REGION-AWARE SCATTERING CONVOLUTION NETWORKS FOR FACIAL BEAUTY PREDICTION

Lingyu Liang[1,2], *Duorui Xie*[1], *Lianwen Jin*[1,*], *Jie Xu*[1], *Mengru Li*[1], *Luojun Lin*[1]

[1] South China University of Technology, Guangzhou, China
[2] The Chinese University of Hong Kong, Shatin, Hong Kong

ABSTRACT

This paper proposes a scattering convolutional network with region-aware facial attributes to obtain a mid-level representation for facial beauty prediction (FBP). Different from the previous works that only focus on the discriminative representation for prediction, this paper also considers the invariant properties of the facial representation that reduces the variances caused by the image transformations, such as rotations and translation. The proposed region-aware scattering convolution network (RegionScatNet) is based on a deep convolution network of scattering transforms (ScatNet) integrated with facial texture and shape features. It consists of three components, including: 1) Region Extraction to obtain the significant facial perception region with implicit shape features using region-aware mask; 2) Attributes Decomposition to separate the extracted region into detail and structure facial layers by a guided filter; 3) Scattering Convolution that computes the roto-translation invariant representation of facial detail and structure for FBP by cascading three-layer wavelets filters and non-linear modulus pooling. The comparisons with related deep learning-based methods illustrate the effectiveness of RegionScatNet for FBP. The evaluations with various prediction model (like SVR and Gaussian process) and with different facial variances (like rotaion) indicate the robustness of the RegionScatNet-based features.

***Index Terms*—** Facial beauty prediction, facial beauty analysis, convolutional networks, invariants, wavelets

1. INTRODUCTION

Facial beauty prediction (FBP) is a significant problem in facial beauty analysis [3, 12, 19]. It is involved with many interesting applications, such as face beautification [7, 9–11], facial makeup synthesis/recommendation [6, 25] and content-based image retrieval [8, 26].

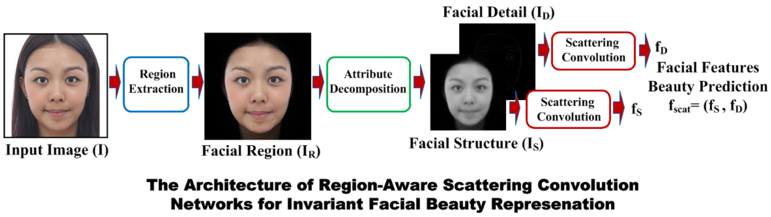


Fig. 1: The region-aware scattering convolution networks (RegionScatNet) consists of three components: region extraction, attributes decomposition and scattering convolution.

FBP has led to ever-growing interests in both machine and learning computer vision communities [12, 14–22], but there are still many open problems to achieve FBP that is consistent to human perception. The first problem is the lack of benchmark database. Since many existing face databases were originally designed for face recognition, evaluation based on these database may introduce large variances. The second problems of FBP is the model to map the facial biometrics into beauty scores. This paper specifically focuses on the construction of a invariant and discriminant facial representation for FBP. To evaluate the methods, experiments were performed on the recent SCUT-FBP benchmark database [24].

FBP can be formulated as a supervised learning task of classification [4, 15, 20, 23], regression [5, 14, 17, 22] or ranking [2, 18]. No matter which type of prediction model is used, facial representation is the core problem for FBP [3, 12, 19]. The facial representation can be obtained by hand-crafted geometric or texture features, such as the geometric ratios and landmark distances [3, 4, 14–17], and the Gabor-/SIFT-like features [1, 2, 18, 29].

Due to the success of deep learning methods for visual recognition [27,28], recent studies used deep neuron networks (DNN) to learn the facial representation for FBP. Xie et al. constructed a six-layer convolutional neuron networks (CNN) for FBP. Gan et al. used a deep self-taught learning method with two-layer convolutional deep belief networks (CBDN) to obtain the facial feature [22]. Wang et al. built pairs of autoencoders networks to obtain visual descriptors for beauty/notbeauty classification [23]. Despite the success of DNN-based

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61472144 and Grant 61502176; in part by GDSTP under Grant 2015B010101004, Grant 2015B010130003, Grant 2015B010131004; in part by the National Key Research & Development Plan of China under Grant 2016YFB1001405; and in part by Fundamental Research Funds for the Central Universities (No.2017BQ058). *Corresponding author: Lianwen Jin (lianwen.jin@gmail.com)

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

FBP, the properties of these networks are not well understood due to the cascaded nonlinear transformation of each layer.

