Automatic Clothing Recognition for the Identification of Fashion Trends

Abstract

Recognition of clothing categories from social media sites can be used for intelligent customer profile analysis, fashion trend analysis, product demand forecasting, or creating a personalized style recommendation system. We present a scalable approach to clothing recognition, given a set of images with plain, uncluttered backgrounds. In this paper, we propose an optimal method for the automatic classification of broad clothing categories, and the subsequent problem of detecting its position within the image. We focused on three machine learning algorithms: SVM, AdaBoost, and Viola-Jones. We used Histogram of Oriented Gradients (HOG) to extract image feature descriptors. We were able to achieve excellent results classifying clothing objects within an image, and fair results detecting the position of the clothing article within the image. For the purpose of this study, we recognize four clothing classes: dresses, skirts, shirts, and pants; however, a larger set of classes may be identified using our method. The techniques we have outlined may serve as a building block for identifying fashion trends and other characteristics for different clothing types across a larger data set.

I. Introduction

The Internet has recently seen an explosion of image data through the emergence of social media websites such as Pinterest, Instagram, and Twitter. An estimated 200 million images are uploaded daily to social media and photo sharing websites. Few companies, however, are using automated techniques to analyze these images to identify customer behaviors and trends. This is particularly true in the image-rich fashion industry, which currently uses human analysis to find the brands, colors, and styles most popular with target markets. We are interested in creating a system that automatically identifies articles of clothing in images for the purpose of identifying fashion trends among targeted markets. The development of such a system would make it feasible for companies to construct a database of currently trending outfits, or get statistics about popular styles and colors. For example, a database on shirts worn by a target market where the majority have a floral pattern may indicate that another floral style shirt may be well-received by this segment. This would significantly improve the quality of current fashion companies' consumer analytics, and thus the effectiveness of their marketing strategies and forecasts.

Although this problem attracts increasing research interests, a clothing recognition program remains challenging for two primary reasons. First, such a program has to filter noise from the environment, for example dealing with object occlusion and complex backgrounds. Second, different types of clothing can have very similar shapes, and it is sometimes difficult for even humans to label (e.g. differentiating a skirt from part of certain dresses as distinct objects). Therefore, the question of whether it is even possible to objectively discern a certain clothing type purely through visual data arises. Recognizing these issues, clothing recognition, and indeed object recognition in general, is largely an unsolved problem and an active area of research. Despite this, there are still gaps in the literature surrounding basic classification across different clothing types.

II. Data Description

A total of 1,878 images were collected from Google (Like.com), Bebe, Pinterest, Chicismo, Chictopia, Abercrombie and Fitch, Land's End, JCPenney, Zara, Kohl's, American Eagle, Charlotte Russe, and Sears. All images had plain backgrounds often of one or two colors. The images contained individuals, both men and women, with at least a portion of their head and legs. Some of the images were full bodied, while others were not. The types of clothing that individuals wore varied from formal to casual and even included nightgowns. Additionally, image sizes ranged between 5 KB to 100 KB. Using a custom built labeling tool, regions of each image were manually labeled with a bounding box for each clothing class. Each data set was comprised of positive examples of its own class and negative examples taken from the other classes.

	Number of Positive Images	Number of Negative Images	Proportion
Dress	380	1498	25.37%
Pants	300	1578	19.01%
Shirt	718	1160	61.90%
Skirt	480	1398	34.33%

Table 1: Proportion of positive to negative images used for training, validation, and test data in each class

Figure 1: Labeling tool for labeling each class in image. A shirt is being labeled in the image above.

II. Problem Definitions and Methods

II. a) Task Definition

Our goal was to test whether HOG features can be effectively used to train a classifier that will distinguish clothing categories. In order to reduce the complexity of the problem and help focus our efforts towards the effective comparison between varied algorithms, we focused on images between 5 KB and 100 KB which contained a single person against a fixed solid/monochromatic background. In these images, we sought to identify which of the four clothing categories (shirt, dress, skirt, shirt) was present in the image, and then find a bounding box for the category within the larger image. Based on our results, we will better understand the viability of extracting characteristics from the identified clothing object, such as colors, styles, and shapes for fashion trend analysis.

