An Hybrid Attention-Based System for the Prediction of Facial Attributes

Souad Khellat-Kihel¹, Zhenan Sun², and Massimo Tistarelli³(✉)

¹ Computer Vision Laboratory, University of Sassari,
Viale Italia 39, 07100 Sassari, Italy

² Center for Research on Intelligent Perception and Computing, National Laboratory
of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences,
Room 1605, Intelligence Bulding, 95 Zhongguancun East Road,
Beijing 100190, P. R. China

³ Computer Vision Laboratory, Department of Biomedical Sciences and Information
Technology, University of Sassari, Viale S. Pietro 43/b, 07100 Sassari, Italy
`tista@uniss.it`

Abstract. Recent research on face analysis has demonstrated the richness of information embedded in feature vectors extracted from a deep convolutional neural network. Even though deep learning achieved a very high performance on several challenging visual tasks, such as determining the identity, age, gender and race, it still lacks a well grounded theory which allows to properly understand the processes taking place inside the network layers. Therefore, most of the underlying processes are unknown and not easy to control. On the other hand, the human visual system follows a well understood process in analyzing a scene or an object, such as a face. The direction of the eye gaze is repeatedly directed, through purposely planned saccadic movements, towards salient regions to capture several details. In this paper we propose to capitalize on the knowledge of the saccadic human visual processes to design a system to predict facial attributes embedding a biologically-inspired network architecture, the HMAX. The architecture is tailored to predict attributes with different textural information and conveying different semantic meaning, such as attributes related and unrelated to the subject's identity. Salient points on the face are extracted from the outputs of the S2 layer of the HMAX architecture and fed to a local texture characterization module based on LBP (Local Binary Pattern). The resulting feature vector is used to perform a binary classification on a set of pre-defined visual attributes. The devised system allows to distill a very informative, yet robust, representation of the imaged faces, allowing to obtain high performance but with a much simpler architecture as compared to a deep convolutional neural network. Several experiments performed on publicly available, challenging, large datasets demonstrate the validity of the proposed approach.

1 Introduction

In the last decade, soft biometric traits have been widely used for person identification because of the robustness to noise, non-intrusiveness, and privacy preser-

vation. In the last years Deep learning approaches have been proposed also to extract soft-biometric attributes from face images. However, the high performance achieved are always paired with the requirement for high computational power and a large dataset for training. Liu et al. [1] proposed a method based on two CNNs which are trained for face localization (LNet) and attributes prediction (ANet). The top network layer, FC is exploited to learn identity-related features, such as the gender and race. Layers C3 and C4 are exploited to extract Identity-unrelated attributes, such as the facial expression, wearing hat and sunglasses. Samangouei et al. [2–4] proposed a CNN architecture suitable for mobile devices, which is based on the analysis of face parts. Recently, Dhar et al. [5] considered the usefulness of the outputs of the internal layers of two deep convolutional networks, Resnet 101 and Inception Resnet v2, for the prediction of facial attributes. Izadi [6] proposed the fusion of the extracted facial attributes with the face image to perform face recognition on a shared CNN architecture. Recently different works [7, 8] proved that the final representation computed by a deep convolutional neural network embeds information not only about identity but also on the head pose and illumination. In this paper we propose to extract information from internal layer corresponding to HMAX network with the purpose of predicting different facial attributes. The HMAX model, which has been developed before deep learning took over in many computer vision problems, demonstrated the feasibility of a biologically-inspired neural architecture for face recognition. The model was tested on several publicly available databases such as LFW, PubFig and SURF-W [9] providing results at the state of the art. In [10], a new C3EF layer, inspired by the ventral and dorsal streams of the visual cortex, has been added to perform view-independent face recognition. Hu [11] proposed a version of the HMAX model, named 'sparse HMAX', addressing the local-to-global structure of the hierarchy, where the S2 bases are learned by sparse coding. In this paper we propose a novel hybrid system based on the HMAX network architecture. The outputs of the internal S2 layer are used as the seeds for extracting interest regions which are then used to generate the feature vector for the classification of the facial attributes. The following issues are addressed:

- How the salient feature points extracted from the HMAX architecture can improve the prediction of facial attributes.
- To which extent the devised system can be applied to predict different kinds of facial attributes.
- What is the robustness of an attentive visual system to variations in head pose, lighting and facial expression.