Recent studies indicate that the multi-layer hierarchical architecture of DNN facilitates to build invariant image representation to discount image transformation [30, 31, 33]. It motivates us to explore the invariant properties of the representation for facial beauty prediction, which is rarely considered in the previous studies in FBP. Inspired by the recent scattering convolution networks (ScatNet) to build a invariant representation for texture categories [30,31] and the attention-based CNN for visual question answering [34], we propose a region-aware scattering convolution networks (RegionScatNet) to build a representation that is both discriminative for facial beauty prediction and invariant to the variances caused by the image transformations, such as rotation and translation.

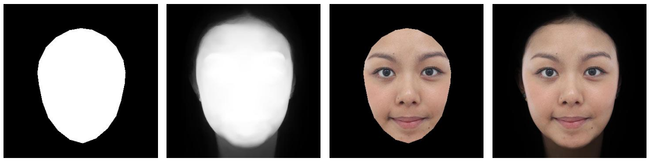
The RegionScatNet is based on a ScatNet integrated with facial texture and shape features. It consists of three components, including region extraction, attributes decomposition and scattering convolution. In region extraction, significant facial region is obtained using a region-aware mask based on the edge-aware label propagation method [10]. Benefited from the region-aware mask, the boundary of the facial region fits the facial shape closely, which implicitly encodes the shape features in the extracted region. In accord with the psychologically studies of the facial beauty perception, attribution decomposition separates the facial region into two layers that contain the facial detail and structure features by a guided filter. Then, a three-layer ScatNet is constructed to computes the roto-translation invariant representation of facial detail and structure by cascading wavelets filters and non-linear modulus pooling. Finally, the extracted RegionScatNet-based features are used to train a prediction model, like SVR, Gaussain regression, for FBP.

In summary, the contributions of this paper include: (i) a new convolution networks architecture, called Region-Aware Scattering Networks (RegionScatNet), that intergraded with region-aware facial attributes decomposition and the scattering transform to obtain facial representation for FBP; (ii) a RegionScatNet-based facial beauty representation that is discriminative for beauty assessment and invariant to the uninformative variances (like rotation).

2. REGION-AWARE SCATTERING CONVOLUTION NETWORKS

The goal of the proposed region-aware scattering convolution networks (RegionScatNet) is to obtain a mid-level representation for FBP, which is discriminative for beauty assessment and invariant to the uninformative variances. It consists of three components, including region extraction, attributes decomposition and scattering convolution, as shown in Fig.1. As opposed to standard convolution networks, the filter-banks of the scattering convolution in RegionScatNet are not learned but are scale and rotation invariant wavelets [30, 31]. The extracted features of RegionScatNet are used to train the predictor for FBP.

2.1. Region Extraction



(a) Init. Mask M' (b) Generated M (c) Region of M' (d) Region of M

Fig. 2: The region extraction using region-aware mask [10].

According to the psychological studies, facial shape, smoothness and lighting are three significant factors that influences the facial beauty perception [13]. It motivates us to extract the corresponding image features and integrate them into the RegionScatNet.

Recent study of attention-based CNN for visual question answering [34] indicates that extracting the salient region of image is significant for the CNN-based image understanding. The region extraction of RegionScatNet aims to extract the significant facial region for FBP. However, facial landmarks detection frequently has large variances for images in the wild, and fails to automatically obtain the facial region that fits the boundary with the facial shape. To tackle this problem, we introduce the region-aware mask based on an edge-aware label propagation model [10, 11] to implement region extraction for FBP.

The region extraction is implemented as $I_R = IM$, where I is an input image, I_R is the output facial region and M is the region-aware mask. The region-aware mask M is generated by a edge-aware label propagation model [10] by minimizing the following energy functional:

$$M = \underset{M}{argmin} \sum_{i} Z_{ii}(M_i - M_i')^2 + \lambda \sum_{i,j} W_{ij}(M_i - M_j)^2,$$

where, M' is the initial mask region detected by facial landmarks, S is a diagonal matrix given by $Z_{ii} = 1$ in the constraint region, otherwise $Z_{ii} = 0$; λ is used to balance the relative weights of the two terms. The weight matrix W_{ij} measures the similarity between the pixels. We set $W_{ij} = \|\mathbf{g}_i - \mathbf{g}_j + b\|^{-1}$, where $\mathbf{g}$ is the log-luminance of the input image to guide edge-aware mask diffusion; $b = 0.001$ is the parameter to control the diffusion property. Benefited from the region-aware mask, the boundary of the facial region fits the facial shape closely, which implicitly encodes the shape features in the extracted region, as shown in Fig. 2.

