A Teacher Classroom Dress Assessment Method Based on a New Assessment Dataset

Ming Fang, Qi Liu, Yunpeng Zhou, Xinning Du, Qiwen Liang, Shuhua Liu *
Northeast Normal University, China
liush129@nenu.edu.cn

Abstract

Proper attire is a professional requirement for teachers and teachers' dress influence students' perceptions of teacher quality. Therefore, evaluating teacher attire can better regulate and improve the teacher's dress. However, the lack of a dataset on teacher attire hinders the development of this field. For this purpose, this paper constructs a Teachers' Classroom Dress Assessment (TCDA) dataset. To our knowledge, it is the first dataset focused on teacher attire. This dataset is entirely from the classroom environment, covering 25 teacher attributes, with a total of 11879 teacher dress samples and sufficient positive and negative examples. Therefore, the TCDA dataset is a challenging evaluation dataset with characteristics such as data diversity. In order to verify the effectiveness of the dataset, this paper systematically explores a new perspective on human attribute information and proposes for the first time a Teachers' Dress Assessment Method (TDAM), aiming to use predicted teacher attributes to scoring the overall attire of each teacher, thereby promoting the development of the teacher's classroom teaching field. The experimental results demonstrate the rationality of the TCDA dataset and the effectiveness of the TDAM method. The dataset and code can be openly obtained at https://github.com/MingZier/TCDA-dataset.

1 Introduction

The teacher's attire has a significant impact on students' cognition and teaching effectiveness. Appropriate dress not only helps to shape students' positive impression of teachers, but also has a positive effect on promoting a good learning atmosphere [Mosca and Buzza, 2013] . Therefore, teachers should dress appropriately in class and conform to the professional norms of teaching. However, in practice, teachers usually have a certain freedom of dress [Marici *et al.*, 2023] . Therefore, in order to regulate the dress of teachers, it is a necessary to check and evaluate them randomly. However, the current evaluation methods of teachers' attire still are mainly manual, which is inefficient and labor-intensive. Therefore, it is urgent to evaluate teachers' attire automatically.

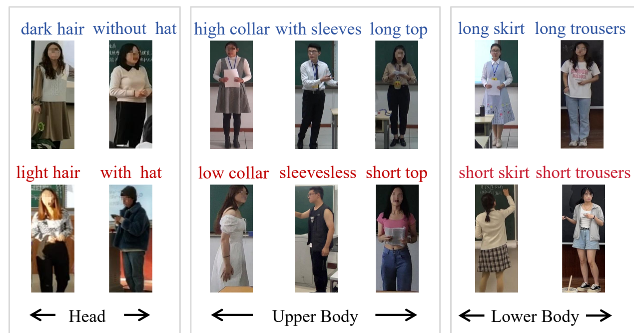


Figure 1: Positive and negative examples in the TCDA dataset. The blue font represents positive attributes, while the red font represents negative attributes.

Pedestrian attribute recognition aims to mine pedestrian attribute information from input images, such as gender, hairstyle, clothing, etc. In recent years, a large number of research have integrated human attribute recognition into other fields, such as pedestrian detection [Tian *et al.*, 2015] , person re-identification [Li *et al.*, 2014] , and social activity analysis [Danelljan *et al.*, 2014] . In this paper, we apply human attribute information to the education for the first time, and propose a method for evaluating teachers' attire. Specifically, we collect pictures of teachers in the classroom and construct a dataset of Teachers' Classroom Dress Assessment (TCDA), which has a total of 11879 samples, each of which is annotated with 25 attribute information. It is worth noting that, on the one hand, as shown in Figure 1, this dataset clearly divides positive and negative samples based on the relevant requirements for teacher attire. On the other hand, as shown in Figure 2, scores are generated based on the positive and negative attributes of the teacher's attire, which to some extent saves labor.

Based on this dataset, we propose a Teachers' Dress Assessment Method (TDAM). This method takes teacher images as input and extracts attribute features through a feature extraction network. Secondly, all the features are fed to the classifier for classification prediction of each attribute,

*Corresponding author

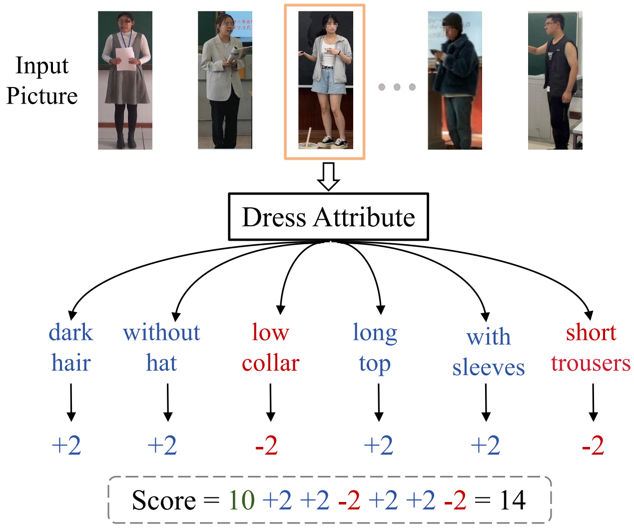


Figure 2: Example of teacher attrie score generation. Specifically, we set the basic score to 10 points, plus 2 points for positive attributes and minus 2 points for negative attributes. Finally, the dress score for the orange framed teacher image is 14 points.

and finally, based on the predicted attributes, a teacher dress score is generated for each teacher image. This method has achieved automated evaluation of teacher attire for the first time, providing a baseline model and bringing significant progress to the field of education.

