Understanding the visual effects of cosmetic products on beauty via Deep Learning

Victor Martin*[1,2], Renaud Séguier[1], Aurélie Porcheron[2], and Frédérique Morizot[2]

[1]CentraleSupelec, Avenue de la Boulaie, 35510 Cesson-Sévigné, France
[2]Chanel, 8 rue du Cheval Blanc, 93500 Pantin, France

Abstract

Most visual properties of a product are used to be evaluated using analytical and instrumental methods while beauty is a holistic perception. Leveraging the recent Deep Learning advances, we have developed an automatic estimation system able to assess perceived age and health from faces imitating the perception of a large group of naive consumers. We introduce a system based on a Convolutional Neural Network and a Ridge Regression that we trained with a dataset composed of women's facial images and their corresponding ratings of age and health. The ratings of age and health were previously obtained with large groups of naïve assessors. As our dataset contains faces wearing a wide variety of make-up, our Neural Network can precisely estimates age and health from women's faces, wearing makeup or not. The system can estimate perceived age and health from faces with a performance of, respectively, 95% and 85% Pearson Correlation, which is better than the average human performance for both tasks observed on our dataset. To validate our approach, we tested the system with new images of women wearing makeup. The makeup was applied to alter facial cues known to influence human perception of beauty; the results showed that our age and health estimation system reacts in the same way as a group of naïve observers. This new method aims to improve modeling of cosmetics effects on beauty. It should contribute to better understand the visual effects of the products on facial appearance and better predict consumer preferences. We anticipate our work to be a starting point for new approaches of cosmetics sensorial performance prediction.

1 Introduction

Vision is one of the five senses upon which a cosmetic product has to perform, since cosmetics aim to improve appearance and beauty in addition to skin comfort. Our objective is to propose a new method that allows predicting consumer preference for a makeup or skincare results in terms of facial beauty and rejuvenative effect.
Most visual properties of a product are used to be evaluated using analytical and instrumental methods while beauty is a global perception. Woman's beauty is related to people perception of her age and health from the face. Women who look

*Corresponding author: Victor Martin. <victor@vmartin.fr>

healthier and younger for their age are more attractive. Leveraging the recent Deep Learning advances, we have developed an automatic estimation system able to assess perceived age and health from faces imitating the perception of a large group of naive consumers.

First, we introduced a system based on a Convolutional Neural Network and Ridge Regressions to estimate age and health from faces with and without makeup (Sec. 2.1), and we compare its performance to human performance on the same task (Sec. 2.2). Secondly, we evaluated the visual effects of cosmetics products on age and health estimation using our system without human intervention (Sec. 3).

2 Age and Health Estimation from faces

2.1 Method Description

Inspired by the functioning of simple and complex cells in the visual cortex (Hubel and Wiesel, 1962; Lindeberg, 2013), the first part of a Convolutional Neural Network is composed of convolution (complex cells) and pooling layers (simple cells). Convolution layers extract specific levels of details in specific directions; an input image is convoluted to several learned filters to generate features maps, each highlighting specific details in the original image. Pooling layers, on the other hand, "compress" generated features maps by taking the maximum activation - in the case of a Max Pooling - in a coarse window. Combinations of these layers allow us to obtain, from an input image, a compact representation insensitive to pose and illumination variations.

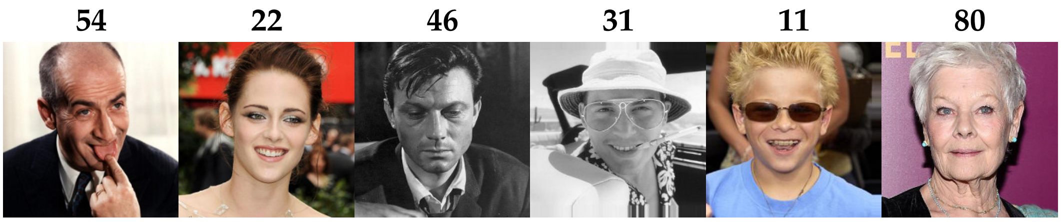

Figure 1: An excerpt of the Internet Movie Database with their corresponding biological age. As we see it above, the database contains faces with large variations in pose, illumination and color distribution. Pictures are resized to 224x224 before training.

