New Insights on Weight Estimation from Face Images

Nelida Mirabet-Herranz, Khawla Mallat* and Jean-Luc Dugelay
Departament of Digital Security, EURECOM, Sophia Antipolis, France

***Abstract*— Weight is a soft biometric trait which estimation is useful in numerous health related applications such as remote estimation from a health professional or at-home daily monitoring. In scenarios when a scale is unavailable or the subject is unable to cooperate, i.e. road accidents, estimating a person's weight from face appearance allows for a contactless measurement. In this article, we define an optimal transfer learning protocol for a ResNet50 architecture obtaining better performances than the state-of-the-art thus moving one step forward in closing the gap between remote weight estimation and physical devices. We also demonstrate that gender-splitting, image cropping and hair occlusion play an important role in weight estimation which might not necessarily be the case in face recognition. We use up-to-date explainability tools to illustrate and validate our assumptions. We conduct extensive simulations on the most popular publicly available face dataset annotated by weight to ensure a fair comparison with other approaches and we aim to overcome its flaws by presenting our self-collected database composed of 400 new images.**

I. Introduction

Human body images encode different types of biometric information. Soft biometrics, unlike hard biometrics, do not have the capacity to differentiate between two different identities because of their lack of robustness and distinctiveness. However, soft biometrics such as gender, height and weight can be useful to improve the quality of different systems [8], [12]. Among those, weight is also an indicator of both physical aspect and health conditions and unlike gender and height, body weight changes during a person's adult life and needs to be periodically measured. Conventional weight measurement techniques require the cooperation of the subject to be measured, which might not be possible during medical emergencies, road accidents or due to different patient disabilities. When non cooperative scenarios occur, visual estimation of the weight of the patient by a health professional is preferred [16] but those estimations might not always be accurate [25].

Nowadays, deep learning technologies provide new solutions to obtain end-to-end learning models that gain knowledge and insights from complex, high-dimensional biomedical data [17]. However, the dominance of black-box models is in conflict with the need for trustworthy and explainable artificial intelligence, fundamental for biometric and health related applications. Assessing trust in a model is crucial when actions are taken in the medical domain and cannot be achieved only through accuracy metrics, a trustworthy model should be evaluated using explainability techniques [22]. Furthermore, face images are a rich source of personal and sensitive data that can be used to support a wide range of applications spanning from biometric recognition to user profiling. It is therefore essential that face images are adequately protected so that they cannot be misused ensuring their use exclusively for the target application. Individuals should be able to have access to privacy preserving systems that allow them to select the attributes to be kept and the attributes to be protected or suppressed from their face [6], [18]. Being able to understand which face image factors are important for a weight estimation model in opposition to a face recognition system, is a first step to performing weight anonymization, avoiding targeted advertisement in social networks that might lead to large-scale eating disorders.

It is well-known that transfer learning helps in the process to improve the performances of AI-based algorithms for solving diverse tasks in computer vision. Even so, an appropriate transfer learning strategy needs to be applied to the training process in order to maximize the network's performance. In this work, we decrease up to 15% with respect to the state-of-the-art results of the weight estimation error with our transfer learning strategy. Our results are reported on the largest publicly available database, up to the authors' knowledge, for weight estimation from face images: the VIP_attribute database composed of 1026 celebrities. The capacity of machines to estimate weight from faces remains nevertheless questionable. This is why, we extend our study to better understand differences between face recognition and weight estimation in terms of face cropping, gender splitting and hair occlusions, factors of minor relevance to current face recognition systems. In addition, since celebrities might not be representative of the general weight population, we introduce our Prisoners database, composed of a group population with a weight distribution different than the VIP_attribute. Motivated by the above, we contribute to the literature at three levels: 1) We present our transfer learning protocol for a Resnet50 architecture delivering a relative improvement of 11,4% and 15,3% on the test set Mean Absolute Error (MAE) for a gender-mixed and gender-based weight estimator networks respectively on the VIP_attribute. Besides, we conduct experiments on our new proposed Prisoners database. 2) We give new insights about weight estimation from faces by explaining the network's weight predictions using the explainability approaches LIME [22] and SHAP [15] thus validating a gender-based model selection by showing how the most relevant face regions for estimating the weight differ by gender. 3) We demonstrate that the optimal image pre-

This work has been partially supported by the European CHIST-ERA program via the French National Research Agency (ANR) within the XAIface project (grant agreement CHIST-ERA-19-XAI-011).