2 Prediction of Facial Attributes

Most of the time celebrities or familiar people are remembered because of their special hair style, accessories or even clothes. This daily life concept is exploited by soft biometrics, or general visual attributes. These attributes can add significant information to face images and are quite robust to image degradation and

changes in appearance. For this purpose, the internal S2 layer of the HMAX is used to detect the most salient points on the subject's face. The Linear Binary Pattern feature extractor is exploited to build a local description of the image texture around the selected points. The feature vectors computed for each salient point are concatenated to produce a global feature vector to characterize the face image. The obtained feature vector is fed to a SVM binary classifier to predict several visual attributes. In Fig. 1 the general architecture of the proposed framework is shown.

Fig. 1. Proposed hybrid system for the prediction of facial attributes.

2.1 The Hierarchical HMAX Network

HMAX is an hierarchical system that closely follows the organization of visual cortex and builds an increasingly complex and invariant feature representation by alternating between a template matching and max pooling [12]. As the network structure is fixed, a limited number of training examples is required for learning. The computational process is hierarchical and it is also invariant to position, scale and view-point. Along the hierarchy, the size of the receptive fields and the complexity of their optimal stimuli increases. The model consists of four computational layers, where *simple* ‘S’ units alternate with *complex* ‘C’ units.

The first layer S1 in the HMAX network consists of a bank of Gabor filters applied to the full resolution image. The response to a particular filter G , of layer S , at the pixel position (X, Y) is given by:

The first layer of the HMAX model ‘S1’ consists of a bank of Gabor filter, the following steps are implemented in (Fig. 2): The Image (at finest scale) is $[256 * 256 * 1]$. In each image intensity, four Gabor filters are applied over

Fig. 2. General architecture of the HMAX model.

each pixel position. The result of S1 layer (at finest scale) is $[246 \times 246 \times 4]$. The response of a patch of pixels X to a particular S1 filter G is given by:

$$R(X, Y) = \left| \frac{\sum X_i G_i}{\sqrt{\sum X_i^2}} \right| \quad (1)$$

The size of the Gabor filter is 11×11 and it is formulated as:

$$G(x, y) = \exp\left(\frac{-(x^2 \gamma^2 Y^2)}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda} X\right) \quad (2)$$

Where $X = x \cos \theta - y \sin \theta$ and $Y = x \sin \theta + y \cos \theta$. x and y vary between -5 and 5 , and θ varies between 0 and π . The parameters ρ (aspect ratio), σ (effective width), and λ (wavelength) are set to 0.3 , 4.5 and 5.6 , respectively. For the local invariance (C1) layer, a local maximum is computed for each orientation. They also perform a subsampling by a factor of 5 in both the X and Y directions [13]. In the intermediate feature layer (S2 level), the response for each C1 grid position is computed. Each feature is tuned to a preferred pattern as stimulus. Starting from an image of size 256×256 pixels, the final S2 layer is a vector of dimension $44 \times 44 \times 400$. The response is obtained using:

$$R(X, P) = \exp\left(\frac{||X - P||^2}{\sigma^2}\right) \quad (3)$$

The last layer of the architecture is the Global Invariance layer (C2). The maximum response to each intermediate feature over all (X, Y) positions and all scales are calculated. The result is a characteristic vector that will be used for classification. For the implementation of the HMAX model we use the tool proposed in [13]. As the final layer (C2) is the features vector that corresponds to maximums obtained from each S2 output, which are 400 characteristics. These

maximum correspond to a certain locations (best coordinates) that are the maximal responses for each patch and image. These coordinates are accumulated and projected into the original faces images and used as interest points.

2.2 Local Texture Description Based on LBP

LBP is a type of visual descriptor used for classification in computer vision. The idea of the texture extraction using LBP is to give to each pixel a code which depends on the gray scale of its neighbors. The gray scale of the central pixel (i_c) is compared to its neighbors in the following formula:

$$LBP(x_c, y_c) = \sum_{n=0}^P s(i_n - i_c) 2^n \quad X = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (4)$$

The LBP code of the current pixel is produced by concatenating the 8 values to construct a binary code. The center of each window corresponds to the interest points obtained from HMAX.

