Journal of Theoretical and Applied Information Technology

15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ALGORITHM FOR GENERATING DESCRIPTIVE SENTENCES OF THE HUMAN HEAD PARTS BASED ON ENGLISH GRAMMAR

LOAI ALAMRO [1]*, YUHANIS YUSOF[1] , NOORAINI YUSOFF [2]

[1] *School of Computing, Universiti Utara Malaysia, Malaysia*

[2] *Department of Data Science, Universiti Malaysia Kelantan, City Campus, Malaysia*

* Corresponding author's Email: loaialamro@gmail.com

ABSTRACT

Human head exhibits many biological features (attributes) that represent the characteristics of the human head with robust inherent stability and individual variation. These attributes provide important discriminative knowledge about humans, such as gender, age, race, hairstyle, hair color, etc. Recently, several human head attribute classification networks have been proposed. However, these networks do not provide a clear picture of the human head because they predict head attributes in terms of binary values (i.e., 0 or 1) or by their labels (i.e., male, young). Therefore, in this study, a description algorithm was proposed to describe the main characteristics of the human head using the adjective's arrangement rules. The proposed algorithm was reviewed by experts, and the responses of seven experts show that the algorithm follows the adjective's arrangement rules in accordance with the conventions of human language. The experts also found the descriptive sentences acceptable, understandable, and grammatically correct.

Keywords: *Human Attribute Classification, Human Attribute Description, Object Identification, Object Recognition, Deep Learning.*

1. INTRODUCTION

The human head holds many biological features (attributes) that represent the human head characteristics with robust inherent stability and individual variation. These attributes provide significant discriminatory knowledge about humans, such as gender, age, race, hairstyle, hair color, etc. During the past few years, classifying these attributes has attracted significant attention in computer vision and pattern recognition, due to its widespread in many real-world applications, such as face identification and verification [1, 2], person re-identification [3, 4], recommendation systems [5, 6], human-computer interaction and visual assistant systems. Given a human head image, the task of human head attributes classification (HHAC) is to predict multiple attributes of the human head, such as gender, smiling, age, and accessories (Fig. 1 shows some examples of HHAC).

Recently, motivated by the outstanding performance of Convolutional Neural Network (CNN), most state-of-the-art HHAC networks take advantage of CNN to classify human head attributes. The HHAC networks divided into two main groups: single-label classification networks [7, 8] and multi-label classification networks [9-15]. The single-label classification networks [7, 8] employ the CNN network for extracting the features of the human head and then use a classifier such as Support Vector Machines (SVM) [16] to learn the human head attributes. However, these networks consider each attribute of the human head as an independent classification challenge. Therefore, these networks classify the human head attributes separately without considering the correlations between them.

On the other hand, multi-label classification networks [9-15] classify multiple attributes of the human head in an end-to-end CNN network. In these networks, the CNN network is employed as a feature extractor and classifier simultaneously. Specifically, the lower layers (convolutional layers) of CNN are used to extract the features of the human head, while the upper layer (single output layer) is used as classification layer. However, these networks deal with the human head attributes equally through the

training stage without considering the different learning complexities of these attributes.

Besides, some multi-label classification networks [17-20] divide the human head attributes into groups and employ the lower layers of CNN to extract the features of the human head and several upper layers to learn each group of human head attributes. These networks group the human head attributes into several groups according to various criteria to learn the correlations between them. For instance, Hand and Chellappa [17] divide the attributes into nine groups according to their locations on a human head, Han, Jain, Wang, Shan and Chen [18] divide the attributes according to the heterogeneity (i.e., ordinal vs. nominal and holistic vs. local) in terms of data kind and semantic meaning, Cao, Li and Zhang [19] divide the head attributes into four groups (i.e., upper, middle, lower, and whole image) according to their locations on the human head, Mao, Yan, Xue and Wang [20] divide the attributes into two groups: objective categories and subjective categories, and [21] divide the head attributes into five groups (i.e., hair, face, style, accessories and appearance) according to the common characteristics among the attributes. However, the existing HHAC networks [7-15, 17-21] do not provide a clear picture of the human head as they predict the head attributes in terms of binary values (i.e., 0 or 1) or by single labels (i.e., male, young). Despite the importance of this information, it is not sufficient and is not understood because it does not provide detailed information about the individuals in a written depiction. Therefore, the performance of HHAC networks [7-15, 17-21] is limited and inefficient in many real-world applications such as visual assistance systems and robotic vision, as the prediction results are incomprehensible to the end user (especially for people with severe visual impairments).