* Khawla is now with SAP Security Research Labs France.

processing of image faces for weight estimation must take into account factors such as hair occlusions and the shape and size of face cropping bounding boxes in a different manner as usually done for face recognition tasks.

The rest of this paper is organized as follows. In Section II a review of the state-of-the-art methods for remote weight estimation and an explanation of the explainability techniques used in this work are presented. Section III details our training approach for weight estimation, the new proposed Prisoners database and the results achieved on it as well as on the VIP_attribute. The explainability techniques are applied in Section IV as well as the consequent experiments. Finally, we conclude with future research directions in Section V.

II. Related work

In this Section, we present an overview of the state-of-the-art methods for weight estimation from a person's image making a distinction between two types; methods employing full body images or face images as input data. We also present the explainability techniques that we use in our experiments.

Weight Estimation: Despite the fact that traditional measurement approaches involve physical contact, self-diagnostic image-based methods are a trend nowadays due to the spread of high-quality cameras on affordable mobile phones [2]. Previous work on self-diagnostic has mainly focused on Body Mass Index (BMI) estimation from face images [2], [13], [24], however, BMI is a ratio between weight and squared height thus its estimation is prone to accumulate errors. Our focus is on weight since it is the one that varies throughout a person's adult life. The height of a subject is consistent and having it, the problem of estimating BMI is equivalent to estimating the weight.

Weight estimation via a person's image has predominantly focused on full body images and videos. In 2010, Velardo *et al.* studied the feasibility of correctly predicting the weight of a person from anthropometric data accessible from a subject's image [25]. In their study, a multiple regression analysis was applied to the set of anthropometric features obtained from the image. Also using anthropometric data, in [4], Cao *et al.* presented a copula based technique that aimed to reduce the impact of noise on weight estimation from geometric measurements. In 2012, Labati *et al.* proposed the first, to our knowledge, deep learning approach for an image-based weight prediction [14]. They automatically extracted a set of features using image processing from a pair of frame sequences of a walking person. The features were processed with a feed-forward neural network. Recently, Altinigne *et al.* presented a more refined deep learning scheme to improve the weight prediction from body images [1]. A regression model based on Mask R-CNN architecture predicts the weights of a subject by learning a body contour mask and the skeletal joints.

However, little attention has been given to a direct weight estimation from a subject's face image. Only a handful of research studies addressed the problem of automated face-based estimation of weight. In 2018, Dantcheva *et al.* conducted for the first time a study of height, weight and BMI estimation from a single subject's facial image by implementing a ResNet architecture with 50 layers [7]. Additionally, they presented a novel gender-balanced publicly available dataset, VIP_attribute, consisting of 1026 face images of different subjects. Motivated by the previous, in 2019, Haritosh *et al.* addressed the challenge by defining a two-step network composed of a feature extraction model plus a customized artificial neural network [10] consisting of 3 fully connected layers. They reported the performance of their model for various feature extractors on two different datasets for different weight ranges. In 2020, Han *et al.* addressed the impact of the lack of labeled data on a Convolutional Neural Network (CNN) by presenting an auxiliary-task learning framework for weight estimation [9]. To avoid their network suffering from poor performance due to lack of labeled data, they define as their auxiliary tasks the prediction of other features such as age and gender.

Explainability approaches: Nowadays machine learning models are capable of achieving high predictive accuracy becoming a widespread tool for several applications such as image classification. Although in many cases their performance can be compared with human abilities, it is often at the cost of limited explainability. Different approaches have been studied for better balancing this trade-off, with techniques such as filter visualization [27], disentangling CNN representations into other structures such as decision trees [29] or directly addressing the learning of disentangled representations where the middle layers are no longer a black box [5]. Recently, a new research direction is proposed aiming at quantifying the contribution of each input feature to the decision taken by the model. In 2016, Ribiero *et al.* proposed LIME (Local Interpretable Model-Agnostic Explanations), a technique that could explain any classifier predictions by learning an interpretable model around the prediction of a single data instance via input data perturbations [22]. More specifically, LIME modifies a test data instance by altering its input values, in our case image pixels, and observes the impact on the model outputted weight prediction. In 2017, Lundberg and Lee presented SHapley Additive exPlanations (SHAP) [15], a framework based on Shapley values, which refers in game theory to the average of all the marginal contributions to all possible coalitions. In our case, our game is reproducing the weight outcome of the model and our players refer to the image pixels.