II.b) Process & Method

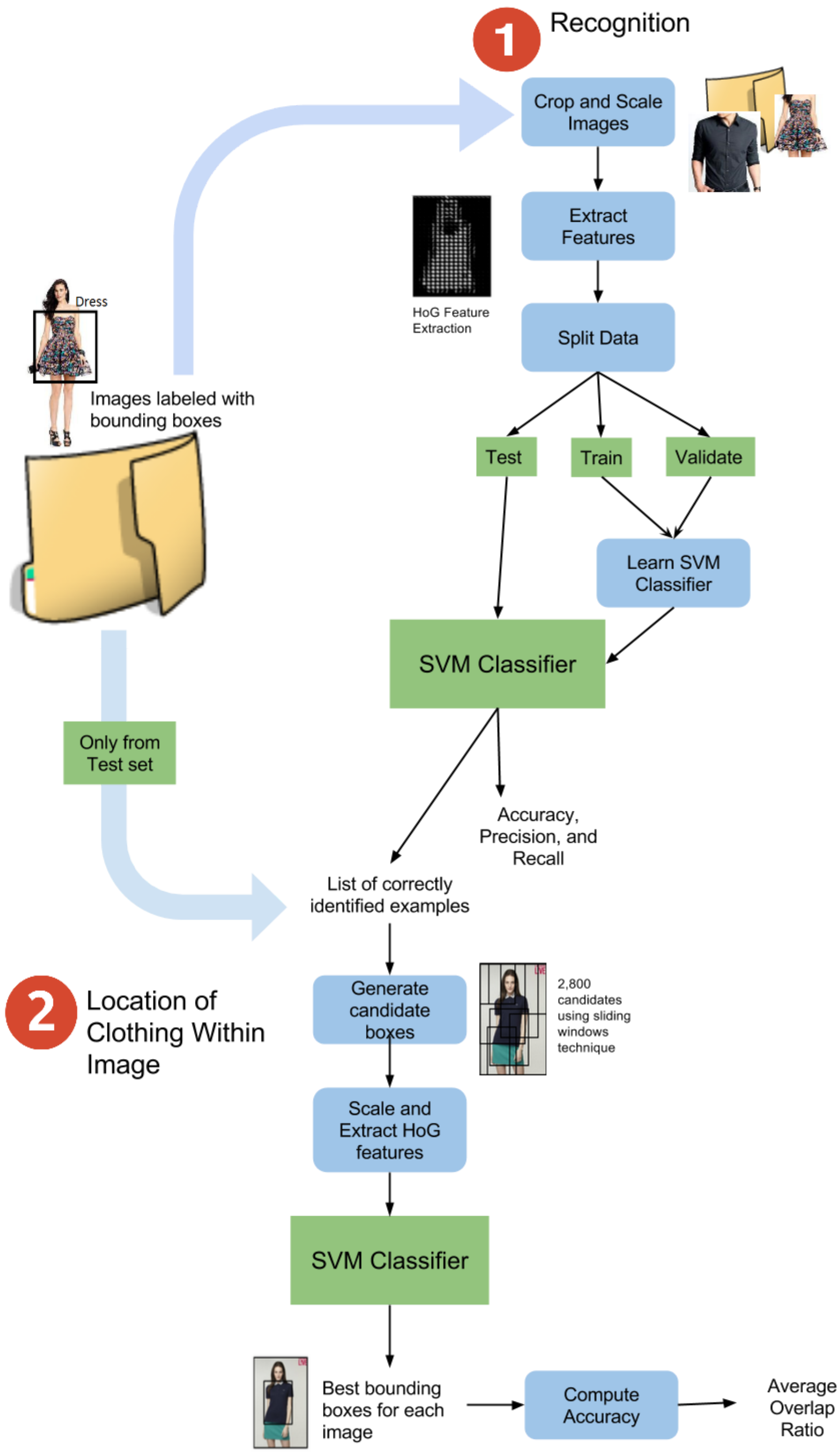

Figure 2: Method Diagram

Generally, our approach can be split into two phases as shown in Figure 2: Method Diagram.

