A Flatter Loss for Bias Mitigation in Cross-dataset Facial Age Estimation

Ali Akbari, Muhammad Awais, Zhenhua Feng, Ammarah Farooq and Josef Kittler
Centre for Vision, Speech and Signal Processing(CVSSP), University of Surrey, Guildford GU2 7XH, UK
Email: ali.akbari@surrey.ac.uk

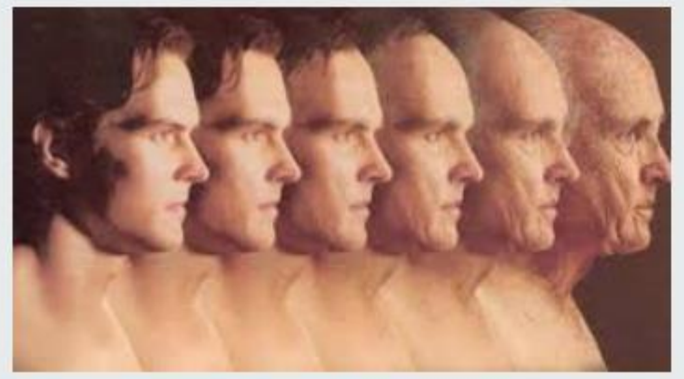
Abstract

The most existing studies in the facial age estimation assume training and test images are captured under similar shooting conditions. However, this is rarely valid in real-world applications, where training and test sets usually have different characteristics. In this paper, we advocate a cross-dataset protocol for age estimation benchmarking. In order to improve the cross-dataset age estimation performance, we mitigate the inherent bias caused by the learning algorithm itself. To this end, we propose a novel loss function that is more effective for neural network training. The relative smoothness of the proposed loss function is its advantage with regards to the optimisation process performed by stochastic gradient descent (SGD). Compared with existing loss functions, the lower gradient of the proposed loss function leads to the convergence of SGD to a better optimum point, and consequently a better generalisation. The cross-dataset experimental results demonstrate the superiority of the proposed method over the state-of-the-art algorithms in terms of accuracy and generalisation capability.

Age Estimation Problem

Age Estimation Problem

Age estimation is the prediction of a persons age based on biometric features extracted from the face.



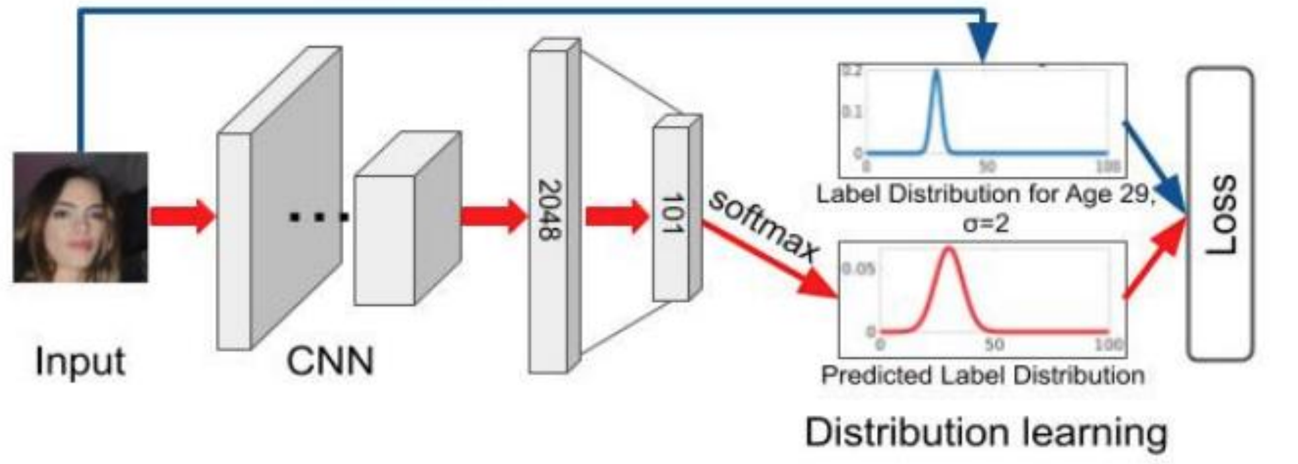
Semantic Similarity

There is semantic similarity between features of adjacent ages. This semantic information should be reflected into the training algorithm.

