Ulster University

Irish Machine Vision and Image Processing Conference 2018

A Minkowski Distance-based Generalisation Method for Improving Centre Loss for Deep Face Recognition

Xin Wei, Hui Wang, Huan Wan and Bryan Scotney

Outline:

1. Introduction
2. The proposed method
3. Conclusion and future works

Ulster University

1. Introduction

1. Introduction

Facial recognition system:

a technology capable of identifying or verifying a person from a digital image or a video frame [1].

1. Wikipedia contributors. (2018, August 27). Facial recognition system. In Wikipedia, The Free Encyclopedia. Retrieved 09:49, August 28, 2018, from https://en.wikipedia.org/w/index.php?title=Facial_recognition_system&oldid=856753855

Ulster University

1. Introduction

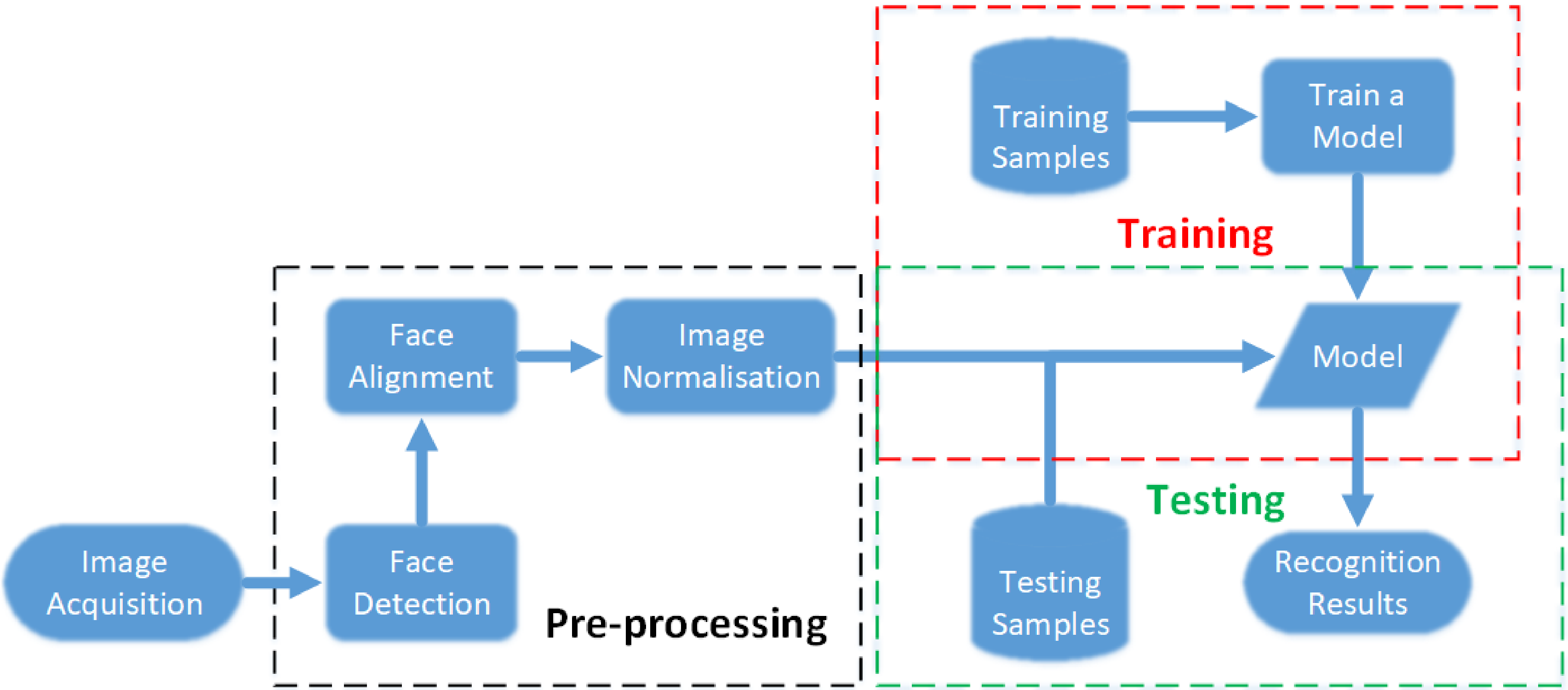


Fig. 1. The basic framework of a face recognition system.