Based on the age estimation method of Rothe et al. (2015), we employed the Convolutional Neural Network VGG-16 pre-trained on the ImageNet database (Simonyan and Zisserman, 2014) to detect 1,000 classes of objects, and trained it on the Internet Movie Database (IMDb) of celebrities (Fig. 1).
We filtered the $\approx 500K$ images to keep only those containing faces with resolution greater than 120x120 pixels, no more than one face detected in each image, and only picture depicting people from 11 to 85 years old. For each picture, we have the date of birth of the celebrity pictured and the date of the photo acquisition, thus we can deduce the biological age of the depicted person.
As shown in Figure 2, the Minimum Absolute Error (MAE) for age estimation decreases during training. With this neural network trained for biological age estimation

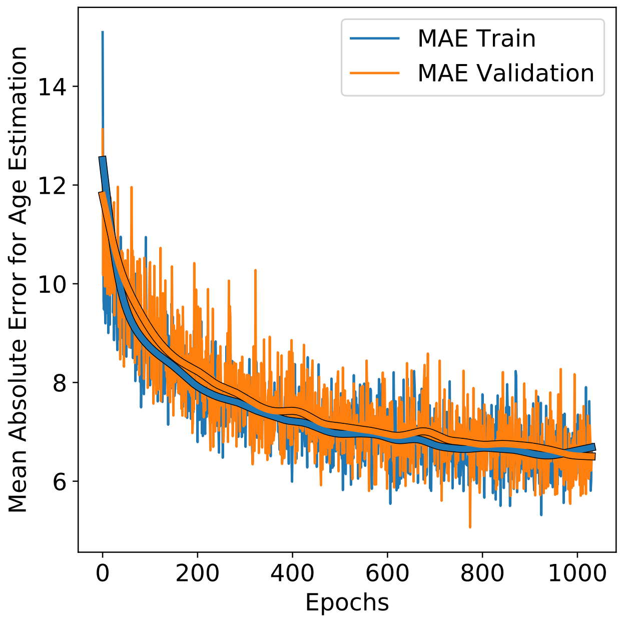

Figure 2: Decrease of the Mean Absolute Error during the training for the train set and the validation set.

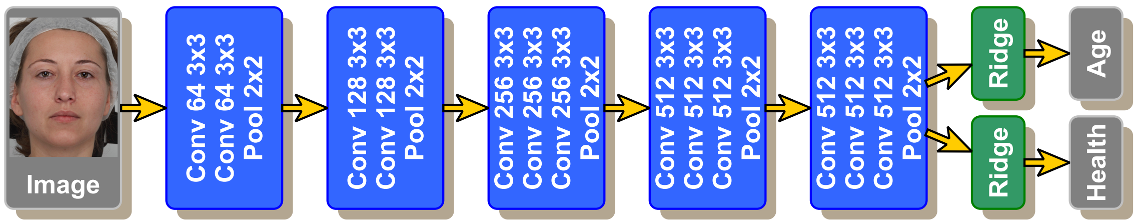

Figure 3: The whole computation chain. The blue part is the Neural Network, computing a representation from an input image. The green part are the two Ridge regressions outputting an age and a health score from the representation.

on faces with large variations, we prune the last multi layer perceptron, and replace it by two parallel Ridge regressions. Afterwards, the first Ridge regression is trained to estimate perceived age on our database, and the second one to estimate perceived health. The final architecture of our system is described in Figure 3.

2.2 Performance Evaluation

2.2.1 Methods

140 images of Caucasian women faces with ages ranging from 20 to 60 years had been rated by 74 Caucasian women raters from the same age range. For every picture, we asked them to evaluate health and to give a score from 0 to 100; 0 being perceived in very bad health and 100 being perceived in very good health. Finally, for each image, we took the average of the 74 ratings to determine a reliable perceived health score. We did the same to obtain a precise perceived age for each face from another database with 300 Caucasian women faces with ages ranging from 20

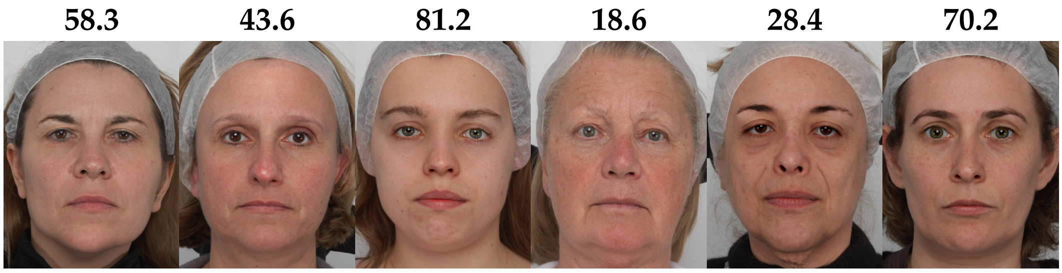

Figure 4: A subsample of our database with their corresponding perceived health scores. Our database contains 140 photos of women faces with a neutral expression in a controlled environment.

to 80 years. Perception of age is a better known field compared to the perception of health. To give the reader an insight on perception of health, we present a sample in Figure 4.
Exploiting the previously described system, we trained the two Ridge regressions in a 140-fold manner to assess its performance.

