TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 18, Số 3 (2021): 572-591
ISSN: 1859-3100
HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 18, No. 3 (2021): 572-591
Website: *http://journal.hcmue.edu.vn*

Review Article

AN OVERVIEW OF FACIAL ATTRIBUTE LEARNING

Phung Thai Thien Trang[1,2*], Fukuzawa Masayuki[3], Ly Quoc Ngoc[1]

[1]University of Science, Vietnam National University Ho Chi Minh City, Vietnam

[2]Saigon University, Ho Chi Minh City, Vietnam

[3]Kyoto Institute of Technology, Japan

[*]Corresponding author: Phung Thai Thien Trang – Email: trangphung@sgu.edu.vn

Received: November 03, 2020; Revised: March 26, 2021; Accepted: 30-3-2021

ABSTRACT

Facial attributes are useful for developing applications such as face recognition, search, and surveillance. They are therefore important for various facial analysis. Many facial attribute learning algorithms have been developed to automatically detect those key attributes over the years. In this paper, we have surveyed some typical facial attribute learning methods. Five major categories of the state-of-the-art methods are identified: (1) Traditional learning, (2) Deep Single Task Learning, (3) Deep Multitask Learning, (4) Imbalanced Data Solver, and (5) Facial Attribute Ontology. They included from traditional learning algorithm to deep learning, along with methods that assist in solving semantic gaps based on ontology and solving data imbalances. For each algorithm of category, basic theories as well as their strengths, weaknesses, and differences are discussed. We also compared their performance on the standard datasets. Finally, based on characteristics and contribution of methods, we present conclusion and future works to solve facial attributes learning. The survey can help researchers gain a quick overview to build future human face applications as well as further studies.

Keywords: deep learning; facial attribute learning; facial attribute ontology; imbalanced data solver; multi-task learning

1. Introduction

Facial attribute learning (FAL) has become an impressive and potential trend for studies on machine learning (ML) to develop applications concerning human face such as predicting facial attributes, face retrieval at attribute level (Kumar et al., 2011), and surveillance environments (Vaquero et al., 2009). Some of the facial attributes are studied such as race (Fu et al., 2014), gender (Eidinger et al., 2014), age (Chen et al., 2014), facial expressions or emotions (Zhang et al., 2018), carrying (jewelry) (Deng et al., 2015), and physical appearance (Zafeiriou et al., 2015).

Cite this article as: Phung Thai Thien Trang, Fukuzawa Masayuki, & Ly Quoc Ngoc (2021). An overview of facial attribute learning. *Ho Chi Minh City University of Education Journal of Science, 18*(3), 572-591.

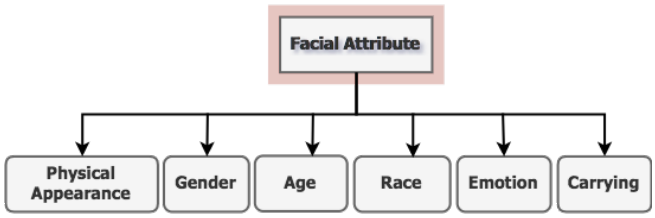
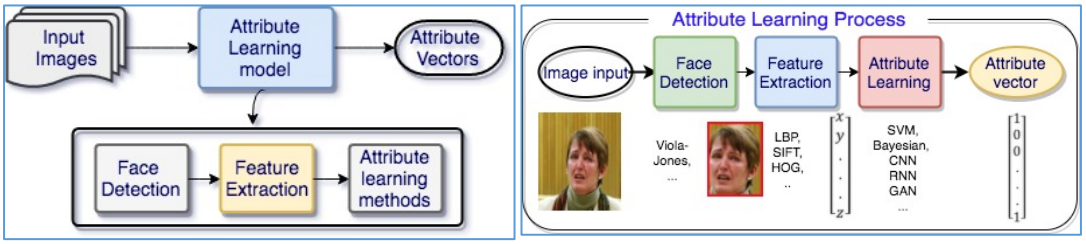


Figure 1. *Facial attribute groups*

The process of attribute learning (AL) involves several main stages such as face detection, feature extraction, and AL methods. Some works used parts of face (such as eyes, nose, or mouth) in an AL model. However, in case of deep learning (DL), some works used the entire face to learn the attributes. This is called an end-to-end deep learning network. The result of AL process is one-hot attribute vector.

The performance of AL is influenced by each stage in the above process. Using the entire face or parts of a face will achieve different performance during the AL stage. Moreover, the feature extraction stage plays a very important role in improving the performance of AL. Finally, the AL method is the most important stage. For example, some studies used parts of face (Zhang et al., 2014), entire face (Jadhav et al., 2016) for AL, hand-crafted features extraction such as SIFT (Lowe, 2004), SURF (Bay et al., 2008), HOG (Dalal, & Triggs, 2005), GICA (Do, & Le, 2009), and deep features (Liu et al., 2016). There are five approaches for building AL methods: (1) Traditional Machine Learning, (2) Deep Single Task Learning, (3) Deep Multi-task learning, (4) Imbalanced Data Solver, and (5) Ontology-based Learning.

Firstly, with the traditional ML approach, SVM is popular and effective algorithm for AL (Kumar et al. 2011). Some works used LDA combined to PCA (Lyons et al., 2000), Bayesian (Everingham et al. 2006), and semi-supervised method (Cherniavsky et al., 2011) for AL.



(a) *(b)*

Figure 2. *Attribute Learning Process. (a) General Attribute Learning Process, (b) Attribute Learning Process in detail*

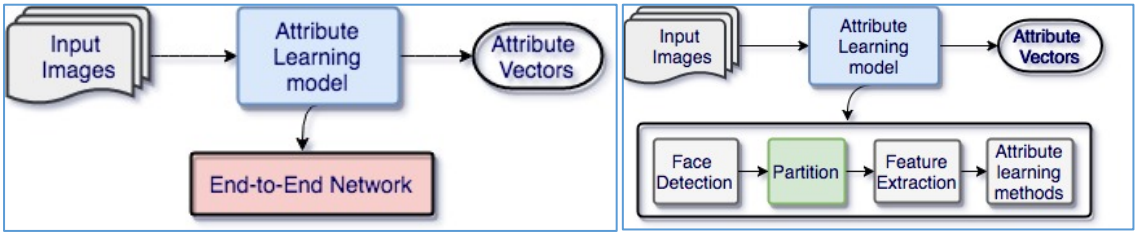


Figure 3*(left).* ***End-to-End deep neural network for Attribute Learning***
Figure 4*(right).* ***A process of attribute learning with part-based***

Secondly, based on the deep single task learning approach, there are many famous works for AL, such as PANDA (Zhang et al., 2014), DeepFace verification (Taigman et al., 2014), LNet+ANet. (Liu et al. 2015), Walk and Learn (Wang et al., 2016), ATNet_GT (Gao et al., 2017), AFFACT (Günther et al., 2017), Off-the-shelf CNN (Zhong et al., 2016), Semantic segmentation network (Kalayeh et al., 2017), General-to-specific learning (Sun, & Yu, 2018), and AttGAN (He et al., 2018).

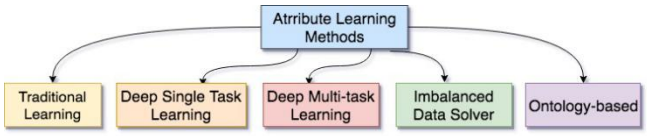


Figure 5. *Attribute learning approaches*

Thirdly, Multitask Learning (MTL) has been proved to be effective in AL, such as MCNN-AUX (Hand et al. 2016), MT-RBM PCA (Ehrlich et al., 2016), DMTL (Han et al., 2017), MOON (Rudd et al., 2016), and FAO (Nguyen et al., 2018).

Fourthly, Imbalanced Data Solver has helped to settle imbalanced data problems in AL. The attributes will be predicted more accurately, the predicted attributes do not tend to favor the majority attribute group. Some works performed Imbalanced Data Solver on fashion image (Ly et al., 2019), on face recognition (P. Wang et al., 2019). In the FAL, typical works are MOON (Rudd et al., 2016), LMLE-KNN (Loy et al., 2017), CRL (Dong et al., 2017), Selective Learning (Hand et al., 2018), CLMLE (Huang et al., 2018), and DCL (Wang et al., 2019).

