

IndoFashion : Apparel Classification for Indian Ethnic Clothes

Pranjal Singh Rajput
Delft University of Technology
pranjal.tudelft@gmail.com

Shivangi Aneja
Technical University of Munich
shivangi.aneja@tum.de



Figure 1: This paper proposes a new dataset with 106K images and 15 classes for cloth categorization for Indian style clothing. Figure shows images from the dataset with the corresponding labels.

Abstract

Cloth categorization is an important research problem that is used by e-commerce websites for displaying correct products to the end-users. Indian clothes have a large number of clothing categories both for men and women. The traditional Indian clothes like “Saree” and “Dhoti” are worn very differently from western clothes like t-shirts and jeans. Moreover, the style and patterns of ethnic clothes have a very different distribution from western outfits. Thus the models trained on standard cloth datasets fail on ethnic outfits. We introduce the first large-scale ethnic dataset of over 106K images with 15 different categories for fine-grained classification of Indian ethnic clothes. We gathered a diverse dataset from a large number of Indian e-commerce websites. We then evaluate several baselines for the cloth classification task on our dataset. We obtain 88.43% classification accuracy. We hope that our dataset would foster research in the development of several algorithms such as cloth classification, landmark detection, especially for ethnic clothes. The dataset is publicly available here ¹.

¹<https://indofashion.github.io/>

1. Introduction

With the increasing use of online shopping for clothing, it has become important for e-commerce websites to correctly identify categories for all the clothing items available on the website. Few datasets like FashionMNIST [17], DeepFashion [12, 15] have been introduced in past few years to identify cloth categories. However, these datasets mostly deal with western attire like shirts, jeans, jackets, shorts, etc. In contrast, our focus in the work is on traditional Indian attire where the cloth categories like Saree, Kurta, Dhoti, etc are very different from western ones, see Figure 1. With this work, we introduce an ethnic fashion dataset, a first of its kind to identify these cloth categories.

To the best of our knowledge, only two datasets Atlas [16] and Generative Indian Fashion [14] deal with Indian ethnic clothes. Although Atlas [16] dataset deals with identifying cloth taxonomy, the dataset is highly imbalanced with less than 40 images for few class categories, make it unsuitable for training deep models. Generative Indian Fashion [14] does not address the cloth classification problem and has very few images (12K) and class categories. Thus, both

these datasets are limited in terms of either dataset size or applicability for cloth classification of Indian ethnic wear.

We notice that no benchmarking is done for cloth categorization of Indian Ethnic clothes. In this work, we propose to address the problem by providing the first large-scale dataset for the task. In addition, we also benchmark several state-of-the-art image classification models to evaluate the performance on the dataset.

In summary, our contributions are as follows:

- To the best of our knowledge, this paper proposes the first automated method to categorize Indian Ethnic clothes.
- We created a large dataset of 106K images with 15 different cloth categories from a variety of e-commerce websites.
- We benchmark our results on several image classification results and analyze their performance.

2. Related Work

2.1. Clothing Datasets

Several datasets [12, 15, 17, 18] have been proposed in the recent past for clothes recognition. FashionMNIST [17] provides a toy dataset of 70K greyscale images of 10 cloth categories, lacking real-world scenarios. DeepFashion [12] provides a very large-scale dataset with dense annotations for a wide variety of cloth-related research problems like cloth category/texture/fabric classification, clothing landmarks, etc. However, the dataset does not contain ethnic clothes. We, however, specifically target Indian ethnic clothes.

2.2. Indian Cloth Datasets

Only a few datasets [16, 14] are available, that contain the Indian ethnic clothes. Atlas [16] contains in total 186k clothing images focusing on product taxonomy classification, consisting of images corresponding to western wear, inner-wears, and ethnic wears. However, the dataset is very imbalanced in the case of some cloth categories. In Generative Indian Fashion [14], the authors focus on modifying the clothing of a person according to a textual description. The dataset is very small consisting of only 12k images. Also, the authors do not address the problem of cloth categorization. On the other hand, IndoFashion dataset consists of 106K images and 15 categories, providing a rich, diverse and balanced dataset, that focuses on ethnic fashion research.

