

Dynamic Graph Representation for Occlusion Handling in Biometrics

Min Ren^{1,2}, Yunlong Wang², Zhenan Sun², Tieniu Tan²

¹University of Chinese Academy of Sciences, ²CRIPAC NLPR CASIA, Beijing, P.R. China
{min.ren, yunlong.wang}@cripac.ia.ac.cn, {znsun, tnt}@nlpr.ia.ac.cn

Abstract

The generalization ability of Convolutional neural networks (CNNs) for biometrics drops greatly due to the adverse effects of various occlusions. To this end, we propose a novel unified framework integrated the merits of both CNNs and graphical models to learn dynamic graph representations for occlusion problems in biometrics, called Dynamic Graph Representation (DGR). Convolutional features onto certain regions are re-crafted by a graph generator to establish the connections among the spatial parts of biometrics and build Feature Graphs based on these node representations. Each node of Feature Graphs corresponds to a specific part of the input image and the edges express the spatial relationships between parts. By analyzing the similarities between the nodes, the framework is able to adaptively remove the nodes representing the occluded parts. During dynamic graph matching, we propose a novel strategy to measure the distances of both nodes and adjacent matrixes. In this way, the proposed method is more convincing than CNNs-based methods because the dynamic graph method implies a more illustrative and reasonable inference of the biometrics decision. Experiments conducted on iris and face demonstrate the superiority of the proposed framework, which boosts the accuracy of occluded biometrics recognition by a large margin comparing with baseline methods.

Introduction

Deep learning methods have gained great success in recent years, especially in the area of computer vision. Convolutional Neural Networks (CNNs) have been widely applied as a powerful tool for image classification and feature extraction (LCun et al. 1998) (Krizhevsky, Sutskever, and Hinton 2012) (Simonyan and Zisserman 2014) (Szegedy et al. 2015) (He et al. 2016) (Gao et al. 2017) (Hu et al. 2018). CNNs are also universally applied to biometrics, including iris recognition (Liu et al. 2016) (Zhao and Kumar 2017) (Zhang et al. 2018), face recognition (Taigman et al. 2014) (Sun, Wang, and Tang 2014) (Schroff, Kalenichenko, and Philbin 2015) (Liu et al. 2017) (Wu et al. 2018) (Deng et al. 2018) and so on.

However, tasks of biometrics are different from natural image classification. They are greatly influenced by illumination, expression, occlusion and so on. Among them, oc-

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

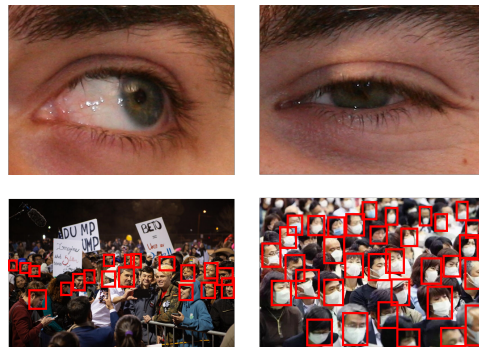


Figure 1: Iris samples and face samples in unconstrained environments.

clusion is the most common challenges problem in unconstrained situations (Jozer et al. 2013) (Azeem et al. 2014) (He et al. 2018) as shown in Figure 1. In these situations, gaps between intra-class samples are significantly enlarged. The generalization abilities of CNNs on occluded cases drops greatly due to the adverse effects of various occlusions.

Masking strategy is commonly seen in literatures for occlusion problems in biometrics. The features of occluded areas are suppressed during matching by masking out the invalid regions. However, detection and labelling of occlusion as masks are error-prone and the errors are accumulated into inaccurate recognition results. More importantly, the features of occluded areas can not be suppressed *after* feature extraction by masks in CNNs which are end-to-end frameworks. And the results of experiments demonstrate that masking the occluded areas *before* feature extraction is harmful.

In this paper, we propose a novel unified framework named Dynamic Graph Representation (DGR) for occluded biometrics. Basically, CNNs are adopted to extract convolutional representations, graph models are leveraged to overcome the occlusion problem. Dynamic graphs are built to remove the nodes of occluded part adaptively. As a result, the similarity between intra-class samples are significantly increased. The discriminative regions and the relationships between the regions are explicitly modeled by graphical blocks in the proposed framework.

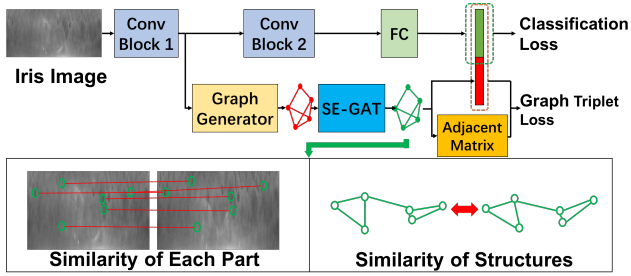


Figure 2: Framework of the proposed method. The similarity of each part of two samples and the similarity of the structures are contained in the Feature Graph.