III. Weight estimation model

In this section, we describe our approach for weight estimation from face images. Because reproducibility is essential for future studies, we give all our model implementation details. Other annex documents such as our Prisoners database and protocol files will be also available upon request. We present as well the experimental results for our gender-mixed and gender-based approaches achieving better results than the state-of-the-art in the public VIP_attribute database. Finally, we introduce our self-collected Prisoners database and conduct intra and cross-database experiments on it.

A. Approach

Image augmentation: Image data augmentation techniques can create Convolutional Neural Networks (CNN) invariant to object location, distortion and image brightness, improving the ability of the network to generalize. In this work, we applied augmentation techniques to the training dataset in order to expand it with new, plausible examples such as variations of the training set images that can be representative of other testing samples to be seen by the model. We use the python library Augmentor [3] to randomly flip and distort the input face images and to change their color, contrast and brightness. By including the mentioned augmentation techniques, the ResNet50 decreased the test MAE from 9.23 to 7.54 kg.

Transfer learning: Transfer learning techniques have been applied in many real-world applications outperforming training an entire CNN with random initialization [19]. In our work, we want to verify that performing transfer learning from an analog task, i.e., age estimation from face images reports high benefits to our task being one step towards a future perfect remote weight estimation. In a CNN, filters operating directly on the input data learn during the training process how to extract low-level features such as edges. When multiple layers are stacked, the network can abstract more complex traits as the model depth is increased. In our work, we keep the first layers of a pre-trained ResNet50, since they extract general facial shapes thus we will not retrain them. Only a last number of layers, the ones more specific for the given task originally age estimation, will be adapted to our current objective.

Therefore, we perform a study of the more suitable number of layers to freeze in our network while performing transfer learning. We also test different types of cost function for the training process. The results can be found in Table I. The choice of freezing the first 20 layers of the ResNet50 architecture trained with the Huber loss function delivers the best network performance. The results for the Huber cost function are consistent with its properties. The presence of outliers in the VIP_dataset presents a challenge to the Mean Squared Error (MSE) cost function since large errors have a high impact on the model training. On the other hand, the Mean Absolute Error (MAE) cost function is very sensitive to local minima.

Gender-based network: Because male and female bodies have different bone mineral and muscle density, their facial appearance differs even when they are of the same weight [9]. We want to demonstrate that weight estimation benefits from gender perception therefore we implement a gender-mixed and two gender-based ResNet50 to prove that in opposition to most state of the art face recognition models, a weight estimator gets highly impacted by a prior gender classification. However, there exist other variables such as muscular density, percentage of water or use of medication that might impact the facial appearance therefore the weight estimation. We choose gender as our factor of study since specific medically annotated databases need to be created to study those characteristics.

TABLE I
EVALUATION OF THE WEIGHT ESTIMATION MODEL IN THE VIP_ATTRIBUTE DATASET FOR A DIFFERENT NUMBER OF FROZEN LAYERS AND COST FUNCTIONS.

	MAE		MSE		HUBER	
# hl	MAE	ρ	MAE	ρ	MAE	ρ
5	9.25	0.67	9.97	0.61	9.37	0.69
10	8.5	0.72	9	0.66	9.62	0.71
15	8.57	0.71	8.63	0.66	8.17	0.72
20	8.23	0.75	7.79	0.75	**7.54**	**0.78**
25	9.15	0.68	8.21	0.69	8.8	0.72
30	9.63	0.68	9.38	0.70	9.28	0.72