2.2. Attributes Decomposition

This stage aims to separate the extracted facial region I_R into facial detail I_D and structure I_S layers, which contains the facial texture and shape features respectively, as shown in Fig.1. The guided filter f_{guided} [35] is used to obtain the facial detail I_S in the extracted region I_R as $I_S = f_{guided}(I_R|I_{L^*}, r, \varepsilon)$, where I_{L^*} is the luminance channel of the input image I in CIELAB color space to guide the edge-preserving smoothing; the parameters are set as $r = 4$ for window radius, $\varepsilon = 0.2^2$ for regularization. Then, the detail layer I_D is obtained by subtracting I_S from I_R, i.e. $I_S = I_R - I_D$.

2.3. Scattering Convolution

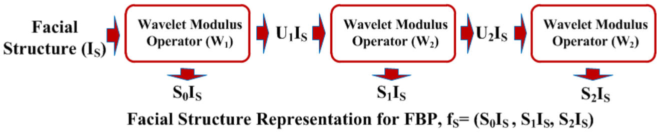

Facial Structure (I_S)
Wavelet Modulus Operator (W_1)
U_1I_S
Wavelet Modulus Operator (W_2)
U_2I_S
Wavelet Modulus Operator (W_2)
S_0I_S
S_1I_S
S_2I_S
Facial Structure Representation for FBP, f_S= (S_0I_S , S_1I_S, S_2I_S)

Fig. 3: The scattering representation $\mathbf{f}_S$ of the facial structure I_S is computed by a cascade of wavelet modulus operator $W_m(m = 0, 1, 2)$, which concatenates the invariant scattering coefficients of each layer, i.e. $\mathbf{f}_S = (S_0I_S, S_1I_S, S_2I_S)$.

We have obtained the salient facial features that implicitly encodes in the image layers of I_S and I_D. Scattering convolution aims to compute the roto-translation invariant representation $\mathbf{f}_S(I_S)$ and $\mathbf{f}_D(I_D)$ of facial structure I_S and detail I_D for FBP. Then, the final facial scattering representation $\mathbf{f}_{scat} = (\mathbf{f}_S, \mathbf{f}_D)$ is used to train the facial beauty predictor.

The scattering convolution networks (ScatNet) is based on scattering transform [32]. The invariant properties of ScatNet facilitates to obtain representation that reduces the uninformative within-class variance during classification, and it has been used for invariant texture discrimination [30,31]. The original ScatNet only offers translation invariant and Lipschitz continuous to deformation [31]. However, face rotation is a common transformation for FBP in the wild. To achieve rotation invariant facial representation, we utilize the roto-translation wavelet structures of [30] in our RegionScatNet, as shown in Fig. 3.

For clarity of presentation, let us only consider the scattering convolution networks of I_S for $\mathbf{f}_S$, while the process of $\mathbf{f}_D$ is the same as $\mathbf{f}_S$. The ScatNet with roto-translation wavelet is a three-layer cascade of wavelet modulus operators W_m, where $m = 0, 1, 2$. Each W_m outputs scattering coefficients S_mI_S and wavelet coefficients $U_{m+1}I_S$ of next layer. Different from the standard convolution networks in deep learning [27, 28], the filter-banks of ScatNet are not learned but are scaled and rotated wavelet that allows the invariant properties [30]. The extracted invariant features $\mathbf{f}_S$ is obtained by concatenates the invariant scattering coefficients of each layer, i.e. $\mathbf{f}_S = (S_0I_S, S_1I_S, S_2I_S)$, as shown in Fig. 3.

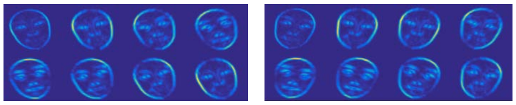


Fig. 4: Visualization of the second layer RegionScatNet feature with facial roll rotation of $-15°$ and $+15°$.