1 Introduction

Face recognition methods

Traditional methods

Deep learning-based methods

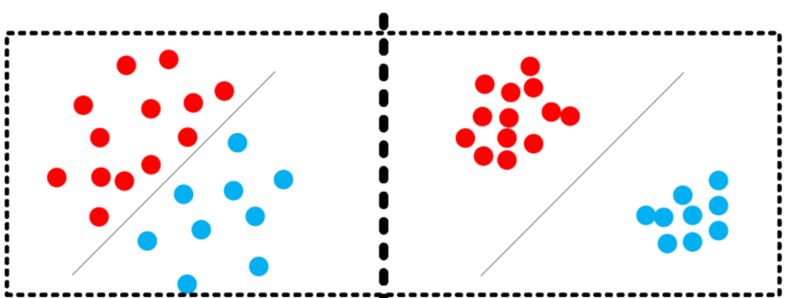
2. The proposed method

-- Minkowski Distance-based Centre loss (MC loss)

2.1 Softmax Loss and Centre Loss

Softmax Loss: $$L_S = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{K} e^{W_j^T f_i + b_j}} \quad (1)$$

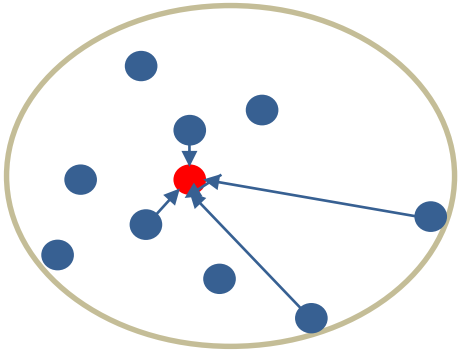
where N is the batch size, K is the class number of a batch, $f_i \in R_d$ denotes the feature of the i th sample belonging to the y_ith class, $W_j \in R_d$ denotes the j th column of the weight matrix W in the final fully connected layer and b_j is the bias term of the j th class.



Separable Features | Discriminative Features

2.1 Softmax Loss and Centre Loss

Centre Loss: $$L_C = \frac{1}{2}\sum_{i=1}^{N} ||f_i - c_{y_i}||_2^2 \qquad (2)$$



where C_{y_i} denotes the class centre of the y_i th class

Total Loss: $$\begin{aligned} L &= L_S + \lambda L_C \\ &= -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{K} e^{W_j^T f_i + b_j}} + \frac{\lambda}{2}\sum_{i=1}^{N} ||f_i - c_{y_i}||_2^2 \qquad (3)\end{aligned}$$

where λ is the hyper-parameter for balancing the two loss functions.

Ulster University

2.2 Minkowski Distance-based Centre loss (MC loss)

The Minkowski distance of order *n* between two points X and Y:

$$D(X,Y) = \left(\sum_{i=1}^{k} |x_i - y_i|^n \right)^{1/n} \quad (4)$$

where $X = (x_1, x_2, \ldots, x_k)$ and $Y = (y_1, y_2, \ldots, y_k) \in R_k$. Therefore the *n*th power of the Minkowski distance of order *n* is:

$$D(X,Y) = \sum_{i=1}^{k} |x_i - y_i|^n \quad (5)$$

Ulster University

2.2 Minkowski Distance-based Centre loss (MC loss)

MC Loss: $$L_M = \frac{1}{n}\sum_{i=1}^{N} ||f_i - c_{y_i}||_n^n \qquad (6)$$

where $n \in R{>}0$. Typically, n is set to be 2, 3, 4, …

Total Loss: $$\begin{aligned} L &= L_S + \lambda L_M \qquad (7) \\ &= -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{K} e^{W_j^T f_i + b_j}} + \frac{\lambda}{n}\sum_{i=1}^{N} ||f_i - c_{y_i}||_n^n \end{aligned}$$

where λ is the hyper-parameters for adjusting the impact of MC loss.

