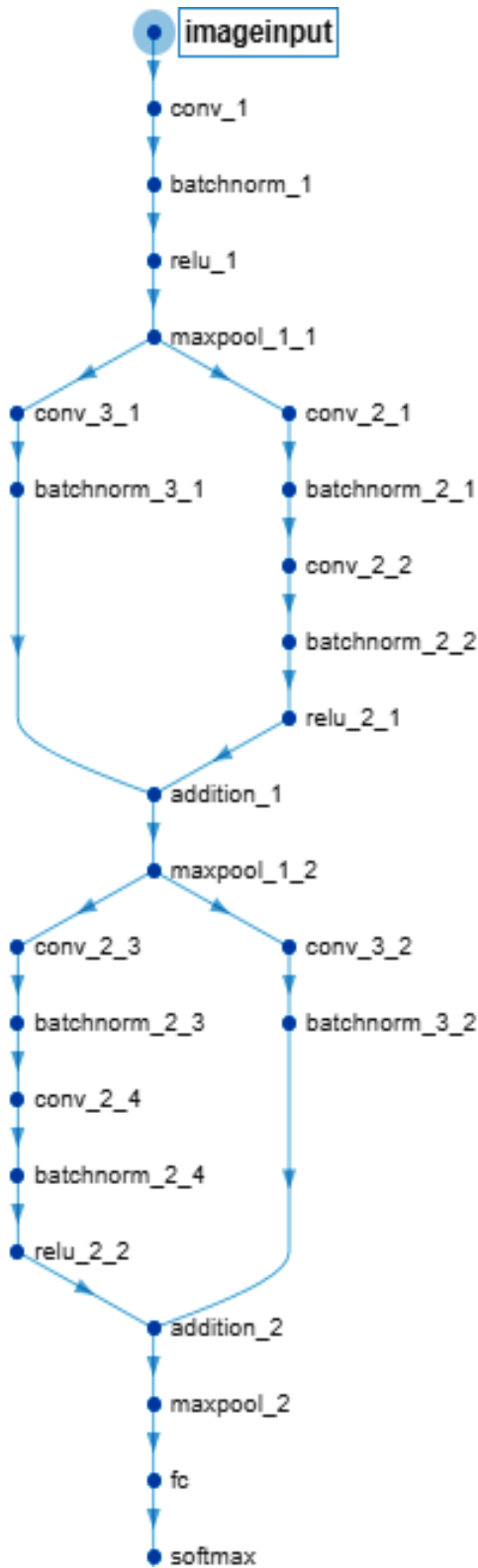


Figure 3: Overview of network connections.

## 4. IMPLEMENTATION AND RESULTS

### 4.1 Cross-spectral Database

The performance of the proposed method has been evaluated on publicly available CASIA NIR-VIS 2.0 [6] database. CASIA is a cross-spectral database in which image is captured in the night, by NIR camera and in the day, by the visual camera. CASIA database has images of different age groups of people. It also has different variations in the pose, illumination, and expressions. The figure shows 1.2 the sample images of the database. In our experiment, we have selected 80 subjects, 720 (9 images per subjects) visual and 720 (9 images per subjects) NIR spectrum images each size of 224X224X3 pixels. We have divided the CASIA dataset into a training set, validation set and test set in the ratio of 7:15:15.

We performed five sets of experimental trials on the dataset by using the same ration of training, validation and test set. Every time random images have taken in the trials, and we also ensure that there no intersection between training, validation and test set. The proposed FaceRes15 model is trained on the training set, and validation is used during training for selecting the best parameter for the network. Hence test set is used to evaluate the performance of the model. Network training performances are shown in figure 4.

### 4.2 Experimental Results

In this section, we present the results obtained from the experiment over the database of face images. In table 1 presents the results obtained in the cross-spectral scenario NIR-Visual. As mentioned previously, Table 1 shows the results obtained in the various experiential trail and mean face recognition accuracy of the proposed model. Figure 5 shows the face images retrieve by the model from the test set.

Table 1: Face Recognition accuracy of proposed model in Cross-Spectral Environment.

