Ying Cai，Meng-long Yang，Jun Li, 2015. Multiclass classification based on a deep convolutional network for head pose estimation. *Frontiers of Information Technology & Electronic Engineering*, **16**(11):930-939. [doi:10.1631/FITEE.1500125]

Multiclass classification based on a deep convolutional network for head pose estimation

Key words: Head pose estimation, Deep convolutional neural network, Multiclass classification

Contact: Ying Cai
E-mail: caiying34@qq.com
ORCID: http://orcid.org/0000-0002-5096-6175

Introduction

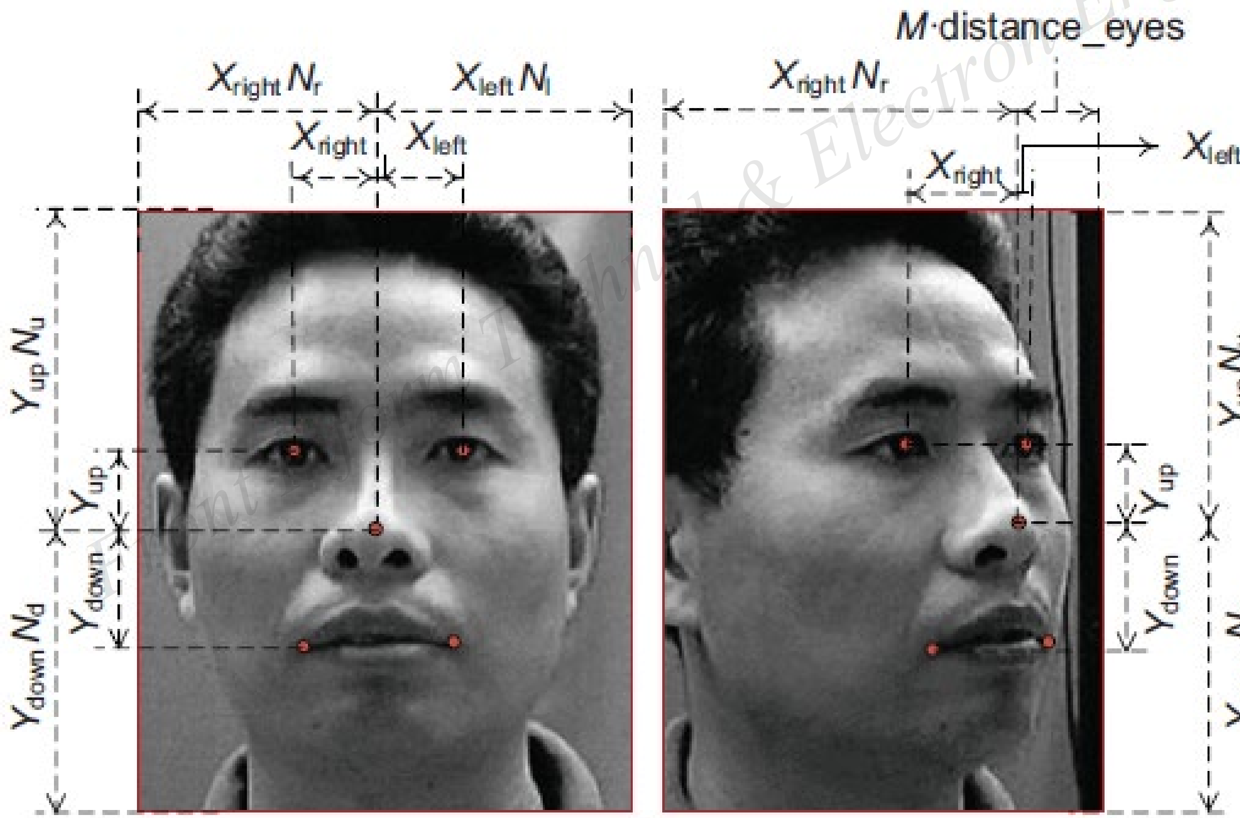
- The problem of head pose estimation has enjoyed substantial attention in the computer vision community.

- Research on neural networks showed that the standard fully connected multi-layer networks can be used as outstanding classifiers if there are good features.

