Variance-guided Attention-based Twin Deep Network for Cross-spectral Periocular Recognition

Sushree Sangeeta Behera, N. B. Puhan

School of Electrical Sciences, IIT Bhubaneswar

National Science Day 11th Research Scholars' Day Feb 2021

Introduction

- We propose an attention-based twin deep convolutional neural network (CNN) with shared parameters to match the periocular images in heterogeneous modality.
- A new variance-based objective function is introduced to guide the network to attend more into the relevant regions of the periocular images.
- The weights of the twin model are learned so as to reduce the intra-class variance and to increase the inter-class variance of the cross-spectral image pairs.

Proposed Framework

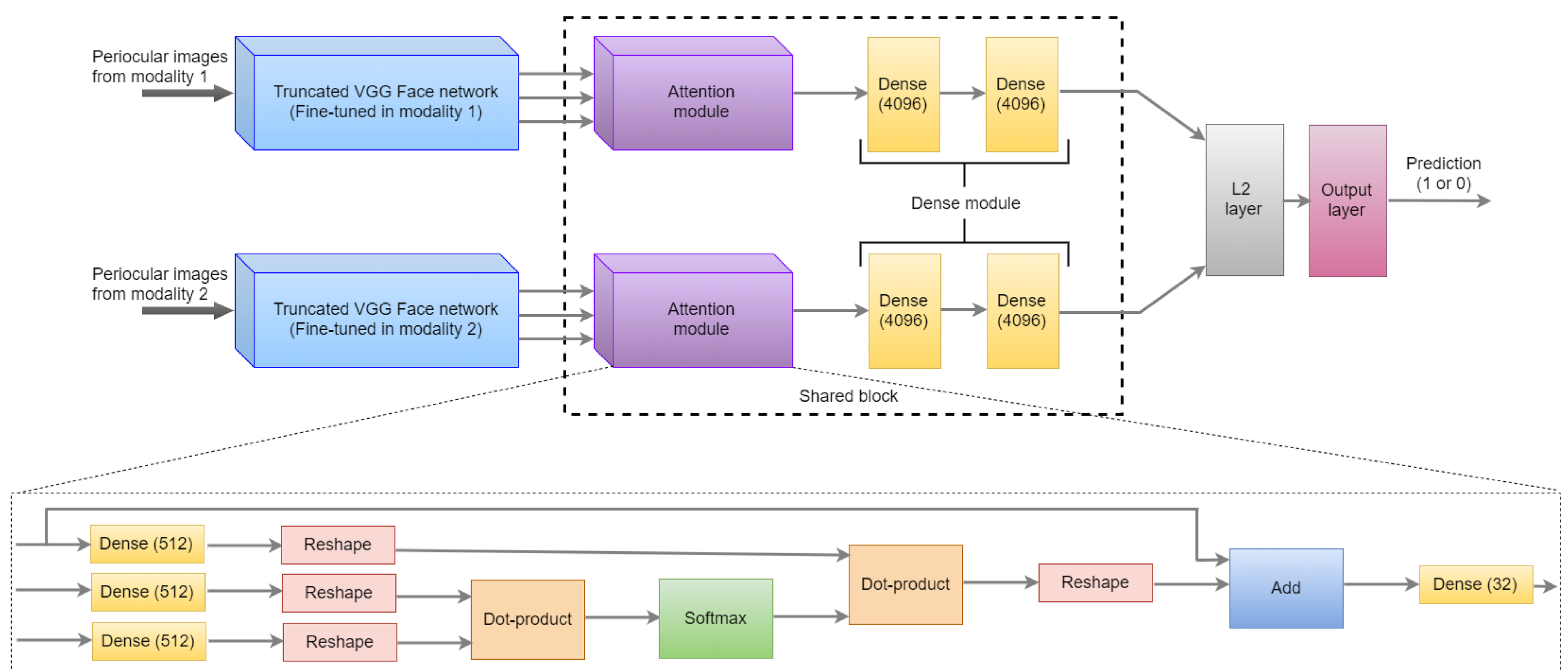


Figure 1: Block diagram of the proposed variance-guided attention-based cross-spectral network (VGACNet).