Besides, graphical representations provide the reasons behind the decisions. The match or non-match nodes in the proposed dynamic graph method provides a more illustrative and reasonable inference of the biometrics decision. For example, the probe face image differs from gallery mainly in left eye and nose regions so it is an imposter.

The structure of the proposed framework is shown in Figure 2. Graphical model and CNNs are fused together in the proposed framework: Conv Block 1 produces the feature maps for Graph Generator to construct graphs which named Feature Graphs. Graph Generator could pick up a subset of spatial nodes inside the feature maps, convey edges to express the spatial relationships between nodes and finally combine them together as Feature Graphs as we defined. A novel deep graph model named Squeeze-and-Excitation Graph Attention Networks (SE-GAT) are proposed in this paper to process information of Feature Graphs. During the matching stage, dynamic graphs are built as shown in Figure 3. The nodes corresponding to the occluded parts are removed automatically by a straightforward and effective strategy.

In general, Feature Graphs can explicitly show us what parts of two samples are considered and the similarities between them. Hence, the underlying reasons results can be perceived.

The major contributions of this paper can be summarized as follows:

1. To the best of our knowledge, this is the first work that applies deep graphical models for biometrics. A novel deep graph model called SE-GAT is proposed as an essential component of the proposed framework.
2. The proposed framework provides a novel strategy to deal with occlusions, and the proposed framework achieves better performance by a large margin comparing with baseline methods. The decisions made by our framework are more convincing and reasonable than vanilla CNN-based methods.
3. The proposed framework possesses superior abilities to narrow the gaps between intra-class samples because of the dynamic graph matching, which significantly improves the generalization capability in partially occluded biometrics.

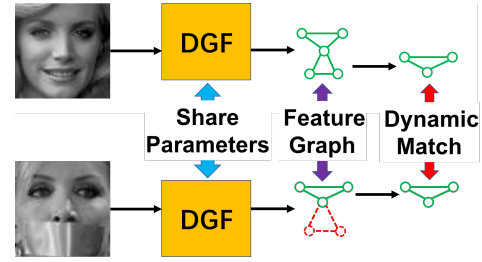


Figure 3: Dynamic matching based on Feature Graphs. The nodes of the occluded parts are removed during matching. Hence, the similarity between intra-class samples are significantly increased.

Related Work

Biometric approaches based on CNNs. Two common traits of biometrics, iris and face recognition, are selected to evaluate the proposed framework. We briefly review researches on these two modalities in this section.

Iris recognition has attracted increasing attention as one of the most accurate and reliable methods for identify authentication. Recently, CNN-based methods for iris recognition are presented. DeepIris is proposed in (Liu et al. 2016) for heterogeneous iris matching. Fully convolutional network (FCN) based model named UniNet is applied for iris recognition in (Zhao and Kumar 2017). UniNet (Zhang et al. 2018) adopts fully convolutional network (FCN) for feature extraction. MaxoutCNNs is proposed for iris and periocular recognition in (Zhang et al. 2018). Occlusion problem is common for iris recognition, almost all existing CNN-based methods adopt the masking strategy.

The first method for face representation based on CNN is proposed by Taigman Yaniv et al. (Taigman et al. 2014). A framework employing multiple convolutional networks is proposed by Yi Sun et al. (Sun, Wang, and Tang 2014). Triplet loss function is applied to get 128-D face embedding representation in (Schroff, Kalenichenko, and Philbin 2015). A light convolutional architecture is proposed for face recognition in (Wu et al. 2018). Topics about feature space of face draw more attention recently. Numerous of methods for better feature embedding are proposed, including SphereFace (Liu et al. 2017), ArcFace (Deng et al. 2018), CosFace (Wang et al. 2018) and so on.

There are also region-based or partial-based models for face recognition (Ou et al. 2018) (Cheheb et al. 2017) (He et al. 2018). However, these approaches need to set apart the occluded regions of face images before feature extraction.

Graphical methods for biometrics. There are some graphical methods for biometrics, including iris (Kerekes et al. 2007), face (Kisku et al. 2009), periocular (Proena and Briceno 2013), ear (Kisku et al. 2009) and hand vein (Horadam et al. 2014). These handcraft methods may provide us useful inspirations, despite these methods can not be blended in deep learning frameworks.

Deep learning approaches on graphs. Graphs are ubiquitous in the real world. Deep learning methods have been extended to graph data recently (Zhang, Cui, and Zhu 2018)

(Zhou et al. 2018). Graph Neural Networks (GNNs) are introduced as a recursive neural networks in (Marco, Gabriele, and Franco 2005) (Scarselli et al. 2009) to deal with graph data. Recently, the generalization of convolutional operation draws increasing attention. Researches in this direction can be roughly categorized as spectral approaches and non-spectral approaches (Zhou et al. 2018). The spectral approaches (JoanBruna et al. 2014) (Kipf and Welling 2016) work with spectral representation of graphs. The non-spectral approaches (Duvenaud et al. 2015) (Atwood and Towsley 2016) (Hamilton, Ying, and Leskovec 2017) define convolutional operation directly on graphs. Graph Attention Networks (GAT) (Velikovi et al. 2017) introduces attention mechanism to graph data.