Trial (s)	Accuracy (in %)
Trial #1	98.25
Trial #2	98.75
Trial #3	98.25
Trial #4	98.75
Trial #5	98.75
Mean Accuracy	98.55

We have also compared our method with the existing state-of-the-art methods in Table 2. They proposed a convolution neural network (CNN) based method for identification, and they have also evaluated their method on the CASIA database, so we have done a direct comparison, our method with an existing method shown in the table. As we can see in table 5.2 proposed approach gives a significant improvement in the recognition rate compared to the existing method.

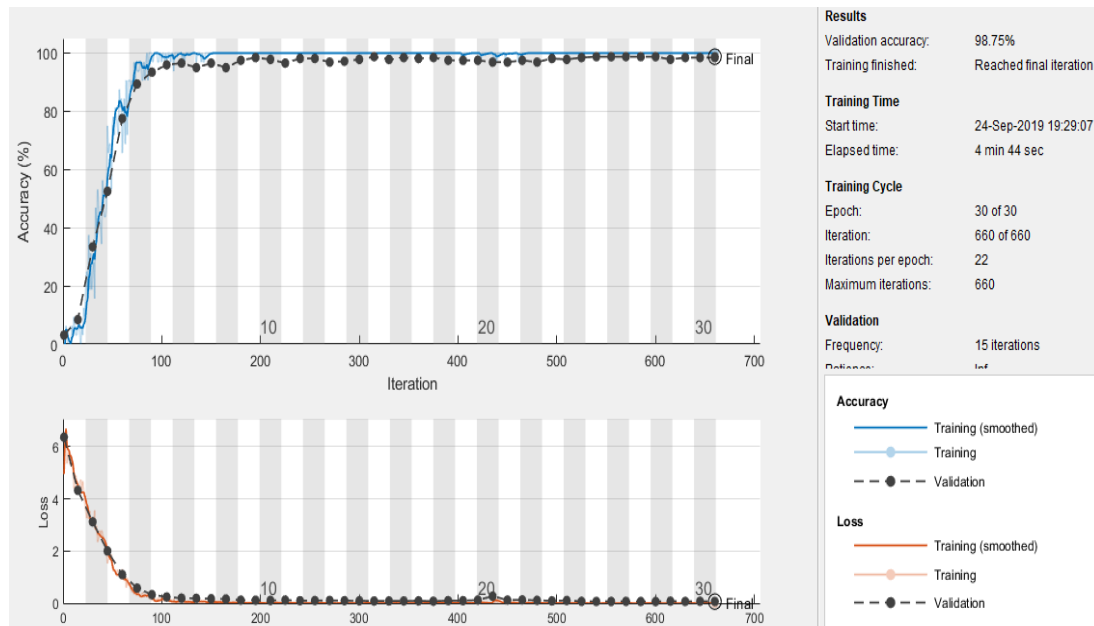


Figure 4: Network training performance.

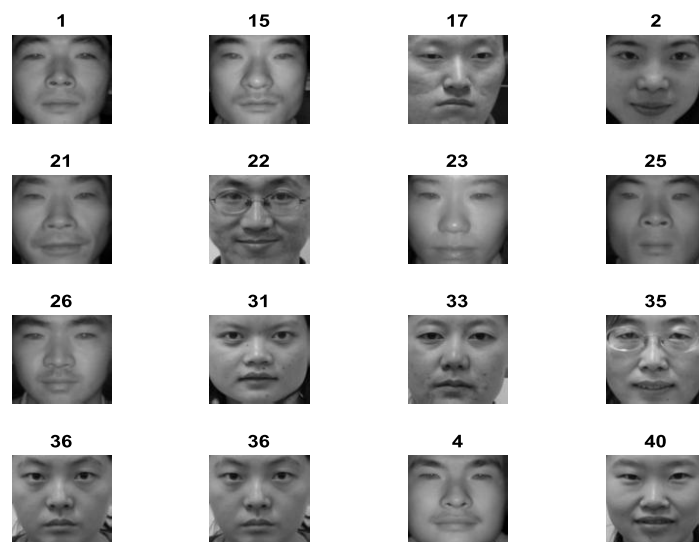


Figure 5: object identification by the proposed model in NIR to the VIS environment.