Age estimation methods:

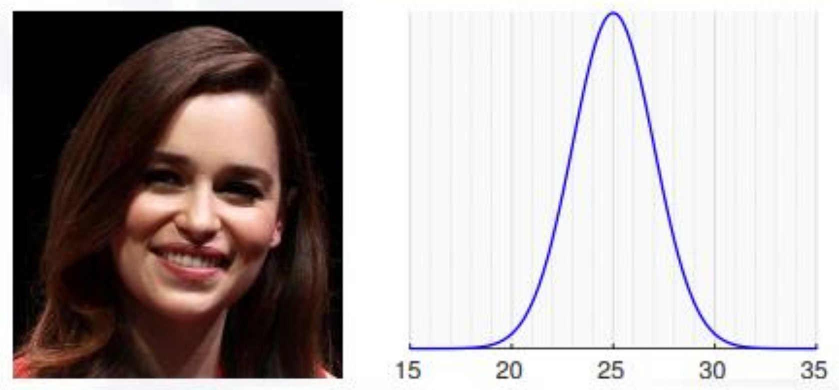
- **Regression:** Scalar labels
- **Classification:** 0/1 labels
- **Ranking:** Ensemble of binary classifiers
- **Label Distribution Learning:** Label distribution

The Proposed Method



Age estimation problem as distribution learning problem

- Due to the similarity between neighbouring ages, a scalar age label is encoded as a label distribution (a set of description degrees) .
- Sum of all description degrees equal to 1.
- The maximum degree is assigned to the corresponding age.



Label distribution for a facial image at the age of 25

Existing Loss Function

Kullback-Leibler divergence (KL)

$$L(\mathbf{p},\mathbf{q}) = \sum_{k=1}^{L} q_k \log\left(\frac{q_k}{p_k}\right)$$

Distribution Cognisant Loss (DC-v1)

$$L(\mathbf{p},\mathbf{q}) = \log\left(1-\alpha\left(1-\sum\nolimits_{k=1}^{L}\sqrt{p_k q_k}\right)\right)/\log(1-\alpha) \quad 0<\alpha<1$$

Distribution Cognisant Loss (DC-v2)

$$L(\mathbf{p},\mathbf{q}) = \sum_{k=1}^{L} |q_k^\alpha - p_k^\alpha|^{\frac{1}{\alpha}} = \sum_{k=1}^{L} q^k \left|1-\left(\frac{p_k}{q_k}\right)^\alpha\right|^{\frac{1}{\alpha}} \quad 0\le\alpha\le 1$$

Main Property

The smoother loss surface provides a better generalisation for the output model trained by that function.