The main contributions of this paper are as follows:

(1) This study constructs a dataset of Teachers' Classroom Dress Assessment (TCDA) with a total of 11879 samples and each sample is annotated with 25 attributes. The background environment of the dataset is from classrooms and it is the first dataset focused on teacher attire, providing a new benchmark dataset for teacher attire evaluation. Therefore, this dataset can boost the research of teacher's attire evaluation.

(2) This paper systematically explores a new world of human attribute information and proposes a method for evaluating teacher attire. This method is based on predictive attributes and can automatically evaluate teacher attire performance, providing the first baseline method in this field.

(3) A large number of experiments are conducted on the TCDA dataset. The experimental results have demonstrated the rationality of the dataset and the effectiveness of TDAM method.

2 Related Work

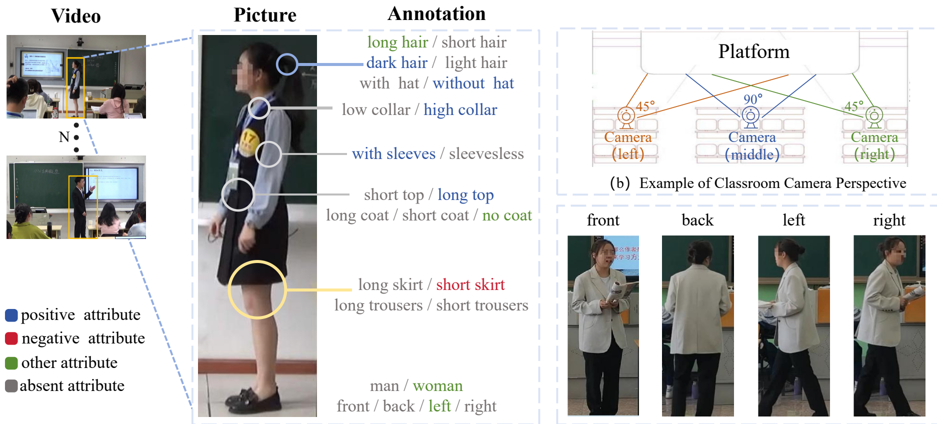
This study adopts attribute recognition for evaluating the teacher's attire. Therefore, we review the related work about teacher's attire evaluation and the development and application of the attribute recognition.

(1) Evaluation of Teachers' Classroom attire. Lightstone et al. [Lightstone *et al.*, 2011] noted that formally dressed teachers were more likely to gain student trust than informal ones. Slabbert's [Slabbert, 2019] research has demonstrated that teachers' dress influence students' perceptions of teacher quality in academic and workplace settings. Kashem et al. [Kashem, 2019] experimentally demonstrated that teacher attire has a significant impact on students' learning and students' attitude in classroom learning. However, current teacher attire evaluation rely mainly on traditional manual methods, such as group evaluation, questionnaires, and self-reflection. In the field of computer vision, the problem of teacher's dress evaluation has not been fully explored. Therefore, this study constructs a dataset for the evaluation of teachers' classroom attributes for the first time, aiming to provide a benchmark for teachers' attire evaluation. At the same time, based on this dataset, we propose a method to prove the validity of this dataset. For the first time, this method realizes the automatic evaluation of teachers' attire.

(2) Pedestrian attribute recognition. Pedestrian attribute recognition is a popular area of research in computer vision, which aims to mine the attributes of a target person in a given person image. Earlier methods [Sudowe *et al.*, 2015; Li *et al.*, 2015; Abdulnabi *et al.*, 2015] took the entire image as input and tried to learn a global attribute representation. However, these methods ignore the fine-grained information. Subsequently, Diba et al. [Diba *et al.*, 2016] proposed DeepCAMP to chunk an image into some blocks and then learn the attribute feature of each image block. Li et al. [Li *et al.*, 2018a] proposed a PGDM algorithm, which used the human pose guidance network for the first time to obtain the key points of the human posture, and then extracted local regional features based on these key points. Moghadam et al. [Moghaddam *et al.*, 2021] mined semantic and spatial information by jointly human semantic parsing and pedestrian attribute recognition. Attention mechanisms [Liu *et al.*, 2017; Sarfraz *et al.*, 2017; Tan *et al.*, 2019] are also widely adopted in pedestrian attribute recognition to locate attribute-related areas and learn to discriminate feature representations. Liu et al. [Liu *et al.*, 2017] proposed HydraPlus-Net with a multi-view attention module to extract pixel-level and semantic-level features, which is conducive to locating fine-grained attributes. Sarfraz et al. [Sarfraz *et al.*, 2017] added a view predictor to the attribute recognition network to estimate the weight of the view. Tan et al. [Tan *et al.*, 2019] employed a multi-task-like approach to simultaneously learn various attentional mechanisms, i.e., parsing attention, labeled attention, and spatial attention, to explore relevant and complementary information.

Despite the significant improvement in recognition performance, the above methods fail to model the potential relationships between attributes. Wang et al. [Wang *et al.*, 2017] proposed a JRL model for exploring attribute context and relevance. Li et al. [Li *et al.*, 2019] considered the complex relationships between attributes and different regions, and proposed a graph reasoning network to jointly model the spatial and semantic relationships of region-region, attribute-attribute, and region-attribute. Tan et al. [Tan *et al.*, 2020] proposed the Joint Learning of Attribute and Contextual relations to solve the task of pedestrian attribute recognition. It consists of two graph modules, called the Attribute Relationship Module and the Contextual Relationship Module, which are used to discover and capture attributes and context relationships.