1. Clothing Recognition:

Images are converted into .jpg format and scaled to a standard size with padding added, if necessary to the left or top sides of the image. For this project, we standardized the images to be 200 by 300. These normalized images are then cropped and divided into labeled zones for each class: dress, shirt, pant, or skirt. Positives and negative examples are then determined. Each image that was identified as positive for a zone, was then classified as negative for the other zones. The background of the cropped image was used as a negative for the classes that its cropped portion was identified as positive. The HOG features of these images are then extracted and put into SVMLight format. Training and validation sets are determined and the best classifier is trained for each class

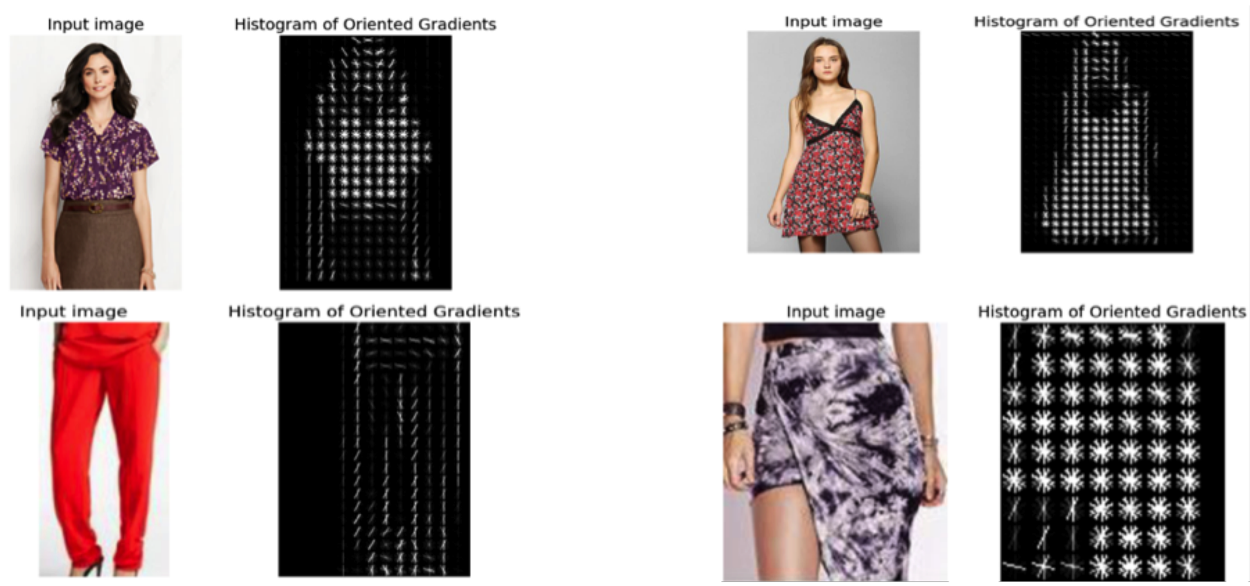

Figure 3: HOG descriptors for each image class

2. Object Location Detection

After we established the clothing type within each image, we searched the image for the location of the clothing item for a subset of the classified images.

We used approximately 2,800 sliding windows on each test image. These windows are then run through the appropriate SVM regression classifier (i.e. dress, shirt, pant, skirt), chosen by whichever clothing type the image was determined to contain in order to determine the best bounding boxes. In order to reduce the variance of the predicted bounding box, the top five bounding boxes are averaged by taking the average for each of the four coordinates. We then compare the human-labeled image with the predicted bounding box, and derive two scores indicating the amount of overlap with the labeled image, and relative size to the labeled image.

II.c) Learning Algorithms

We tried three separate algorithms for object recognition: SVM, Adaboost, and Viola-Jones, all using the popular HOG feature descriptor.

SVM: SVM was chosen because it is commonly used in computer vision tasks for recognizing various different object types. We believe that SVM would be effective given that we were training a limited number of classes, and we were interested in getting high accuracy without regard for response times. We initially started to use SVM classifier, but switched to SVM-regression which was useful for getting a confidence level for each candidate bounding box in the second phase. We got poor results with the linear classifier, and switched to the polynomial kernel and got much better results.

Adaboost: In contrast to SVM, we tried Adaboost, because we were interested in what kind of accuracy numbers we would get for an algorithm that's supposed to be very fast. This also serves a simple baseline implementation.

Viola-Jones: Finally, we attempted to use Viola-Jones algorithm but were not able to tune a good model. We were unable to get past even a single cascade, but we suspect this is because our labels are not tight enough around the clothing since they all contain a nontrivial amount of background and noise.