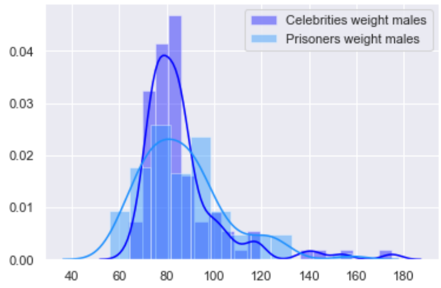
B. Implementation details

We resize the face images to 256×256 and provide them as input for our CNN. To boost the learning process, we apply augmentation techniques to each training image for every training epoch. The ResNet50 structures were implemented in TensorFlow and Keras. We initiated the weights of the network with the filters generated by an age classifier [23], trained in more than 20000 images from the UTKFace dataset [30]. The first 20 layers of the model were frozen, the trainable layers were trained during 10 epochs and the final regression layer during 10 epochs more. Adam optimizer was selected with learning rate set to 0.01. The loss function selected was Huber loss with $\delta = 1$.

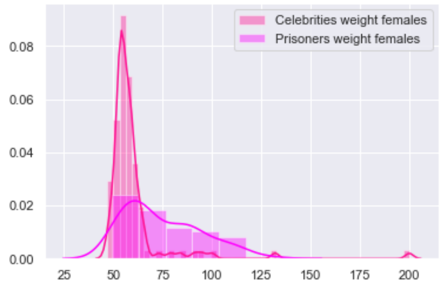
C. Databases

VIP_attribute database: To validate our weight estimation model and enable a fair comparison with future researches, we evaluated our method on the public dataset VIP_attribute consisting of 513 female and 513 male face images of different celebrities (mainly actors, singers and athletes) collected from the WWW. Images are frontal images with different illumination, expression, image quality and resolution conditions. As the authors reported, the weight annotation might not be fully accurate due to inaccurate self-reports or weight fluctuations over time. Other challenges included in this database can be artificial beautification with techniques such as the presence of makeup or plastic surgery. In [7] the authors reported that Viola-Jones algorithm [26] was applied to the images for face detection and an already cropped version of the database is distributed for model comparison purposes. We performed a random gender balanced split of the subjects in training (820 individuals) and testing (206 individuals) subsets.

Prisoners database: The collection of a new database of face images and their associated weight annotation is motivated by the limited number of these publicly available databases and the small weight variations considered in them. To the authors' knowledge, the VIP_attribute dataset is the largest face database annotated with weight. Nonetheless, celebrities might not be representative of the population. Intending to contribute to the literature with a new test group consisting of subjects with a different weight distribution, we proposed the Prisoners database. This database is composed



Weight distribution of the male subjects.



Weight distribution of the female subjects.

Fig. 1. Comparison between the weight distribution of the VIP_attribute and Prisoners database.

of 400 (304 male and 98 female) face images annotated by their associate age, height, weight, ethnicity, gender, eyes color and hair color, collected from the Polk County Jail official webpage[1]. In Figure 1 a comparison between the weight distributions of the individuals of the VIP_attribute and Prisoners databases for the female and male subjects are presented. In opposition to the VIP_attribute, the reliability of the weights associated with each prisoner is high. Their weight was computed at the same moment the face picture was taken, therefore guarantying the accuracy of their annotation and decreasing the uncertainty on our final prediction. We performed a random split of the subjects in training (320 individuals) and testing (80 individuals) subsets.

D. Experimental results

We performed three experiments: 1) Train a single network with the full training set, 2) Train two separate networks splitting our training and testing sets per gender and 3) Perform cross-database and intra-database experiments with the Prisoners dataset.

In Table II, we present a comparison between other weight estimators and our proposed approach. The results demonstrate how with our approach the MAE on the test set for a gender-mixed and gender-based models are significantly reduced, achieving a relative improvement of 11,4% and of 15,3% respectively compared to [7]. In particular, for the female subjects our model achieves an error of 4.24 kg, nearly half the one reported by [7]. In opposition, [7] reports a MAE of $8,06$ and $8,25$ for the female and male

[1]https://www.polksheriff.org/detention/jail-inquiry

TABLE II

PERFORMANCE COMPARISON BETWEEN OUR APPROACH AND THE SoA ON THE VIP_ATTRIBUTE DATASET.