(3) Person Re-identification. In the past few years, hu-



(a) TCDA Dataset Annotation Information

(c) Examples of Different Viewpoints

Figure 3: An overview of the TCDA dataset. (a) shows the annotation information of the TCDA dataset. (b) shows the viewpoint position of the camera in the classroom during the recording process. (c) shows examples of different perspectives.

man attribute information has been widely used in pedestrian re-identification tasks to improve the detection accuracy of this task. Schhumann et al. [Schumann and Stiefelhagen, 2017] trained an attribute classifier and applied it to the person re-identification model. Wang et al. [Wang *et al.*, 2018] introduced a transferable Joint Attribute-Identity Deep Learning for simultaneously learning an attribute-semantic and identity-discriminative feature representation space to improve the performance of re-identification. Rather than directly using attributes to predict scores, Lin et al. [Lin *et al.*, 2019] proposed a re-weight mechanism of correlations between attributes. Zhao et al. [Zhao *et al.*, 2019] proposed an attribute-driven method for feature disentangling and frame re-weighting. Zhang et al. [Zhang *et al.*, 2019] proposed a multi-branch model, based on a part-based attribute-aware network, to take advantage of the person's reID and attribute information. Considering that incorrect attribute annotation will affect the performance of re- identification, Zhang et al. [Zhang *et al.*, 2020] proposed a novel Attribute Attentional Block, which can be integrated into any backbone network or framework. Chai et al. [Chai *et al.*, 2022] proposed a novel network architecture Attribute Salience Assisted Network to assist in pedestrian re-identification tasks in video. Yin et al. [Yin *et al.*, 2020] first analyzed the effect of attribute information on re-identification. On this basis, they proposed a Recognition Network (IRN) and an Attribute Recognition Network (ARN). Among them, ARN can accurately calculate the attribute similarity between pedestrians to promote the identification of IRN. Chen et al. [Chen *et al.*, 2021] proposed a post-hoc method called Attribute-guided Metric Distillation (AMD). In AMD, it is possible to generate quantitative contributions to attributes and visualize the most discriminative attribute attention maps. Wang et al. [Wang *et al.*, 2023] proposed a novel attribute-guided Transformer (AiT), which explicitly utilizes pedestrian attributes as a semantic prior knowledge for discriminant representation learning. Zhang et al. [Zhang *et al.*, 2023] combined human attributes and designed a cyclic heterogeneous graph convolutional network, which effectively fused cross-modal pedestrian information through intra-graph and inter-graph interactions, resulting in a robust pedestrian representation.

Summarily, the human attribute information has been widely used in various research fields, and remarkable progress has been made in recent years. Education issues are getting more and more attention, and the application of deep learning technology in the field of education is becoming increasingly important. However, to date, no scholar has applied human attribute information to the education fields. Therefore, this paper introduces human attribute information into the education for the first time, aiming to study the teachers' dress code by analyzing the attribute information of teachers, so as to promote the development of the field of education.

3 Dataset

In order to improve the overall professional image of the teachers, this article studies how to evaluate teacher's dress codes and thereby constructs a Teacher Classroom Dress Assessment (TCDA) dataset for the first time. The overall overview of this dataset is shown in Figure 3. To our knowledge, this is the first attempt to apply human attribute information to the field of education. Next, we will provide a detailed introduction to the construction process of TCDA.

3.1 Attribute Label Generation

The existing public pedestrian attribute datasets [Liu *et al.*, 2017; Deng *et al.*, 2014; Li *et al.*, 2018b] come from some common scenarios, such as outdoor streets and indoor shopping centers. However, for specific tasks such as applying attribute information to evaluate teacher dress, the aforementioned dataset in general scenarios is not sufficient to meet requirements. Therefore, we specifically collect videos from various classroom scenarios and define 25 annotation lables to meet the teacher's dress evaluation, in order to improve the overall image of teachers. The attribute annotation information is shown in Figure 3 (a).

Specifically, according to the requirements for teachers' dress in [Wu and Cai, 2019; Gao, 2013] , we define a series of dress attributes aimed at more comprehensively capturing the characteristics of teacher dress. Table 1 shows the attribute categories and positive and negative sample distributions of the TCDA dataset. Considering the limitations of the scene, the teacher's body is often partly obscured by the lectern or the teacher interacts with students and so on. Therefore, as shown in Figure 3 (c), we select teacher's viewpoint information from four directions, namely facing forward (F), facing backward (B), facing left (L), and facing right (R). Finally, we carefully capture images of each teacher from these four perspectives and treat them as independent samples. In this way, each teacher with a suit includes 25 teacher attributes including 4 viewpoints

3.2 Video Collection

When evaluating the teacher's dress, it is essential to keep the balance of positive and negative samples. However, due to privacy issues, it is difficult to obtain a sufficient negative samples of teacher's dress from online or offline channels. In view of this, this paper takes different strategies to collect positive and negative samples of the TCDA dataset to ensure the category balance.

The positive samples of the TCDA dataset come from videos in various classroom scenarios. To ensure the diversity of the dataset, we carefully select teacher lecture videos from different disciplines, classrooms, and diverse teaching environments, including but not limited to classroom teaching, teaching assessment and teaching skills competitions at different levels. A total of 2354 video files are collected.