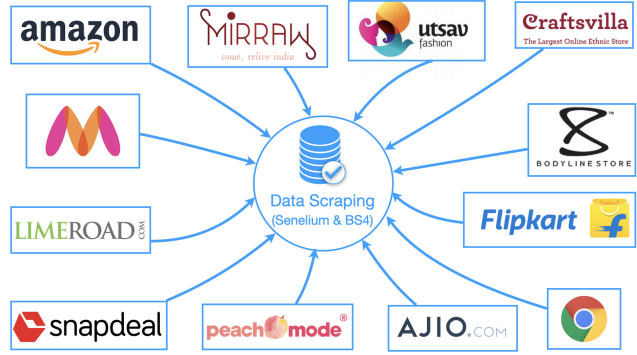


Figure 2: For our dataset, we gathered images from e-commerce websites and google image search as shown in figure. To scrape the images custom webscrapers based on beautifulsoup [?] and selenium webdriver [?] were used.

3. IndoFashion Dataset

IndoFashion is a unique fashion dataset consisting of a variety of Indian ethnic clothes. It is one of its kind of dataset and to the best of our knowledge, this is the first large-scale dataset for ethnic cloth classification. It contains 106K images of 15 different categories covering the most famous Indian ethnic clothes. The images are collected from google image search and several Indian e-commerce websites [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], capturing different styles, colors, and patterns. Every image in the dataset is annotated with a label that represents its class category. The distribution of the dataset per category is shown in Figure 3. Some example images per class label are shown in Figure 1.

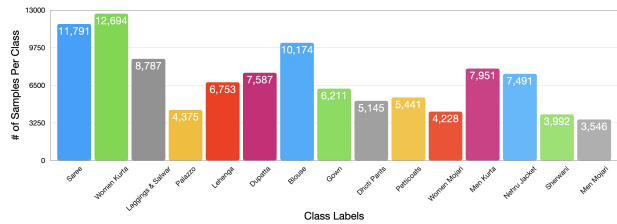


Figure 3: Figure shows category-wise frequency distribution of the 15 cloth categories in our proposed dataset.

3.1. Dataset Collection

We gathered our dataset primarily from famous Indian e-commerce websites such Flipkart [5], Amazon [2], Myntra [8], etc, refer Figure 2. We curate our dataset in two steps: First, we used our custom scrapers to gather images corresponding to each of the categories in the dataset. Second, the dataset is cleaned by removing the duplicate im-

ages from the dataset. In the end, several images per class category is obtained with varying styles and colors.

3.2. Dataset Statistics

Our dataset consists of 106K images and 15 unique cloth categories. For the cloth classification task, we split the dataset into training, validation, and test set. For a fair evaluation, we ensure equal distribution of the classes in the validation and the test set consisting of 500 samples per class. We spent a total of 25 hours to collect, cleanup, and restructure the dataset. Our proposed dataset consists of 15 cloth categories as documented in Table 1 and dataset split is shown in Table 2.

Gender	Categories
Women	Saree, Women Kurta, Leggings & Salwar, Palazzo, Lehenga, Dupatta, Blouse, Gown, Dhoti Pants, Petticoats, Women Mojari
Men	Men Kurta, Nehru Jacket, Sherwani, Men Mojari

Table 1: The table lists the class categories for both men and women in the dataset.

Split	No. of Images
Train	91,166
Validation	7,500
Test	7,500

Table 2: Statistics of our IndoFashion dataset.

3.3. Comparison with other Indian ethnic datasets

We compare our dataset against other existing Indian ethnic cloth datasets in Table 3. Overall our dataset contains 106,166 images which is larger than any other existing dataset for ethnic clothes. Note that, Atlas dataset [16], categorizes *Men Suits* and *Jackets* as ethnic wear, which are **not** considered ethnic clothes. Similarly, Jain *et al.* [14] incorrectly categorize *Casual Shirts*, *Shirts*, *T-Shirts* as ethnic wear. This incorrect categorization makes these datasets unsuitable for classification of ethnic clothes.