In the first layer of the ScatNet computes S_0I_S and U_1I_S with the first wavelet modulus operator W_1, where $W_1I_S = (S_0I_S, U_1I_S)$. S_0I_S is first computed by averaging I_S by convoluting a rotation invariant gaussian ϕ_J:

$$S_0I_S(x) = I_S * \phi_J(x) = \sum_y I_S(y)\phi_J(x - y), \qquad (1)$$

where we set the parameter $J = 5$ in this paper.The vector of wavelet coefficients U_1 is computed with Morlet wavelets and non-linear $\mathbf{L}^1(\mathbb{R}^2)$ norm:

$$U_1I_S(x) = \{|I_S * \psi_{\theta,j}|\}_{\theta,j}, \qquad (2)$$

where, ψ is a Morlet wavelet given by $\psi(x) = \alpha(e^{iu\xi} - \beta)e^{|x|^2/(2\sigma^2)}$ with $\sigma = 0.85$, $\xi = 3\pi/4$ and $\beta \ll 1$; θ, j are variables of angles and scales.

The second and third layer shares the same wavelet modulus operator $W_2 = W_3$, which computes the average S_1I_S of U_1I_S and S_2I_S of U_2I_S on the roto-translation group. The visualization of the RegionScatNet feature is shown in Fig. 4 and more details about the ScatNet can refer to [30–32].

Despite the invariant properties of scattering convolution, the direct usage of ScatNet fails to perform well in the experiments of FBP. However, the RegionScatNet integrated with region-aware facial attributes decomposition obtains competitive performance for beauty prediction, which indicates the effectiveness of the facial region extraction and attribute decomposition structures.

3. EXPERIMENTS AND ANALYSIS

We evaluated the performance of RegionScatNet for FBP on the recently proposed SCUT-FBP benchmark database [24], which contains 500 Asian female faces with beauty scores ($score \in [1, 5]$) labeled by 75 raters. The ground-truth beauty scores is the average of the 75 scores, and the Pearson Correlation is used to measures the performance between the ground-truth and the predicted result as in [18–20]. To reduce the sample variances, all the methods were tested in 5-folds cross validation.

3.1. Evaluations of RegionScatNet

The extracted RegionScatNet-based features were used to train different models for FBP, such as linear regression (LR), Support Vector Regression (SVR) and Gaussian Processes (GP). To better evaluate the effectiveness of RegionScatNet, all the prediction model uses the same parameter settings. Furthermore, images with different rotation were generated to evaluate the invariant properties of the RegionScatNet.

Basic Evaluations of RegionScatNet. Since the ScatNet structures has been proved to be efficient for invariant texture discrimination, we directly used the ScatNet [30] with LR, GP and SVR for FBP. The results are set as the baseline for our evaluations. Note that the ScatNet used here only contains the roto-transition patch scattering component of [30] to build the rotation invariant features.

Table 1: Comparisons of the ScatNet [30] and our RegionScatNet for FBP. 'LR' is Linear Regression; 'GP' is Gaussian Process for regression; 'SVR' is Support Vector Regression.

Model	LR	GP	SVR
ScatNet [30]	54.22	63.74	67.10
RegionScatNet	68.23	79.01	83.47

To evaluate the effectiveness of our methods specifically designed for FBP, we compared the RegionScatNet and the baseline ScatNet [30] with LR, GP and SVR prediction models, as shown in Table 1. The results indicate that despite the success of ScatNet for texture classification, it fails to perform well for FBP due to the lack of facial beauty priors. By contrast, the RegionScatNet with region-aware facial attributes achieved considerable improvement.

Robustness to Face Rotation. To evaluate the robustness of the RegionScatNet to the roll rotation, experiments were performed using samples with different rotations (ranging from $[-45°, +45°]$), as shown in Table 2 and Fig. 5. All the model were trained using the non-rotation samples, and tested on the samples with different rotation transformations. All the prediction models used the same parameter settings to better evaluate the invariant properties of the RegionScatNet and the visualization of the feature is shown in Fig. 4.

Table 2: Evaluations of RegionScatNet with different image rotations and prediction model for FBP.

Rota.	$-45°$	$-35°$	$-30°$	$-25°$	$-15°$
LR	49.92	50.11	50.13	50.32	50.73
GP	75.62	75.92	76.16	75.97	75.86
SVR	74.74	75.13	75.67	75.61	75.79
Rota.	$+15°$	$+25°$	$+30°$	$+35°$	$+45°$
LR	53.59	55.21	55.61	56.02	57.84
GP	76.76	77.03	77.24	77.83	78.10
SVR	78.02	77.42	77.26	77.37	77.42

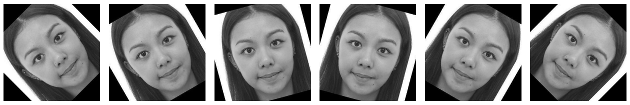


Fig. 5: Facial rotation transformation with the rotation angles of $[-45°, -30°, -15°, +15°, +30°, +45°]$, respectively.