	MAE_f	MAE_m	MAE_{all}	ρ
Dantcheva *et al.* [7]	**-**	-	8,51	0,75
Dantcheva *et al.* [7]	8,06	**8,25**	8,15	0,77
Han *et al.* [9]	-	-	7,20	-
Ours gender-mixed	-	-	7,54	**0,78**
Ours gender-based	**4,24**	9,59	**6,91**	**0,78**

TABLE III

RESULTS OF THE INTRA-DATABASE AND CROSS-DATABASE EXPERIMENTS ON THE VIP_ATTRIBUTE AND PRISONERS DATABASES.

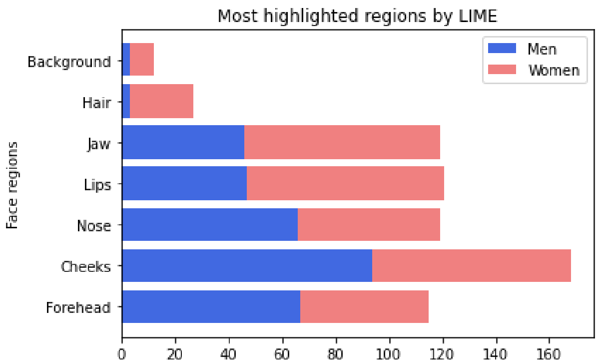
	MAE	ρ	PAP (%)
Train: Prisoners, Test: Prisoners	9,79	0,44	46,25 %
Train: VIP, Test: Prisoners	10.20	0,42	43,75 %
Train: Prisoners, Test: VIP	12,32	0,41	36,58 %
Train: VIP, Test: VIP	7,54	0,78	62,43%

subjects. The VIP_attribute weight distribution has a mean of $\mu = 72.60$ and a standard deviation (std) of $\sigma = 21.94$. The weights of female and male subjects follow a distribution of mean $\mu = 58.34$ and std $\sigma = 11.02$ and mean $\mu = 86.93$ and std $\sigma = 21.00$ respectively. Our results, $4,24$kg (female) and $9,59$kg (male) are more consistent with regards to the weight distribution. Our gender-based models, achieve the lowest MAE of 6.91 kg on the VIP_attribute indicating that training two separate networks, helps the model to learn the appropriate features for weight prediction.

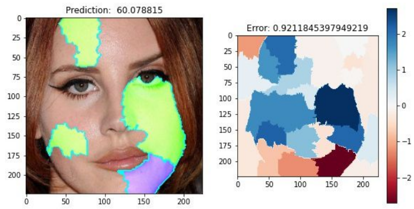
In Table III the results of the experiments conducted on the Prisoners dataset are displayed. The Percentage of Acceptable Predictions (PAP) was introduced by [25] and it represents the percentage of the prediction whose error is smaller than 10% of the initial weight, i.e. a reasonable error in medical applications. Considering as baseline the network trained and tested on the above-mentioned, it is inferable that the ResNet50 trained on the VIP_attribute can generalize its predictions for an unknown population such as the prisoners. When no prisoners are given to the model during the training process, the test MAE experiments a slight increase of $0,41$ kg. However, we observed how the model trained on the Prisoners database is not capable to extrapolate its knowledge to the VIP_attribute population being the test MAE increased of $4,78$ kg. All the prisoners' pictures were taken in a controlled environment with the same light and background conditions while the celebrities' images were collected from different sources. Deep learning models extrapolate better knowledge when a variety of training images and bigger dataset sizes are provided [20].

IV. EXPLAINABILITY STUDY OF CLUE FACTORS ON WEIGHT ESTIMATION FROM FACES

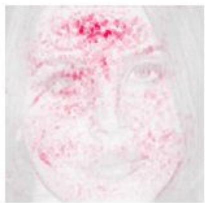
In this Section, we use up-to-date explainability tools to illustrate and validate our hypothesis. We present how gender-splitting, image cropping and hair occlusion are important factors to be considered when remotely estimating the weight from faces.



(a) Count of the most contributive regions after applying LIME.



(b) LIME example



(c) SHAP example

Fig. 2. Explainability approaches applied to the VIP_attribute dataset.