For the negative samples, we invite 50 normal students to record videos of the simulated teacher's dress. The record locations are chosen in the classrooms with ample lighting and traditional classroom scenes. In order to enhance the robustness of the dataset, we select shooting angles as much as possible when we record the classroom teaching videos, namely front the classroom, behind the classroom, 45° behind the left of the classroom, and 45° behind the right of the classroom,as shown in Figure 3 (b). By this way, we collected 1295 negative sample videos, providing a foundation for balancing the dataset.

Finally, we have invited 6 master students majoring in education to manually capture images of teachers from all directions throughout the video according to the specific environment of the classroom and the teacher's activities. Specifically, the six master students are divided into two groups. One group is responsible for capturing screenshots of the teacher's images of various scenes, and another group is responsible for verifying these captured images, removing unqualified images such as severe occlusion and overexposure, so as to ensure that each sample meets the quality specifications. In the end, 11879 samples are obtained with a resolution of 246×506. Table 1 lists the details of the TCDA dataset.

It is worth noting that this paper fully considers ethical issues and uses RetinaFace [Deng *et al.*, 2019] for encoding the teachers' faces in video clips. At the same time, the coded images will be manually verified in order to ensure that on the one hand, the RetinaFace algorithm accurately protects the privacy information of all teachers, on the other hand, it does not obscure the teacher's attributes. Furthermore, for privacy reasons, all participants for recording negative samples have signed the statement of information consent.

Class	Description
Number of positive samples	6679
Number of negative samples	5200
Video resolution	246×506
Viewpoints	facing front (F), facing back (B), facing left (L) and facing right (R)
Positive attire	dark hair, without hat, high collar, with sleeves, long top, long shirt, long trousers
Negative attribute	light hair, with hat, low collar, sleevesless, short top, short shirt, short trousers
Other attribute	long hair, short hait, man, woman, long coat, short coat, no coat

Table 1: Details of the TCDA dataset.

Dataset Split	#Samples	#Positive samples	#Negative samples	Proportion
Training set	9505	5345	4160	80%
Validation set	1187	667	520	10%
Test set	1187	667	520	10%
All	11879	6679	5200	100%

Table 2: Details of the training, validation and test sets.

3.3 Dataset Split

This study randomly divides the entire dataset into training, validation, and test sets in an 8:1:1 ratio based on commonly dataset partitioning rules. As shown in Table 2.

3.4 Dataset Characteristics

Face to face real classroom. In the traditional face-to-face education environment, classroom teaching has always been considered a core component of the school education system. As organizers, guides, and knowledge imparter of teaching, teachers' appearance and dress style also have a significant impact on the teaching atmosphere. However, traditional attribute recognition mainly focuses on generic attribute, which limits its applicability in teacher attire. To compensate for this shortcoming, the dataset focuses on identifying specific attributes of teacher attire. In addition, unlike traditional general attribute datasets, this dataset focuses on collecting real teacher teaching images in classroom environments to more accurately evaluate the teacher's dressing style.

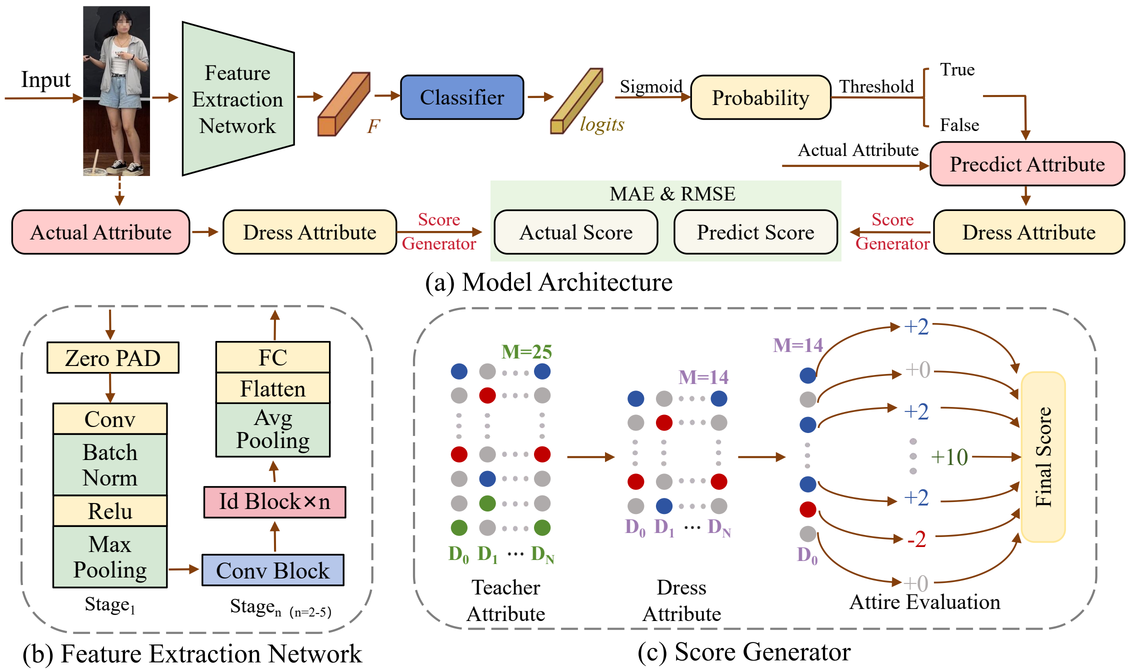


Figure 4: Overview of TDAM. (a) is the overall architecture, and Actual Score is the ground truth of the attributes of each teacher's attire. (b) shows the feature learning network ResNet50. (c) shows the score generator, where M represents the number of attributes, N represents the number of samples.