Dynamic Graph Representation

In this section we introduce the framework of Dynamic Graph Representation (DGR). Two crucial components of DGR: Graph Generator and SE-GAT are described separately. After that, a novel loss function designed for the training of Dynamic Graph Feature Learning is shown. Finally, dynamic graph matching strategy is described.

In the proposed framework, images are firstly sent to a convolutional block (Conv Block 1) as shown in Figure 2. The block consists of several convolutional layers and pooling layers. There are two branches after Conv Block 1. The upper branch in Figure 2 contains another convolutional block (Conv Block 2) and fully connected layers. The two convolutional blocks and the fully connected layers constitute a common CNNs pipeline and any kinds of convolutional neural network can be incorporated into this branch. Typically, global features are extracted at the top of the upper branch.

Another branch starts at the feature maps extracted from the first convolutional block (Conv Block 1). A graph is generated from the feature maps by the Graph Generator. Each node of the graph contains a feature vector and the weights of edges express the relationships between nodes. We call it Feature Graph. Then, Feature Graph is sent to SE-GAT which is a hierarchical feature extractor for graph. SE-GAT is a novel structure based on Graph Attention Networks (GAT) and the details will be shown latter.

Two kinds of loss functions are applied during training. The global feature generated by the fully connected layers is sent to cross-entropy loss function which measures the classification loss. The Feature Graph generated by SE-GAT is sent to a novel graph triplet loss function. The novel loss function provides a new way to measure the similarity of two graphs. Semantic similarity and relationship similarity are both taken into consideration.

Graph Generator

There are two steps to generate Feature Graph from feature map as shown in Figure 4. The first step is generating the nodes, i.e. regression the spatial location of nodes and sampling from the feature map. The second step is generating the edges, i.e. generating the adjacent matrix according to relationships of nodes.

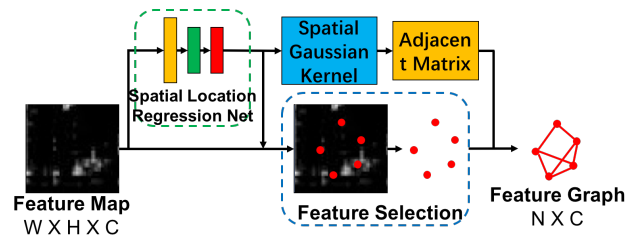


Figure 4: Structure of Graph Generator.

Spatial location regression of nodes are realized by a light network which called Spatial Location Regression Net as shown in Figure 4. This network consists of two convolutional layers and two fully connected layers, which transforms a $W \times H \times C$ feature map into N spatial coordinates in $N \times 2$ format by regression. Spatial coordinates are regressed explicitly. Feature vector of each node is sampled from the feature map by bilinear interpolation according to the spatial coordinates:

$$\mathbf{f}_{i,j}(c) = \mathbf{F}(i, j, c) \quad (1)$$

where \mathbf{f} is the feature vector, \mathbf{F} is the feature map, i, j are the spatial coordinates of the node and they are real numbers, c is the index of feature vector and $c \in \{1, 2, \dots, C\}$, C is the number of channels of the feature map. Bilinear interpolation is not shown in Equation 1 for brevity.

The adjacent matrix is generated by Gaussian kernel function according to the spatial position of nodes:

$$\mathbf{M}_{adj}(a, b) = \begin{cases} 0 & \|\mathbf{n}_a - \mathbf{n}_b\|_2 > R, \\ \exp(-\frac{\|\mathbf{n}_a - \mathbf{n}_b\|_2^2}{2R^2}) & \|\mathbf{n}_a - \mathbf{n}_b\|_2 < R. \end{cases} \quad (2)$$

where \mathbf{M}_{adj} is the $N \times N$ adjacent matrix, N is the number of nodes, $\mathbf{n}_a, \mathbf{n}_b$ are the spatial coordinates of two nodes, R is the scale of receptive field of each node. According to Equation 2, if the receptive fields on input image of two nodes are overlapped, edge will be generated between them. And the larger the overlapped area is, the larger the weight of the edge is.

After the two steps, $N \times C$ Feature Graph is generated from $W \times H \times C$ feature map, where N is the number of nodes. The Spatial Location Regression Net selects the most important locations of feature map as nodes of the Feature Graph. Spatial Gaussian kernel function establishes relationships between nodes with overlapped receptive fields.

SE-GAT

SE-GAT is a new network based on Graph Attention Networks (GAT) (Velikovi et al. 2017). It consists of two kinds of layers: squeeze-and-excitation (SE) layer and graph attention (GAT) layer.