The effective description of humans can be of great benefit to visually impaired and blind people by giving them access to visual information, facilitating communication, enriching their social experiences, and enabling them to better navigate the world. Therefore, there is an urgent need to develop a description algorithm that increases the effectiveness of visual assistance systems by providing the visually impaired with descriptions of people in the environment. However, how to generate the description of the human head and convert it into grammatically correct sentences remains a challenging problem. Therefore, this study proposed an algorithm that generates descriptive sentences based on English grammar. The proposed description algorithm processes the binary outputs of the classification network [21] and generates understandable textual descriptions. The description algorithm uses the adjective arrangement rules to organize the classification outputs in a grammatically correct sentence.

The main contributions of this study are summarized as follows:

- A novel description algorithm that generates descriptive sentences based on English grammar to describe the different parts of the human head.
- A standardised framework for using English grammar to describe people, which makes it possible to generate sentences that are structured in a logical and comprehensible way.

2. PROPOSED ALGORITHM

CelebA and LWFA are datasets for human head attributes classification, where they are used for training and testing DL networks to classify human characteristics such as blonde hair, smiling, wearing a necklace, etc. The CelebA and LWFA datasets have large diversities and rich annotations of human head attributes, including attributes of hair, face, style, accessories, and appearance. In these datasets, the attributes are represented by binary values, where each attribute takes one of just two possible values (i.e., 0 or 1). For example, when a human's gender is predicted, the prediction result will be (0) if the human is male and (1) if the human is female. Therefore, existing DL networks predict those attributes in the form of binary values (i.e., 0 or 1) or by their labels (i.e., male, young, etc.). This limits the effectiveness of these networks in real-world applications such as visual assistance systems and robotic vision a s discussed previously. Therefore, in this study, a description algorithm was proposed to process the binary outputs of our previous HHAC network known as Multi-Output Convolutional Neural Network for Automatic Human Head Attributes Classification (MOCNN-HHAC) [21].

The description algorithm generates descriptive sentences about human head based on the binary outputs of MOCNN-HHAC. The CelebA and LWFA datasets involve 40 binary attribute annotations that classify different parts of the human head. These attributes vary between the attributes of gender, external appearance, face appearance, and accessories. Therefore, in order to generate

sentences that describe humans correctly, this study proposed to classify these attributes into eight (8) groups; gender, appearance, hair, eyebrows & eyes, nose & mouth, face, facial hair, and accessories, as presented in Table 1. Moreover, in the English language, these attributes denote adjectives which are usually organized in a specific order. Therefore, to generate grammatically correct descriptive sentences, these attributes should be arranged according to English grammar. The English grammar dealing with attribute structures and sentence structures are known as "adjectives arrangement rules" [22, 23]. The adjective arrangement rules [22, 23] arrange adjectives (attributes) according to the order depicted in Table 2. The description algorithm has four main functions: 1) assign the prediction output of the MOCNN-HHAC to the corresponding labels, 2) determine a human gender, 3) order the labels according to the adjectives arrangement rules, and 4) generate the description sentences. First, the proposed algorithm processes the prediction output vector (POV) of MOCNN-HHAC by converting each binary value in POV to the corresponding label. Second, the algorithm determines the human gender and stores it in the gender variable, then the algorithm stores the pronoun compatible with the gender (i.e., he or she) in the output variable. This is followed by checking the availability of the labels of appearance hair, eyebrows & eyes, nose & mouth, face, facial hair and accessories, and stores them in the relevant variables. Then, the algorithm orders the labels in each variable according to the adjective's arrangement rules. Finally, the algorithm stores all labels in the output variable to generate the description sentence. Figure 1 shows samples of description sentences, while the steps of the proposed description are depicted in Algorithm 1.