Multiple educational scenario. This dataset contains real teacher teaching videos from multiple scenarios, aiming to provide a rich variety of teaching situations and teacher attire samples. The raw data was collected from multiple sources, including videos from the 6th National Teacher Competition, daily teaching videos from multiple middle schools, and special classroom environments and attributes recorded by normal students with volunteer teaching experience. The dataset design takes into account the diversity of scenarios and the characteristics of teachers at different levels. The National Teacher Competition provides teaching scenarios for high-level teachers, the daily teaching videos in middle schools reflect conventional teaching situations, and the simulated teaching of normal students provides samples for special environments and attributes. This comprehensive data collection method helps to provide more comprehensive and realistic educational scenarios, providing a more challenging dataset for the training and evaluation of teacher dress attribute recognition algorithms.

Multiple disciplinary and perspective. In order to eliminate the influence of different subject teacher attire styles on teacher evaluation, the TCDA dataset includes teaching videos of teachers in 9 high school subjects: Chinese, Mathematics, English, Politics, History, Geography, Physics, Chemistry, and Biology. In addition, considering the limitations of the scene, such as the fact that the teacher's body is often obstructed by the lectern, this dataset selected four directions as the teacher's viewpoint information. Therefore, our dataset not only has disciplinary richness and diverse perspectives, but also can provide more comprehensive teaching information.

4 Methodology

Most of the previous work related to human attributes involved classifying and identifying attributes, without scoring human clothing, which limited its applicability in the field of teacher teaching evaluation. Therefore, in order to evaluate teacher attire, based on the TCDA dataset, this paper proposes a Teachers' Dress Assessment Method (TDAM), and its overall architecture is shown in Figure 4 (a).

Firstly, the model takes teacher images with attribute labels as input to the entire network, utilizes the feature extraction network to extract the teacher attribute features of each image. Then, TDAM uses a feature classifier to classify and predict each attribute, and finally generates a teacher dress score for each teacher image based on the predicted attributes. Below is a detailed description of the method.

4.1 Feature Extraction Network

Using teacher images as input for the feature extraction network aims to obtain all attire information for each teacher. Specifically, for a given dataset $\mathrm{D} = \{(\mathrm{X_i}; \mathrm{y_i}), \mathrm{i} = 1, 2, 3, \ldots, \mathrm{N}\}$, $\mathrm{X_i}$ represents the $i - th$ teacher image, and N represents the number of samples. Using the feature learning network to predict multiple attributes $\mathrm{y_i} \in \{0, 1\}^{\mathrm{M}}$ of teacher images to obtain corresponding attribute features, where 0 and 1 indicate the absence and presence of corresponding attributes in the teacher image, and M represents the number of attributes.It is worth noting that this paper uses ResNet50 as the backbone network to extract image features, and its network architecture is shown in Figure 4 (b).

4.2 Feature Classifier

Next, use a feature classifier to classify and predict each attribute feature obtained in Section 4.1. Inspired by [Li *et al.*, 2018b; Guo *et al.*, 2019; Tang *et al.*, 2019], this article formulates teacher attribute prediction as a multi-label classification task and employs multiple binary classifiers with sigmoid functions [Li *et al.*, 2015]. In order to handle the imbalanced label distribution in different attribute categories, we adopt the Binary cross-entropy loss function to optimize the entire framework, which is defined as follows:

$$\mathrm{Loss} = \frac{1}{\mathrm{N}}\Sigma_{\mathrm{i}}^{\mathrm{N}}\Sigma_{\mathrm{j}}^{\mathrm{M}}\omega_{\mathrm{j}}(\mathrm{y_{i,j}}\log(\mathrm{p_{i,j}}) + (1-\mathrm{y_{i,j}})\log(1-\mathrm{p_{i,j}})) \tag{1}$$

$$\omega_j = \begin{cases} e^{1-r_j}, & y_{i,j} = 1 \\ e^{r_j}, & y_{i,j} = 0 \end{cases} \tag{2}$$

Among them, $\mathrm{p_{i,j}} = \sigma(\mathrm{z_{i,j}})$ in formula 1 is the predicted probability of the classifier's output $z_{\mathrm{i,j}}$, while $\sigma(\mathrm{z_{i,j}}) = 1/(1+\mathrm{e}^{-\mathrm{x}})$ is the sigmoid function. In formula 2, r_j represents the positive sample ratio of the $j-th$ attribute in the training set.