2.2.2 Results

As we can see on Figure 5 and 6, we can achieve a good performance on our dataset with a scarce amount of data. Using mean absolute error MAE, coefficient of determination R^2 and Pearson correlation PC, Table 1 shows that our system estimates age and health more accurately than an average human working on the same dataset. The estimated age and estimated health are 95.9% and 87.2% correlated, respectively, to the perceived ratings. In the right parts of Figures 5 and 6, the great variability of human ratings is shown. In average, humans ratings are 66.0% and 64.7% correlated for age and health, respectively, to the averaged ratings.

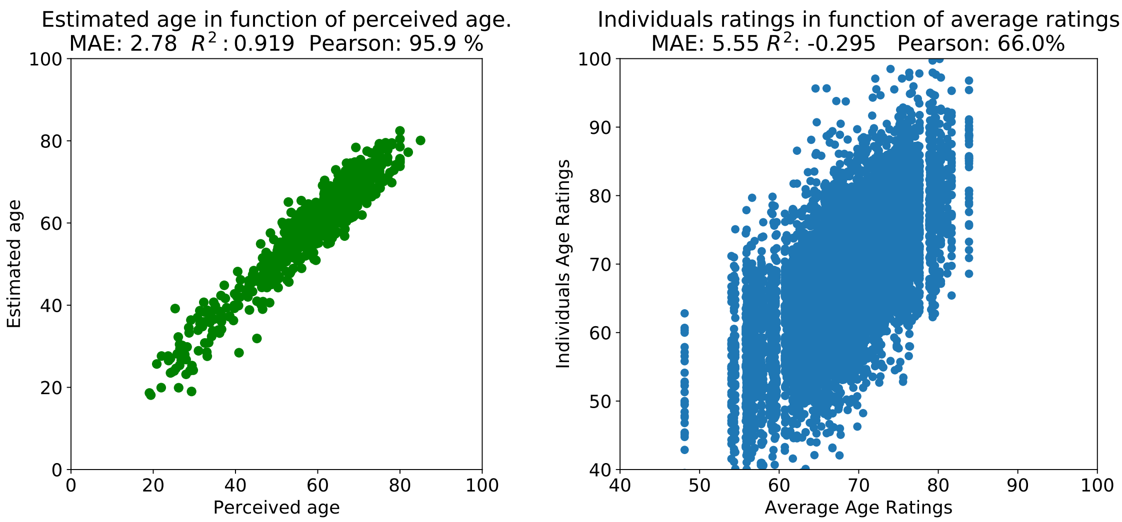

Figure 5: Left graph: The predictions of our system compared to the perceived age scores, which is the average age ratings from humans. Right graph: all individual ratings from humans in function of average ratings. This 2^{nd} graph shows the relatively high variance of human ratings for each image.