Finally, using ontology in AL to support semantic of images is also common such as in a study by Nguyen et al. (2018), Ly et al. (2019), Ly et al. (2020).

The rest of this paper is organized as follows: Works of Traditional ML are reviewed in Section **2**. Deep Learning is presented in Section **3**. Deep Multitask Learning is reviewed in Section **4**. Imbalanced Data Solver is discussed in Section **5**. Ontology-based methods are investigated in Section **6**. Datasets in reviewed works are described in Section **7**. The performance of the reviewed works will be discussed in section **8**. Finally, the conclusion is drawn in Section **9**.

2. Traditional methods for attribute learning

In computer vision, attributes are considered as the bridge (intermediate level) between the low-level features and the high-level semantics of the image (Feris et al., 2017) because they support object recognition or object detection by object attributes more easily. For example, instead of searching for zebras, it is easier to find a black and white striped pattern as in a study by Ferrari and Zisserman (2007) or a spotted dog is replaced by a black dot pattern as in several studies (Farhadi et al., 2009; Lampert et al., 2013; Grauman, & Leibe, 2011; Parikh, & Grauman, 2011) Russakovsky and Fei-Fei (2012) discovered the blue lizard with its "green" and "long" attribute.

Some typical algorithms of AL will be then presented based on the traditional learning (i.e. do not use DL) such as SVM, Bayesian, and some other methods such as Hidden Markov and LDA.

2.1. SVM

Before the era of DL, SVM has been the most prominent method for AL. A study by Kumar, Member, Berg, Belhumeur, and Nayar (2011) is a famous study of the human face attributes. The author used a manual method to divide the face image into ten parts, then chose different methods such as SIFT, HoG, Color to extract features of each part. They then used SVM to train facial attributes. Besides, the system allowed searching with keywords like "Asian woman with eyeglasses." With 73 attributes on the LFW dataset with high accuracy, it has created challenges for the next works. Following Kumar, Bor Chun Chen et al. (2014) also used SVM to train age attributes and received high accuracy on the Morph and CACD datasets. Chen et al. (2014) tried to use SVM and local description to identify face attribute vectors for predicting age and gender from names.

2.2. BAYESIAN

Besides SVM, Bayesian is a popular algorithm for training attributes based on probability and statistical methods.

Everingham and Zisserman (2006), with the Bayesian and Adaboost methods, trained attributes such as eye and non-eye, achieved high accuracy on the FERET dataset. D. Chen et al.(2012) combined Bayesian and EM (Expectation–Maximization) for face recognition on the LFW dataset. Demirkus et al. (2015) used Bayesian to train hair and gender attributes. In particular, the experiments have yielded 100% accuracy of the hair attribute on the LFW dataset. Gao et al. (2015) used Bayesian Naive for learning aesthetic attributes. Bayesian and PCA (Chen et al. 2016) combination also showed good results.

The Bayesian method is less popular than the SVM method, but the method has contributed significantly to AL in the statistical learning approach.

2.3. Other methods

Chen et al. (2013) combined many feature extraction methods such as HOG, Color, Gabor, and LBP. Then they used adaboost for voting in training some attributes such as age, gender, and race. B. Chen et al. (2013) proposed two methods "attribute enhanced sparse codeword" and "attribute embedded inverted indexing" for AL on LFW and Pubfig to improve MAP for large-scale face retrieval. Lin et al. (2014) used the Latent Human Topics (LHT) for AL. Using Markov Random Field for AL about age and expression (Liao et al., 2014) was gained a high result. The table below shows some contributions in traditional learning approach.

Table 1. *Traditional methods for attribute learning*

Method Name	Attribute Learning	Datasets	Accuracy (Avg, %)
Face Tracer (Kumar et al., 2011)	SVM,73	LFW, LFWA, CelebA	83.62, 73.9, 81.1
(Demirkus et al., 2015)	Bayesian, Hair, gender	LFW	89.25
LFDA (An et al. 2015)	FDA+LPP,8, Age, Gender Race	LFW, PubFig	MAP-19.3,21.8
(Alorf et al. , 2018)	SVM, Eye/non-eye	LFWA, CelebA,	80.3.,95.6,
(Chen et al. 2013),	Automatic Training, adaboost, 6	Flickr,	76.0
(Chen et al. 2013)	sparse codewords,40	LFW, Pubfig	MAP: 43.5%
(Lin et al. 2014)	Latent Human Topics,	LFWA,6	85.0
(Liao et al. 2014)	MRF Groupwise	LFW, age, emotion	86.73

In the next section, we will present DL methods which is the most important contribution of attribute learning.

3. Deep Single task learning for attribute learning

Since 2012, DL has created a new wave for ML. Especially, Deep learning has proved that it could achieve the higher performance on almost tasks in Computer Vision.

There are some typical deep neuron networks (DNN) such as (1) Convolutional Neural Network (CNN), (2) Recurrent Neural Network (RNN), (3) Generative Adversarial Network (GAN), and (4) Combination of multiple networks. They have their own power in each data domain.

Deep Learning has provided many powerful methods and contributed significantly to AL. We will review some typical DNNs for AL.

3.1. CNN

Deep CNN was born in 1998 from a study by LeCun et al. (1998), but until 2012, AlexNet created a new wave for ML. There are two kinds for deep AL methods: (1) End-to-End CNN and (2) Many CNNs. Each architecture will show its own strong points.

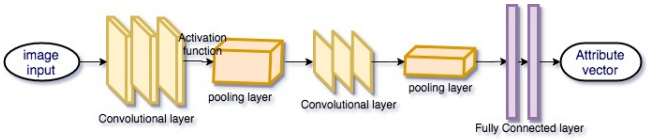


Figure 6. *A typical CNN model for attribute learning*

Zhang et al. (2014) proposed PANDA (Pose Align Networks for Deep Attribution), a Deep CNN to classify attributes based on parts of the face image. The work achieved State-of-the-art performance on CelebA and LFWA. The work by Z. Liu et al. (2015) used many CNN networks (LNet+ANet) in combination with SVM for AL. They reported their experimental results on 40 attributes. The study outperformed the state-of-the-art methods on CelebA and LFWA datasets. To understand pedestrians, Wang et al. (2016) proposed a "Walk

and Learn" model for learning attribute representation by exploring contexts from geographic location and weather conditions. The model using CNN for AL gained the state-of-the-art performance on CelebA and LFWA datasets, and the highest result for lipstick attribute on LFWA dataset. Günther et al. (2017) used ResNet to develop Alignment-Free Facial Attribute Classification Technique (AFFACT) for attribute classification on CelebA dataset. In (Zhong et al., 2016), authors used the CNN Off-The-Shelf and SVM features to train 40 attributes on CelebA and LFWA datasets. In (Jadhav et al., 2016) the authors constructed the deep attributes using VGG-CNN and achieved the better results than (Liu et al., 2015) using ANet + LNet for some attributes such as gender, mustache, chubby.

Using bounding boxes for face detection, authors (He et al., 2017) used VGG-16 for analysis of facial attributes (eyeglasses, smiles, kisses) on the CelebA and LFWA datasets to achieve an average error comparable to the works such as Face Tracer, LNet+ANet (Liu et al., 2015), Walk and Learn (Wang et al., 2016), MOON (Rudd et al., 2016). In (Kalayeh et al., 2017), authors used semantic segmentation networks for predicting facial attributes. They perform training 40 attributes on LFWA and CelebA datasets. Especially, it gained the highest accuracy for gray hair attribute on LFWA.