4.3 Score Generator

Finally, after attribute classification, a score generator is used to generate teacher attire scores for each teacher image, aiming to automate the evaluation of teacher attire. As shown in Figure 4 (c), we set the basic score to 10 points, and when positive and negative attributes are identified, corresponding bonus and minus points are given. Specifically, On the one hand, given the image sample $\{\mathrm{X_i}, \mathrm{i} \in \mathrm{N}\}$, the actual score S_{act}^{i} is obtained based on $\mathrm{y_i} \in \{0,1\}^{\mathrm{M}}$. On the other hand, the output of the classifier is used to calculate the prediction probability of each attribute using the sigmoid function. The calculation formula is as follows:

$$p_{i,j} = P_r(Y = \mathrm{y_{i,j}}|\mathrm{X_i}) = \frac{1}{1+e^{-w_j^T\mathrm{X_i}}} \tag{3}$$

Where $p_{i,j}$ is the probability of the $j-th$ attribute of the $i-th$ image, and $\mathbf{w}_j$ is the weight of the $j-th$ attribute. In this paper, we set predefined confidence thresholds τ. When $p_{i,j}$ is greater than τ, $p_{i,j}$ set the predicted attribute label l_{pre} to $True$, when $p_{i,j}$ is less than τ, the predicted attribute label l_{pre} is set to $False$, and the calculation is as follows:

$$l_{pre} = \begin{cases} True, & p_{i,j} \geq \tau \\ False, & p_{i,j} < \tau \end{cases} \tag{4}$$

Next, obtain the prediction score S_{pre}^{i} for the $i-th$ image based on the predicted attribute label l_{pre} and the groundtruth y_i. Finally, the effectiveness of score generation is evaluated through Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

5 Experiment

5.1 Implementation Details

The TDAM proposed in this paper adopts the PyTorch framework and carries on end-to-end training. The hardware platform is NVIDIA GeForce RTX A6000, and the operating system is Ubuntu 20.04.4. Specifically, this method is trained using Stochastic Gradient Descent algorithm with a momentum of 0.9 and a weight decay of 0.0005. The optimizer is Adam. The initial learning rate is 0.0001, and the Platform learning rate scheduler is used with reduction factor of 0.1. Batch size set to 64. Total epoch number of training is 100.

5.2 Evaluation Metrics

This article uses Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as evaluation indicators. Among them, MAE represents the average absolute error between the predicted value and the true value. RMSE represents the root mean square error between the predicted and true values, used to measure the deviation between the predicted and true values, and this indicator can better reflect the performance of the assessment model. The smaller the values of MAE and RMSE, the better the result. The calculation formula is as follows:

$$\mathrm{MAE(x,y)} = \frac{1}{\mathrm{n}}\sum\nolimits_{\mathrm{i}=1}^{n}|\mathrm{x}_i - \mathrm{y}_i| \tag{5}$$

$$\mathrm{RMSE(x,y)} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i}-\mathrm{y_i})^2} \tag{6}$$

Where n denotes the sample size, x denotes the true value and y denotes the predicted value.

5.3 Experimental Results

Performance of baselines. This paper evaluates the performance of the baseline model on the TCDA dataset using teacher images with attribute labels as inputs to TDAM. Table 3 shows the performance of seven baseline models on the TCDA dataset. Among them:

ResNet [He *et al.*, 2016] is a deep neural network architecture. This model solves the problems of vanishing and exploding gradients by introducing residual connections, which enables the network to learn feature representations more deeply. ResNet18, ResNet34, ResNet50, and ResNet101 are ResNet for layers 18, 34, 50, and 101, respectively.

TResNet [Ridnik *et al.*, 2021] is an improved ResNet architecture that enhances feature extraction capabilities by introducing new modules and optimizing residual connections. TResNet-M is the medium TresNet.

Vision Transformer (ViT) is a visual model based on Transformer. Among them, Vision Transformer-small (Vit-s) is the smaller scale Vision Transformer model, and Vision Transformer-base (ViT-b) is the basic version of the Vision Transformer model.

It can be intuitively seen from Table 3 and Figure 5 that: (1)The performance of the seven baseline models is relatively stable, which to some extent proves the rationality and stability of the TCDA dataset. (2) By increasing the depth of

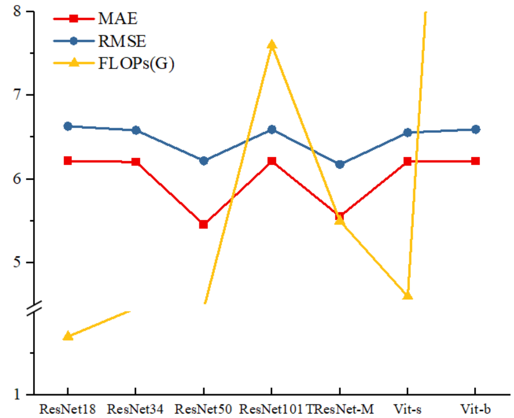


Figure 5: Line graph of computational complexity and performance of seven baseline models.

Dataset	Backbone	MAE	RMSE	FLOPs(G)
TCDA	ResNet18	6.2172	6.6320	**1.7**
	ResNet34	6.2054	6.5838	3.6
	ResNet50	**5.4579**	6.2199	3.8
	ResNet101	6.2121	6.5925	7.6
	TResNet-M	5.5556	**6.1759**	5.5
	Vit-s	6.2098	6.5560	4.6
	Vit-b	6.2138	6.5943	17.6

Table 3: Performance comparison of baseline model on TCDA dataset.

ResNet to 101 layers, both MAE and RMSE began to increase. The reason behind this may be that the increase in parameter count in the feature extraction network led to overfitting, similar to Vit-s and Vit-b. (3)ResNet50 and TResNet-M achieve the best performance under MAE and RMSE metrics, respectively, indicating that these two baseline models can capture more teacher attribute information. Due to the comparable performance and lower computational complexity achieved by ResNet50 compared to TResNet-M, we use ResNet50 as the feature extraction network for TDAM method.