III.a) Methodology

We primarily compare SVM to the baseline Adaboost model, and use the McNemar's test to compare the results.

	Clothing Recognition	Locating Clothing Item	McNemar's Test
SVM	✓	✓	Adaboost
Adaboost	✓	-	SVM

Table 1: Methodology Summary

We use the same dataset for each algorithm, however, we do not use Adaboost to locate the clothing item within the image, because the classifier does not provide a confidence level to compare the candidate bounding boxes.

We are unable to compare our results to related work in the literature, because this is the first of its kind study that is attempting to classify a few set of broad clothing types. To our knowledge, the next best result came from [6] where over 15 clothing classes were classified on cluttered backgrounds with a best accuracy of 41%.

Model Selection

We ran our data using two different kernels, polynomial kernels and linear kernels. While both kernels performed reasonably well, we chose the least complex model that yielded the highest accuracy.

Polynomial kernels did perform marginally better for all classes and the best j parameters corresponded to the ratio of negative to positive examples.

III.b) Results and Discussion

Overall, both classifiers performs well, with most accuracy, precision, recall, and f-scores in the 80% to 90% range.

Our SVM classifier performs as well as Adaboost in classifying clothing objects in images. Applying McNemar's test to the results from AdaBoost and SVM led to a p value of 0.372 and chi square value of 0.372 with degree 1. It is very unlikely that there is a significant difference between the results from AdaBoost and SVM, but both algorithms were so accurate that it's difficult to tell. Because both algorithms use the same features, it makes sense that they would give similar results. This indicates that our algorithm can be improved by increasing the number and type of features more than changing the algorithm that processed those features.

We also found that balancing the data set across the four classes significantly improved the recall. For example, we originally had one-third the number of dress and skirt examples, and were getting very low recall rates. By equalizing the number of skirts and dresses, the recall improved to mid-90% range.

		Accuracy	Precision	Recall	F-Score
SVM	Dress	97.96	97.30	94.74	96.00
	Pants	95.56	78.30	91.67	84.46
	Shirt	96.06	97.20	97.20	97.20
	Skirt	96.45	100	88.30	93.79
Adaboost	Dress	97.27	92.2	95.30	93.72
	Pants	95.2	92.6	90.10	91.33
	Shirt	92.3	95.4	92.30	93.82
	Skirt	96.7	93.9	95.30	94.59

Table 2: Accuracy, Precision, Recall for SVM and AdaBoost algorithms

We were able to create bounding boxes around the clothing items. Although we had fair accuracy, there is much more room for improvement, especially for pants and skirts. For example, the pants bounding boxes are very thin and narrow (see Appendix I). This is probably from the nature of the training data, where a lot of the images were cut off above the model's knee level. Accuracy for skirts was also quite low, and we believe there was not enough differentiation in the HOG features between skirts and shirts.

	Average Overlap Ratio	Sample Size
Dress	1.153	72.00
Pants	2.889	45.00
Shirt	1.39	41.00
Skirt	2.16	43.00

Table 3: Overlap Ratio across the classes

In Table 3, we can see that the average overlap of the predicted bounding boxes align well with the labeled bounding box for dresses and shirts. A value of “1” indicates a perfect overlap, values under “1” indicate the predicted boxes have a smaller overlap region, and values over “1” indicate the predicted bounding boxes encapsulate the labeled region. In this case, we see that on average the predicted bounding boxes for skirts are twice as large the labeled bounding box.

IV. Related Work

The majority of related work focus is on specific clothing classes such as coats or dresses and it is only recently that generic clothing recognition is being confronted. The approach we present in this paper tackles on four category-level classes that have the ability to be dwindled down into special clothing subclasses.

In a paper from eBay Labs [5], the team was able to provide fine-grained attribute annotations specific to coat and jacket styles. Examples include the material the coat was made of, the collar style, fitting, length, and fastener style. They also noted a matrix of attributes cooccurrences, where it was shown that some pairs of attributes have higher occurrence rates. An example is where a clothing with leather-like features is likely to have a zip fastener, versus a button fastener.