In segment-based approach, Upal Mahbub et al. (Mahbub et al., 2018) proposed two-steps neural network-based approach to detect facial attributes. (Sun et al., 2018) used many VGG-16s to build general to specific AL. It has achieved the highest accuracy for blonde hair attribute on LFWA. Olivia Wiles (Wiles et al., 2019) suggested Fab-Net, a self-supervised without any label for AL by using encode-decode architecture and Curriculum Strategy. In (Ahmed, & B, 2019), authors suggested an age estimation method by using Bayesian Optimization and CNN on three large-scale datasets: MORPH, FG-NET and FERET. The results showed that using Bayesian Optimization for CNN outperformed the state-of-the-art methods on FG-NET and FERET datasets. In (Dornaika et al., 2020), authors suggested a method for age estimation by using CNN on FG-NET, and MORPH II datasets and gained State-of-the-art performances.

Deep learning with CNN has contributed significantly to improve the performance of AL model. CNN is a strength in the development of AL and most of works have used CNN. However, there are also some studies using RNN, GAN, and some other deep neuron networks for AL.

3.2. RNN, GAN and other deep networks

Recurrent Neural Network (RNN) is a type of Neural Network. Different from feedforward network (such as feedforward CNN), RNN includes loops of hidden layers in the network and tries to remember the information of the previous step for entering the next iteration step (Zhang et al., 2019). Because of this difference, RNN is appropriate model for problems with sequential data in time and spatial such as text, speech and video analysis (Sundararajan et al. 2018). The variants of RNN are Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) which solve the vanishing gradient problem (Zhang et al., 2019).

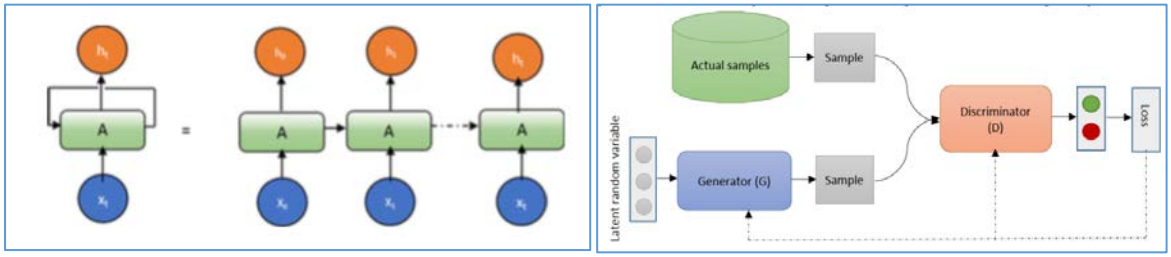


Figure 7*(left). Conceptual diagram for Generative Adversarial Networks (GAN) (Alom et al., 2019a)*

Figure 8 *(right).* ***An unrolled RNNs*** (Alom et al., 2019b)

In AL, RNN has some contributions in aging and emotion prediction. We can divide it into three groups for RNN for AL: (1) Emotion prediction, (2) Aging prediction and (3) other facial attributes learning. In RNN-based emotion learning, Wang Xiaohua et al. (2019) suggested Bi-RNN for facial emotion learning (valence, arousal) by using Deep RNN with two-level attention combining MTL. (Pini et al., 2017) proposed multimodal DL architecture with LSTM and CNN for emotion recognition in video. Using DL for emotion from text and image (Illendula, & Sheth, 2019). Nicolaou et al. (2011) used the bidirectional Long Short-Term Memory neural networks (BLSTM-NNs) and Support Vector Machines for Regression (SVR) for emotion prediction on Sensitive Artificial Listener Database (SAL- DB). Deng et al. (2018) suggested a deep bidirectional Long Short Term Memory (DBLSTM) network for emotion prediction by using LSTMs in layers. In RNN-based aging, Wang et al. (2018) used seven RNNs for aging decentralized faces. Jang et al. (2018) used LSTMs to recognize face attributes such as facial expressions, age, and gender on benchmark datasets: Multi-PIE, CK+, and Adience. Due to strong points of RNN is suitable for sequential data in time and spatial, RNN is effective for some FAL such as emotion and aging.

Difference of CNN and RNN, GAN is a model for generating adversarial samples for robust discriminative learning model. As the name “Generative Adversarial Networks” (GAN), it is invented by Goodfellow in 2014 (Alom et al., 2019a), it can be considered that Discriminator (D) and Generator (G) two players playing the min-max game. In the FAL, GAN devoted a new approach in generating facial attribute data and predicting for prediction facial attributes such as emotion, age, gender and race. In the survey (Zheng et al., 2018), GAN was method of facial attribute manipulation (FAM) group. An emergent work of (He et al., 2018) used GAN to build AttGAN and gained the state-of-the- art on the CelebA dataset for training 40 attributes. Many works used GAN for prediction facial attributes such as emotion, age, gender and race.

In emotion prediction, (Bozorgtabar et al., 2019) proposed Learn to synthesize and synthesize to learn (LSSL) for seven facial expressions classification by using GAN. Min Kyu Lee et al. suggested RcGAN (Lee et al., 2019) for emotion AL on CK+ and MMI.

In age prediction, Penghui Sun et al. Proposed SADAL (Penghui et al., 2019) by using GAN for age estimation on FG-NET and MORPH. (Duong et al. 2017) used GAN for age-invariant face recognition. In (Zhang et al., 2017), authors used GAN with two adversarial networks are imposed on the encoder and generator for Age Progression/Regression. In

(Yang et al., 2018), authors proposed Pyramid Architecture of GANs to learn how to deal with face aging.

In age, gender, and race estimation, (Mirjalili et al., 2020) suggested PrivacyNet by using GAN based Semi-Adversarial Networks (SAN) model for predicting age, gender and race attributes on five datasets such as CelebA, MORPH, MUCT, RaFD and UTK-face. In (Wan et al., 2018), they built a fine-grained, multi-faceted learning to create faces of age, gender and ethnicity. In (Akbir et al., 2019), authors suggested a method for race classification by using GAN and transfer learning and gained outperforming on CelebA dataset. In (Gauthier, 2014), the authors built a conditional GAN to create a toddler face. (Tai et al., 2018) used Conditional CycleGAN for attribute-guided and identity-guided face image generation.

3.3. Combining methods

Combining the learning methods could increase the performance of AL system. Most of works used DL for feature learning and using SVM or combining some deep networks such as CNN, RNN, and GAN for attribute classification. Using SVM for attribute training, (Alorf et al., 2018) used SVM for training attributes and RootSIFT for extracting and representing features. The high results are achieved for the open / close eye classification on the CEW and ZIU datasets, and classification of the presence or absence eyeglasses in high precision on LFWA and CelebA. The work achieved the accuracy rate equal to LNet + ANet (Liu et al., 2015) and better than Tracer (Kumar et al., 2011) and Panda (Zhang et al., 2014). In the topic (Li et al., 2017), they built multi-modal 2D + 3D to recognize facial expressions with deep-fused CNN. Using CNN-based Exemplar-SVM (Jadhav et al., 2016) to identify faces with hair and gender attributes. A combination of LDA, CNN and SVM (Tian et al., 2017) has classified gender. In (Tzirakis et al., 2016), the authors have built an end-to-end network for emotional recognition based on CNN and RNN. Also using CNN and RNN (Kahou et al., 2015) for emotion recognition in video. Huang et al. (Huang et al., 2016) proposed a DL network for classification of imbalances by CNN and KNN. In (Haque et al., 2018) the authors used CNN and LSTM models to identify pain recognition. In (Yang et al., 2018), learning face age progression was performed by combining GAN and CNN. In (Li et al., 2021), authors combined VAE and GAN for facial image editing model that can effectively generate high-quality facial images with diverse specified attributes such as hair color, gender, age etc. A new work for the approach combining traditional learning and DL is in (Liang et al., 2021). The authors used a handcrafted feature Gabor surface feature (GSF) for Patch Attention Layer (PAL) in their CNN to achieve State-of-the-art in facial expression recognition on the CK+, Oulu-CASIA, and JAFFE datasets. Using effective transfering learning to build CNN model for orthognathic surgery on 3D face image gain high results (Lin et al., 2021). Besides general FAL, a survey about face hallucination presented a comprehensive review of DL techniques in face super-resolution (Jiang et al., 2021).