- The proposed model takes a pair of periocular images (from two different modalities) at its input and decides whether they belong to a same class or two different classes.
- The architectures of the top and bottom sub-networks are identical to each other. Each of these sub-networks contains a pre-trained VGGFace CNN, fine-tuned and truncated, followed by a shared block.

$$
\begin{aligned}
CNN_{trunc} \equiv & I(224,224,3) \longrightarrow 2\times conv(64) \longrightarrow Maxpool \\
& \longrightarrow 2\times conv(128) \longrightarrow Maxpool \longrightarrow 3\times conv(256) \\
& \longrightarrow Maxpool \longrightarrow 3\times conv(512) \longrightarrow Maxpool \\
& \longrightarrow 3\times conv(512) \longrightarrow Maxpool,
\end{aligned}
\tag{1}
$$

- The attention module transforms the input features into 8 smaller feature subspaces and then computes multiple scalar products to train the weights in order to attend these subspaces.
- Scaled dot product is employed to calculate the alignment scores which are then passed through the softmax function to compute the attention coefficients.

$$
\begin{aligned}
\mathbf{F}^{c,n}_{m,att} &= Att_{MH}(\mathbf{T}^{c,n}_{m,a}, \mathbf{T}^{c,n}_{m,b}, \mathbf{T}^{c,n}_{m,c}) \\
&= \left[\mathbf{T}^{c,n}_{m,c} \odot \left\{ softmax\left(\frac{\mathbf{T}^{c,n}_{m,a} \odot \mathbf{T}^{c,n}_{m,b}}{\sqrt{k}}\right)\right\}\right] + \mathbf{T}^{c,n}_{m,c}
\end{aligned}
\tag{2}
$$

$$
softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{p} e^{x_j}} \tag{3}
$$

- The output feature of the attention module is then fed to the input of the dense module.
- The feature vectors from both modalities are obtained from the final dense layer of this module and fed to the L2 layer which computes the difference between them using $L2$ norm.

$$
Diff(\mathbf{f}^{c,n}_{m_1,shared}, \mathbf{f}^{c,n}_{m_2,shared}) = \left[\mathbf{f}^{c,n}_{m_1,shared} - \mathbf{f}^{c,n}_{m_2,shared}\right]^2 \tag{4}
$$

Variance-guided objective function

- The new objective function takes both the categorical cross entropy (CCE) loss function and the variance based loss function into account using a scalar multiple μ whose value lies between 0 and 2.

$$
J_{var,CCE} = \begin{cases} J_{var} + J_{CCE}, & \mu = 1 \\ \frac{1}{2}[\mu \cdot J_{var} + (1-\mu)\cdot J_{CCE}], & \mu \neq 1, 0 \leq \mu \leq 2 \end{cases} \tag{5}
$$

- The term J_{var} refers to the variance based loss function defined as:

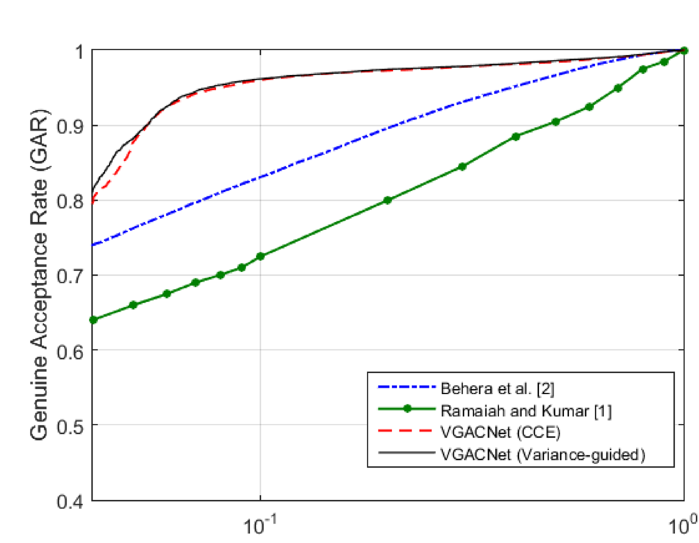
$$
\begin{aligned}
J_{var} = & \; l \cdot |\mathbb{E}[\mathbf{f}^2_{diff}] - \mathbb{E}[\mathbf{f}_{diff}]^2| \\
& + (1-l)\cdot [|\mathbb{E}[\mathbf{f}^2_{diff}] - \mathbb{E}[\mathbf{f}_{diff}]^2|]^{-1}
\end{aligned}
\tag{6}
$$