2.3 Binary Classification with Support Vector Machines

The SVMs are groups of learning techniques that are designated to solve problems of discrimination, i.e. to decide to which class a pattern belongs, or of regression, i.e. to predict the numerical value of a variable. The success of this method is defined by solid mathematical bases. The main objective of the SVM is like the perceptron principle but it consists not only into finding an hyper plan that separates perfectly the classes but also to find the optimal one that can separate perfectly the classes by maximizing the margin. They project the data in space of characteristics by using non-linear functions. In this space it builds the optimal hyper plan which separates the transformed data. The principal idea is to build a linear separation surface in the space of the characteristics which corresponds to a non-linear surface in the space entry (Figure 3 presents this non-linear transformation).

The Support Vector Machines approach passes through two steps: The train that consists of searching an optimal hyper plan of separation by maximizing the margin, with the resolution of a quadratic program and determination of

Fig. 3. The non linearity.

the Lagrange multipliers [14]. The test, after the determination of the Lagrange multipliers, it applies the decision function to the test examples to determinate the class [14]. The classification is conducted by using the SVM-KM toolbox [15] and by considering a Gamma value of $1e-7$ and penalty parameter of the error term $C = 100$ while we use Gaussian kernel.

3 Experimental Results

Several publicly available large datasets have been used for testing the proposed architecture. The CelebFaces Attributes Dataset (CelebA) [1] is a large-scale face attributes dataset with more than 200K celebrity images each is notated with 40 attributes. The images in CelebA dataset include large variations in appearance such as pose and background. It contains 10.177 identities having 20 images in average. The CelebA does not overlap with LFW dataset identities. We also use the LFW dataset [16] which is a large dataset, real-world face dataset consisting of 13.000 images of faces collected from internet. These images are taken in completely uncontrolled situations. This dataset contains variations in pose, lighting, expression, camera, imaging conditions. CelebA and LFW were intensively used in the recent proposed work in the litterature for the aim to predict facial attributes. The PubFig dataset [19] has been used to test the sensitivity to variations in pose, illumination and facial expression. The PubFig dataset is a large, real-world face dataset (including both celebrities and politicians) consisting of 58797 images of 200 subjects collected from the internet. The PubFig dataset is both larger and deeper, on average 300 images per individual than previous seen dataset. These images are taken in completely uncontrolled situations. This database contains variations in pose, lighting, expression, camera, imaging conditions. The PubFig dataset is similar to LFW dataset. However, the PubFig dataset has enough examples per each subject.

Experiment 1: The first experiment consists of the facial attributes prediction using the Labeled Faces in the Wild (LFW) and the CelebFaces dataset. Table 1 and Table 2 represent the facial attributes prediction results using CelebA and LFW accordingly. In this experiment we propose also to compare our proposed system (internal layer) for attributes prediction with the top layers corresponding to HMAX, VGG, Alexnet and ResNet-50. Alexnet is a convolutional neural network that is trained on more than a million images from the ImageNet database [19]. The network is 8 layers deep and it can classify the images into 1000 object categories [20,21]. The network has an image input size of 227-by-227. ResNet-50 is also a CNN framework trained with the images from ImageNet [19]. It is 50 layers deep with an input size of 224-by-224. But unlike Alexnet, Resnet-50 layers are organised in residual blocks. With each block encompassing of at-least 3 CNN layers (1×1 ; 3×3 ; and 1×1 convolutions) followed by a shortcut connection. VGG is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 16 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil,

and many animals. As a result, the network has learned rich feature representations for a wide range of images [20]. This experiment is for the aim to compare between the proposed system that is based on internal layer and the final layers corresponding to the HMAX, VGG, Alexnet and Resnet-50.

Experiment 2: In the second experiment we propose to compare the obtained results on LFW and CelebA faces with some recent obtained results in the state of the art such as FaceTracer [17], PANDA [18], LNet+ANet [3] and Shared CNN proposed in [8]. Table 3 and Table 4 represent the results on LFW and CelebA datasets respectively. The aim of this experiment is to compare our proposed hybrid architecture with recently proposed systems in the literature based on DNNs.