Deep learning is a breakthrough in ML in general and AL in particular. The projects are diverse and plentiful, using traditional methods such as SVM, Bayesian to DL networks CNN, RNN, GAN and have combined the methods together and give high accuracy.

However, there are still challenges with learning facial attributes such as photos partially obscured, inner-group correlation, outer-group correlation, imbalanced data. Therefore, AL need to be studied for further improvement. The table II below will show survey of typical works with accuracy performance.

Table 2. *Deep learning methods for attribute learning*

Method Name	Attribute Learning	Datasets	Accuracy (Avg %)
PANDA	CNN	LFWA, CelebA	81, 85
LNet+Anet	CNN, 40	CelebA, LFWA	87, 84
Walk and Learn	CNN, 40	LFWA, CelebA	86.6, 88.7
(Zhong et al., 2016)	CNN+SVM, 40	CelebA, LFWA	89.80, 85.90
ATNet_GT	CNN, 40	CelebA	90.18
AFFACT	CNN, 40	CelebA	91.01
(Kalayeh et al., 2017)	CNN, 40	LFWA, CelebA	87.1, 91.2
(Sun, & Yu, 2018)	VGG-16, 40	LFWA, CelebA	87.1, 91.6
AttGAN(He et al., 2018)	GAN, Encoder-Decoder, 40	CelebA	90.98
Deep Attribute (Gupta et al. 2017)	CNN, Gender	LFW-S	71.2
FATAUVA-Net(Chang et al. 2017)	Deep, 10	CelebA	91.34
FaceNet2ExpNet (Ding et al. 2017)	Emotion Net: 2 stage, CNNs, 8	TFD, CK+, OULU-CAS	88.9, 98.6, 87.71
FAAN(Chan et al. 2017)	ResNet, gender, ethnicity, hair, eyeglasses	MPIE, CASIA-WebFace	91
(Mahbub et al., 2018)	CNN, 40	CelebA LFWA	90.72, 84.02
AFFAIR(Li et al., 2018)	CNN, 40	MTFL, LFWA, CelebA	86.55,83.01, 79.63
D^2AE(Liu et al., 2018)	encoder-decoder, 40	LFWA, CelebA	83.16, 87.82
FAb-Net (Wiles et al., 2019)	GAN, 8	EmotioNet	76.4
(Bozorgtabar et al., 2019)	GAN, 7	Oulu-CASIA	87.40
(Akbir et al. 2019)	GAN, Race	Celeb-A	91.00
(Xiaohua et al., 2019)	RNN, Valence, Arousal	AffectNet	0.48
RcGAN(Lee et al., 2019)	GAN, emotions	CK+, MMI	97.93,82.86
SADAL (Penghui et al., 2019)	GAN, Age	FG-NET MORPH	(MAE) 3.67,2.75
DLOB(Ahmed, & B, 2019),	CNN, Bayesian, Age	MORPH, FG-NET FERET	MAE: 3.01, 2.88, 1.3
Age Estimation (Dornaika et al., 2020)	CNN, Age	FG-NET MORPH II	MAE:3.05 2.74
PrivacyNet (Mirjalili et al., 2020)	SAN+GAN, gender, age, race	CelebA, MORPH, MUCT, RaFD, UTK-face	ERR: 10%-20%

From the above survey, DL proves that it is the most attractive trend for AL. The top work is in (Sun et al., 2018). With high accuracy, it wins all in any approach from simple to complex AL in human face images.

In section V, we will review AL based on Multi-task learning.

4. Attribute learning based on deep multi-task learning

Multi-Task Learning (MTL) is a learning paradigm in ML (Zhang et al., 2018) for leveraging and helping to improve the performance by decreasing number of parameters in the process learning and to exploit inner-group, outer-group correlations between attributes. It has two stages with (1) shared stage and (2) separated stage. Besides, parallel and distributed techniques can be used in MTL models.

In (Zhang et al., 2018), the survey about MTL separated two main approaches for MTL methods: (1) feature-based approach and (2) parameter-based approach. The feature-based approach included feature transformation and feature selection approach. The parameter-based approach can be classified as: low-rank approach, task clustering approach, task relation learning approach, and decomposition approach. In DL, however, the overview of (Sebastian et al., 2017) divided MTL methods into two kinds: (1) hard-parameter sharing and (2) soft-parameter sharing.

In AL, most of MTL methods have followed the feature-based approach and some are using parameter-sharing approach. MTL has devoted many contributions for AL by reducing parameters in learning. In (Rudd et al., 2016), Rudd et al. built a MOON is the typical example for this approach in AL model, based on combining CNN and objective optimization with imbalanced training data and supporting domain adaptation.

Difference from MOON method (Rudd et al., 2016), without using CNN, Max Ehrlich et al. (Ehrlich et al., 2016) used RBM for building a deep neural network by combining RBM and PCA methods. Their experiment achieved state-of-the-art performance with 40 attributes on the CelebA dataset. Most of works have used feature learning in shared stage. (Hand et al., 2016) proposed MCNN-AUX, using a multi-tasking deep neural network (MCNN) in conjunction with an auxiliary network (AUX) to achieve the state-of-the-art performance on the CelebA and LFWA with high accuracy on CelebA dataset. Modifying AlexNet, (Han et al., 2017) built DMTL (deep multi-tasking learning) based on CNN to estimate heterogeneous facial attributes. The network is not only success on 40 facial attributes but also obtaining high results on age, race, and gender attributes on the CelebA, MORPH II, and FotW datasets. Trying to optimizing weighted loss function, in (Wang et al., 2017) and (He, et al., 2017) proposed a framework for facial attribute prediction based adaptively weighted loss on Mega dataset by using MTL on CNN. (Wang et al., 2017) also used MTCNN for AL on imbalanced data.

Using many networks for building AL system, in (Cao et al., 2018), Jiajiong Cao et al. suggested a method for attribute representation, PS-MCNN (Partially Shared Multi-task Convolutional Neural Network), including four Task Specific Networks (TSNets) and one

Shared Network (SNet) are connected by Partially Shared (PS) for boost the performance on CelebA and LFWA datasets. Using cascade net to build method for predicting face attributes, (Li et al., 2017) proposed a deep multi-task cascaded network by using CNN and stochastic gradient descent (SGD). Following curriculum learning, (Fanhe et al., 2019) suggested KL-MTL combining curriculum learning, MTL and Inception V3 for AL on CelebA. With 74 facial attributes learning, (Luigi Celona et al., 2018) built the method based multi-task learning and convolutional neural network to estimate attributes. The net included a gating mechanism effectively for gaining high results. Only focusing smile and gender attributes, (Fan et al., 2019) suggested a face attributes prediction method using multi-task learning architecture to boost the performance and gained the state-of-the-art methods on FotW and LFWA datasets. Focus on age, gender, and race attributes, (Hsieh et al., 2017) used MTL and CNN for building a method for human attributes learning, with training on CASIA-Webface dataset and testing on LFW. In (Wang et al., 2021), based on adversarial learning, authors built a Multi-task network with two modules (1) recognizer R and (2) discriminator D for face analyses and gain State-of-the-art on LFWA and celebA with 40 attributes. Some typical AL methods based on MTL are showed as follows:

Table 3. *Deep Multitask learning methods for attribute learning*

Method Name	Attribute Learning	Datasets	Accuracy Avg (%)
MCNN-AUX	CNN, 40	CelebA, LFWA	91.29
MT-RBM PCA	RBM, PCA, 40	CelebA	86.98
MOON (Rudd et al., 2016)	CNN	CelebA	90.94
DMTL (Han et al., 2017)	CNN, age, gender, race	**CelebA, LFWA**	**92.60**
(Hsieh et al., 2017)	CNN, Gender, age	LFW (gender, age), Adience	93.66, 93.48, 86.7
(He, & Wang, et al., 2017)	CNN	CelebA, LFW	91.80, 73.07
Ya Li (Li et al., 2017)	CNN,40	LFWA, CelebA	87.8, 90.2
(Taherkhani et al. 2018)	CNN, MTL, 40	CelebA, MegaFace	88.98, 78.82
KT-MTL(Fanhe et al., 2019)	Inception v3, 40	CelebA	92.19
(Fan et al., 2019)	CNN, Smile, gender	LFWA	Smile:91.13 gender:92.49
(Celona et al., 2018)	ResNet-101, LSTM, 74, Age, gender	Adience, LFWA, CelebA	90.7 ± 1.7
(Wang et al., 2021)	CNN, GAN, 40	**LFWA, CelebA**	**91, 96**

In the table above, using deep multi task learning for AL approach,(Wang et al., 2021) is the best method for AL on CelebA dataset. But DMTL (Han et al., 2017) is the best method on both datasets LFWA and CelebA with 40 facial attributes.