3.2. Comparisons with Related Methods

We compared our method with the handcrafted features and the deep learning-based methods, as shown in Table 3 and Table 4. The results indicate that the RegionScatNet is not only superior to handcrafted features (like Eigenface, LBP, and Gabor), but also the deep learning methods (like PCAnet [37], RBFnet [28], MLP [28] and C2 [36]) for FBP.

We also compared our RegionScatNet with the CDBN-based model of Gan et al. [22]. In the experiments on SCUT-FBP database [24], Gan's CDBN-based feature obtain 0.83 Pearson correlation in the training set, but its performance drops to 0.49 in the testing set. Since the faces of SCUT-FBP database are not aligned as that in [22], it illustrates that the facial representation of RegionScatNet is more robust than the previous methods for facial beauty prediction.

Table 3: Comparisons between hand-crafted features with LR, GP, SVR and CNN prediction with model.

Model	LR	GP	SVR	CNN
Eigenface	0.11	0.13	0.16	0.40
LBP	0.30	0.29	0.28	0.71
Gabor	0.72	0.69	0.75	0.80
RegionScatNet	0.68	0.79	0.83	-

Table 4: Comparisons between deep learning methods with PCAnet [37], RBFnet [28], MLP [28] and C2 [36] for FBP.

Model	PCAnet	RBFnet	MLP	C2	Ours
Pearson Corr.	0.54	0.57	0.71	0.79	0.83

4. CONCLUSIONS

This paper introduces a deep convolutional neural networks, called region-aware scattering convolution networks (RegionScatNet), for facial beauty prediction. The RegionScatNet is integrated with region-aware facial attributes decomposition and the scattering transform. It obtains facial representation that is both discriminative for beauty assessment and invariant to the variances caused by transformation like rotation. State-of-the-art facial beauty assessment results are obtained on the recent SCUT-FBP benchmark database that specifically designed for facial beauty analysis.

There is, however, some limitation in the current work. Firstly, it is hard to interpret the discriminant characteristics of the CNN-based feature in some case, but the visualization of our wavelet-based network (shown in Fig. 4) may provide a cue to understand the property of the feature. Secondly, the small-scale SCUT-FBP face database constrains the predictor to Asian female faces, which motivates us to construct a large-scale benchmark database in the future work.

5. REFERENCES

[1] Y. Chen, H. Mao, L. Jin, “A novel method for evaluating facial attractiveness”, *IEEE Proc. ICALIP*, pp. 1382-1386, 2010.

[2] H. Altwaijry and S. Belongie, “Relative ranking of facial attractiveness,” *IEEE Workshop on WACV*, pp. 117-124, 2013.

[3] H. Gunes, “A survey of perception and computation of human beauty,” *Proc. of J-HGBU*, pp. 19-24, 2011.

[4] Y. Eisenthal, G. Dror and E. Ruppin, “Computational facial attractiveness prediction by aesthetics-aware features,” *Neural Computation*, vol. 18, pp. 119-142, 2006.

[5] Y. Mu, “Computational facial attractiveness prediction by aesthetics-aware features,” *Neurocomputing*, vol. 99, pp. 59-64, 2013.

[6] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz and H. Seidel, “Computer-suggested facial makeup,” *Comput. Graph. Forum*, vol. 30, no. 2, pp. 485-492, 2011.

[7] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, “Data-driven enhancement of facial attractiveness,” *ACM Trans. Graph.*, pp. 28:1-10, 2008.

[8] L. Marchesotti, N. Murray and F. Perronnin, “Discovering beautiful attributes for aesthetic image analysis,” *Int. J. Comput. Vis.* vol. 113, pp. 246-266, 2015.

[9] J. Li, X. Chao, L. Liu, X. Shu, and S. Yan, “Deep face beautification,” *Proc. of the 23rd ACM international conference on Multimedia*, pp. 793-794, 2015.

[10] L. Liang, L. Jin, and D. Liu, “Edge-aware label propagation for mobile facial enhancement on the cloud,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 27, no. 1, pp. 125-138, 2017.

[11] L. Liang, L. Jin, and X. Li, “Facial skin beautification using adaptive region-aware mask,” *IEEE Trans. on Cybernetics*, vol. 44, no. 12, pp. 2600-2612, 2014.