Experiment 3: In the third experiment, we test the PubFig dataset as it contains a large number of images per each subject. We study the efficiency of our proposed system on pose variation, faces under different illumination and expression. As the PubFig dataset contains images from the web a lot of them are not available, for this reason we construct a database composed of 54 subjects with 8942 images. Table 5 represents the obtained results which were compared also with Alexnet, VGG and Resnet-50 features.

From the obtained results (Table 1, Table 2), one can see clearly that the LFW dataset obtain better results with our hybrid system comparing to the CelebA and this is mainly due to the fact that we used CelebA faces in nonuniform order for the identities because the CelebA is composed of thousand of identities with twenty images per identity in average, these twenty images are dispatched randomly in the data; however, in the LFW database all the images corresponding to the same subject are in a successive order. Our proposed system demonstrate a comparable efficiency with the pretrained Deep Neural Networks (VGG, Alexnet and Resnet-50) in the same time surpass the obtained results with the features obtained from the final layer C2 corresponding to HMAX. Some facial attributes such as ‘Attractive’, ‘Gray Hair’, ‘Male’, ‘Mustache’ were well predicted using our model comparing to the pretrained DCNNs. In addition, our proposed biological system shows good performances comparing to the proposed models in the state of the art [3, 8, 17, 18] specially on identity facial attributes. These identity-based facial attributes such as ‘gender’, ‘hair color’, ‘nose’ and ‘lips shape’, ‘chubby’ and ‘Blad’ can add meaningful information for face identification. Additionally, the proposed hybrid attention-based system achieve a comparable results with the final layers of VGG, AlexNet, ResNet-50 and HMAX. Another advantage is that our model use different locations to predict the whole attributes faces. However, LNet-ANets [1] use different layers to predict different kind of attributes, also [3] consider the fusion of different and specific regions from the face to detect a specific facial attribute. In the latest case we may not find all facial regions available specially with different poses on mobile devices. Even though the PubFig is very challenging database with variation in pose, illumination and expression, our proposed approach can distinguish between frontal and dark lighting images.

Table 1. Facial attributes prediction accuracies on LFW.

Attribute	HMAX	VGG	Alexnet	Resnet	Ours
5 o Clock Shadow	79.5	93	95.5	97.5	95
Arched Eyebrows	91	98.5	98.5	99.5	96
Attractive	76	95	96	93.5	96.5
Bags Under Eyes	86	92	95.5	96	91
Bald	95	99	98	98.5	95
Bangs	83	96.5	97	97.5	90
Big Lips	66	92.5	92.5	89.5	88
Big Nose	94.5	99.5	99	99.5	88
Black hair	99.5	99.5	99.5	99.5	86
Blond hair	100	100	99.5	100	93
Blurry	72.5	97	96	97	92
Brown hair	99.5	100	99.5	99.5	89
Bushy Eyebrows	84.5	95	96	96	86
Chubby	96.5	97.5	99.5	99.5	90
Double Chin	88.5	94.5	98.5	99	96
Eyeglasses	74.5	93	95	95	92
Goatee	84.5	97.5	97.5	97.5	91
Gray Hair	45.5	32.5	27.5	24	91
Heavy Makeup	99	100	98	99.5	91
High Cheekbones	72.5	96	96	95	88
Male	55	56.5	57	64	91
Mouth Slightly Open	45	46.5	46	34.5	91
Mustache	70	85.5	87.5	88	91
Narrow Eyes	81.5	83	91	88.5	88
No Beard	85.5	95.5	98.5	98.5	91
Oval face	88	98.5	97.5	97.5	91
Pale Skin	99	99.5	99.5	100	89
Pointy Nose	71	93	92	95	90
Receding Hairline	66.5	89.5	91	77.5	90
Rosy Cheeks	99	99	99	99	91
Sideburns	80	93.5	96.5	96	93
Smiling	88	96.5	96.5	97	94
Straight Hair	87	95.5	98.5	96.5	93
Wavy hair	75.5	88.5	94.5	94.5	89
Wearing Earrings	86.5	96.5	96	97	92
Wearing Hat	68	91	93.5	89.5	88
Wearing Lipstick	91	99	98.5	97.5	90
Wearing Necklace	72	93	98	97	88
Wearing Necktie	99	98	99.5	99.5	93
Young	100	100	100	100	88

Table 2. Facial attributes prediction accuracies on CelebA.