Squeeze-and-excitation layer is a generalization of squeeze-and-excitation block (Hu et al. 2018). As shown in Figure 5, the average energy of each channel of nodes is calculated by the global pooling operation:

$$\mathbf{z}(c) = F_{sq}(\mathbf{G}) = \frac{1}{N} \sum_{i=1}^N \mathbf{f}_i(c) \quad (3)$$

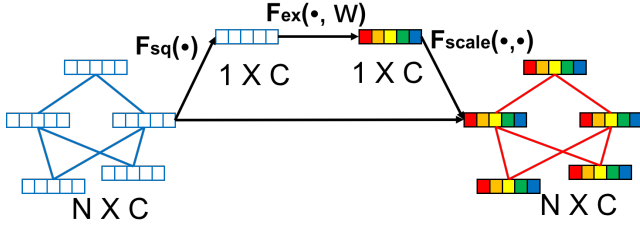


Figure 5: Structure of SE layer.

where \mathbf{z} is the squeezed vector, \mathbf{G} is the Feature Graph, N is the number of nodes, c is the index of the feature vector of each node and $c \in \{1, 2, \dots, C\}$. Two fully connected layers are used to make use of the global information:

$$\mathbf{s} = F_{ex}(\mathbf{z}, \mathbf{W}) = \delta_2(\mathbf{W}_2 \delta_1(\mathbf{W}_1 \mathbf{z})) \quad (4)$$

where \mathbf{s} is the scale vector, \mathbf{W}_1 , \mathbf{W}_2 are parameters of the two fully connected layers, δ_1 , δ_2 are ReLU function. The scale vector \mathbf{s} is used to re-scale each channel of the Feature Graph:

$$\mathbf{f}'_i(c) = \mathbf{s}(c) \mathbf{f}_i(c) \quad (5)$$

where \mathbf{f}' is the re-scaled feature vector, c is the index of feature vector and $c \in \{1, 2, \dots, C\}$, i is the index of nodes and $i \in \{1, 2, \dots, N\}$.

The graph attention (GAT) layer is a modified variant of the layer in (Velikovi et al. 2017). The weights of different edges of the Feature Graph is different rather than the same. The formulation of the layer is given as follows:

$$\mathbf{f}'_a = \delta(\alpha_{ab}(a, b) \mathbf{W} \mathbf{f}_b) \quad (6)$$

$$\alpha_{ab} = \frac{\exp(\delta(\mathbf{M}_{adj}(a, b) e_{ab}))}{\sum_k \exp(\delta(\mathbf{M}_{adj}(a, k) e_{ab}))} \quad (7)$$

$$e_{ab} = \mathbf{w}_{att}^T [\mathbf{W} \mathbf{f}_a || \mathbf{W} \mathbf{f}_b] \quad (8)$$

where $\mathbf{f}' \in \mathbb{R}^C$ is the output vector, $\mathbf{f} \in \mathbb{R}^C$ is the input vector, δ is nonlinear activation function, α_{ab} is the attention value between node a and node b , $\mathbf{W} \in \mathbb{R}^{C \times C}$ and $\mathbf{w}_{att} \in \mathbb{R}^{2C}$ are the parameters of the layer, \mathbf{M}_{adj} is the adjacent matrix of input graph, $||$ is the concatenation operation.

The framework of SE-GAT is shown in Figure 6. Residual structure has been proved to be effective for CNNs (He et al. 2016). We extend it to graph model in SE-GAT. There are four layers, two of them are SE layer and the rest are GAT layer as shown in Figure 6. Then the output is added to the input before a dimension reduction layer. The dimension reduction layer reduces the dimension of feature vectors of nodes:

$$\mathbf{f}'_i = \delta(\mathbf{W}_{C' \times C} \mathbf{f}_i) \quad (9)$$

where $\mathbf{W}_{C' \times C}$ is the parameter matrix, δ is nonlinear activation function, C' is the dimension of output feature vectors. Two kinds of features are separated from the output Feature Graphs of SE-GAT: feature vectors which express the semantic feature, adjacent matrix which expresses the relationship feature.

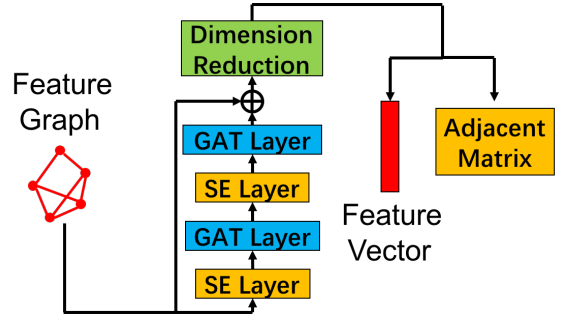


Figure 6: Framework of SE-GAT.

Graph Triplet Loss Function

We describe a new approach to measure the similarity between two Feature Graphs firstly. Then, the formula of the novel loss function is given.

The output of DGR contains two kinds of information: semantic information in the feature vectors and spatial relationship information in the adjacent matrixes of Feature Graphs. Hence, the similarity measure of two Feature Graphs consists of two terms:

$$S = S_{fea} + S_{adj} \quad (10)$$

where S_{fea} is the semantic similarity, S_{adj} is the relationship similarity.

$$S_{fea} = \cosine(\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2) \quad (11)$$

where \cosine represents the cosine similarity, $\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2$ are the concatenation of all feature vectors of Feature Graphs and feature vectors generated by fully connected layers.

$$S_{adj} = \|\mathbf{M}_{adj}^1 - \mathbf{M}_{adj}^2\|_2 \quad (12)$$

where $\mathbf{M}_{adj}^1, \mathbf{M}_{adj}^2$ are the adjacent matrix of Feature Graphs.