Performance of different classifiers. Table 4 shows the performance comparison of Cosine and Linear classifiers on the TCDA dataset. Among them, Linear classifier typically use linear functions to classify input data, and the Cosine classifier classifies by calculating the cosine similarity between the input data and each category. It can be seen that both in MAE and RMSE metrics, the performance of the Linear classifier is better than that of the Cosine classifier. Therefore, this paper uses a Linear classifier to classify attribute features.

Performance of different thresholds. This paper uses a linear binary classifier to classify and predict teacher attributes. In the experiment, it is found that when the threshold τ is set to below 0.5, the model experiences oscillation. This may be because in this situation, the model tends to classify samples as positive categories, which increases the false positive rate, meaning that some negative samples are misclassified as positive samples. This can lead to unstable performance of the model on training data, which may result in performance fluctuations between different epochs. Table 5 shows when τ above 0.5, we compare the performance of TDAM with different thresholds on the TCDA dataset. It can be seen that when $\tau = 0.5$, the performance is best. Therefore, this article sets the threshold to 0.5 for predicting the classification of teacher attributes.

Method	Classifier	MAE	RMSE
TDAM	Cosine	6.1751	6.9737
	Linear	**5.4579**	**6.2199**

Table 4: Performance comparison of different classifiers on TCDA dataset.

Method	Threshold	MAE	RMSE
TDAM	$\tau = 0.5$	**5.4579**	**6.2199**
	$\tau = 0.6$	6.2121	6.5966
	$\tau = 0.7$	6.2290	6.6155
	$\tau = 0.8$	6.2239	6.6152
	$\tau = 0.9$	6.2424	6.6378

Table 5: Performance comparison of different thresholds τ on TCDA dataset.

In summary, after a large number of experiments, it has been proven that the baseline method for assessing teacher attire proposed in this paper uses ResNet50 as a feature learning network, a binary linear classifier to classify and predict teacher attributes, and when the threshold is set to 0.5, the best performance is achieved. This method aims to score teacher attire more accurately and efficiently.

6 Conclusion

In this article, for the first time, a dataset called "Teachers' Classroom Dress Assessment" (TCDA) is constructed, which includes a collection of teachers dressed properly and improperly images in the classroom, with a total of 11879 samples. This dataset is the first teacher dress evaluation dataset. In addition, TCDA dataset covers a total of 25 attributes including 4 different viewpoints, making it a challenging assessment dataset with data diversity. To demonstrate the validity of this dataset, we propose a baseline method TDAM for assessing teacher attire, aimed at evaluating attire based on teacher attribute information. This is also the first automated method for attire evaluation. We hope that our work will help teachers dress appropriately and enhance the overall professionalism of the teaching staff. At the same time, we also hope to attract more scholars' attention to the application of deep learning in the field of education, and promote the development of this field.

Acknowledgments

This work is supported by the project of National Natural Science Foundation of China under the Grant 62277009, the project of Jilin Provincial Science and Technology Department under the Grant 20220201140GX.

Contribution Statement

The contributions of Ming Fang and Qi Liu to this paper were equal.

References

[Abdulnabi *et al.*, 2015] Abrar H Abdulnabi, Gang Wang, Jiwen Lu, and Kui Jia. Multi-task cnn model for attribute prediction. *IEEE Transactions on Multimedia*, 17(11):1949–1959, 2015.

[Chai *et al.*, 2022] Tianrui Chai, Zhiyuan Chen, Annan Li, Jiaxin Chen, Xinyu Mei, and Yunhong Wang. Video person re-identification using attribute-enhanced features. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7951–7966, 2022.

[Chen *et al.*, 2021] Xiaodong Chen, Xinchen Liu, Wu Liu, Xiao-Ping Zhang, Yongdong Zhang, and Tao Mei. Explainable person re-identification with attribute-guided metric distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11813–11822, 2021.

[Danelljan *et al.*, 2014] Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost Van de Weijer. Adaptive color attributes for real-time visual tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1090–1097, 2014.

[Deng *et al.*, 2014] Yubin Deng, Ping Luo, Chen Change Loy, and Xiaoou Tang. Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 789–792, 2014.

[Deng *et al.*, 2019] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*, 2019.

[Diba *et al.*, 2016] Ali Diba, Ali Mohammad Pazandeh, Hamed Pirsiavash, and Luc Van Gool. Deepcamp: Deep convolutional action & attribute mid-level patterns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3557–3565, 2016.

[Gao, 2013] Yueting Gao. A brief discussion on the dressing of university teachers. *Science and Technology Information*, (14):91, 2013.

[Guo *et al.*, 2019] Hao Guo, Kang Zheng, Xiaochuan Fan, Hongkai Yu, and Song Wang. Visual attention consistency under image transforms for multi-label image classification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 729–739, 2019.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[Kashem, 2019] Mohammad Abul Kashem. The effect of teachers' dress on students' attitude and students' learning: Higher education view. *Education Research International*, 2019:1–7, 2019.

[Li *et al.*, 2014] Annan Li, Luoqi Liu, Kang Wang, Si Liu, and Shuicheng Yan. Clothing attributes assisted person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):869–878, 2014.

[Li *et al.*, 2015] Dangwei Li, Xiaotang Chen, and Kaiqi Huang. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 111–115. IEEE, 2015.

[Li *et al.*, 2018a] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Pose guided deep model for pedestrian attribute recognition in surveillance scenarios. In *2018 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2018.