Attribute	HMAX	VGG	Alexnet	Resnet	Ours
5 o Clock Shadow	85.4	88.2	88.2	88.2	91.5
Arched Eyebrows	96.8	77.2	77.6	76	60.5
Attractive	71.2	77.6	77.2	75.6	59.5
Bags Under Eyes	71.8	79.6	79.8	78.6	74
Bald	97.4	97.4	97.4	97.4	99.5
Bangs	81.8	87.2	87	88.8	82.5
Big Lips	64.6	75.2	75.4	74.6	60
Big Nose	70.4	78	77.4	77	74
Black hair	69.6	75.6	75.6	77.6	85
Blond hair	80.8	86.8	84.4	88	80.5
Blurry	93.8	94	93.6	93.6	99.5
Brown hair	71.4	78	78.4	77.4	81
Bushy Eyebrows	78	83.4	83.4	83	86
Chubby	92.6	95.8	95.4	95.4	96.5
Double Chin	94.4	95.6	95	95.4	97.5
Eyeglasses	91.4	97.8	97.6	96	97.5
Goatee	91.2	92.4	92	92.4	97.5
Gray Hair	93.6	94.2	93.8	94	98
Heavy Makeup	75.2	82.2	80.6	83.8	59.5
High Cheekbones	62.8	73	68.2	73.8	64.5
Male	73.6	79.6	81	84	83.5
Mouth Slightly Open	55.6	62	62.4	66.2	50
Mustache	94.8	95.6	95.6	95.2	96.5
Narrow Eyes	82.4	89.2	89	89.2	87
No Beard	81.4	84.4	83.4	84.6	91
Oval face	64.4	71.2	71.4	71	94.5
Pale Skin	95.6	96.8	96.2	96.6	98.5
Pointy Nose	66.6	72.6	72.2	71	67
Receding Hairline	91.2	94.8	95	94.6	95.5
Rosy Cheeks	92.4	94.2	94	94.2	85
Sideburns	92.6	93.6	93.2	93.6	96
Smiling	62.6	71	68.2	75.2	58
Straight Hair	73.6	81	81.2	81	84
Wavy hair	66.4	66.8	65.8	71.8	64.5
Wearing Earrings	72.4	82	80.6	82.2	57
Wearing Hat	95	97	96.4	97.4	99
Wearing Lipstick	75	82.8	80.8	85.8	74
Wearing Necklace	79	86.4	85.6	86.2	77
Wearing Necktie	86.4	91.4	91.2	91.4	97.5
Young	73.6	80	79.6	76.8	81

Table 3. Comparison of the attribute prediction frameworks on LFW

	Bald	Big Lips	Big Nose	Chubby	Male	Black Hair	Blond Hair
FaceTracer	77	68	73	67	84	76	88
PANDA	84	73	79	69	92	87	94
LNets-ANet	88	75	81	73	94	90	97
Shared CNN	90	65	75	62	94	80	88
Our model	91	88	86	96	91	93	92

Table 4. Comparison of the attribute prediction frameworks on CelebAFaces

	Bald	Big Lips	Big Nose	Chubby	Male	Black Hair	Blond Hair
FaceTracer	89	64	74	86	91	70	80
PANDA	96	67	75	86	97	85	93
LNets-ANet	98	68	78	91	98	88	95
Shared CNN	97	67	79	91	99	75	84
Our model	99.5	60	74	96.5	83.5	85	80.5

Table 5. Comparison of the attribute prediction on PubFig database

	Pose	Lighting	Expression
HMAX features	60	61	50.3
VGG features	70.40	70.7	48.1
Alexnet features	70.8	69.7	49.7
Resnet features	71	69.9	51.5
Ours	70.9	70.9	51.4

4 Conclusion

In this paper a visual attention-based system have been proposed to predict the facial attributes. This hybrid system allows a biological hierarchical network ‘HMAX’ to look into a particular salient regions of the input faces in the same time reduce the complexity by discarding the irrelevant information. These regions were introduced to LBP with the aim of extracting texture feature around these interest points extracted with HMAX. This proposed framework shows a promising results comparing to Deep Neural Networks architectures. The success of the hybrid architecture is due not only to the biological vision perception but also to the possibility and flexibility of this approach to learn and treat small amount of data and predict different facial attributes. By surpassing these issues we can solve the main problem of DCNNs that require large amount of data for training. The proposed approach can add a good impact on face recognition as it can predict the most challenging real-world scenario (different background,

pose variation, illumination, expression) that can degrade significantly the face recognition performances.