5. Imbalanced data solver for attribute learning

In survey on imbalance data solver (He et al., 2009) divided it into 4 groups: (1) sampling methods, (2) cost-sensitive learning methods, (3) kernel-based learning methods, and (4) active learning methods. But by difference of naming, in survey on DL with class imbalance (Johnson et al., 2019), Justin M. Johnson presented two method groups: (1) data level methods and (2) algorithm level methods, each group with traditional algorithm and deep algorithm.

Typical works about imbalance data learning with fashion image (Ly, Do, & Nguyen, 2019), for face recognition (Wang et al., 2019), and for facial attribute learning MOON(Rudd et al., 2016), LMLE-KNN (Loy et al. 2017), CRL (Dong et al. 2017), Selective Learning (Hand et al., 2018), CLMLE (Huang et al., 2018), DCL (Wang et al., 2019).

In (Rudd et al., 2016), Rudd et al. built a MOON attribute learning model based on CNN with imbalanced training data based on weighted loss function on each attribute learning, and achieved high accuracy on the CelebA dataset. In (Loy et al., 2017), C. Huang et al. suggested LMLE-KNN with learning deep approach by using LMLE for feature representation and KNN for classification on imbalanced data. Using triplet-header hinge loss in learning is to margin intra-class and inter-class. In (Dong et al., 2017), Qi Dong et al. built CRL (Class Rectification Loss) to improve minority class learning by using deep features with updated batch-balance in imbalanced multi-label attributes. CRL gained high accuracy on the CelebA data set in cost-sensitive learning approach. In (Hand et al., 2018) Emily M. Hand et al. proposed Selective Learning for imbalance learning by using AttCNN. With sampling the data in order to balance the positive and negative labels, it achieved high result on CelebA, LFWA in facial attribute prediction with multi-labels. In (Huang et al., 2018), Chen Huang et al. suggested CLMLE (Cluster-based Large Margin Local Embedding), a deep imbalanced learning for face recognition and attribute prediction, by using LMLE and KNN algorithm. In (Wang et al., 2019), Yiru Wang et al. proposed DCL (Dynamic Curriculum Learning), a unified framework for human attribute analysis. With two-level curriculum schedulers: sampling scheduler and loss scheduler, it gained state-of-the-art performance on CelebA dataset. In (Hupont, & Fernández, 2019), Isabelle Hupont et al. suggested DemogPairs for race and gender recognition on imbalanced dataset with a new validation set with 10.8K facial images and 58.3M identity verification pairs for Asian, Black and White females and males. The experiments gained the state-of-the-art on deep face recognition models (such as SphereFace, FaceNet and ResNet50).

Beside accuracy, recall, precision, and F1-score, Matthews correlation coefficient (MCC) is a measurement for binary classification proposed by Matthews (Matthews, 1975). In data imbalance context, MCC is strong tool for imbalanced learning. Especially, with applying AL on Fashion dataset, in (Ly et al., 2019), the authors have proposed the imbalanced data solver based on MCC to address imbalanced data that has contributed

effectively to increasing the quality of linking object ontology to raw data without adjusting network architecture and data augmentation. Using MCC for FAL (Xiao et al., 2019). The table IV showed some typical works for AL methods deal to imbalanced data.

Table 4. *Attribute learning methods deal to imbalanced data*

Method Name	Attribute Learning	Dataset	Accuracy (%)
MOON	VGG-16	CelebA, **LFW**	90.94, **84.73 ± 1.99**
(Loy et al., 2017)	CNN, KNN	CelebA	84.00
(Dong et al., 2017)	CNN	CelebA	85-86
SL (Hand et al., 2018)	AttCNN	LFWA, **CelebA**	73.03, **90.97**
(Huang et al., 2018)	LMLE, KNN	CelebA	88.78
(Wang et al., 2019)	ResNet50	CelebA	89.05
(Hupont et al. 2019)	CNN, Race, gender	CWF, VGGFace VGGFace2	94.00

In the table above, MOON (Rudd et al., 2016) is the best method on LFW dataset and SL (Hand et al., 2018) is the best method on CelebA dataset.

6. Using ontology for attribute learning

Ontology is a tool to represent knowledge for sharing and reusing (Gruber, 1993). Especially, it is useful to support semantic of images in building applications about image retrieval and image interpretation (Ly et al., 2019) (Hudelot, 2008; Maillot, 2005; Mezaris, Kompatsiaris, & Strintzis, 2004). In AL, ontology supports to build the hierarchical semantic tree for facial attributes (Contreras et al. 2010), (Nguyen, Ly, & Phung, 2018), (Bashar et al. 2007). In (Contreras et al., 2010), Garcıa-Rojas et al. built an ontology for facial representation for emotional face expression profiles.

Combining deep imbalanced data learning and ontology, Hung M Nguyen et al. (Nguyen et al., 2018) proposed a facial attribute ontology (FAO) for AL and achieved high accuracy with 40 facial attributes on CelebA and LFWA datasets.

Focusing aesthetics, (Xu et al., 2018) suggested Bio-FAO (Bio-inspired Facial Aesthetic Ontology) and combined with CNN to predict facial aesthetic. Using fuzzy reasoning, (Contreras et al., 2010) proposed an emotion recognition method based on ontology. Some popular works about AL based on ontology presented as follows:

Table 5. *Attribute learning based on ontology*

Method Name	AttributeLearning	Dataset	Accuracy
FAO (Nguyen et al., 2018)	CNN, ontology, 40	CelebA, LFWA	85.68%
Bio-FAO(Xu et al., 2018)	CNN, aesthetics	JAFFE, FaceWarehouse	72.1%
(Contreras et al., 2010)	fuzzy reasoning,	MMI (FDPs,5 Emotion)	70-90%

The next section will review some popular datasets used in the AL methods.

7. Datasets

Most of the work on AL has been experimented on LFW and CelebA datasets. In addition, there are Fgnet, Morph and CACD datasets for some works on human face detection.

Table 6. *Datasets for attribute learning*

No.	Name of datasets	No. of images	No. of Candidates
1	LFW/LFWA	13,233	5,749
2	CelebA	202,599	10,177
3	Pubfig	58,797	200
4	FGNet	1,002	82
5	MORPH	55,134	13,618
6	CACD	163,446	2,000

8. Results

The best method in traditional learning is the work of Kumar (Kumar et al., 2011) with SVM, Hog, and Color feature to learn 73 attributes with high performance. However, it also shows a disadvantage of using hand-crafted features, which is not good in automatic learning. The best with DL, MTL, imbalance data solver and ontology on CelebA dataset with 40 facial attributes will be presented in the table below:

Table 7. *Top attribute learning methods*

Method Name	Attribute Learning	Datasets	Accuracy Avg %	group
Face Tracer	SVM, 73 attributes	LFW, LFWA, CelebA	83.62, 73.9, 81.1	Traditional learning
(Sun, & Yu, 2018)	VGG-16	CelebA	91.6	Deep learning
(Kalayeh et al., 2017)	CNN	**LFWA,** CelebA	**87.1,** 91.2	Deep learning
DMTL(Han et al., 2017)	CNN	CelebA, LFWA	92.60	MTL
(Wang et al., 2021)	CNN, GAN	**LFWA, CelebA**	**91, 96**	MTL
MOON (Rudd et al., 2016)	VGG-16	CelebA, **LFW**	90.94,**84.73** % ± 1.99	Imbalanced data
SL (Hand et al., 2018)	AttCNN	LFWA, **CelebA**	73.03,**90.97**	Imbalanced data
FAO (Nguyen et al., 2018)	CNN, Ontology	CelebA, LFWA	85.68	Ontology

The table VII shows that deep multitask learning is the best method for attribute learning on LFWA and CelebA datasets.