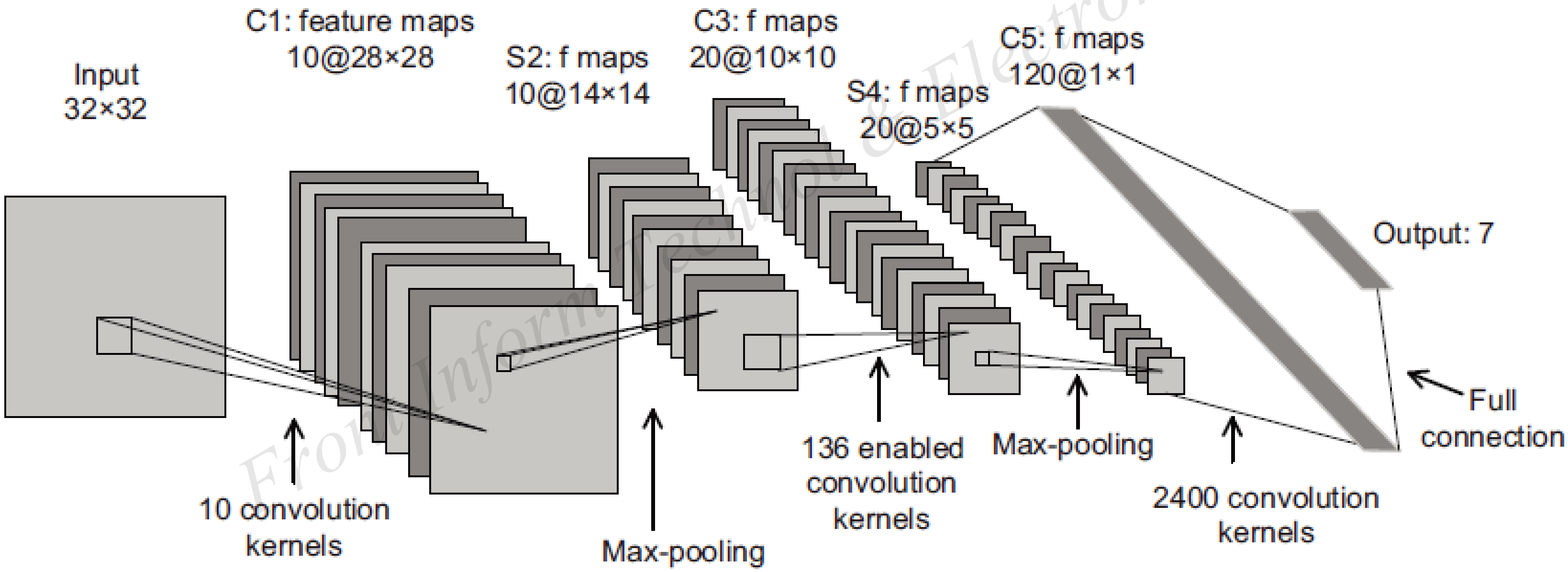
- We find that the deep convolutional neural network (DCNN) performs well on many visual tasks, because spatial topology and shift-invariant local features are well captured (LeCun *et al.*, 1998). We consider that appropriate DCNN architecture and an effective image preprocess will produce good performance on head pose estimation.