The graph triplet loss is:

$$L = \max\{0, m + S_{anc-neg} - S_{anc-pos}\} \quad (13)$$

where $S_{anc-neg}$ is the similarity between anchor sample and negative sample, $S_{anc-pos}$ is the similarity between anchor sample and positive sample, m is a positive margin. We set $m = 1$ during training.

Dynamic Graph Match

For two Feature Graphs, G_A and G_B , the mean of cosine similarities of corresponding nodes are calculated:

$$s_{gate} = \frac{1}{N} \sum_{i=1}^N \cosine(\mathbf{f}_i^A, \mathbf{f}_i^B) \quad (14)$$

where \mathbf{f}_i^A and \mathbf{f}_i^B are the feature vectors of corresponding nodes, N is the number of nodes in each Feature Graph. If the similarity of a node pair is less than s_{gate} , these two nodes are removed from the two Feature Graphs:

$$G_A^{Dyn}, G_B^{Dyn} = Re(G_A, G_B, s_{gate}) \quad (15)$$

where Re represents the remove operation according to the s_{gate} . The dynamic graphs G_A^{Dyn}, G_B^{Dyn} are built and the similarity of G_A^{Dyn}, G_B^{Dyn} is calculated according to the approach described above.

The node pairs with lower similarities are removed because the variances of occluded parts are much larger than the visible parts. And the gate given by the mean of similarities provides a straightforward and effective way to build dynamic graphs.

Experiments

Two common modalities of biometrics, iris and face, are selected to evaluate the proposed method.

Experiments of Iris Recognition

Protocols and Databases A simple novel convolutional architecture is used for iris experiments. The Conv Block 1 contains four convolutional layers and the Conv Block 2 contains one convolutional layer. The number of nodes of Feature Graph is 32.

For all iris images, the preprocessing procedure contains three steps: 1) eye detection by Haar-like Adaboost detectors (Viola and Jones 2004), 2) iris boundaries localization using method in (He et al. 2009), 3) iris normalization by rubber sheet model (Daugman 1993). Iris images are normalized to a rectangle with 128×256 resolution.

Four databases are used for experiments: (1) ND CrossSensor Iris 2013 Dataset-LG4000. It contains 29,986 iris samples from 1,352 classes. (2) CASIA Iris Image Database V4-Distance. Images of this database are acquired from 3 meters away. It contains 2,446 iris samples from 284 classes. (3) CASIA-Iris-M1-S2. Images of this database are acquired by mobile devices. It contains 6,000 iris samples from 400 classes. (4) CASIA Iris Image Database V4-Lamp. This database contains 16,212 iris samples from 819 classes. A lamp was turned on/off close to the subject to introduce elastic deformation under different illumination conditions.

The images of left eyes are selected for training and the images of right eyes for testing.

Comparisons to State-of-the-Art The proposed method is compared with the state-of-the-art approaches, including log-Gabor (Masek 2003), Ordinal Measures (OMs) (Sun and Tan 2009), UniNet (Zhao and Kumar 2017) and MaxoutCNNs (Zhang et al. 2018).

The results are shown in Figure 7 and Table 1. Significant improvements from proposed method can be found on all four databases.

Feature Graph of a iris sample from the CASIA Iris Image Database V4-Lamp is visualized in Figure 8. The nodes avoid the eyelid which is useless for iris recognition as we can see. Almost all nodes locate around the boundary of pupillary zone and ciliary zone. The locations of nodes indicate that the network “thinks” this area is the most important part of iris for recognition.

Occluded Iris Recognition Occluded iris recognition experiment is launched on the test database of the ND CrossSensor Iris 2013 Dataset-LG4000. Special area of iris

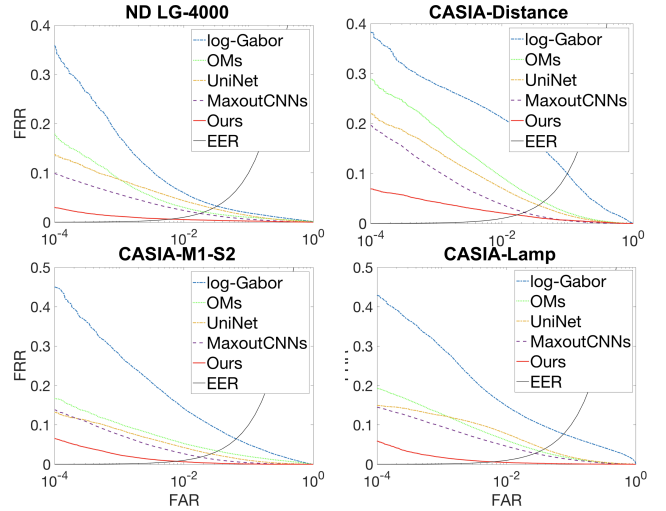


Figure 7: DET curves of iris recognition experiments.

