International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 671-675 www.ijaem.net ISSN: 2395-5252

From Street Photos to Fashion Trends Leveraging User Provided Noisy Labels for Fashion Understanding

Ramyashree H S, Chethana Rao M, Ashrita Kumari Chauhan
Guidance Of: Mrs Suchitra Devi
Department Of Computer Science and Engineering
SAMBHRAM INSTITUTE OF TECHNOLOGY
BENGALURU-560097

Date of Submission: 08-05-2023 Date of Acceptance: 20-05-2023

I. INTRODUCTION

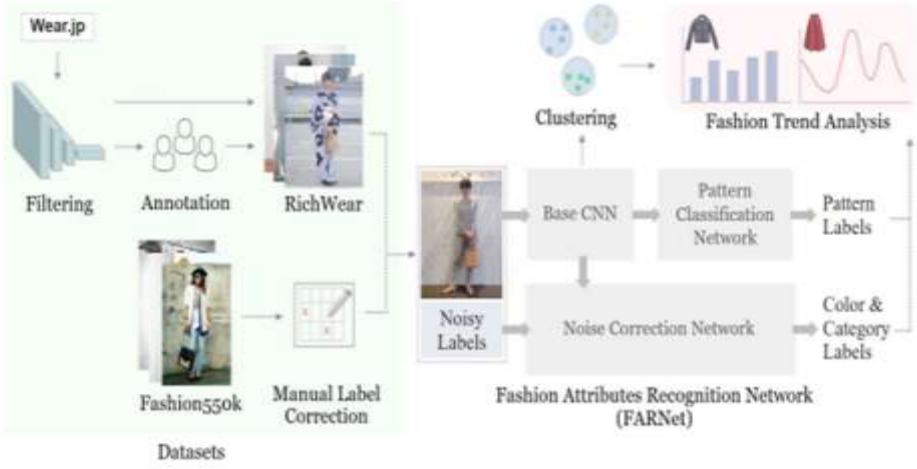
As interest has increased in the possible relationships between artificial intelligence (AI) and fashion, more and more approaches are being proposed for fashion recognition and understanding .

These street photos on social media sites provide much-needed data for AI research. At the same time, the large-scale street images have led researchers to analyze street fashion . We have observed that most datasets used for street fashion research are collected from social media sites based in the United States and Europe, and these images are mainly related to American and European street styles. Few datasets focus on Asian street styles.

Moreover, fashion data collected from the internet usually contain user-provided labels that are inconsistent with the images and are referred to as noisy labels. Training deep learning models directly on noisy labels may result in degraded predictive performance Moreover, there are rich clothing attributes in a street fashion image, such as colors, categories (e.g., shirt and pants), and patterns

To address these issues, we created RichWear, a new fash_ion dataset containing 322,198 high-quality street fashion images with noisy labels from the Asian fashion website WEAR. The dataset focuses on the street styles in Japan and other Asian areas from 2017 to 2019. To improve fashion recognition, we propose the Fashion Attributes Recognition Network (FARNet) that includes a Noise Correction Network and a Pattern Classification Network on top of a Convolutional Neural Network (CNN) image feature extractor.

The two main networks of FARNet are trained jointly to simultaneously predict clothing colors, categories, and patterns based on the input image and its noisy labels.



International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 671-675 www.ijaem.net ISSN: 2395-5252