In [3], Kalantidis et al. presented an approach to a clothing recommendation system using pose estimation and clustering to segment promising image regions and image retrieval techniques to retrieve visually similar products. Pose estimation was done by detecting a person and their given body parts. In order to achieve faster speed, more focus was put towards image regions likely to contain clothing, thus they calculated a prior probability map of clothing appearance, of which was applied clothing masks to detect clothing. Classification was then based on nearest neighbor matching given the segmented images represented as binary vectors.

Some work has been primarily focused on creating an efficient and automatic pipeline for recognizing and classifying people’s clothing in natural scenes. In [6], gathering data was automated using web crawlers and random forests were used to reduce noisy images derived from the webcrawler. They outperformed an SVM baseline with 41.38% vs 35.07% average accuracy. Their process first includes an upper body detection algorithm being applied to the image. Then they densely extract a number of features including HOG, SURF, LBG, Self-Similarity, and color information. These were used as input for a Random Forest type classification and SVM’s attribute classification.

V. Future Work

Long-term improvements for the classifier would allow for the identification of articles of clothing even when those articles are in front of busy, multicolored backgrounds or occluded. We could use computer vision algorithms such as canny edge detection to first find the person or person in the image, before separating the background from the person. Also, articles of clothing should be identifiable even in rotated pictures or pictures in which the subject is not upright using more detailed shape recognition. One possible method for resolving this is to use Hough shape approximation for abstract shapes. With these improvements, everyday images from Facebook could be processed.

A finer grained model that takes into account additional features besides HOG could improve accuracy. HOG is limited because it depends solely on the gradient information of an image and not other important attributes such as color. Sometimes, the gradient information in an image is ambiguous, and the same gradient may correspond to different curves. We found this to be the case in certain images, where the head was mistaken as a possible article of clothing, perhaps because the two locks of hair resembled a person's arms and the face resembled a shirt. Therefore, only using gradient information may not be enough to classify an image. But if additional features were used, such as an algorithm that detected skin tones, combined with face detection to find the body, the bounding boxes detected would be more accurate. Additionally, shape detection using the adaptive hough transform could help provide a second opinion in the search and classification of the articles of clothing. Most pieces of clothes have a general shape to them, for example loose dresses often look triangular versus a loose shirt would look more rectangular. Therefore, when locating or classifying an image based on gradients or color fails, shape may provide a viable solution.

Although our classifier performs well, one can also incorporate logic in the model that accounts for mutually exclusive clothing items. For example, dresses are not typically worn with pants, skirts are not worn with pants, and dresses are not worn with skirts.

One can also see how this model can serve as a baseline for more sophisticated models and different subtypes for each class. For example, once a shirt is identified different types of shirts can be classified, such as 'polos' or 'button-downs.'

Once the articles of clothing are identified, color and texture analysis can be done to find more relevant fashion information. Gabor filters can be used for the texture analysis. This type of analysis could simplify the job of human fashion trend analysts.

VI. Conclusion

A high accuracy in detecting the basic articles of clothing suggests that a similar effort with more specific types of clothing might be effective. HOG features seem to be effective in distinguishing different types of clothing. Also, our bounding boxes can narrow down where in the image an article of clothing appears, which could be useful in preprocessing images before running a more complicated analysis on the fashion of the person depicted.

APPENDIX I: Bounding Boxes

Figure 4: Top bounding box results for shirts (row 1), pants (row 2), dresses (row 3), and skirts (row 4)

APPENDIX II: Resources

[1] "Shape Detection Using the Adaptive Hough Transform."
http://link.springer.com/chapter/10.1007/978-3-642-83325-0_8

[2] “Histograms of Oriented Gradients for Human Detection”
http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf

[3] “Getting the Look: Clothing Recognition and Segmentation for Automatic Product Suggestions.” http://www.image.ece.ntua.gr/papers/774.pdf

[4] “Efficient Graph-Based Image Segmentation”
http://www.cs.cornell.edu/~dph/papers/seg-ijcv.pdf

[5] “Style Finder: Fine-Grained Clothing Style Recognition and Retrieval”
http://labs.ebay.com/wp-content/uploads/2013/06/cvpr2013_style_detection_mv3.pdf

[6] “Apparel Classification with Style.”
http://www.vision.ee.ethz.ch/publications/papers/proceedings/eth_biwi_00974.pdf