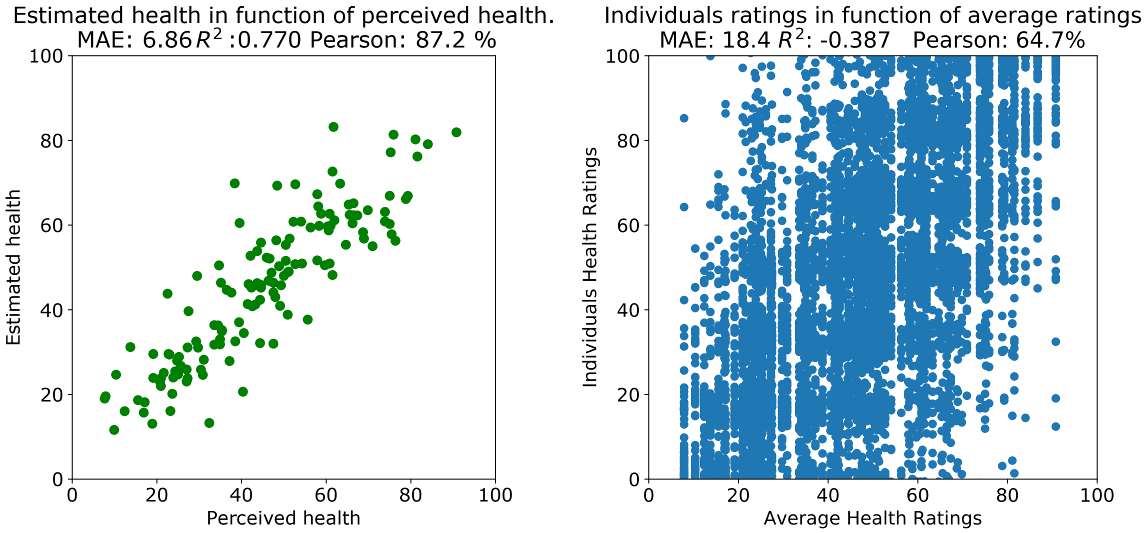

Figure 6: We did the same analysis than for age estimation in Figure 5. Left graph: The predictions of our system compared to the perceived health scores, which is the average health ratings from humans. Right graph: all individual ratings from humans in function of average ratings.

Table 1: Performance of our age and health estimation system compared to human performance.

	Age			Health		
	MAE	PC	R^2	MAE	PC	R^2
System	**2.78**	**95.9%**	**0.919**	**6.86**	**87.2%**	**0.770**
Average Human	5.55	66.0%	-0.295	18.4	64.7%	-0.387

3 Evaluation of cosmetics effect on Age and Health estimation

3.1 Methods

We disposed of facial photographs of 32 women taken under controlled illumination with a neutral pose. Each woman was made up by a makeup artist, to obtain 7 independent conditions:

1. without makeup
2. with concealer and foundation
3. the precedent condition where blush was added
4. the precedent condition where lipstick and lip liner were added
5. the condition 3 where eye liner, eye shadow, mascara, and eyebrow pencil were added
6. with full face makeup where the makeup artist where asked to give a natural look

7. with full face makeup where the makeup artist where asked to give a intense look

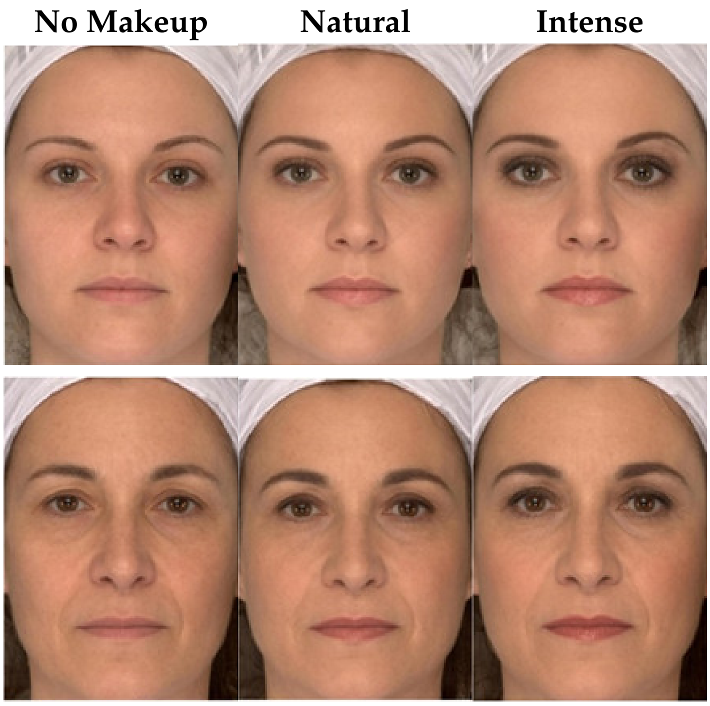

Figure 7: Exemple of 2 faces under 3 conditions: no makeup, with makeup to give a natural look, and with makeup to give an intense look. These faces are average faces of women in their 30's (first line) and in their 50's (second line).

The Figure 7 shows 3 conditions for two average faces for 2 age ranges, 30 years and 50 years.
Leveraging the previously described system, first we estimated the ages and health scores from faces without makeup (condition 1). After that, we estimated the ages and health scores from faces in one of the 6 other conditions. Finally, we computed an average difference between estimates to account the effect of a specific condition.