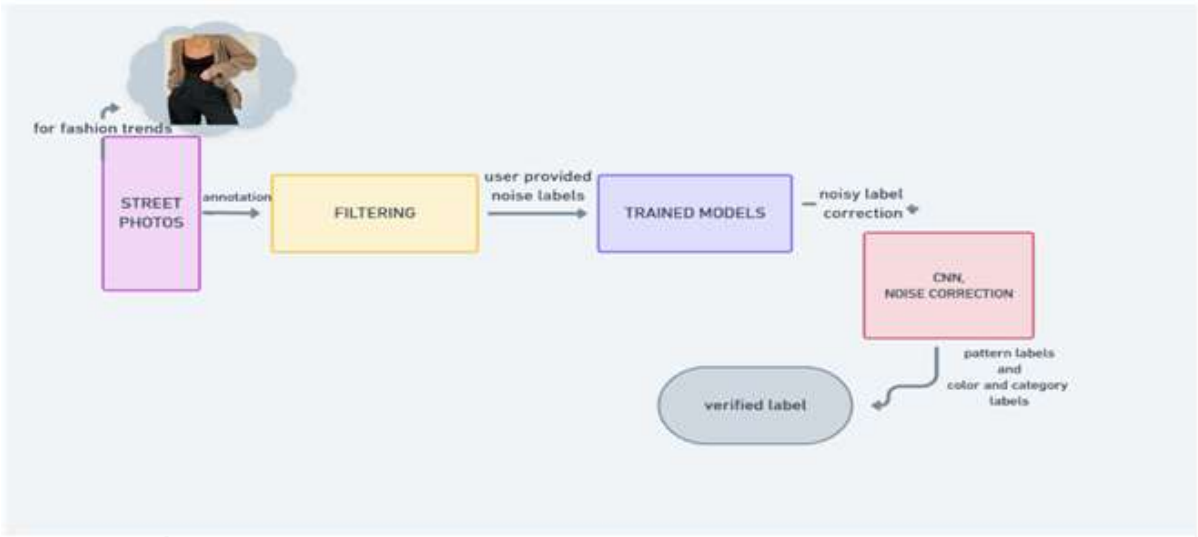
MOTIVATION

There is increased interest in using street photos to understand fashion trends. Though street photos usually contain rich clothing information, there are several technical challenges to their analysis.

First, street photos collected from social media sites often contain user-provided noisy labels, and training models using these labels may deteriorate prediction performance. Second, most existing methods predict multiple clothing attributes individually and do not consider the potential to share knowledge between related tasks.

In addition to these technical challenges, most fashion image datasets created by previous studies focus on American and European fashion styles. To address these technical challenges and understand fashion trends in Asia, we created RichWear, a new street fashion dataset containing 322,198 images with various text labels for fashion analysis.

So aim of our project is Street photos collected from social media sites often contain user-provided noisy labels, and training models using these labels may deteriorate prediction performance. To create RichWear, a new street fashion dataset containing 322,198 images with various text labels for fashion analysis.



Literature Survey

There is a large amount of literature that develops deep learning and computer vision approaches to image understanding and fashion recognition. Our literature review focuses on three streams of related studies: (1) fashion attribute recognition, (2) street fashion analysis, and (3) street fashion datasets.

(1) fashion attribute recognition

Fashion attribute recognition aims to identify one or more fashion-related attributes according to input images. Recent works have shown growing interest in training machines for effective visual recognition on large-scale image datasets.

(2) street fashion analysis

Street fashion does not originate from studios or runways, but from real-life streetwear. As more internet users share street photos on social media, street fashion has gradually become a driving force for fashion change and fashion design . Nevertheless, unlike online shopping images and runway images with standing models and simple backgrounds, street fashion images usually contain people in different poses with various backgrounds, a difference that increases the difficulty of image recognition.

(3) street fashion datasets

Several street fashion studies created new datasets to facilitate model training and evaluation. These datasets each contain a different number of street fashion images with text labels or other types of annotations dependent on the original research purposes. The most commonly-used text labels are related to clothing attributes, such as category, color, and sleeve length, or to fashion styles, such as ethnic, casual, and fairy.

PROBLEM STATEMENT

1.Street photos collected from social media sites often contain user-provided noisy labels, and training models using these labels may deteriorate prediction performance.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 671-675 www.ijaem.net ISSN: 2395-5252

2.To create RichWear, a new street fashion dataset containing 322,198 images with various text labels for fashion analysis.
3.To propose the Fashion Attributes Recognition Network (FARNet) based on the multi-task learning framework to improve fashion recognition. We propose the Fashion Attributes Recognition Network, recognize three types of clothing attributes, including colors, categories, and patterns, in noisy-labeled images.

ALGORITHM

CNN stands for Convolutional Neural Network, which is a type of deep neural network that is primarily used for image classification and recognition. It is inspired by the structure of the human visual system, where the input image is processed in a hierarchical manner.