It is the determination that MTL is the best effective and efficient method for FAL on LFWA and CelebA datasets with 40 attributes.

9. Summary and conclusion

In this article, we have tried to examine typical works of AL with the approaches: traditional learning, deep single task learning, deep multi-task learning, imbalance data

solver, and Ontology-based. The results of the works have revealed table VII, SVM is the best method among traditional learning approaches, and CNN is the most important architecture at other approaches.

With the purpose is to increase the accuracy, there are some ideas for future trends of AL. First, it is to try to improve learning methods from traditional learning to DL such as inferring analogous attributes by one-shot method (Chen, & Grauman, 2014), using training and test augmentation techniques (Günther et al., 2017), using multi-modal (the 2-D texture and 3-D shape) and Attributed Relational Graph (ARG) for recognize faces with expressions (Mahoor et al., 2008), investigating challenges such as illumination, pose normalization, and automatic attribute category for efficient attribute prediction (Wang et al., 2017), solving challenges such as lighting conditions or direction, eye gaze, hair features, facial expression, mouth opening, sharpness and group portraits (Davis, 2014), using ontology and GAN on massive datasets (Nguyen et al., 2018). Second, it is to investigate relationships of attributes such as discovering automatically relative attributes (Parikh, & Grauman, 2011), exploiting facial attributes dependencies for robust gender recognition and analyzing the best features for multi-view and multi-attribute facial demography estimation (Ryu et al., 2017), discovering the relationship among different attributes (Zheng et al., 2018), being interested in exploring the kind of attributes that are useful for improving face recognition (Jadhav et al., 2016).

❖ ***Conflict of Interest:*** *Authors have no conflict of interest to declare.*

❖ ***Acknowledgement:*** *This research is funded by Viet Nam National University Ho Chi Minh City (VNUHCM) under grant no. B2018-18-01.*

REFERENCES

Ahmed, M., & B, S. V. (2019). *Optimization for Facial Age Estimation* (Vol. 2). Springer International Publishing. https://doi.org/10.1007/978-3-030-27272-2

Akbir, K., & Mahmoud, M. (2019). Considering race a problem of transfer learning. *Proceedings - 2019 IEEE Winter conf WACVW 2019*, 100-106.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Asari, V. K. (2019a). A state-of-the-art survey on deep learning theory and architectures. *Electronics.*

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S.,… Asari, V. K. (2019b). A state-of-the-art survey on deep learning theory and architectures. *Electronics.*

Alorf, A., & Abbott, A. L. (2018). In defense of low-level structural features and SVMs for facial attribute classification: Application to detection of eye state, Mouth State, and eyeglasses in the wild. *IEEE International Joint Conf on Biometrics, IJCB 2017*, *2018-Janua*, 599-607.

An, L., Zou, C., Zhang, L., & Denney, B. (2015). Scalable attribute-driven face image retrieval. *Neurocomputing*, *172*, 215–224. https://doi.org/10.1016/j.neucom.2014.09.098

B, Y. L., Tai, Y., & Tang, C. (2018). *Attribute-Guided Face Generation Using Conditional CycleGAN* (Vol. 3951). Springer International Publishing. https://doi.org/10.1007/11744023

Bashar, R., Kang, S. K., Dawadi, P. R., & Rhee, P. K. (2007). A Context-Aware Statistical Ontology Approach for Adaptive Face Recognition. Convergence of Bioscience and Information Technologies, Jeju, Korea (South), 2007, 698-703. doi: 10.1109/FBIT.2007.112

Bozorgtabar, B., Rad, M. S., Ekenel, H. K., & Thiran, J.-P. (2019). Learn to synthesize and synthesize to learn. *Computer Vision and Image Understanding*, *185*(June 2018), 1-11.

Cao, J., Li, Y., & Zhang, Z. (2018). Partially Shared Multi-task Convolutional Neural Network with Local Constraint for Face Attribute Learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4290-4299.

Celona, L., Bianco, S., & Schettini, R. (2018). Fine-grained face annotation using deep Multi-Task CNN. *Sensors (Switzerland)*, *18*(8). https://doi.org/10.3390/s18082666

Chan, J.-S., Hsu, G.-S. (Jison), Shie, H.-C., & Chen, Y.-X. (2017). Face recognition by facial attribute assisted network. *ICIP*, 3825-3829.

Chang, W.-Y., Hsu, S.-H., & Chien, J.-H. (2017). FATAUVA-Net: An Integrated Deep Learning Framework for Facial Attribute Recognition, Action Unit Detection, and Valence-Arousal Estimation. *2017 IEEE Conference on CVPRW*, 1963-1971.

Chen, B., Chen, Y., Kuo, Y., Hsu, W. H., & Member, S. (2013). Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords. *IEEE Transactions on Multimedia*, *15*(5), 1163-1173.

Chen, D., Cao, X., Wang, L., Wen, F., & Sun, J. (2012). Bayesian face revisited: A joint formulation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7574 LNCS*(PART 3), 566-579.

Chen, D., Cao, X., Wipf, D., Wen, F., & Sun, J. (2016). An Efficient Joint Formulation for Bayesian Face Verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(1), 32-46, https://doi.org/10.1109/TPAMI.2016.2533383

Chen, H., Gallagher, A. C., & Girod, B. (2014). The Hidden Sides of Names—Face Modeling with First Name Attributes. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, *36*(9), 1860-1873. https://doi.org/10.1109/TPAMI.2014.2302443

Chen, Y. Y., Hsu, W. H., & Liao, H. Y. M. (2013). Automatic training image acquisition and effective feature selection from community-contributed photos for facial attribute detection. *IEEE Trans. Multimed.*, *15*(6), 1388-1399. https://doi.org/10.1109/TMM.2013.2250492

Contreras, R., Starostenko, O., Alarcon-Aquino, V., & Flores-Pulido, L. (2010). Facial feature model for emotion recognition using fuzzy reasoning. *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, *6256 LNCS*, 11-21.

Demirkus, M., Precup, D., Clark, J., & Arbel, T. (2015). Hierarchical Spatio-Temporal Probabilistic Graphical Model with Multiple Feature Fusion for Estimating Binary Facial Attribute Classes in Real-World Face Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8828*(FEBRUARY 2014), 1-22.

Ding, H., Zhou, S. K., & Chellappa, R. (2017). FaceNet2ExpNet : Regularizing a Deep Face Recognition Net for Expression Recognition, 118-126. https://doi.org/10.1109/FG.2017.23

Do, T. T., & Le, T. H. (2009). Facial feature extraction using geometric feature and independent component analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5465 LNAI*, 231-241.

Dong, Q., Gong, S., & Zhu, X. (2017). Class Rectification Hard Mining for Imbalanced Deep Learning. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 1869-1878. https://doi.org/10.1109/ICCV.2017.205

Dornaika, F., Bekhouche, S. E., & Arganda-Carreras, I. (2020). Robust regression with deep CNNs for facial age estimation: An empirical study. *Expert Syst. Appl.*, *141*.

Duong, C. N., Quach, K. G., Luu, K., Le, T. H. N., & Savvides, M. (2017). Temporal Non-volume Preserving Approach to Facial Age-Progression and Age-Invariant Face Recognition. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 3755-3763. https://doi.org/10.1109/ICCV.2017.403

Ehrlich, M., Shields, T. J., Almaev, T., & Amer, M. R. (2016). Facial Attributes Classification Using Multi-task Representation Learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 752-760.

Everingham, M., & Zisserman, A. (2006). Regression and classification approaches to eye localization in face images. *7th International Conference on Automatic Face and Gesture Recognition FGR06, pages*, 441-448. https://doi.org/10.1109/FGR.2006.90

Fan, D., Kim, H., Kim, J., Liu, Y., & Huang, Q. (2019). Multi-task learning using task dependencies for face attributes prediction. *Appl. Sci.*, *9*(12).