[12] A. Laurentini and A. Bottino, “Computer analysis of face beauty: A survey,” *Computer Vision and Image Understanding*, vol. 125, pp. 184-199, 2014.

[13] I. Stephen, M. Law Smith, M. Stirrat, and D. Perrett, “Facial skin coloration affects perceived health of human faces,” *International Journal of Primatology*, vol. 30, no. 6, pp. 845-857, 2009.

[14] A. Kagian, G. Dror, T. Leyvand, D. Cohen-Or, and E. Ruppin, “A humanlike predictor of facial attractiveness,” *Proc. of NIPS*, pp. 649-656, 2006.

[15] H. Mao, L. Jin and M. Du, “Automatic classification of Chinese female facial beauty using Support Vector Machine,” *Proc. of IEEE SMC*, pp. 4842-4846, 2009.

[16] D. Zhang, Q. Zhao and F. Chen, “Quantitative analysis of human facial beauty using geometric features,” *Pattern Recognition*, vol. 44, no. 4, pp. 940-950, 2011.

[17] J. Fan, K. P. Chau, X. Wan, L. Zhai and E. Lau, “Prediction of facial attractiveness from facial proportions,” *Pattern Recognition*, vol. 45, pp. 2326-2334, 2012.

[18] H. Yan, “Cost-sensitive ordinal regression for fully automatic facial beauty assessment,” *Neurocomputing*, no. 129, pp. 334-342, 2014.

[19] D. Zhang, F. Chen and Y. Xu, *Computer Models for Facial Beauty Analysis*, Springer International Publishing Switzerland, 2016.

[20] W. Chiang, H. Lin, C. Huang, L. Lo, and S. Wan, “The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D Moir features,” *Pattern Recognition*, vol. 47, no. 3, pp. 1249-1260, 2014.

[21] S. Kalayci, H. K. Ekenel, and H. Gunes, “Automatic analysis of facial attractiveness from video,” *Proc. of ICIP*, pp. 4191-4195, 2014.

[22] J. Gan, L. Li, Y. Zhai and Y. Liu, “Deep self-taught learning for facial beauty prediction,” *Neurocomputing*, no. 144, pp. 295-303, 2014.

[23] S. Wang, M. Shao and Y. Fu, “Attractive or not? Beauty prediction with attractiveness-aware encoders and robust late fusion,” *ACM Multimedia*, pp. 805-808, 2014.

[24] D. Xie, L. Liang, L. Jin, J. Xu and M. Li, “SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception,” *Proc. of IEEE SMC*, pp. 1821-1826, 2015.

[25] L. Liu, J. Xing, S. Liu, H. Xu, X. Zhou, and S. Yan, “Wow! you are so beautiful today!,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 11, no. 1s, p. 20, 2014.

[26] N. Murray, L. Marchesotti L, and F. Perronnin, “AVA: A large-scale database for aesthetic visual analysis,” *Proc. of CVPR*, pp. 2408-2415, 2012.

[27] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” *Nature*, vol. 251, pp. 436-444, 2015.

[28] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[29] J. Whithehill and J. Movellan, “Personalized facial attractiveness prediction,” *IEEE Proc. 8th Int. Conf. Aut. Face & Gesture Reco.*, pp. 1-7, 2008,

[30] L. Sifre and S. Mallat, “Rotation, scaling and deformation invariant scattering for texture discrimination,” *Proc. of CVPR*, pp. 1233-1240, 2013.

[31] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872-1886, 2013.

[32] S. Mallat, “Group Invariant Scattering,” *Communications in Pure and Applied Mathematics*, vol. 65, no. 10. pp. 1331-1398, 2012.

[33] T. Poggio, J. Mutch, F. Anselmi, L. Rosasco, J.Z. Leibo, and A. Tacchetti, “The computational magic of the ventral stream: sketch of a theory,” MIT-CSAIL-TR-2012-035, December 2012.

[34] K. Chen, J. Wang, L. Chen, H. Gao, W. Xu, and R. Nevatia, “ABC-CNN: An attention based convolutional neural network for visual question answering,” *arXiv preprint arXiv:1511.05960.*

[35] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397-1409, 2013.

[36] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T.Poggio, “Robust object recognition with cortex-like mechanisms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, pp. 411-426, 2007.

[37] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Yi Ma, “PCANet: A simple deep learning baseline for image classification?.” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017-5032, 2015.