Table 2: Comparison of the proposed approach with the existing method

Method	Verification Accuracy (%)
X. Liu et al. [14]	91.03
H. Shi et al. [15]	89.91
W. Hu et al. [16]	97.3
C. Peng et al. [17]	96.5
Proposed method	98.55
Mean Accuracy	98.55

## 5. CONCLUSION

Near-infrared to visual face matching very challenging problem and many researchers are proposed their algorithm for NIR to VIS matching for face recognition and got promising results in terms of recognition accuracy. In this research work, we have proposed yet another simple algorithm based on deep learning model for cross-spectral matching. The 26-layered deep residual network is capable of the extracted discriminative features from the face images and also able to learn the common feature of the subject in the cross-spectral for the matching. To trained the proposed model we have applied both VIS and NIR images with corresponding labels of the publicly available CASIA 2.0 NIR-VIS face dataset. The proposed method produced significant improvement in GAR. Our method gives recognition accuracy of 98.55 %.

## 6. REFERENCES

- [1] C. Sanderson, *Biometric Person Recognition: Face, Speech and Fusion*. VDM Publishing, 2008.
- [2] S. Ouyang, T. Hospedales, Y. Song, and X. Li. A survey on heterogeneous face recognition: Sketch, infra-red, 3d, and low-resolution. In arXiv preprint arXiv:1409.5114, 2014.
- [3] G. H. K. and S. T. Inter-modality face sketch recognition. In *Multimedia and Expo, IEEE International Conference on*, pages 224–229, 2012.
- [4] B. Klare and A. K. Jain. Sketch-to-photo matching: a feature-based approach. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7667, page 1, 2010.
- [5] S. Z. Li, S. R. Chu, Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *PAMI*, 29:627-639, 2007.
- [6] S. Z. Li, D. Yi, Z. Lei, and S. Liao. The cassia nir-vis 2.0 face database. In *Computer Vision and Pattern Recognition Workshops, IEEE International Conference on*, pages 348–353, 2013.
- [7] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
- [8] B. Klare and A. Jain, “Heterogeneous face recognition: Matching NIR to visible light images,” in *Proc. 20th ICPR*, Aug. 2010, pp. 1513–1516.
- [9] S. Liu, D. Yi, Z. Lei, and S. Li, “Heterogeneous face image matching using multi-scale features,” in *Proc. 5th IAPRICB*, Mar./Apr. 2012, pp. 79–84.
- [10] D. Goswami, C. H. Chan, D. Windridge, and J. Kittler, “Evaluation of face recognition system in heterogeneous environments (visible vs NIR),” in *Proc. IEEE ICCVW*, Nov. 2011, pp. 2160–2167.
- [11] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Li, “Heterogeneous face recognition from local structures of normalized appearance,” in *Proc. 3rd Int. Conf. ICB*, Jun. 2009, pp. 209–218.
- [12] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Li, “Partial face matching between near infrared and visual images in MBGC portal challenge,” in *Proc. 3rd Int. Conf. ICB*, Jun. 2009, pp. 733–742.
- [13] T. Ojala, M. Pietikainen, and T. Maenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [14] Xiaoxiang Liu, Lingxiao Song, Xiang Wu, and Tieniu Tan, “Transferring Deep Representation for NIR-VIS Heterogeneous Face Recognition,” *Int. Conf. on Biometrics (ICB)*, pp. 1-8, 2016
- [15] Hailin Shi, Xiaobo Wang, Dong Yi, Zhen Lei, Xiangyu Zhu, and Stan Z. Li, Cross-Modality Face Recognition via Heterogeneous Joint Bayesian, *IEEE SIGNAL PROCESSING LETTERS*, VOL. 24, NO. 1, JANUARY 2017
- [16] Hu, Weipeng, and Haifeng Hu. "Disentangled Spectrum Variations Networks for NIR-VIS Face Recognition." *IEEE Transactions on Multimedia* (2019).
- [17] Peng, Chunlei, et al. "Re-ranking High-Dimensional Deep Local Representation for NIR-VIS Face Recognition." *IEEE Transactions on Image Processing* (2019).