Fanhe, X., Guo, J., Huang, Z., Qiu, W., & Zhang, Y. (2019). Multi-task learning with knowledge transfer for facial attribute classification. *Proc. IEEE Int. Conf. Ind. Technol.*, *2019-Febru*, 877-882. https://doi.org/10.1109/ICIT.2019.8755180

Gao, Z., & Wang, S. (2015). Multiple Aesthetic Attribute Assessment by Exploiting Relations Among Aesthetic Attributes, 575-578.

Gauthier, J. (2014). Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter Semester 2014*.

Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human - Computer Studies*, *43*(5-6), 907-928.

Günther, M., Rozsa, A., & Boult, T. E. (2017). AFFACT - Alignment Free Facial Attribute Classification Technique. *Fg*, 90-99.

Gupta, N., Gupta, A., Joshi, V., Subramaniam, L. V., & Mehta, S. (2017). Deep Attribute Driven Image Similarity Learning Using Limited Data. *Proceedings - 2017 IEEE International Symposium on Multimedia, ISM 2017*, *2017-Janua*, 146-153.

Han, H., Jain, A. K., Shan, S., & Chen, X. (2017). Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach. *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit.*, *8828*(c), 1-14. https://doi.org/10.1109/TPAMI.2017.2738004

Hand, E. M., Castillo, C., & Chellappa, R. (2018). Doing the best we can with what we have: Multi-label balancing with selective learning for attribute prediction. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 6878-6885.

Hand, E. M., & Chellappa, R. (2016). Attributes for Improved Attributes: A Multi-Task Network for Attribute Classification, 8057–8058. Retrieved from http://arxiv.org/abs/1604.07360

Haque, M. A., Bautista, R. B., Noroozi, F., Kulkarni, K., Laursen, C. B., Irani, R.,… Moeslund, T. B. (2018). Deep Multimodal Pain Recognition : A Database and Comparison of Spatio-Temporal Visual Modalities. *IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 250-257. https://doi.org/10.1109/FG.2018.00044

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *Ieee Transactions On Knowledge And Data Engineering*, *21*(9), 1263-1284.

He, K., Fu, Y., & Xue, X. (2017). A Jointly Learned Deep Architecture for Facial Attribute Analysis and Face Detection in the Wild. Retrieved from http://arxiv.org/abs/1707.08705

He, K., Wang, Z., Fu, Y., Feng, R., Jiang, Y. G., & Xue, X. (2017). Adaptively weighted multi-task deep network for person atribute classification. *2017 ACM Multimed. Conf.*, 1636-1644.

He, Z., Zuo, W., Member, S., Kan, M., Shan, S., Member, S., & Chen, X. (2018). AttGAN : Facial Attribute Editing by Only Changing What You Want, 1-16.

Hsieh, H.-L., Hsu, W., & Chen, Y.-Y. (2017). Multi-task learning for face identification and attribute estimation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2981-2985.

Hsieh, H.-L., Hsu, W., & Chen, Y.-Y. (2017). *Multi-task learning for face identification and attribute estimation*, *1*, 2981-2985.

Huang, C., Li, Y., Loy, C. C., & Tang, X. (2016). Learning Deep Representation for Imbalanced Classification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6.

Huang, C., Li, Y., Loy, C. C., & Tang, X. (2018). Deep Imbalanced Learning for Face Recognition and Attribute Prediction, 1-14. Retrieved from http://arxiv.org/abs/1806.00194

Hudelot, C. (2008). Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms, 280.

Hupont, I., & Fernández, C. (2019). DemogPairs: Quantifying the impact of demographic imbalance in deep face recognition. *Proc. - 14th IEEE Int. Conf. FG 2019*.

Illendula, A., & Sheth, A. (2019). Multimodal emotion classification. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, *2*, 439-449.

Jadhav, A., Namboodiri, V. P., & Venkatesh, K. S. (2016). Deep Attributes for One-Shot Face Recognition. *ECCV Workshops*, (3), 516-523. https://doi.org/10.1007/978-3-319-49409-8_44

Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep Learning-based Face Super-resolution: A Survey. Retrieved from http://arxiv.org/abs/2101.03749

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0192-5

Kahou, S. E., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015). Recurrent neural networks for emotion recognition in video. *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, 467-474.

Kalayeh, M. M., Gong, B., & Shah, M. (2017). Improving Facial Attribute Prediction using Semantic Segmentation, 6942-6950. https://doi.org/10.1109/CVPR.2017.450

Kumar, N., Member, S., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2011). Describable Visual Attributes for Face Verification and Image Search, 1-17.

Lee, M. K., Choi, D. Y., & Song, B. C. (2019). Facial expression recognition via relation-based conditional generative adversarial network. *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction*, 35-39.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86, 2278-2324.

Li, D., Zhang, M., Zhang, L., Chen, W., & Feng, G. (2021). A novel attribute-based generation architecture for facial image editing. *Multimedia Tools and Applications*, *80*(4), 4881-4902.

Li, H., Sun, J., & Xu, Z. (2017). Multimodal 2D + 3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network, *9210*(c), 1-16.

Li, J., Zhao, F., Feng, J., Roy, S., Yan, S., & Sim, T. (2018). Landmark free face attribute prediction. *IEEE Transactions on Image Processing*, *27*(9), 4651-4662.

Li, Y., Wang, Q., Nie, L., & Cheng, H. (2017). Face Attributes Recognition via Deep Multi-Task Cascade. *Proc. 2017 Int. Conf. Data Mining, Commun. Inf. Technol. - DMCIT '17*, 5-9.

Liang, X., Xu, L., Liu, J., Liu, Z., Cheng, G., Xu, J., & Liu, L. (2021). Patch attention layer of embedding handcrafted features in CNN for facial expression recognition. *Sensors*

Liao, S., Shen, D., & Chung, A. C. S. (2014). A Markov Random Field Groupwise Registration Framework for Face Recognition, *36*(4).

Lin, C.-H., Chen, Y.-Y., Chen, B.-C., Hou, Y.-L., & Hsu, W. (2014). Facial Attribute Space Compression by Latent Human Topic Discovery. *Proc. ACM Int. Conf. Multimed. - MM '14*,

Lin, H. H., Chiang, W. C., Yang, C. T., Cheng, C. T., Zhang, T., & Lo, L. J. (2021). On construction of transfer learning for facial symmetry assessment before and after orthognathic surgery. *Computer Methods and Programs in Biomedicine*, *200*.

Liu, Y., Wei, F., Shao, J., Sheng, L., Yan, J., & Wang, X. (2018). Exploring Disentangled Feature Representation Beyond Face Identification, 2080-2089.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of the IEEE International Conf on Computer Vision*, *2015 Inter*, 3730-3738.

Loy, C. C., Luo, P., & Huang, C. (2017). *Deep Learning Face Attributes for Detection and Alignment*. https://doi.org/10.1007/978-3-319-50077-5

Ly, N. Q., Do, T. K., & Nguyen, B. X. (2019). Large-scale coarse-to-fine object retrieval ontology

and deep local multitask learning. *Computational Intelligence and Neuroscience*, *2019*.

Ly, N. Q., Cao, H. N.M., Nguyen, T. T (2020). Person Re-Identification System at Semantic Level based on Pedestrian Attributes Ontology. *International Journal of Advanced Computer Science and Applications* (IJACSA), *11*(2), 2020.

Mahbub, U., Sarkar, S., & Chellappa, R. (2018). Segment-based Methods for Facial Attribute Detection from Partial Faces, 1-13. Retrieved from http://arxiv.org/abs/1801.03546

Maillot, N. (2005). *Ontology Based Object Learning and Recognition*.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA - Protein Structure*, *405*(2), 442-451.

Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2004). An ontology approach to object-based image retrieval, II-511-514. https://doi.org/10.1109/icip.2003.1246729

Mirjalili, V., Raschka, S., & Ross, A. (2020). PrivacyNet: Semi-Adversarial Networks for Multi-attribute Face Privacy, 1-3. Retrieved from http://arxiv.org/abs/2001.00561

Nguyen, H. M., Ly, N. Q., & Phung, T. T. T. (2018). Large-Scale Face Image Retrieval System at attribute level based on Facial Attribute Ontology and Deep Neuron Network.

