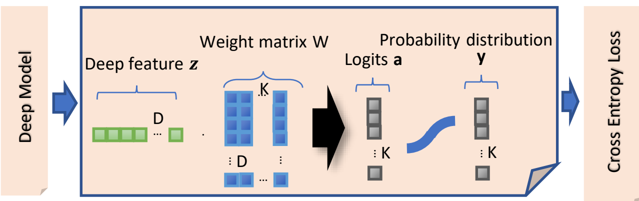
Cam-softmax for discriminative deep feature learning

Tamás Süveges, Stephen McKenna
University of Dundee, Nethergate, Dundee DD1 4HN, UK
Email: t.suveges@dundee.ac.uk, s.j.z.mckenna@dundee.ac.uk

1 Abstract

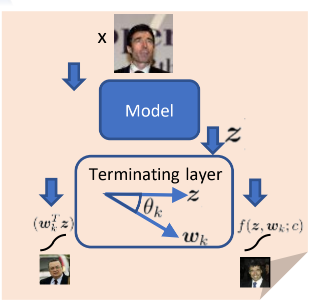
Deep convolutional neural networks are widely used to learn feature spaces for image classification tasks. We propose cam-softmax, a generalisation of the final layer activations and softmax function, that encourages deep feature spaces to exhibit high intra-class compactness and high inter-class separability. We provide an algorithm to automatically adapt the method's main hyperparameter so that it gradually diverges from the standard activations and softmax method during training. We report experiments using CASIA-Webface, LFW, and YTF face datasets demonstrating that cam-softmax leads to representations well suited to open-set face recognition and face pair matching. Furthermore, we provide empirical evidence that cam-softmax provides some robustness to class labelling errors in training data, making it of potential use for deep learning from large datasets with poorly verified labels.

2 Background



A deep neural network trained to map input images to class labels via a hidden representation, z, which is then mapped to a probability distribution. A softmax function is commonly used for normalization.

3 Cam-softmax



$$y_k = \frac{e^{a_k}}{\sum_j e^{a_j}}$$

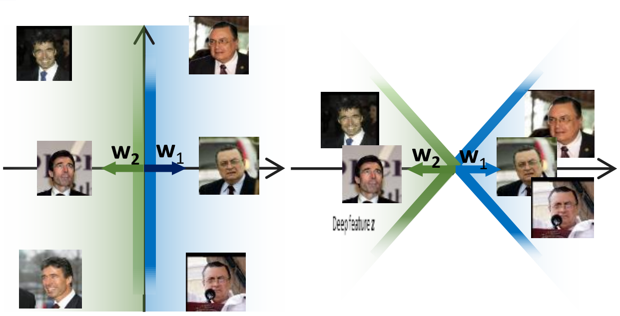
$$a_k = \mathbf{1}(k \neq t)(\boldsymbol{w}_k^T \boldsymbol{z}) + \mathbf{1}(k = t)(f(\boldsymbol{z}, \boldsymbol{w}_k; c))$$

$$f(\boldsymbol{z}, \boldsymbol{w}_k; c) = \frac{(\cos\theta_k + 1)^{g(c)}}{2^{(g(c)-1)}} - 1$$

c is a free parameter and cam-softmax defaults to softmax when c=pi/2.

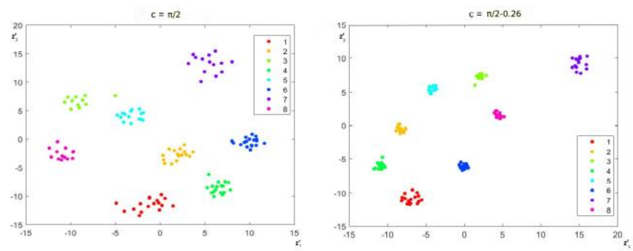
The function g ensures: $f(\boldsymbol{z}, \boldsymbol{w}_k; c) = 0$ when $\theta_k = c$

4 Softmax vs cam-softmax



Unnecessarily spread   Well concentrated

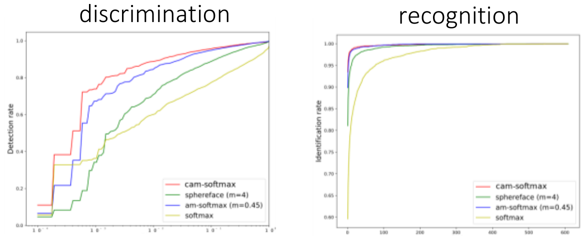
5 Effect on feature space



softmax   cam-softmax

6 Open-set face recognition

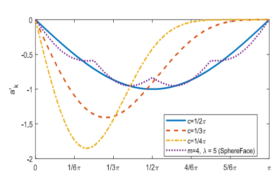
Known vs unknown discrimination   Known identity recognition

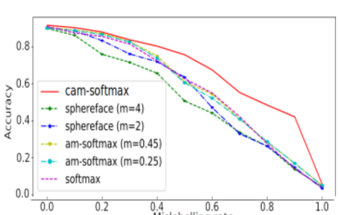


FAR   Rank

7 Labelling noise

The shape of the derivative attenuates gradient for potentially mislabelled examples.





Fashion-MNIST examples

This research received funding from the UK Engineering and Physical Sciences Research Council (EPSRC grant EP/N014278/1)

University of Dundee   DUNDEE