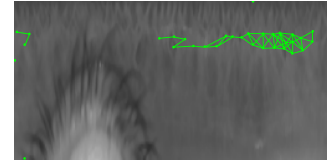


Figure 8: Visualization of the Feature Graph of a iris sample from the CASIA Iris Image Database V4-Lamp.

samples are covered by random noise to simulate the occluded situations in Figure 9. Two kinds of occlusion situations shown in Figure 9 are selected randomly for a special iris sample pair while the percentage of covered area remains. Note that all models used in this experiment are not trained on the occluded database.

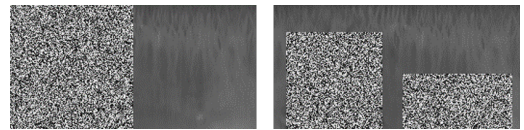


Figure 9: Two situations of occluded iris samples.

The results of 30% occluded are shown in Figure 10 and Table 2. The performance of the proposed method under occluded situations reduces slightly while the other method become worse seriously. The results indicate that the propose method has desired generalization ability for occluded situations.

Additional experiment are conducted to evaluate the influence of the covered area size. Performance of the proposed method under different percentages of covering are provided in Figure 11.

Experiments of Occluded Face Recognition

Protocols and Database Light CNN (Wu et al. 2018) is adopted to build the architecture for face recognition. The Light CNN-9 version which contains 9 convolutional layers is selected. The Conv Block 1 consists of the first 5 convolu-

Table 1: False reject rates (FRR) and equal error rates (EER) of iris recognition experiments.

	ND-LG4000		CASIA-Distance		CASIA-M1-S2		CASIA-Lamp	
	FRR@FAR=0.01%	EER	FRR@FAR=0.01%	EER	FRR@FAR=0.01%	EER	FRR@FAR=0.01%	EER
log-Gabor	35.89%	3.25%	38.30%	9.96%	45.02%	6.35%	42.94%	7.64%
OMs	17.94%	2.07%	28.87%	4.31%	16.67%	3.35%	19.32%	3.30%
UniNet	13.75%	2.80%	21.93%	3.63%	13.18%	2.64%	14.93%	3.88%
MaxouCNNs	9.90%	1.77%	19.61%	2.21%	13.83%	1.81%	14.46%	2.77%
Ours	3.02%	0.62%	6.94%	1.71%	6.57%	0.76%	5.92%	0.61%

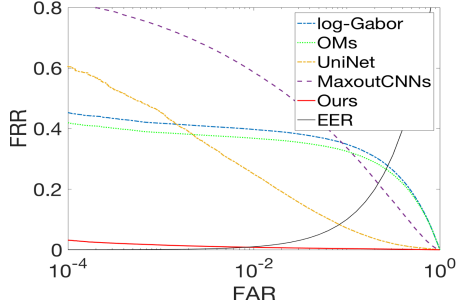


Figure 10: DET curves under occluded situations. The percentage of occluded area is 30%.

Table 2: FRR and EER of occluded situations. The percentage of occluded area is 30%.

	FRR@FAR=0.01%	EER
log-Gabor	45.29%	27.26%
OMs	41.94%	26.29%
UniNet	60.66%	8.31%
MaxoutCNNs	81.75%	21.79%
Ours	3.22%	0.85%

tional layers, and the Conv Block 2 consists of rest 4 convolutional layers. The number of nodes of Feature Graph is 64. Light CNN-9 is selected as the baseline method for comparison naturally.

The CASIA-WebFace (Yi et al. 2014) is adopted as the training database. The CASIA-WebFace is a public database which contains 494,414 face images from 10,575 subjects. All face images in training database are converted to gray-scale images before normalized to 128×128 resolution according to landmarks. Horizontal mirror operation is conducted for data augmentation because faces have nearly symmetric structure.

Occluded Face Verification A simulated occluded face database named Occluded-LFW which is based on the Labeled Faces in the Wild (LFW) database is used for evaluation. LFW contains 13,233 images from 7,749 individuals. After the same pre-processing of the training images, special area of face images are covered by random noise to simulate the occluded situations during this experiment as shown in Figure 12. Four kinds of occlusion situations shown in Figure 12 are selected randomly for a special face image pair while the percentage of covered area remains. In order to evaluate the generalization ability of the proposed method,

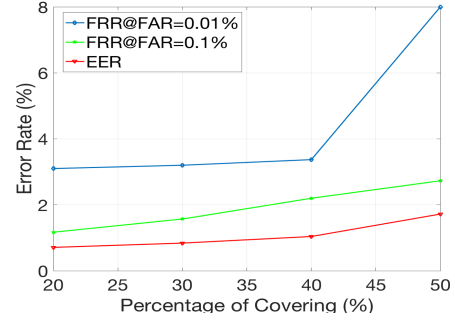


Figure 11: Performance of the proposed method under different percentages of occlusion.

the models in this experiment are not tuned on the occluded database.

We follow the Labeled Faces in the Wild (LFW) benchmark protocol¹, where 3,000 positive pairs and 3,000 negative pairs of images are selected for face verification. One of the face image of each pair is from the Occluded-LFW database and the other is from the LFW database.