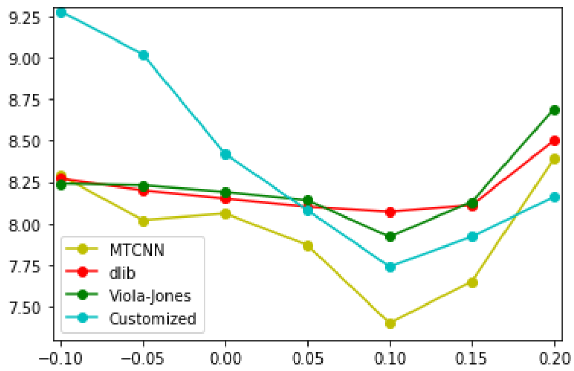


Fig. 3. MAE in kg of the VIP_attribute test set for various face detectors and cropping margins.

TABLE IV

MAE AND MAPE FOR DIFFERENT HAIRSTYLES

Hairstyle	# of subjects	MAE	MAPE
Bold - Short bold	11	11.32	12.98
Short	96	8.73	11.12
Medium	26	5.47	8.00
Long - Long volume	72	6.12	8.92
Facial Hair			
Clean	136	7.40	10.17
Moustache - Goatee	11	7.47	8.84
Beard	47	8.30	10.16
Fringe	11	6.10	9.28

Model explainability: In Fig. 2 we present representations of the output of LIME (b) and SHAP (c) when those explainability techniques are applied to the face images. In (b), the highlighted green areas represent the image regions contributing to an increase of the weight and opposite to them, the purple ones constitute the portions decreasing the weight value. In (c), the red dots represent meaningful players for the game outcome. The explainability techniques do not assess the validity of the result, instead, they give complementary information on which image areas were most significant for the prediction. Nearly all the highlighted pixels for both methods lay in the face skin areas of the image, excluding regions such as background or eye pupil, reinforcing our model trust since meaningful parts of the image were taken into account. We made a count of the most returned face areas by LIME across the test set and presented them in Fig. 2 (a) for the male and female subjects. We observed how in all cases, the cheek area was the most highlighted feature. We also noticed how different areas were highlighted for males and females, as is the case of the jaw for women confirming that the model focuses on different face regions depending on gender. Finally, we also counted the times that a non facial region was highlighted. The background was not often considered relevant information while the hair areas were wrongly taken into account in more cases.

Face detection margins. While assessing trust in the predictions through LIME, we discover that significant regions for our weight estimation model such as the face contour (forehead, cheeks and jaw) are usually excluded from face cropping algorithms since eyes contain the most meaningful information for face recognition tasks [11]. Indeed, the VIP_attribute dataset [7] is distributed in an already cropped version, narrow bounding box and excluding in most cases parts of the face contour (forehead, cheeks and jaw). We contacted the authors of the VIP_attribute dataset which provided us with the original version. Therefore, we evaluated whether different croppings, specially the ones considering larger face areas, will lead to a more accurate prediction as suggested by the explainability approaches. In our experiment, we have trained and tested our network for 4 different face cropping methods: Viola-Jones [26], Multi-Cascade CNN (MTCNN) [28], the dlib python package and a customized cropping. We defined our face cropping by considering the highest, lowest, and furthest at the left and furthest at the right facial landmarks computed by the dlib landmark detector. The results presented in Fig. 3, represent the MAE (y-axis) for different cropping margins (x-axis). Fig. 3 highlights the benefit of an increased margin, specifically of 0.1 (10% increase of the original bounding box), specially for the rectangular face croppings (MTCNN and customized) whose output is more adapted to the face shape. Nevertheless, large croppings include a high amount of hair and background regions increasing the network's MAE.

Hairstyle. Another occlusion factor when considering a face image is hair. LIME highlighted in some cases hair areas as relevant as those can be present above the forehead and cheeks. We extended the annotation of the VIP_attribute database by adding for every subject annotations of their hairstyle, presence and type of facial hair and presence of glasses thus making possible further studies of those categories. The new metadata is based on the one proposed by [21] and will be available upon request to the authors. Table IV presents the MAE and Mean Absolute Percentage Error (MAPE) per category. Some categories such as "Bold-Short bold" are underrepresented so the high values can be due to the presence of outliers. But in the case of facial hair, we can observe how the presence of Fringe as an occlusion is not as meaningful for the prediction as the presence of a beard which increases the error by more than 2 kg on average.