1.**Data Preparation**: This step involves collecting and preparing the dataset. The dataset is usually split into training, validation, and testing sets. Each image is also preprocessed by resizing and normalizing to reduce variations in the input data.

2.**Convolutional Layers**: The first few layers in a CNN are usually convolutional layers. These layers perform feature extraction by applying a set of filters to the input image. Each filter applies a convolution operation to the input image and produces a feature map.

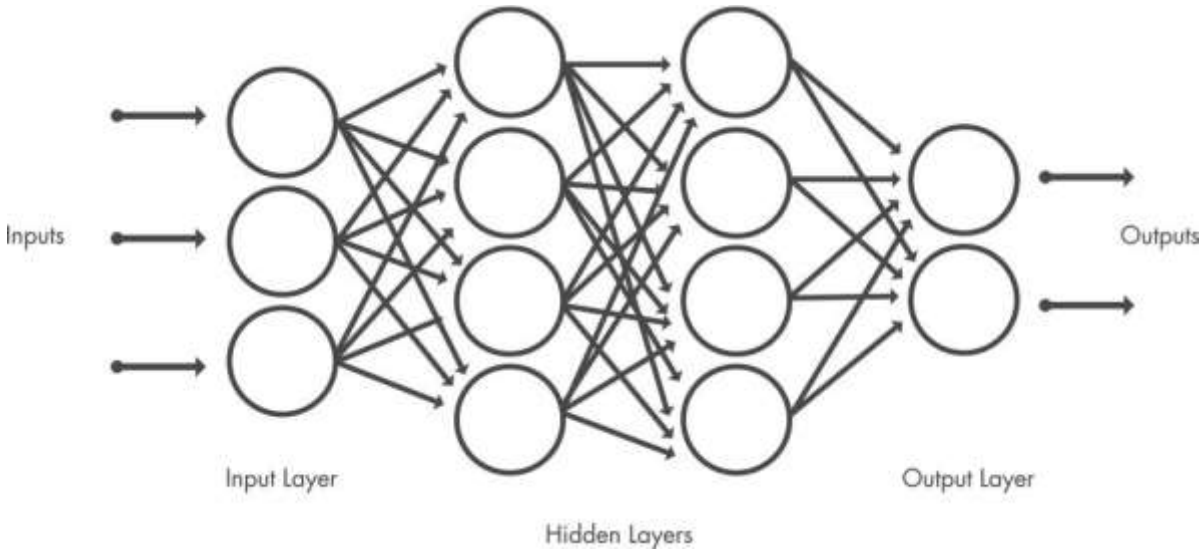
3.**Pooling Layers**: After the convolutional layers, pooling layers are added to reduce the spatial size of the feature maps. Max pooling is a common technique used in pooling layers.

4.**Flatten Layer**: The output of the last pooling layer is flattened into a one-dimensional vector to be fed into a fully connected neural network.

5.**Fully Connected Layers**: The flattened vector is then passed through several fully connected layers, which perform the classification task. These layers use activation functions such as ReLU or sigmoid to produce output probabilities for each class.

6.**Output Layer**: The final fully connected layer produces the output probabilities for each class.

7.**Training**: The CNN is trained by repeating the above steps on the training set. The validation set is used to monitor the performance of the model and to prevent overfitting.



Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression analysis. The goal of SVM is to find the optimal hyperplane in a high-dimensional space that maximally separates the data into different classes.

1.**Data preparation**: The first step is to prepare your data for SVM. This involves collecting and cleaning data, and then splitting it into training and testing sets.

2.**Feature selection**: You need to identify the features that will be used to classify the data. The quality of the features you choose can greatly affect the performance of the SVM algorithm.