4. Experiments

We evaluate several state-of-the-art Image Classification models for our task of classifying Indian ethnic clothes. All our models use an ILSVRC 2012-pretrained ResNet [13]-18, 50, and 101 backbone. All numbers reported in the

Dataset	#Images	#Classes	Avg images per class
Atlas dataset [16]	81,813	17	4,812
Jain <i>et al.</i> [14]	12,304	12	1,025
IndoFashion(Ours)	106,166	15	7,077

Table 3: Quantitative comparison of various Indian ethnic cloth datasets. IndoFashion dataset contains the largest number of images, in total as well as per category.

paper are on the test split of the dataset. We use Precision, Recall, and F1-Score to measure the performance of our models.

4.1. Pre-processing and Augmentations

We first pre-process all the images in the dataset by resizing the images to 128x256 pixels before feeding them into the network.

We also apply various types of augmentation to the images during training to mitigate over-fitting and improve the model performance. Note that we apply the augmentations to the training images only and are not applied during inference. We document these augmentations below:

1. Geometric Transformation: Images are geometrically transformed by flipping them horizontally and rotating by a small angle of 10 degrees randomly. Images are geometrically transformed during training with a probability of 50%.
2. Color Transformation: In this method, we apply color jitter by varying the hue and saturation by a factor of 0.2.

4.2. Hyperparameters

We run all our experiments on an Nvidia GeForce GTX 2080 Ti card. We use a learning rate of 1e-3 with Adam optimizer for our model with decay when validation loss plateaus for 5 consecutive epochs, early stopping with the patience of 25 successive epochs on validation loss.

5. Results

We visualize the confusion matrix of our model in Figure 4. In confusion matrix, it is observed that visually similar classes like *Women Kurta* and *Gowns* are difficult to distinguish using our trained model.

5.1. How Data Augmentation helps?

We analyze the effect of augmentations mentioned in Section 4.1 in Table 4. We plot the confusion matrix of our best performing model in Figure 4. We notice that applying

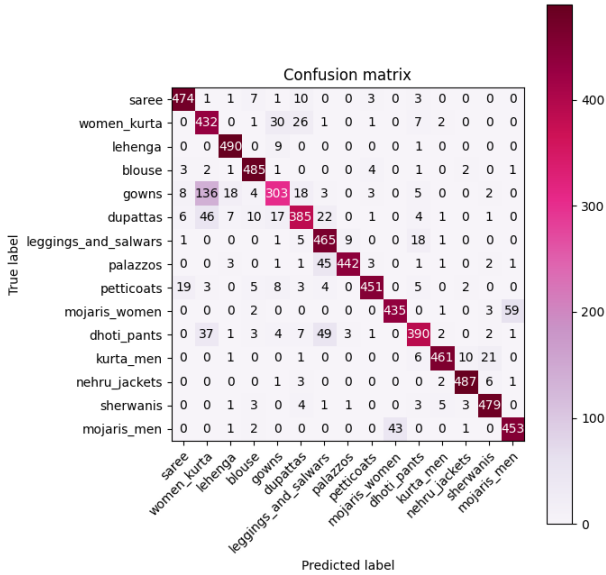


Figure 4: Figure show the Confusion Matrix of our best performing model, Resnet-50(J+F), where (J+F)=Jitter and Flipping. Note that the class Women Kurta is often confused with Gowns, Mojari Men with Mojari Women since these pair of classes are visually similar to each other.

all these augmtations together gives us the best performing model.

Model	Precision	Recall	F1-Score
R-18(NA)	86.9	86.12	86.11
R-18(F)	88.3	87.41	87.44
R-18(J)	87.83	87.24	87.24
R-18(J+F)	88.62	87.52	87.46
R-50(NA)	87.95	86.94	86.93
R-50(F)	88.85	87.81	87.84
R-50(J)	87.78	87.04	87.04
R-50(J+F)	88.93	88.42	88.38
R-101(NA)	87.66	86.82	86.86
R-101(F)	89.35	88.56	88.5
R-101(J)	86.86	86.3	86.29
R-101(J+F)	88.63	87.97	87.94

Table 4: The table illustrates the effect of applying different augmentations during training. To avoid clutter, following notations are used: R=ResNet, (J)=Jitter augmentation, (F)=Flipping, and (J+F)=Jitter and Flipping together, (NA)=No Augmentation.