References

1. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 3730–3738 (2015). <https://doi.org/10.1109/ICCV.2015.425>
2. Samangouei, P., Chellappa, R.: Convolutional neural networks for attribute-based active authentication on mobile devices. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, pp. 1–8 (2016)
3. Samangouei, P., Patel, V.M., Chellappa, R.: Attribute-based continuous user authentication on mobile devices. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, pp. 1–8 (2015)
4. Samangouei, P., Patel, V.M., Chellappa, R.: Facial attributes for active authentication on mobile devices. *Image Vis. Comput.* **58**, 181–192 (2017). ISSN 0262-8856
5. Dhar, P., Bansal, A., Castillo, C., Gleason, J., Phillips, P.J., Chellappa, R.: How are attributes expressed in face DCNNs?, arXiv preprint [arXiv:1910.05657](https://arxiv.org/abs/1910.05657) (2019)
6. Izadi, M.R.: Feature Level Fusion from Facial Attributes for Face Recognition, arXiv preprint [arXiv: 1909.13126](https://arxiv.org/abs/1909.13126), September 2019
7. Khellat-Kihel, S., Lagorio, A., Tistarelli, M.: Foveated vision for deepface recognition. In: Nyström, I., Hernández Heredia, Y., Milián Núñez, V. (eds.) CIARP 2019. LNCS, vol. 11896, pp. 31–41. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33904-3_3
8. O’Toole, A.J., Castillo, C.D., Padre, C.J., Hill, M.Q., Chellappa, R.: Face space representations in deep convolutional neural networks. *Trends Cogn. Sci.* **22**(9), 1364–6613 (2018)
9. Liao, Q., Leibo, J.Z., Poggio, T.: Learning invariant representations and applications to face verification. In: NIPS 2013 Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, vol. 2, pp. 3057–3065 (2013)
10. Esmaili, S., Maghooli, K., Motie Nasrabadi, A.: C3 effective features inspired from ventral and dorsal stream of visual cortex for view independent face recognition. *Adv. Comput. Sci.* 1–9 (2016). ISSN 2322-5157
11. Hu, X., Zhang, J., Li, J., Zhang, B.: Sparsity-regularized HMAX for visual recognition. *PLoS ONE* **9**(1), e81813 (2014). <https://doi.org/10.1371/journal.pone.0081813>
12. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999). <https://doi.org/10.1038/14819>
13. HMAX toolbox. <http://maxlab.neuro.georgetown.edu/hmax.html>
14. Ayat, N.: Sélection automatique de modèle dans les machines à vecteurs de support: application à la reconnaissance d’images de chiffres manuscrits. Thèse de doctorat électronique, Montréal, École de technologie supérieure (2004)
15. Canu, S., Grandvalet, Y., Guigue, V., Rakotomamonjy, A.: SVM and Kernel Methods Matlab Toolbox. Perception Systemes et Information, INSA de Rouen, Rouen, France (2005)

16. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision, Kyoto, pp. 365–372 (2009). <https://doi.org/10.1109/ICCV.2009.5459250>
17. Kumar, N., Belhumeur, P., Nayar, S.: FaceTracer: a search engine for large collections of images with faces. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5305, pp. 340–353. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88693-8_25
18. Zhang, N., Paluri, M., Ranzato, M., Darrell, T., Bourdev, L.: PANDA: pose aligned networks for deep attribute modeling. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 1637–1644 (2014). <https://doi.org/10.1109/CVPR.2014.212>
19. <http://www.cs.columbia.edu/CAVE/databases/pubfig/>
20. ImageNet. <http://www.image-net.org>
21. Krizhevsky, A., Ilya, S., Geoffrey, E.H.: ImageNet classification with deep convolutional neural networks. In: NIPS 2012 Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097–1105 (2012)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.