3.**Training the SVM model**: The next step is to train the SVM model using the training data. The SVM algorithm tries to find the best hyperplane that separates the data into classes. The hyperplane is chosen in such a way that it maximizes the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 671-675 www.ijaem.net ISSN: 2395-5252

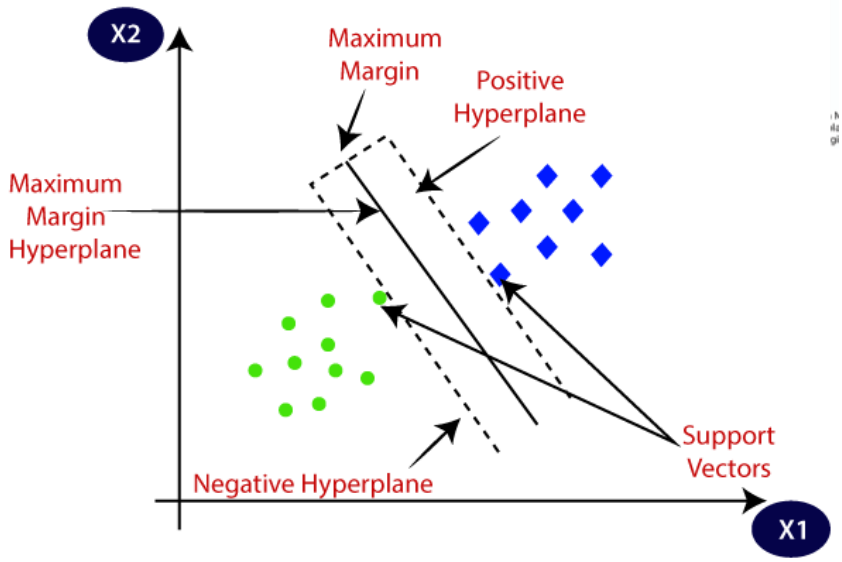
margin, i.e., the distance between the hyperplane and the closest data points of each class.

4.**Choosing a kernel function**: A kernel function is used to transform the input data into a higher dimensional space where it can be separated more easily. There are several kernel functions to choose from, such as linear, polynomial, and radial basis function (RBF).

5.**Tuning hyperparameters**: There are several hyperparameters that need to be tuned in order to improve the performance of the SVM model. These include the regularization parameter (C), the kernel parameter (gamma), and the degree of the polynomial kernel (d).

6.**Testing the SVM model**: Once the SVM model is trained, it can be used to classify new data. The performance of the SVM model is evaluated on the test set by calculating metrics such as accuracy, precision, recall, and F1-score.

7. **Deployment**: Finally, the SVM model can be deployed in a production environment where it can be used to classify new data.



II. FINDINGS

- This study aimed to explore Asian street fashion by creating a novel fashion recognition architecture and a large-scale image dataset
- For fashion recognition, we have proposed a multi-task neural network, FARNet, which recognize multiple clothing attributes. This network facilitates our street fashion exploration in the large-scale dataset collected from a social media site. . It achieves better performance (71.33%) than the compared baselines (55.46%–59.79%)

In future work, we plan to incorporate product and brand information to further refine the fashion trend analysis. We are interested in the mercurial popularity of products andbrands, a deeper understanding that may help us predict the rise and fall of a particular product, brand, or style

III. CONCLUSION

This study aimed to explore Asian street fashion by creating a novel fashion recognition architecture and a large-scale image dataset with user-provided noisy labels, and it contributes to the existing literature in three areas. First, this study developed a new street fashion dataset named RichWear, which contains 322,198 street fashion images with upload date, users' gender and country, clothing brands, and user-created hashtags. In addition to user-provided noisy labels, we also created a 4,368-image subset with expert-verified labels for three types of clothing attributes.

In particular, RichWear focuses on street styles in Japan and other Asian areas, providing a good data source for Asian fashion understanding. For fashion recognition, we have proposed a multi-task neural network, FARNet, which can leverage noisy labels and simultaneously recognize multiple clothing attributes. This network facilitates our street fashion exploration in the large-scale dataset collected from a social media site.