It is observed that the performance improves by implementing either geometrical or color augmentation alone.

However, applying both augmentations together boosts the performance of the model. Also, we noticed that increasing the model size from resnet-18 to resnet-50 improves the performance further. However, increasing the model size to more bigger resnet-101 backbone does not show any improvements. Our best performing model is resnet-50 achieving a precision of 88.93%. So we conclude that very large and complex models are not required for our dataset.

Dataset Size	Images	Model Accuracy		
		R-18	R-50	R-100
Ours (10%)	9K	77.19	79.65	79.84
Ours (20%)	18K	80.2	82.96	82.93
Ours (50%)	45K	85.59	85.55	76.36
Ours (100%)	91K	87.52	88.43	87.97

Table 5: Ablation with variations in number of training images w.r.t. our full corpus of 91K images. Using all available data achieves the best performance. Here, R=ResNet.

5.2. How training data size affects?

We experiment with different percentage of training data size (w.r.t. our full data of 91K; refer to Table 5 for details). For these experiments, we evaluate all three Resnet-18, 50 and 101 backbones and apply color and geometric augmentation during trainig. We observed that a larger training set improves the performance of all the model significantly. For instance, training with full-dataset (91K images) improves the classification accuracy by an absolute 10.33% (from 77.19% to 87.52%) compared to a model trained only with 10% of the dataset (9K images) even for the small ResNet-18 model, thus benefiting from the large diversity in the dataset.

6. Conclusion and Future Work

In this work, we have introduced the first large-scale dataset for Indian ethnic cloth classification. Our dataset is diverse consisting of images gathered from a large number of Indian e-commerce websites. We train several image classification models to benchmark our dataset. We hope that our new dataset, which we will publish along with this work, will lead to the development of several algorithms (cloth classification, clothing landmark detection, etc) specifically for ethnic clothes.

Even though we achieve 88.43% of classification accuracy, more research towards developing better classification models can incentivize cloth categorization. The dataset can be improved by balancing the number of images per category. In future, we plan to extend the dataset by introducing more categories from other ethnic groups.

References

- [1] Ajio. <https://www.ajio.com/>. Accessed: 2021-03-24.
- [2] Amazon.in. <https://www.amazon.in/>. Accessed: 2021-03-24.
- [3] bodylinestore. <https://www.bodylinestore.com/>. Accessed: 2021-03-24.
- [4] Craftsvilla. <https://www.craftsvilla.com/>. Accessed: 2021-03-24.
- [5] Flipkart. <https://www.flipkart.com/>. Accessed: 2021-03-24.
- [6] Limeroad. <https://www.limeroad.com/>. Accessed: 2021-03-24.
- [7] Mirraw. <https://www.mirraw.com/>. Accessed: 2021-03-24.
- [8] Myntra. <http://myntra.com/>. Accessed: 2021-03-24.
- [9] Peachmode. <https://peachmode.com/>. Accessed: 2021-03-24.
- [10] Snapdeal. <https://www.snapdeal.com/>. Accessed: 2021-03-24.
- [11] Utsav fashion. <https://www.utsavfashion.com/>. Accessed: 2021-03-24.
- [12] Yuying Ge, Ruimao Zhang, Lingyun Wu, X. Wang, X. Tang, and Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5332–5340, 2019.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [14] Harshil Jain, Rohit Patil, Utsav Jethva, Ronak Kaoshik, Shaurya Agarawal, Ritik Dutta, and Nipun Batra. *Generative Fashion for Indian Clothing*, page 415. Association for Computing Machinery, New York, NY, USA, 2021.
- [15] Z. Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1096–1104, 2016.
- [16] Venkatesh Umaashankar, Girish Shanmugam S, and Aditi Prakash. Atlas: A dataset and benchmark for e-commerce clothing product categorization, 2019.
- [17] H. Xiao, K. Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *ArXiv*, abs/1708.07747, 2017.
- [18] X. Zou, X. Kong, W. Wong, C. Wang, Y. Liu, and Y. Cao. Fashionai: A hierarchical dataset for fashion understanding. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 296–304, 2019.