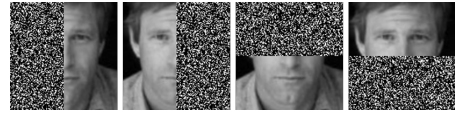


Figure 12: Four situations of occluded face samples.

Experiments with two different percentages of occluded area are launched: 30% and 50%. The results are shown in Figure 13 and Table 3. The performance of the proposed method is better than the baseline method obviously on both two percentages of occluded area.

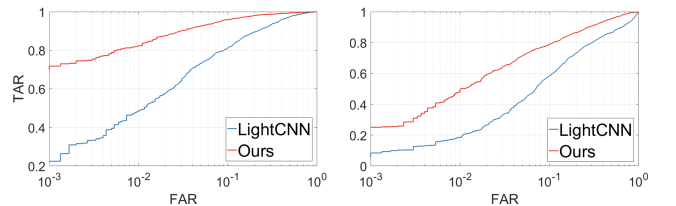


Figure 13: ROC curves of occluded face verification. The percentage of occluded area of the left figure is 30%. The percentage of occluded area of the right figure is 50%.

¹<http://vis-www.cs.umass.edu/lfw/pairs.txt>

Table 3: Results of occluded face verification on the Occluded-LFW.

	TPR@FTR=0.1%	100%-EER
Percentage of occlusion: 30%		
LCNN-9	22.47%	85.97%
Ours	71.8%	93.63%
Percentage of occlusion: 50%		
LCNN-9	8.47%	76.47%
Ours	25.19%	84.23%

Feature Graphs of a pair of face images are visualized in Figure 14. The first row shows the 10 pairs of nodes with highest similarity scores. The second row shows the 10 node pairs with lowest similarity scores. All the 10 pairs nodes with highest similarity scores lie in the left area of the right eye which indicates that the network “thinks” this is the most similar part of the two images. This example proves that we can read out the reasons for the results shown by the network from the Feature Graph.

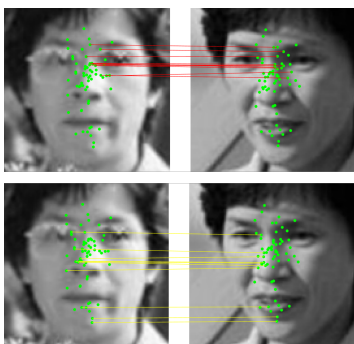


Figure 14: Matched nodes of a pair of images. The first row shows the 10 node pairs with highest similarity scores. The second row shows the 10 node pairs with lowest similarity scores.

Experiments on Masking Strategy

In this section, experiments on masking strategy are launched for comparison. Experiments of this section are launched on the test database of the ND CrossSensor Iris 2013 Dataset-LG4000. The percentage of occluded area of each sample is 50%. Two frameworks are used to evaluate the masking strategy: the proposed method and the backbone, which is a normal CNNs, of it. In the situation without masking strategy, the occluded samples are sent to networks directly. In the situation with masking strategy, occluded areas of samples are masked by precise masks, which eliminate the influence of inaccuracy of masks, before sent to networks.

The results are shown in Figure 15. The performances of situations with masking strategy are worse than the performances without masking strategy in both two frameworks as we can observe in Figure 15. The reasons may be that the masks disturb the feature extraction. The results indicate that covering occluded areas of input images by masks is harmful to deep learning framework.

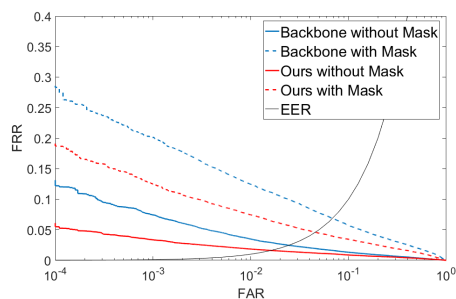


Figure 15: Comparisons to masking strategy. The percentage of occluded area is 50%.

Parameter Analysis

The number of nodes of Feature Graph is a important hyper-parameter of the proposed framework. To evaluate the influence of the number of nodes, we conduct experiments on different number of nodes. The test database of the ND CrossSensor Iris 2013 Dataset-LG4000 are selected to launch the experiments.

The results are shown in Table 4. The dimension of feature vectors are also listed. Too many nodes will increase computational cost and some distractive information is included. On the other hand, too few nodes fail to achieve good accuracy. An analysis on the graph nodes/connections will be added.

Table 4: Results on the ND-LG4000

Num Nodes	FRR@FAR=0.01%	EER	Feature Dimension
8	4.51%	0.93%	512
32	3.01%	0.62%	1280
128	4.56%	0.73%	4352

Conclusion

We propose a novel deep learning framework called Dynamic Graph Representation (DGR) for occlusion handling in biometrics in this paper. Dynamic graphs are adopted to overcome the occlusion situations. A novel deep graph model is proposed for processing of the Feature Graph. The excellent performance and superior generalization ability are demonstrated by extensive experiments on multiple modality databases. The main idea of this paper is straightforward, and we believe that there is much room for improvements. Further exploration of this work should focus on the better graph generation strategy and the better method for dynamic graph building.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 61427811, U1836217, 61603391). Special thanks to Ran He, Qi Li and Yuhao Zhu.