The Noise Correction Network in FARNet is based on an existing network, but it more effectively corrects for noisy labels. It achieves better performance (71.33%) than the compared baselines (55.46%–59.79%). Moreover, our empirical results show that MTL, when compared to STL, can noticeably improve generalization performance of attribute recognition. Finally, by using both supervised and unsupervised learn_ing methods, we have documented interesting street

fashion trends in Asia from 2017 to 2019. We have also observed significant seasonal dynamics for men's and women's street styles that have not been explored in previous studies. In future work, we plan to incorporate product and brand information to further refine the fashion trend analysis. We are interested in the mercurial popularity of products andbrands, a deeper understanding that may help us predict the rise and fall of a particular product, brand, or style.

REFERENCES

[1]. N. Liu, S. Ren, T.-M. Choi, C.-L. Hui, and S.-F. Ng, "Sales forecasting for fashion retailing service industry: A review," Math. Problems Eng., vol. 2013, Nov. 2013, Art. no. 738675.

[2]. Y. Guan, Q. Wei, and G. Chen, "Deep learning based personalized recom_mendation with multi-view information integration," Decis. Support Syst., vol. 118, pp. 58–69, Mar. 2019.

[3]. M. F. Hashmi, B. K. K. Ashish, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "FashionFit: Analysis of mapping 3D pose and neural body fit for custom virtual try-on," IEEE Access, vol. 8, pp. 91603–91615, 2020.

[4]. X. Gu, Y. Wong, P. Peng, L. Shou, G. Chen, and M. S. Kankanhalli, "Understanding fashion trends from street photos via neighbor-constrained embedding learning," in Proc. 25th ACM Int. Conf. Multimedia, Mountain View, CA, USA, Oct. 2017, pp. 190–198.

[5]. N. Inoue, E. Simo-Serra, T. Yamasaki, and H. Ishikawa, "Multi-label fash_ion image classification with minimal human supervision," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Venice, Italy, Oct. 2017, pp. 2261–2267.

[6]. L. Lo, C.-L. Liu, R.-A. Lin, B. Wu, H.-H. Shuai, and W.-H. Cheng, "Dressing for attention: Outfit based fashion popularity prediction," in Proc. IEEE Int. Conf. Image Process. (ICIP), Taipei, Taiwan, Sep. 2019, pp. 3222–3226.

[7]. Y. Ma, J. Jia, S. Zhou, J. Fu, Y. Liu, and Z. Tong, "Towards bet_ter understanding the clothing fashion styles: A multimodal deep learn_ing approach," in Proc. Assoc. Advancement Artif. Intell. (AAAI), San Francisco, CA, USA, 2017, pp. 38–44.

[8]. K. Matzen, K. Bala, and N. Snavely, "StreetStyle: Exploring world_wide clothing styles from millions of photos," 2017, arXiv:1706.01869. [Online]. Available: http://arxiv.org/abs/1706.01869.

[9]. U. Mall, K. Matzen, B. Hariharan, N. Snavely, and K. Bala, "GeoStyle: Discovering fashion trends and events," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, South Korea, Oct. 2019, pp. 411–420.

[10]. M. Takagi, E. Simo-Serra, S. Iizuka, and H. Ishikawa, "What makes a style: Experimental analysis of fashion prediction," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Venice, Italy, Oct. 2017, pp. 2247–2253.

[11]. W.-L. Hsiao and K. Grauman, "Learning the latent 'Look': Unsuper_vised discovery of a style-coherent embedding from fashion images,"in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, Oct. 2017, pp. 4203–4212.

[12]. E. Simo-Serra, S. Fidler, F. Moreno-Noguer, and R. Urtasun, "Neuroaes_thetics in fashion: Modeling the perception of fashionability," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 869–877.

[13]. K. Yamaguchi, M. H. Kiapour, and T. L. Berg, "Paper doll parsing: Retrieving similar styles to parse clothing items," in Proc. IEEE Int. Conf. Comput. Vis., Sydney, NSW, Australia, Dec. 2013, pp. 3519–3526.

[14]. K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, "Parsing clothing in fashion photographs," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI, USA, Jun. 2012, pp. 3570–3577.

[15]. S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu, "ModaNet: A large_scale street fashion dataset with polygon annotations," in Proc. 26th ACM Int. Conf. Multimedia, Seoul, South Korea, 2018, pp. 1670–1678.