V. Conclusion and Future Works

Weight is a soft biometric trait useful for daily health assessment. Weight estimation from a single facial shot is also the first step toward weight attribute anonymization. Little attention has been given to remote weight estimation from face images in the literature, presenting the existing methods several kg of error. In this study, we move one step closer to make deep learning models accurate on remote weight estimation from faces decreasing the test error 15,3% thus improving the state-of-the-art on the public VIP_attribute dataset. We present the new Prisoners database composed of individuals with a different weight distribution with respect to existing databases and we conduct intra and cross-database experiments on it. We assess the performance of the model with traditional metrics and through explainability techniques presenting three crucial factors when building automatic weight estimators from faces 1) gender, 2) cropping bounding box, 3) facial hair occlusions. A deeper study on the important face regions for the weight estimation will be further explored by occluding different areas and visualizing with explainability tools how the model predicts without being provided the occluded region. Furthermore, the study of other influencing factors such as muscular density and the use of medication should be considered, being not possible with the current publicly available databases.

References

[1] C. Y. Altinigne, D. Thanou, and R. Achanta. Height and weight estimation from unconstrained images. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2298–2302. IEEE, 2020.

[2] M. L. Barr, G. Guo, S. E. Colby, and M. D. Olfert. Detecting body mass index from a facial photograph in lifestyle intervention. *Technologies*, 6, 2018.

[3] M. D. Bloice, P. M. Roth, and A. Holzinger. Biomedical image augmentation using augmentor. *Bioinformatics*, 35(21), 2019.

[4] D. Cao, C. Chen, D. Adjeroh, and A. Ross. Predicting gender and weight from human metrology using a copula model. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 162–169. IEEE, 2012.

[5] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 2180–2188, 2016.

[6] S. Chhabra, R. Singh, M. Vatsa, and G. Gupta. Anonymizing k-facial attributes via adversarial perturbations. *arXiv preprint arXiv:1805.09380*, 2018.

[7] A. Dantcheva, F. Bremond, and P. Bilinski. Show me your face and i will tell you your height, weight and body mass index. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3555–3560. IEEE, 2018.

[8] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51, 2011.

[9] D. Han, J. Zhang, and S. Shan. Leveraging auxiliary tasks for height and weight estimation by multi task learning. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2020.

[10] A. Haritosh, A. Gupta, E. S. Chahal, A. Misra, and S. Chandra. A novel method to estimate height, weight and body mass index from face images. In *Twelfth International Conference on Contemporary Computing (IC3)*. IEEE, 2019.

[11] E. Hjelmas and J. Wroldsen. Recognizing faces from the eyes only. In *In Proceedings of the 11th Scandinavian Conference on Image Analysis*. Citeseer, 1999.

[12] A. K. Jain, S. C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. In *International conference on biometric authentication*, pages 731–738. Springer, 2004.

[13] M. Jiang, Y. Shang, and G. Guo. On visual bmi analysis from facial images. *Image and Vision Computing*, 89:183–196, 2019.

[14] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti. Weight estimation from frame sequences using computational intelligence techniques. In *2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, pages 29–34. IEEE, 2012.

[15] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777, 2017.

[16] S. Menon and A.-M. Kelly. How accurate is weight estimation in the emergency department? *Emergency Medicine Australasia*, 17(2):113–116, 2005.

[17] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

[18] V. Mirjalili, S. Raschka, and A. Ross. Privacynet: semi-adversarial networks for multi-attribute face privacy. *IEEE Transactions on Image Processing*, 29:9400–9412, 2020.

[19] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[20] L. Perez and J. Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[21] H. Proenca and J. C. Neves. Soft biometrics: Globally coherent solutions for hair segmentation and style recognition based on hierarchical mrfs. *IEEE Transactions on Information Forensics and Security*, 12(7):1637–1645, 2017.

[22] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[23] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE international conference on computer vision workshops*, 2015.

[24] H. Siddiqui, A. Rattani, D. R. Kisku, and T. Dean. Al-based bmi inference from facial images: An application to weight monitoring. In *2020 IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2020.

[25] C. Velardo and J.-L. Dugelay. Weight estimation from visual body appearance. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2010.

[26] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[27] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[28] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[29] Q. Zhang, Y. Yang, H. Ma, and Y. N. Wu. Interpreting cnns via decision trees. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6261–6270, 2019.

[30] S. Y. Zhang, Zhifei and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.