Experimental Results

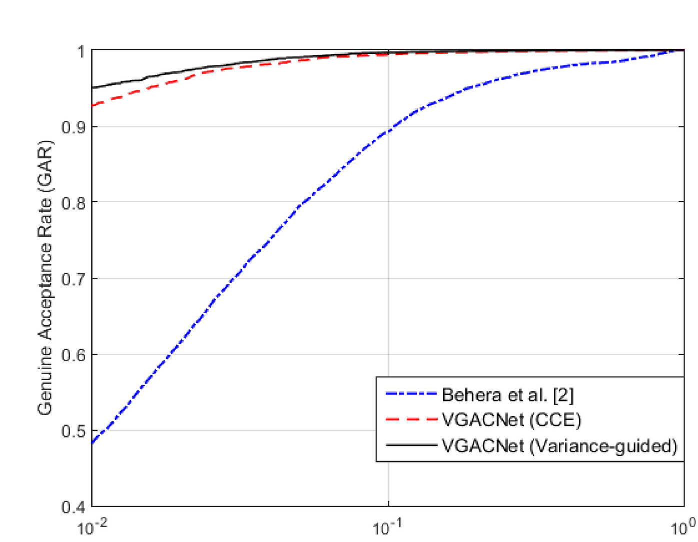
- Performance has been evaluated on three publicly available periocular image datasets namely, IIIT Delhi multi-spectral periocular (IMP) dataset, PolyU cross-spectral iris database and cross-eyed periocular database.
- Verification performance measures: ROC curves, EER and GAR at 0.1 FAR.

Table 1: Performance of the proposed VGACNet on all three evaluation datasets using both CCE loss function (J_{CCE}) and variance-guided objective function ($J_{var,CCE}$).

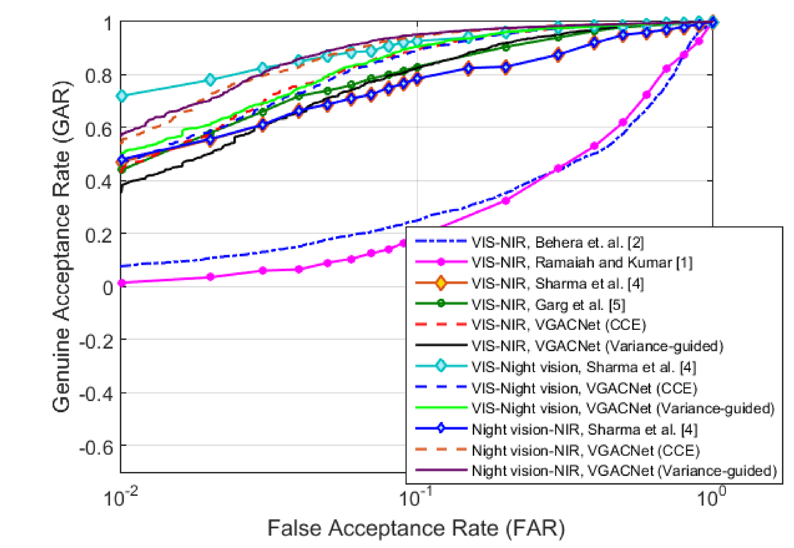
Database	Modality	Objective function	Accuracy (%)	Precision (%)	Recall (%)	AUC	GAR (%) at 10 % FAR	EER (%)
PolyU	VIS-NIR	J_{CCE}	94.60	91.50	93.93	0.97	96.01	6.51
		$J_{var,CCE}$	94.72	91.65	94.12	0.97	96.17	6.38
Cross-eyed	VIS-NIR	J_{CCE}	91.76	66.71	95.59	1.00	99.42	2.64
		$J_{var,CCE}$	99.00	93.02	94.67	1.00	99.70	2.36
IMP	VIS-NIR	J_{CCE}	67.47	51.45	83.17	0.96	89.47	10.27
		$J_{var,CCE}$	75.07	51.79	85.21	0.94	82.49	13.59
IMP	VIS-Night vision	J_{CCE}	71.81	51.67	85.47	0.95	89.37	10.26
		$J_{var,CCE}$	72.77	51.73	86.15	0.96	90.62	9.71
IMP	Night vision-NIR	J_{CCE}	71.48	51.66	85.60	0.97	94.33	7.62
		$J_{var,CCE}$	72.97	51.74	86.05	0.97	95.17	7.06