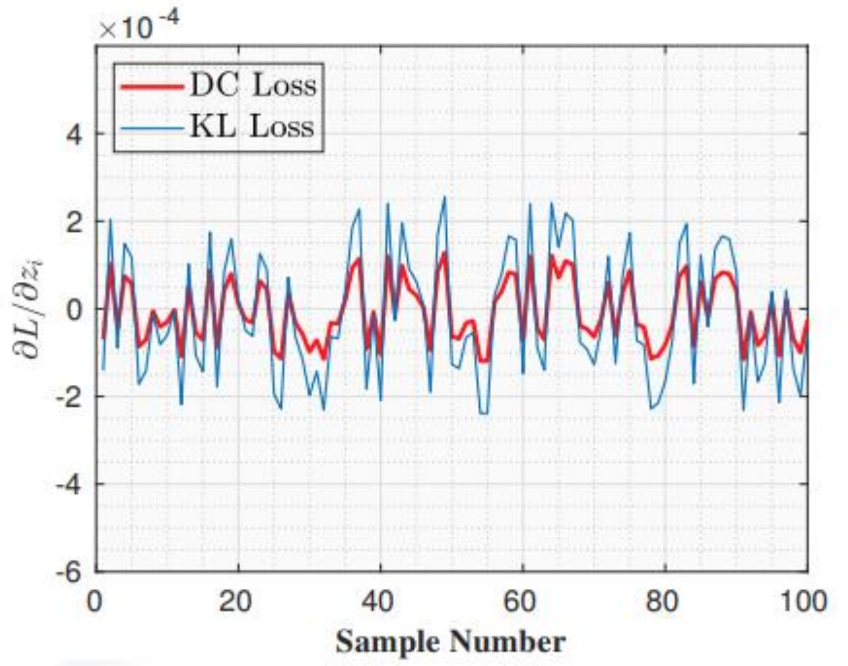
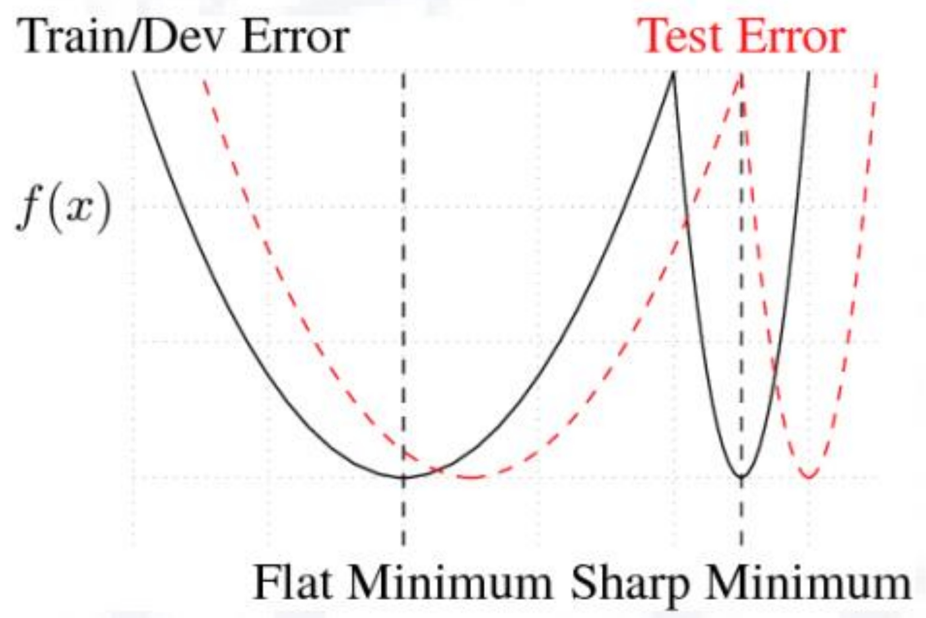


Figure: Behaviour of gradient of the KL loss and the proposed DC loss.

A Conceptual Sketch of Flat and Sharp Minima.



Main Property

It is well known that the flat minimum could help to improve the generalisation capability.

Numerical Results

Train Database

- **Our BAG:** 200,123 images, crawled from Internet

Test Databases

- **FGNET:** Images with different lighting condition
- **MORPH:** Four ethnicities
- **FACES:** Six expressions
- **SC-ROT:** Images with different pose
- **SC-SUR:** Images with different quality

Architecture

- VGG16

Evaluation

- Cross database evaluation

Cross-database Evaluation (MAE & CS) on the Target Databases

	FG-NET		MORPH		FACES		SC-FACE		Average	
Method	MAE	CS(%)	MAE	CS(%)	MAE	CS(%)	MAE	CS(%)	MAE	CS(%)
Human	4.70	69.5	6.30	51.0	NA	NA	NA	NA	5.50	60.25
Microsoft	6.20	53.80	6.59	46.00	-	-	-	-	6.39	49.90
DEX	3.57	78.94	6.54	53.38	6.59	50.83	6.19	65.05	5.86	59.50
AGEn	3.53	79.78	6.40	53.97	6.34	52.40	6.12	65.21	5.72	60.60
DLDL	3.24	81.54	6.01	57.36	6.11	55.60	6.52	60.64	5.55	61.98
CE-MV	3.34	80.44	6.22	55.60	6.25	54.63	6.23	64.38	5.62	61.84
DLDL-v2	3.35	81.44	5.80	57.30	5.92	56.68	6.52	61.61	5.48	62.77
Proposed	**3.26**	**81.57**	**5.69**	**58.83**	**5.92**	**57.45**	**5.41**	**67.90**	**5.07**	**66.43**

Conclusion

In this paper, we addressed the cross-dataset age estimation problem which is more realistic, as well as demanding than the conventional intra-set performance testing. The age estimation problem was modelled as distribution learning problem. A novel loss function was then proposed to improve the generalisation performance of the system by mitigating the inherent bias of the trained model. The ability of our proposed loss function to mitigate bias is directly related to its relative flatness which improves the accuracy in unseen (cross-dataset) scenarios. The superiority of our proposed approach is confirmed by the experimental results.

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021