3.2 Results

The influence of makeup conditions on estimated age and health is presented in Table 2. The effect of makeup appears to be low on estimated age but stronger on estimated health.
Only condition 3 shows a significant and positive influence on estimated age, while all conditions except condition 3 impact significantly and positively the estimation of health.
In the light of the literature we might explain the result on age estimation by recent findings showing differential effects of makeup on perceived age according to the age of the women (Dayan et al., 2015; Russell et al., 2018). Russell et al. (2018) found that 40 - and especially 50-year-old-women did appear significantly younger when wearing makeup, 30-year-old women looked no different in age with or without makeup, while 20-year-old women looked older with makeup. Based on these results, we conducted a supplementary analysis by splitting our set of faces in four

age groups, women in their 20s, 30s, 40s and 50s, with 8 women by age groups. The results are shown on Figure 8. The graphs show similar pattern of results with our system compared to the results observed with human perception in Russell et al. (2018). We tested whether differential effects of makeup also exist on the estimation of health according to age, these results are shown on Figure 9. The pattern of results shows positive effects of makeup on health estimation that vary according to age groups; however the effect was significant only for women in their 30s. The small number of faces by age group might explain the non significant effects of makeup on the other age groups.

Table 2: Influence of different makeup conditions on the estimation of age and health. Non significant values are grayed out.

	Age		**Health**	
	Delta Score	p-value	Delta Score	p-value
2. Foundation	1.02	0.07	1.97	0.03
3. Blush	1.98	0.003	0.85	0.41
4. Lipstick	1.13	0.05	3.72	0.002
5. Eye shadow	-0.40	0.52	2.96	0.003
6. Natural	-0.088	0.86	3.85	0.002
7. Intense	0.13	0.84	3.23	0.002

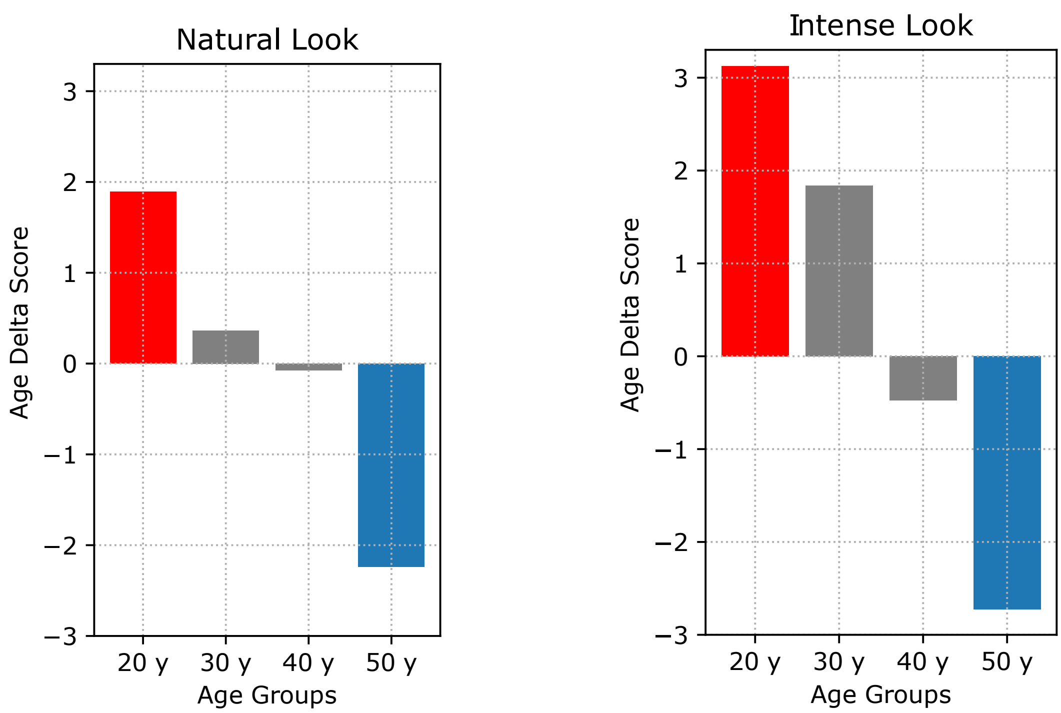

Figure 8: We measure the average difference of scores for makeup giving a natural look (left) and makeup giving an intense look (right) for four age groups. Non significant variations are grayed out.