Our method of crop face

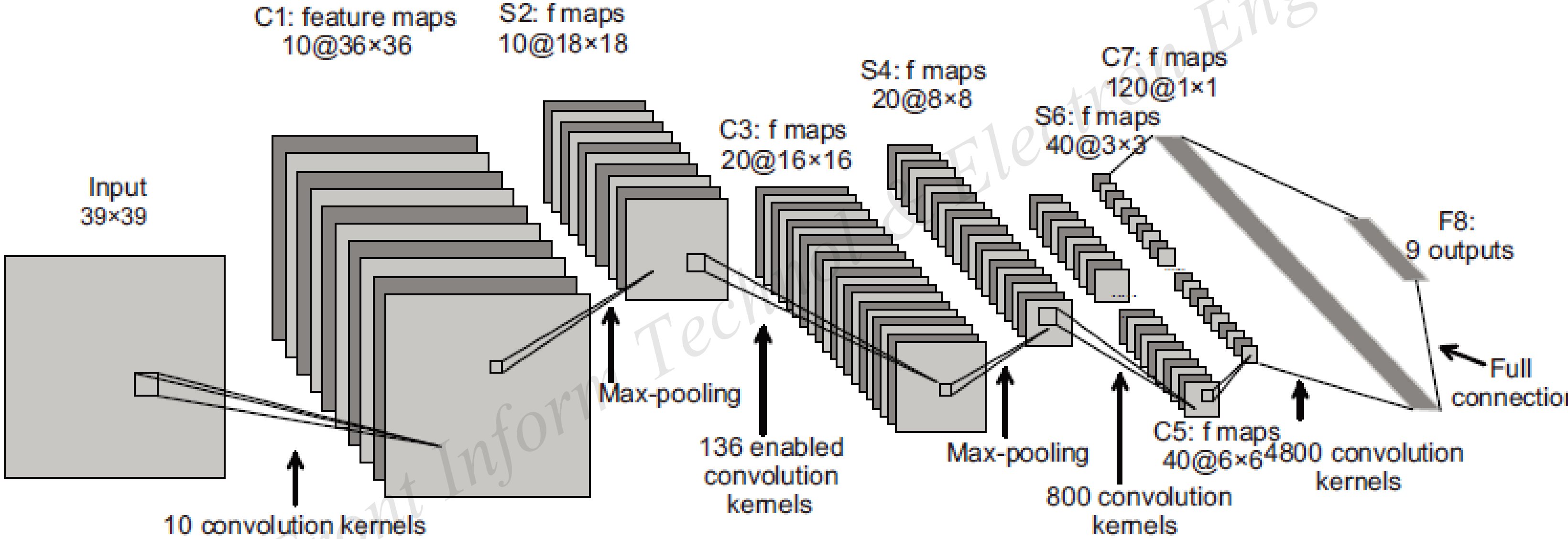
(the red frame is the crop face box)



Our first network of DCNN



Our second network of DCNN



Estimation results (1)

Table 1 The average results for different kinds of training image sets

Training set	CA (%)	SMSE	NE	NEA	NENA
Original images	96.4	48.78	25	18	7
CFFI	93.2	39.03	47	47	0
CWFI	96.4	26.68	25	23	2
Scaled images	97.4	15.05	18	18	0
Shifted images	98.0	11.32	14	14	0
SSI	98.4	9.82	11	11	0

CA: classification accuracy; SMSE: summation of mean squared error; NE: number of all error classifications; NEA: number of error classifications which occur between adjacent categories; NENA: number of error classifications which occur between nonadjacent categories; CFFI: cropped facial feature images; CWFI: cropped whole face images; SSI: scaled and shifted images

Estimation results (2)

Table 2 The results using six and eight-layer convolutional networks on nine poses

Network architecture	NTS	CA (%)	SMSE	NE	NEA	NENA
6-layer network	196 950	97.54	18.4	22	21	1
8-layer network	196 950	98.29	17.5	15	15	0
8-layer network	291 486	98.51	16.3	13	13	0

NTS: number of training samples

Estimation results (3)

Table 3 The results on the CMU and CUBIC FacePix databases using seven poses

Database	Training set	CA (%)	SMSE	NE	NEA	NENA
FacePix	Original images	81	4.2	4	3	1
	CWFI	90.5	5.9	2	2	0
	SSI	100	1.5	0	0	0
	WFI	100	2.7	0	0	0
CMU	Original images	100	2.22	0	0	0
	CWFI	93.9	6.4	3	2	1
	SSI	100	3.05	0	0	0

Estimation results (4)

Table 4 The results of our method compared with three state-of-the-art methods

Method	Classification accuracy (%)	Number of test samples	Number of training samples
B. Ma. LGBP	97.14	466	934
C. Huang. VRF+LDA	97.23	466	934
Our experiment-1	98.47	466	136 900
B. Ma. LBIF+SVM	94.57	1400	2800
Our experiment-2	97.17	1400	410 700

Conclusions

- In this paper, we propose an effective method for head pose classification based on DCNN. We design two appropriate DCNN architectures to compare their classfication ability in all kinds of situations. Before training, an effective way has been proposed to preprocess images. To obtain a powerful classifier, we adopt the shift and scale strategies to complete the preparation of training images.

- Our classifier performs well on three different databases. As the results show, our method significantly improves the classification accuracy compared with state-of-the-art methods.