Penghui, S., Hao, L., Xin, W., Zhenhua, Y., & Wu, S. (2019). Similarity-aware deep adversarial learning for facial age estimation. *Proc. - IEEE Int. Conf. Multimed. Expo*, *2019-July*.

Pini, S., Ahmed, O. Ben, Cornia, M., Baraldi, L., Cucchiara, R., & Huet, B. (2017). Modeling Multimodal Cues in a Deep Learning-based Framework for Emotion Recognition in the Wild. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*.

Rudd, E. M., Günther, M., & Boult, T. E. (2016). MOON: A mixed objective optimization network for the recognition of facial attributes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9909*

Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks, (May). Retrieved from http://arxiv.org/abs/1706.05098

Sun, Y., & Yu, J. (2018). General-to-specific learning for facial attribute classification in the wild. *J. Vis. Commun. Image Represent.*, *56*, 83-91. https://doi.org/10.1016/j.jvcir.2018.09.003

Sundararajan, K., & Woodard, D. L. (2018). Deep learning for biometrics: A survey. ACM Computing Surveys, *51*(3). https://doi.org/10.1145/3190618

Taherkhani, F., Nasrabadi, N. M., & Dawson, J. (2018). A Deep Face Identification Network Enhanced by Facial Attributes Prediction, 666-673.

Tian, Q., Arbel, T., & Clark, J. J. (2017). Deep LDA-Pruned Nets for Efficient Facial Gender Classification. https://doi.org/10.1109/CVPRW.2017.78

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). End-to-End Multimodal Emotion Recognition using Deep Neural Networks, *14*(8), 1-9.

Wan, L., Wan, J., Jin, Y., Tan, Z., & Li, S. Z. (2018). Fine-grained multi-attribute adversarial learning for face generation of age, gender and ethnicity. *Proceedings - 2018 International Conference on Biometrics, ICB 2018*, 98-103. https://doi.org/10.1109/ICB2018.2018.00025

Wang, J., Cheng, Y., & Feris, R. S. (2016). Walk and Learn: Facial Attribute Representation Learning from Egocentric Video and Contextual Data.

Wang, P., Su, F., & Zhao, Z. (2017). Joint Multi-Feature Fusion and Attribute Relationships for Facial Attribute Prediction, 3-6.

Wang, P., Su, F., Zhao, Z., Guo, Y., Zhao, Y., & Zhuang, B. (2019). Deep class-skewed learning for face recognition. *Neurocomputing*, *363*, 35-45.

Wang, S., Yin, S., Hao, L., & Liang, G. (2021). Multi-task face analyses through adversarial learning. *Pattern Recognition*, *114*, 107837. https://doi.org/10.1016/j.patcog.2021.107837

Wang, Y., Gan, W., Yang, J., Wu, W., & Yan, J. (2019). Dynamic Curriculum Learning for Imbalanced Data Classification, (2), 5017-5026. http://arxiv.org/abs/1901.06783

Wang, Z., He, K., & Fu, Y. (2017). Multi-task Deep Neural Network for Joint Face Recognition and Facial Attribute Prediction. *ICMR'17*, 365-374.

Wiles, O., Sophia Koepke, A., & Zisserman, A. (2019). Self-supervised learning of a facial attribute embedding from video. *British Machine Vision Conference 2018, BMVC 2018*.

Xiao, T., Tsai, Y.-H., Sohn, K., Chandraker, M., & Yang, M.-H. (2019). Adversarial Learning of Privacy-Preserving and Task-Oriented Representations. http://arxiv.org/abs/1911.10143

Xiaohua, W., Muzi, P., Lijuan, P., Min, H., Chunhua, J., & Fuji, R. (2019). Two-level attention with two-stage multi-task learning for facial emotion recognition. *J. Vis. Commun. Image Represent.*, *62*, 217-225. https://doi.org/10.1016/j.jvcir.2019.05.009

Xu, M., Chen, F., Li, L., Shen, C., Lv, P., Zhou, B., & Ji, R. (2018). Bio-Inspired Deep Attribute Learning Towards Facial Aesthetic Prediction. *IEEE Transactions on Affective Computing*.

Yang, H., Huang, D., Wang, Y., & Jain, A. K. (2018). Learning Face Age Progression : A Pyramid Architecture of GANs. *CVPR*, 31-39.

Zhang, N., Paluri, M., Ranzato, M. A., Darrell, T., Bourdev, L., & Berkeley, U. C. (2014). PANDA : Pose Aligned Networks for Deep Attribute Modeling.

Zhang, Y., & Yang, Q. (2018). A Survey on Multi-Task Learning, 1-20.

Zhang, Z., Song, Y., & Qi, H. (2017). Age Progression / Regression by Conditional Adversarial Autoencoder, 5810-5818.

Zheng, X., Guo, Y., Huang, H., Li, Y., & He, R. (2018). A Survey to Deep Facial Attribute Analysis. Retrieved from http://arxiv.org/abs/1812.10265

Zhong, Y., Sullivan, J., & Li, H. (2016). Leveraging mid-level deep representations for predicting face attributes in the wild. *Proceedings - ICIP*, *2016-Augus*, 3239-3243.

TỔNG QUAN VỀ PHƯƠNG PHÁP HỌC THUỘC TÍNH MẶT NGƯỜI

Phùng Thái Thiên Trang[1,2*], Fukuzawa Masayuki[3], Lý Quốc Ngọc[1,2]

[1]Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam

[2]Trường Đại học Sài Gòn, Việt Nam

[3]Học viện Kỹ thuật Kyoto, Nhật Bản

[*]Tác giả liên hệ: Phùng Thái Thiên Trang – Email: trangphung@sgu.edu.vn

Ngày nhận bài: 03-11-2020; ngày nhận bài sửa: 26-3-2021; ngày duyệt đăng: 30-03-2021

TÓM TẮT

Thuộc tính mặt người là thông tin hữu ích cho việc xây dựng các ứng dụng như nhận dạng, tìm kiếm và giám sát khuôn mặt người. Do đó, chúng rất quan trọng đối với các nhiệm vụ phân tích khuôn mặt khác nhau. Nhiều thuật toán học thuộc tính khuôn mặt người đã và đang được phát triển để tự động phát hiện các thuộc tính trong nhiều năm qua. Trong bài báo này, chúng tôi khảo sát một số phương pháp điển hình về học thuộc tính khuôn mặt người. Chúng tôi chia ra năm loại chính của các phương pháp: (1) Học truyền thống, (2) Học sâu đơn nhiệm, (3) Học sâu đa nhiệm, (4) Giải quyết vấn đề mất cân bằng dữ liệu và (5) Thuộc tính khuôn mặt dựa vào phả hệ tri thức. Các phương pháp bao gồm từ học truyền thống đến học sâu, cùng với các phương pháp hỗ trợ giải quyết bài toán lỗ hổng ngữ nghĩa dựa trên phả hệ tri thức và giải quyết sự mất cân bằng dữ liệu. Đối với mỗi phương pháp trong mỗi loại, chúng tôi thảo luận về các lí thuyết cơ bản cũng như điểm mạnh, điểm yếu và sự khác biệt của chúng. Chúng tôi cũng so sánh hiệu suất của chúng trên bộ dữ liệu tiêu chuẩn. Cuối cùng, dựa trên đặc điểm và đóng góp của các phương pháp, chúng tôi đưa ra kết luận và hướng nghiên cứu trong tương lai để giải quyết vấn đề học thuộc tính khuôn mặt. bài khảo sát này sẽ giúp các nhà nghiên cứu có góc nhìn tổng quan nhanh để xây dựng các ứng dụng khuôn mặt người trong tương lai cũng như các nghiên cứu mới.

Từ khóa: học sâu; học thuộc tính mặt người; học đa nhiệm; sự mất cân bằng dữ liệu; phả hệ tri thức