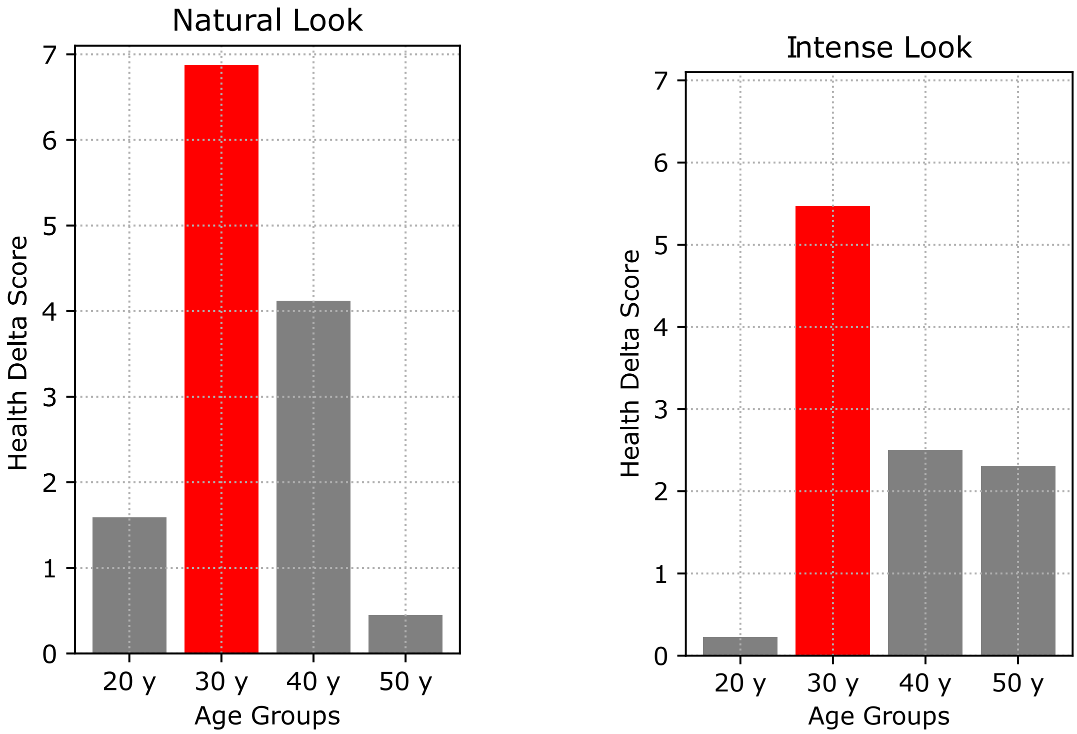

Figure 9: We did an analysis similar to the one in Figure 8, but for health estimation. We measure the average difference of health scores for makeup giving a natural look (left) and makeup giving an intense look (right) for four age groups. Non significant variations are grayed out.

4 Conclusion

We introduced a new system able to automatically estimate age and health from faces outperforming humans doing the same task. We used this system to estimate the effects of wearing different kinds of makeup on human perception, without human intervention.

This new method aims to improve modeling of cosmetics effects on beauty. It should contribute to better understand the visual effects of the products on facial appearance and better predict consumer preferences.
As an improvement direction,this work could be extended to other ethnic groups, in order to evaluate the different effects of cosmetics in various ethnicities.
We anticipate our work to be a starting point for new approaches of cosmetics sensorial performance prediction.

References

Steven H. Dayan, Katherine Cho, Mary Siracusa, and Selika Gutierrez-Borst. Quantifying the impact cosmetic make-up has on age perception and the first impression projected. *Journal of drugs in dermatology: JDD*, 14(4):366–374, April 2015. ISSN 1545-9616.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.2, January 1962. ISSN 0022-3751. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359523/.

Tony Lindeberg. A computational theory of visual receptive fields. *Biological Cybernetics*, 107(6):589–635, December 2013. ISSN 0340-1200, 1432-0770. doi: 10.1007/s00422-013-0569-z. URL https://link.springer.com/article/10.1007/s00422-013-0569-z.

Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 10–15, 2015. URL http://www.cv-foundation.org/openaccess/content_iccv_2015_workshops/w11/html/Rothe_DEX_Deep_EXpectation_ICCV_2015_paper.html. bibtex: rothe_dex_2015.

Richard Russell, Carlota Batres, Sandra Courrèges, Gwenaël Kaminski, Frédérique Soppelsa, Frédérique Morizot, and Aurélie Porcheron. Differential effects of makeup on perceived age. *Under Revision*, 2018.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. URL https://arxiv.org/abs/1409.1556. bibtex: simonyan_very_2014.