[Li *et al.*, 2018b] Dangwei Li, Zhang Zhang, Xiaotang Chen, and Kaiqi Huang. A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios. *IEEE transactions on image processing*, 28(4):1575–1590, 2018.

[Li *et al.*, 2019] Qiaozhe Li, Xin Zhao, Ran He, and Kaiqi Huang. Visual-semantic graph reasoning for pedestrian attribute recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8634–8641, 2019.

[Lightstone *et al.*, 2011] Karen Lightstone, Rob Francis, and Lucie Kocum. University faculty style of dress and students' perception of instructor credibility. *International Journal of Business and Social Science*, 2(15), 2011.

[Lin *et al.*, 2019] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. *Pattern recognition*, 95:151–161, 2019.

[Liu *et al.*, 2017] Xihui Liu, Haiyu Zhao, Maoqing Tian, Lu Sheng, Jing Shao, Shuai Yi, Junjie Yan, and Xiaogang Wang. Hydraplus-net: Attentive deep features for pedestrian analysis. In *Proceedings of the IEEE international conference on computer vision*, pages 350–359, 2017.

[Marici *et al.*, 2023] Marius Marici, Remus Runcan, Iasmina Iosim, and Alexandra Haisan. The effect of attire attractiveness on students' perception of their teachers. *Frontiers in Psychology*, 13:1059631, 2023.

[Moghaddam *et al.*, 2021] Mahnaz Moghaddam, Mostafa Charmi, and Hossein Hassanpoor. Jointly human semantic parsing and attribute recognition with feature pyramid structure in efficientnets. *IET Image Processing*, 15(10):2281–2291, 2021.

[Mosca and Buzza, 2013] Joseph B Mosca and John Buzza. Clothing and the affects on a teacher's image: How students view them. *Contemporary Issues in Education Research*, 6(1):59–66, 2013.

[Ridnik *et al.*, 2021] Tal Ridnik, Hussam Lawen, Asaf Noy, Emanuel Ben, Baruch Gilad Sharir, and Itamar Friedman. Tresnet: High performance gpu-dedicated architecture. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1399–1408, 2021.

[Sarfraz *et al.*, 2017] M Saquib Sarfraz, Arne Schumann, Yan Wang, and Rainer Stiefelhagen. Deep view-sensitive pedestrian attribute inference in an end-to-end model. *arXiv preprint arXiv:1707.06089*, 2017.

[Schumann and Stiefelhagen, 2017] Arne Schumann and Rainer Stiefelhagen. Person re-identification by deep learning attribute-complementary information. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 20–28, 2017.

[Slabbert, 2019] Anke Slabbert. Lecturer dress code and student perceptions of lecturer professional attributes. *Journal of Psychology in Africa*, 29(2):176–181, 2019.

[Sudowe *et al.*, 2015] Patrick Sudowe, Hannah Spitzer, and Bastian Leibe. Person attribute recognition with a jointly-trained holistic cnn model. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 87–95, 2015.

[Tan *et al.*, 2019] Zichang Tan, Yang Yang, Jun Wan, Hanyuan Hang, Guodong Guo, and Stan Z Li. Attention-based pedestrian attribute analysis. *IEEE transactions on image processing*, 28(12):6126–6140, 2019.

[Tan *et al.*, 2020] Zichang Tan, Yang Yang, Jun Wan, Guodong Guo, and Stan Z Li. Relation-aware pedestrian attribute recognition with graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 12055–12062, 2020.

[Tang *et al.*, 2019] Chufeng Tang, Lu Sheng, Zhaoxiang Zhang, and Xiaolin Hu. Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4997–5006, 2019.

[Tian *et al.*, 2015] Yonglong Tian, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Pedestrian detection aided by deep learning semantic tasks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5079–5087, 2015.

[Wang *et al.*, 2017] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Attribute recognition by joint recurrent learning of context and correlation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 531–540, 2017.

[Wang *et al.*, 2018] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2275–2284, 2018.

[Wang *et al.*, 2023] Zhe Wang, Jun Wang, and Junliang Xing. Attribute-guided transformer for robust person re-identification. *IET Computer Vision*, 2023.

[Wu and Cai, 2019] Yunhui Wu and Yang Cai. A brief discussion on dress etiquette for university teachers. *Light Industry Science and Technology*, 35(10):122–123, 2019.

[Yin *et al.*, 2020] Jianyuan Yin, Zheyi Fan, Shuni Chen, and Yilin Wang. In-depth exploration of attribute information for person re-identification. *Applied Intelligence*, 50:3607–3622, 2020.

[Zhang *et al.*, 2019] Yan Zhang, Xusheng Gu, Jun Tang, Ke Cheng, and Shoubiao Tan. Part-based attribute-aware network for person re-identification. *IEEE Access*, 7:53585–53595, 2019.

[Zhang *et al.*, 2020] Jianfu Zhang, Li Niu, and Liqing Zhang. Person re-identification with reinforced attribute attention selection. *IEEE Transactions on Image Processing*, 30:603–616, 2020.

[Zhang *et al.*, 2023] Haoyu Zhang, Meng Liu, Yuhong Li, Ming Yan, Zan Gao, Xiaojun Chang, and Liqiang Nie. Attribute-guided collaborative learning for partial person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Zhao *et al.*, 2019] Yiru Zhao, Xu Shen, Zhongming Jin, Hongtao Lu, and Xian-sheng Hua. Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4913–4922, 2019.