3. EVALUATION

The use of expert reviews is an important method for examining the design process of a proposed algorithm, as it has been acknowledged in previous research [24, 25]. Consequently, the present study utilized expert reviews to verify that the proposed algorithm could generate sentences that describe human characteristics according to English grammar (adjective arrangement rules). The verification process aimed to confirm that the steps taken in designing the algorithm were accurate. There were three main points to be verified; the order of opinion and fact adjectives, the acceptability and understanding of the generated sentences, and the consistency of the algorithm's performance with human descriptions. Experts who had expertise and experience in the academic field of English were identified for the verification process, based on the characteristics outlined in previous studies by Rogers and Lopez [26] and Hallowell and Gambatese [27].

4. ANALYSIS AND RESULTS

The use of expert reviews is an important method to investigate the design process of a proposed algorithm, as found in previous research [24, 25]. Therefore, in the present study, expert reviews were used to verify whether the proposed algorithm can generate sentences that describe human characteristics according to English grammar (adjective arrangement rules). The verification process aimed to confirm that the steps taken in the development of the algorithm were correct. There were three main points to be checked: the order of opinion and fact adjectives, the acceptability and understanding of the generated sentences and the consistency of the algorithm's performance with human descriptions. For the verification process, experts who had expertise and experience in the academic field of English were selected based on the characteristics described in previous studies by Rogers and Lopez [26] and Hallowell and Gambatese [27].

Seven experts in English were identified and a face-to-face meeting were conducted with them. All experts received a feedback form that includes: 1) instructions, 2) a demographic profile, 3) samples of images with attached descriptive sentences, 4) a table with the ordering of adjectives arrangement rules, and 5) evaluation questions. The were seven experts with various academic background related to English Five of the experts have a Ph.D., while the other two have a master's degree. The majority of the experts (5 out of 7) have more than 20 years of experience, with the remaining two experts having between 11 to 19 years of experience. The experts also have different academic positions, with four being Associate Professors, two being Senior Lecturers, and one being a professor. Overall, these characteristics suggest that the experts are highly educated and experienced in their respective fields, which may make them valuable sources of expertise in research and academia, which is considered sufficient [26, 27]. The review sessions involved several activities include:

1. The researcher presented an overview of the study and the proposed steps for designing the description algorithm to the experts.
2. The experts reviewed the implementation steps of the description algorithm and created sentences that were included in related documents. They also had the opportunity to ask questions for clarification.
3. The experts provided response by completing the feedback form.
4. The researcher has taken the experts' suggestions and updated the algorithm accordingly.

The description algorithm was verified for its validity, i.e., whether the description algorithm implemented in this study conforms to the adjective arrangement rules in English. Table 4 contains the responses of seven experts who were interviewed and asked to answer a series of questions about the adjective arrangement rules in English and the possibility of using it to describe human head. They were also asked about the validity of the descriptive sentences generated by the description algorithm. For each question, the experts were asked to rate their level of agreement on a Likert scale from "strongly agree" to "strongly disagree".

The first question asked whether the adjective arrangement rules can be used to describe a human head. All experts agreed with this statement (5 of 7 strongly agreed). This indicates that the experts believe that the established adjective arrangement rules can be effectively applied to describe a human head. The second and third questions asked whether the feedback form contained the correct order for the adjective's "opinion" and "fact." The majority of experts agreed or strongly agreed that the feedback form contains the correct order of adjectives for both types, indicating that the adjectives order is correct and consistent with established rules.

The fourth and fifth questions asked whether the generated descriptive sentences in English were acceptable and understandable and whether they were grammatically correct. The majority of experts (6 out of 7) strongly agreed that the sentences were both acceptable and grammatically correct, while one expert was neutral on each question. This indicates that the sentences were well worded and appropriate for describing a human head. Finally, the sixth question asked whether the description algorithm follows the adjective arrangement rules consistently with the human description style. The majority of the experts (6 out of 7) agreed with this statement, and two experts even strongly agreed. This indicates that the description algorithm accurately follows the adjective arrangement rules, as it describes human effectively and consistently with human language conventions.