(a)



(b)



(c)

Figure 2: ROC curve comparisons of the proposed approach, VGACNet, with the state-of-the-art methods on (a) PolyU, (b) Cross-eyed, and (c) IMP datasets.

Table 2: Performance comparison in terms of GAR (%) at 10% FAR and EER (%) of the VGACNet (using both CCE loss function (J_{CCE}) and variance-guided objective function ($J_{var,CCE}$)) with existing methods.

Database (Modality)	Method	Performance measure GAR (%) at 10 % FAR	EER (%)
PolyU (VIS-NIR)	Ramaiah and Kumar [1]	73.20	-
	Behera [2]	83.12	13.87
	VGACNet (J_{CCE})	96.01	6.51
	VGACNet ($J_{var,CCE}$)	**96.17**	**6.38**
Cross-eyed (VIS-NIR)	Behera [2]	89.27	10.36
	Sequeira [3]	-	0.82
	VGACNet (J_{CCE})	99.42	2.64
	VGACNet ($J_{var,CCE}$)	99.70	2.36
IMP (VIS-NIR)	Ramaiah and Kumar [1]	18.35	-
	Behera [2]	25.03	45.84
	Sharma[1] [4]	47.08	-
	Garg[2] [5]	82.97	-
	VGACNet (J_{CCE})	**89.47**	**10.47**
	VGACNet ($J_{var,CCE}$)	82.49	13.59
IMP (VIS-Night vision)	Sharma[1] [4]	71.93	-
	VGACNet (J_{CCE})	89.37	10.26
	VGACNet ($J_{var,CCE}$)	**90.62**	**9.71**
IMP (Night vision-NIR)	Sharma[1] [4]	48.21	-
	VGACNet (J_{CCE})	94.33	7.62
	VGACNet ($J_{var,CCE}$)	**95.17**	**7.06**

[1] *GAR (%) at 1% FAR.*

[2] *The model is trained on the cropped periocular images from the huge CASIA NIR-VIS 2.0 face database and tested on the IMP database.*

Conclusion

- We propose a twin deep CNN, along with an innovative attention module guided by variance loss information to address the problem of matching periocular images in challenging cross-spectral environment.
- The experiments and evaluations on three benchmark databases show that the proposed VGACNet architecture gives comparable performance results with the state-of-the-art methods.

Publications

1. Behera, S. S., Mandal, B., and Puhan, N. B. (2020, February). Twin Deep Convolutional Neural Network-based Cross-spectral Periocular Recognition. In 2020 National Conference on Communications (NCC) (pp. 1-6). IEEE.
2. Behera, S. S., Mishra, S. S., Mandal, B., and Puhan, N. B. (2020). Variance-guided attention-based twin deep network for cross-spectral periocular recognition. Image and Vision Computing, 104, 104016.

References

[1] N. P. Ramaiah and A. Kumar. On matching cross-spectral periocular images for accurate biometric identification. In *8th IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, pp. 1–6, Sept 2016.

[2] S. Behera, et al. Periocular recognition in cross-spectral scenario. In *IEEE Int. Joint Conf. on Biometrics*, pp. 681–687, 2017.

[3] A. Sequeira, et al. Cross-eyed 2017: Cross-spectral iris/periocular recognition competition, In *IEEE Int. Joint Conf. on Biometrics*, pp. 725–732, 2017.

[4] A. Sharma, et al. On cross spectral periocular recognition. In *IEEE Int. Conf. on Image Processing*, Sept 2014.

[5] R. Garg, et al. Heterogeneity aware deep embedding for mobile periocular recognition, In *9th IEEE Int. Conf. on Biometrics Theory, Applications and Systems*, pp. 1–7, 2018.