2.2 Minkowski Distance-based Centre loss (MC loss)

Indicator: $\left(\frac{|f_{ik}-c_k|}{|f_{jk}-c_k|}\right)^n$,

where c_k is the kth feature of a class centre, f_{ik} is the kth feature of sample i, f_{jk} is the kth feature of sample j. Sample i is on the edge of the class while sample j is near the class centre, and $\frac{|f_{ik}-c_k|}{|f_{jk}-c_k|} > 1$.

When n = 2, MC loss deteriorates to Centre loss.

When n > 2, $\left(\frac{|f_{ik}-c_k|}{|f_{jk}-c_k|}\right)^n$ is greater than it was in the Centre loss.

Ulster University

2.2 Minkowski Distance-based Centre loss (MC loss)

Algorithm 1 shows the basic learning steps in the CNNs with the proposed **Softmax + MC loss**.

Algorithm 1 Learning algorithm in the CNNs with the proposed Softmax + MC loss.

Input: Training samples $\{f_i\}$, initialised parameters θ_C in convolution layers, parameters W in the final fully connected layer, and initialised class centres $\{c_j | j = 1, 2, ..., n\}$. Learning rate μ^t, hyperparameter λ, learning rate of the class centres α and the number of iteration $t \leftarrow 1$.

Output: The parameters θ_C.

1: **while** not converge **do**
2: Calculate the total loss by $L^t = L^t_S + L^t_M$.
3: Calculate the backpropagation error $\frac{\partial L^t}{\partial f^t_i}$ for each sample i by $\frac{\partial L^t}{\partial f^t_i} = \frac{\partial L^t_S}{\partial f^t_i} + \lambda \frac{\partial L^t_M}{\partial f^t_i}$.
4: Update W by $W^{t+1} = W^t - \mu^t \frac{\partial L^t_S}{\partial W^t}$.
5: Update c_j for each centre j by $c^{t+1}_j = c^t_j - \alpha \Delta c^t_j$.
6: Update θ_C by $\theta^{t+1}_C = \theta^t_C - \mu^t \sum^N_i \frac{\partial L^t}{\partial f^t_i} \frac{\partial f^t_i}{\partial \theta^t_C}$.
7: $t \leftarrow t + 1$.
8: **end while**

Ulster University

2.2 Minkowski Distance-based Centre loss (MC loss)

Table 1: Verification performance of state-of-the-art methods on LFW and YTF datasets.

Methods	Source	Images	LFW(%)	YTF(%)
Deep Face [Taigman et al., 2014]	CVPR	4M	97.35	91.4
DeepID2+ [Sun et al., 2015]	CVPR	N/A	99.47	93.2
Fusion [Taigman et al., 2015]	CVPR	500M	98.37	N/A
FaceNet [Schroff et al., 2015]	CVPR	200M	99.63	95.1
Baidu [Liu et al., 2015]	arXiv	1.3M	99.13	N/A
Centre Loss [Wen et al., 2016]	ECCV	0.7M	99.28	94.9
Multibatch [Tadmor et al., 2016]	NIPS	2.6M	98.20	N/A
Aug [Masi et al., 2016]	ECCV	0.5M	98.06	N/A
SphereFace [Liu et al., 2017]	CVPR	0.5M	99.42	95.0
Range Loss [Zhang et al., 2017]	ICCV	1.5M	99.52	93.7
Softmax loss	N/A	3M	99.43	94.9
Softmax loss + Centre loss	N/A	3M	99.50	95.1
Softmax loss + MC loss (Proposed)	N/A	3M	99.57	95.3

Ulster University

3. Conclusion and future works

Conclusion

- We proposed a Minkowski distance-based generalisation method for improving Centre loss for deep face recognition.
- Experimental results on the LFW image dataset and the YTF video dataset show that the proposed method is highly competitive even compared with the state-of-the-art methods.

Future works

- Design better loss functions to further enhance the discriminative ability of features.
- Explore the relationship between the network structures and the loss functions, and find a stronger network structure.

Ulster University

Question and Answer