In summary, the responses of the seven experts indicate that the adjective arrangement rules can be applied to describe a human head, that the feedback form contains the correct order for both opinion and fact adjectives, and that the description algorithm follows the adjective arrangement rules consistently with the conventions of human language. The experts also found the descriptive sentences acceptable, understandable and grammatically correct.

5. CONCLUSION

This paper presents a new description algorithm that generates descriptive sentences by processing the binary output of the MOCNN-HHAC. This algorithm generates descriptive sentences based specifically on English grammar (adjective's arrangement rules) to describe humans. The algorithm was reviewed using the expert review method, in which seven English language experts reviewed the algorithm in a face-to-face interview. The responses from seven experts show that the algorithm follows the adjective's arrangement rules in accordance with the conventions of human language. The experts also found the descriptive sentences generated acceptable, understandable and grammatically correct.

Table 1: Groupings of Human Head Attributes Used in Human Description Algorithm

Gender	Appearance	Hair	Eyebrows & Eyes	Nose & Mouth	Face	Facial Hair	Accessories
Male	Attractive	Bald	Arched Eyebrows	Big Lips	Double Chin	5o'clock Shadow	Eyeglasses
	Chubby	Bangs	Bushy Eyebrows	Big Nose	High Cheekbones	Goatee	Heavy Makeup
	Smiling	Black Hair	Bags Under Eyes	Mouth Open	Oval Face	Mustache	Wearing Earrings
	Young	Blond Hair	Blurry	Pointy Nose	Rosy Cheeks	No Beard	Wearing Hat
		Brown Hair	Narrow Eyes		Pale Skin	Sideburns	Wearing Lipstick
		Gray Hair					Wearing Necklace
		Receding Hairline					Wearing Necktie
		Straight Hair					
		Wavy Hair					

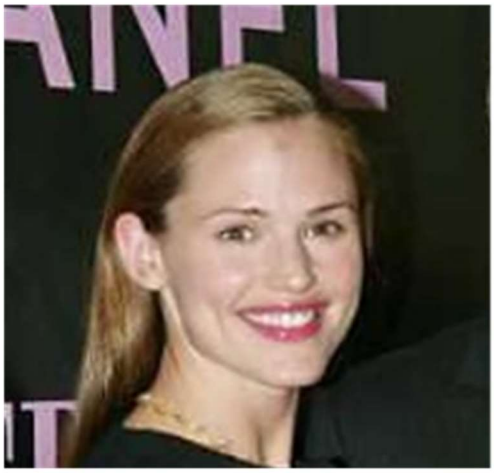
Table 2: Adjectives Arrangement Rules

	Rule
1	Opinion adjectives: (Smiling + Attractive + Chubby + Young)
2	Fact adjectives: (Size + Shape + Age + Color + Nationality + Material + Noun)
3	Opinion adjectives must precede Fact adjectives.

He is a smiling, chubby male. He has gray hair and a receding hairline. He has narrow eyes with bags under eyes. He has a big nose with a mouth slightly open. He has an oval face, high cheekbones, a double chin with pale skin.

She is a smiling female. She has wavy gray hair. She has arched eyebrows. She has a big nose with a mouth slightly open. She has high cheekbones. She is wearing lipstick and earrings.

He is an attractive, chubby male. He has brown hair and a receding hairline. He has bushy eyebrows and narrow eyes. He has a pointy nose with a mouth slightly open. He has a double chin with pale skin. He is wearing eyeglasses and a necktie.



She is a smiling, attractive female. She has wavy blond hair. She has arched eyebrows and narrow eyes. She has big lips, a pointy nose with a mouth slightly open. She has an oval face, high cheekbones, rosy cheeks, and a double chin. She is wearing lipstick and a necklace.