References

- Atwood, J., and Towsley, D. 2016. Diffusion-convolutional neural networks. *Computer Science*.
- Azeem, A.; Sharif, M.; Raza, M.; and et al. 2014. A survey: Face recognition techniques under partial occlusion. 11(1): 1–10.

- Cheheb, I.; Al-Maadeed, N.; Al-madeed, S.; Bouridane, A.; and Jiang, R. 2017. Random sampling for patch-based face recognition. In *International Workshop on Biometrics & Forensics*.
- Daugman, J. G. 1993. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 15(11):1148–1161.
- Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2018. Arcface: Additive angular margin loss for deep face recognition. *arXiv preprint arXiv:1801.07698*.
- Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gomez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; and Adams, R. P. 2015. Convolutional networks on graphs for learning molecular fingerprints. In *International Conference on Neural Information Processing Systems*.
- Gao, H.; Liu, Z.; Maaten, L. V. D.; and et al. 2017. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* 4700–4708.
- Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems* 1024–1034.
- He, Z.; Tan, T.; Sun, Z.; and Qiu, X. 2009. Toward accurate and fast iris segmentation for iris biometrics. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 31(9):1670–1684.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition.
- He, L.; Li, H.; Zhang, Q.; and Sun, Z. 2018. Dynamic feature matching for partial face recognition. 791–802.
- Horadam, K.; Arakala, A.; Davis, S.; and Lajevardi, S. M. 2014. Hand vein authentication using biometric graph matching. *Biometrics Iet* 3(4):302–313.
- Hu, J.; Shen, L.; Albanie, S.; Sun, G.; and Wu, E. 2018. Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* 7132–7141.
- JoanBruna; Zaremba, W.; Szlam, A.; and Lecun, Y. 2014. Spectral networks and locally connected networks on graphs. *Computer Science*.
- Jozer, B.; Feder, M.; Omelina, L.; Milos, O.; and Pavlovicova, J. 2013. Face recognition under partial occlusion and noise. In *Eurocon*.
- Kerekes, R.; Narayanaswamy, B.; Thornton, J.; Savvides, M.; and Kumar, B. V. K. V. 2007. Graphical model approach to iris matching under deformation and occlusion.
- Kipf, T. N., and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kisku, D. R.; Mehrotra, H.; Gupta, P.; and Sing, J. K. 2009. Probabilistic graph-based feature fusion and score fusion using sift features for face and ear biometrics. *Proc Spie* 7443:744306–744306–12.
- Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*.
- Liu, N.; Zhang, M.; Li, H.; Sun, Z.; and Tan, T. 2016. Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters* 82(2):154–161.
- Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; and Song, L. 2017. Sphereface: Deep hypersphere embedding for face recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* 212–220.
- LCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.
- Marco, G.; Gabriele, M.; and Franco, S. 2005. A new model for learning in graph domains. In *IEEE International Joint Conference on Neural Networks*.
- Masek, L. 2003. Recognition of human iris patterns for biometric identification. *The University of Western Australia* 2.
- Ou, W.; Li, G.; Yu, S.; Xie, G.; Ren, F.; and Tang, Y. 2018. Robust discriminative nonnegative patch alignment for occluded face recognition. *Pattern Recognition Letters* 107: 41–49.
- Proena, H., and Briceno, J. 2013. Periocular biometrics: Constraining the elastic graph matching algorithm to biologically plausible distortions. *Biometrics Iet* 3(4):167–175.
- Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2009. The graph neural network model. *IEEE Transactions on Neural Networks* 20(1):61.
- Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision & Pattern Recognition*.
- Simonyan, K., and Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.7923*.
- Sun, Z., and Tan, T. 2009. Ordinal measures for iris recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 31(12):2211.
- Sun, Y.; Wang, X.; and Tang, X. 2014. Deep learning face representation from predicting 10,000 classes. In *IEEE Conference on Computer Vision & Pattern Recognition*.
- Szegedy, C.; Liu, W.; Jia, Y.; and et al. 2015. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition* 1–9.
- Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. 2014. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision & Pattern Recognition*.
- Velikovi, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks.
- Viola, P. A., and Jones, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision* 57(2):137–154.
- Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Li, Z.; Gong, D.; Zhou, J.; and Liu, W. 2018. Cosface: Large margin cosine loss for deep face recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* 5265–5274.
- Wu, X.; He, R.; Sun, Z.; and Tan, T. 2018. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics & Security* 13(11):1–1.
- Yi, D.; Lei, Z.; Liao, S.; and Li, S. Z. 2014. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.
- Zhang, Q.; Li, H.; Sun, Z.; and Tan, T. 2018. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics & Security* PP(99):1–1.
- Zhang, Z.; Cui, P.; and Zhu, W. 2018. Deep learning on graphs. *arXiv preprint arXiv:1812.04202*.
- Zhao, Z., and Kumar, A. 2017. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *IEEE International Conference on Computer Vision*, 3829–3838.
- Zhou, J.; Cui, G.; Zhang, Z.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2018. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*.