Figure 1: Samples of Human Description

Algorithm 1: Human Description

1) Convert the binary values in the classification vector of the MOCNN-HHAC network into corresponding labels for the human head parts.
2) Determine the gender of a human and set it to the gender variable (GE).
3) Check for appearance labels in the classification vector, move the available labels to the appearance variable (AP), and then prioritize these labels based on ordering rules.
4) Check for hair labels in the classification vector, move the available labels to the appearance variable (HA), and then prioritize these labels based on ordering rules.
5) Check for eyebrow and eye labels in the classification vector, move the available labels to the appearance variable (EE), and then prioritize these labels based on ordering rules.
6) Check for nose and mouth labels in the classification vector, move the available labels to the appearance variable (NM), and then prioritize these labels based on ordering rules.
7) Check for face labels in the classification vector, move the available labels to the appearance variable (FA), and then prioritize these labels based on ordering rules.
8) Check for facial hair labels in the classification vector, move the available labels to the appearance variable (FH), and then prioritize these labels based on ordering rules.
9) Check for accessory labels in the classification vector, move the available labels to the appearance variable (AC), and then prioritize these labels based on ordering rules.
10) Generating a description sentence by concatenating the variables (GE, AP, HA, EE, NM, FA, FH, AC).

Table 3: Summarizes The Experts' Background, Including Their Gender, Age, Education, Academic Position, And Years Of Experience.

Expert	Gender	Age	Education	Organization	Position	Experience Years
A	Male	> 50	Ph.D.	University	Professor	> 20
B	Male	> 50	Ph.D.	University	Associate Professor	> 20
C	Female	31 to 40	Ph.D.	University	Associate Professor	11 to 19
D	Male	> 50	Master's	University	Associate Professor	> 20
E	Male	41 to 50	Ph.D.	University	Associate Professor	11 to 19
F	Male	31 to 40	Master's	University	Senior Lecturer	11 to 19
G	Female	41 to 50	Master's	University	Senior Lecturer	> 20

Table 4: Results for Description Algorithm Verification

Question	Expert A	Expert B	Expert C	Expert D	Expert E	Expert F	Expert G
Can the adjective arrangement rules be used to describe human head?	Strongly Agree	Agree	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Strongly Agree
Are the opinion adjectives correctly ordered?	Agree	Strongly Agree	Agree	Strongly Agree	Agree	Agree	Agree
Are the fact adjectives correctly ordered?	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Agree	Strongly Agree	Agree
Are the description sentences acceptable and understandable in English?	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Neutral	Agree	Strongly Agree
Are the description sentences grammatically, correct?	Strongly Agree	Strongly Agree	Strongly Agree	Strongly Agree	Neutral	Strongly Agree	Strongly Agree

ACKNOWLEDGMENTS

This research was supported by the Ministry of Higher Education (MoHE) of Malaysia through Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UUM/02/1).

REFERENCES

[1] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification." pp. 365-372.

[2] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, "Describable Visual Attributes for Face Verification and Image Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, no. 10, pp. 1962-1977, 2011.

[3] S. Khamis, C.-H. Kuo, V. K. Singh, V. D. Shet, and L. S. Davis, "Joint Learning for Attribute-Consistent Person Re-Identification," *Computer Vision - ECCV 2014 Workshops.* pp. 134-146.

[4] L. Wu, S. Chunhua, and A. v. d. Hengel, "Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification," *Pattern Recognition,* vol. 65, pp. 238-250, 2017/05/01/, 2017.

[5] G. J. Qi, C. Aggarwal, Q. Tian, H. Ji, and T. Huang, "Exploring Context and Content Links in Social Media: A Latent Space Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 5, pp. 850-862, 2012.

[6] G.-J. Qi, X.-S. Hua, and H.-J. Zhang, "Learning semantic distance from community-tagged media collection," in Proceedings of the 17th ACM international conference on Multimedia, Beijing, China, 2009, pp. 243–252.

[7] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev, "PANDA: Pose Aligned Networks for Deep Attribute Modeling." pp. 1637-1644.

[8] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild." pp. 3730-3738.

[9] E. M. Rudd, M. Günther, and T. E. Boult, "MOON: A Mixed Objective Optimization Network for the Recognition of Facial Attributes," *Computer Vision – ECCV 2016.* pp. 19-35.

[10] K. He, Z. Wang, Y. Fu, R. Feng, Y.-G. Jiang, and X. Xue, "Adaptively Weighted Multi-task Deep Network for Person Attribute Classification," in Proceedings of the 25th ACM international conference on Multimedia, Mountain View, California, USA, 2017, pp. 1636–1644.

[11] H. Guo, X. Fan, and S. Wang, "Human attribute recognition by refining attention heat map," *Pattern Recognition Letters,* vol. 94, pp. 38-45, 2017/07/15/, 2017.

[12] U. Mahbub, S. Sarkar, and R. Chellappa, "Segment-Based Methods for Facial Attribute Detection from Partial Faces," *IEEE Transactions on Affective Computing,* vol. 11, no. 4, pp. 601-613, 2018.
[13] M. Xu, F. Chen, L. Li, C. Shen, P. Lv, B. Zhou, and R. Ji, "Bio-Inspired Deep Attribute Learning Towards Facial Aesthetic Prediction," *IEEE Transactions on Affective Computing,* vol. 12, no. 1, pp. 227-238, 2018.
[14] N. Zhuang, Y. Yan, S. Chen, and H. Wang, "Multi-task Learning of Cascaded CNN for Facial Attribute Classification." pp. 2069-2074.
[15] S. Huang, X. Li, Z.-Q. Cheng, Z. Zhang, and A. Hauptmann, "GNAS: A Greedy Neural Architecture Search Method for Multi-Attribute Learning," in Proceedings of the 26th ACM international conference on Multimedia, Seoul, Republic of Korea, 2018, pp. 2049–2057.
[16] A. M. Andrew, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods by Nello Christianini and John Shawe-Taylor," *Robotica,* vol. 18, no. 6, pp. 687-689, 2000.
[17] E. M. Hand, and R. Chellappa, "Attributes for improved attributes: A multi-task network for attribute classification," *arXiv preprint arXiv:1604.07360*, pp. 1-16, 2016.
[18] H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen, "Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 11, pp. 2597-2609, 2018.
[19] J. Cao, Y. Li, and Z. Zhang, "Partially Shared Multi-task Convolutional Neural Network with Local Constraint for Face Attribute Learning." pp. 4290-4299.
[20] L. Mao, Y. Yan, J. Xue, and H. Wang, "Deep Multi-task Multi-label CNN for Effective Facial Attribute Classification," *IEEE Transactions on Affective Computing*, pp. 1-1, 2020.
[21] L. Alamro, Y. Yusof, and N. Yusoff, "Multi-Output Convolutional Neural Network for Automatic Human Head Attributes Classification," *International Journal of Intelligent Engineering Systems,* vol. 15, no. 5, pp. 219-231, 2022.
[22] R. Murphy, B. Viney, and M. Craven, *English Grammar In Use with Answers and CD ROM: A Self-study Reference and Practice Book for Intermediate Students of English*: Cambridge university press, 2004.
[23] J. Praninskas, "Rapid review of english grammar for students of english as a second language," 1961.
[24] M. Komuro, and N. Komoda, "An explanation model for quality improvement effect of peer reviews." pp. 1159-1164.
[25] K. E. Wiegers, "Seven truths about peer reviews," *Cutter IT Journal,* vol. 15, no. 7, pp. 31-37, 2002.
[26] M. R. Rogers, and E. C. Lopez, "Identifying critical cross-cultural school psychology competencies," *Journal of school psychology,* vol. 40, no. 2, pp. 115-141, 2002.
[27] M. R. Hallowell, and J. A. Gambatese, "Qualitative research: Application of the Delphi method to CEM research," *Journal of construction engineering management,* vol. 136, no. 1, pp. 99-107, 2010.