Applying Transfer Learning to Recognize Clothing Patterns Using a Finger-Mounted Camera

Lee Stearns[1], Leah Findlater[2], Jon E. Froehlich[2]
[1]University of Maryland, College Park; [2]University of Washington, Seattle
lstearns@umd.edu; {leahkf, jfroehli}@uw.edu

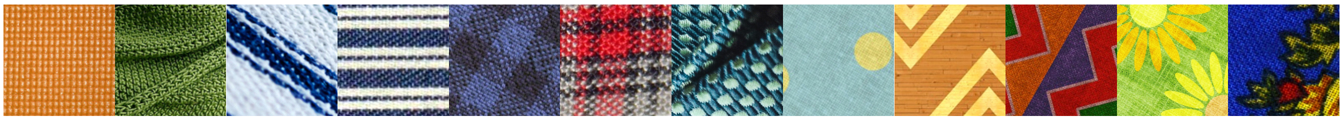

Figure 1. Examples of the six classes in our fabric pattern dataset: solid, striped, checkered, dotted, zigzag, and floral.

ABSTRACT
Color identification tools do not identify visual patterns or allow users to quickly inspect multiple locations, which are both important for identifying clothing. We are exploring the use of a finger-based camera that allows users to query clothing colors and patterns by touch. Previously, we demonstrated the feasibility of this approach using a small, highly-controlled dataset and combining two image classification techniques commonly used for object recognition. Here, to improve scalability and robustness, we collect a dataset of fabric images from online sources and apply transfer learning to train an end-to-end deep neural network to recognize visual patterns. This new approach achieves 92% accuracy in a general case and 97% when tuned for images from a finger-mounted camera.

Author Keywords
Visually impaired users; wearables; pattern recognition

ACM Classification Keywords
Human-centered computing → Accessibility technologies

INTRODUCTION
While color identification tools for users with visual impairments are widely available (e.g., [1,5]), they do not identify visual patterns or allow users to quickly inspect multiple locations—both of which are important for recognizing clothing [2]. Showing promise for more advanced clothing pattern identification, Yuan *et al.* [12–14] developed a system to identify 4 patterns and 11 colors in images captured with a mobile phone or head-mounted camera. Blind users responded positively to the system, although more detailed identification was desired. The interaction was also inefficient, requiring the user to hold out the clothing in front of them and use speech input to individually capture each still image to be classified.

ASSETS '18, October 22–24, 2018, Galway, Ireland

ACM ISBN 978-1-4503-5650-3/18/10.
https://doi.org/10.1145/3234695.3241015

In contrast, we are investigating a finger-mounted camera approach that allows users to move their finger across an article of clothing, combining tactile information with continuous audio description of the fabric's appearance (Figure 2). In previous work [7], we captured 520 images across 29 articles of clothing using this finger-mounted approach. We combined two complementary image features commonly used for object recognition [3] to train a support vector machine, achieving high classification accuracy (99.4%). While useful for demonstrating feasibility, the dataset was highly controlled, which risks overfitting, and the training process was not easily scalable.

In this paper, we build a larger and more varied dataset of images from online sources (Figure 1), which should allow our system to identify previously *unseen* fabrics—for example, to support shopping for new clothes. To assess whether this Internet-based dataset can be used to identify patterns in images collected with our finger-mounted system, we apply transfer learning to train an end-to-end deep neural network and test with the previously collected finger-mounted camera images. Our contributions include: (i) a novel dataset of 77,052 fabric images across six visual patterns; (ii) results of applying transfer learning to train a convolutional neural network for classifying fabric patterns; (iii) a preliminary system to describe clothing appearance in realtime using images from a finger-mounted camera. Because this work is still at an early stage, we also discuss open questions and challenges important for future systems.

DATA COLLECTION AND DATASETS
Existing texture datasets include textures that can easily be distinguished by touch or that are not relevant to clothing. For example, the *Describable Textures Dataset* [3] includes *braided* and *frilly*, and our prior dataset [7] includes *denim*,

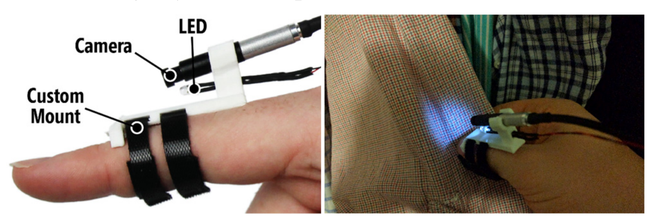

Figure 2. Prototype wearable camera and LED system, the same used in previous work [7]: (a) Close-up view of system, (b) Identifying an article of clothing.

knitted, and *lacelike*. While automatic identification of these textures may be useful to avoid misclassifications, in general they are not necessary to assist blind users. Instead, we selected six common visual patterns that are difficult or impossible to distinguish by touch alone: *solid*, *striped*, *checkered*, *dotted*, *zigzag*, and *floral* (Figure 1).

To create our dataset, we added the word "fabric" after each class name and downloaded the top 1000 search results from Google Images.[1] After one author manually removed erroneous results and cropped others as necessary (*e.g.*, to remove logos or background imagery), the dataset contained between 317 and 584 images per class (2764 images total). We augmented this data using a standard process to increase the training set size and improve robustness [10], rotating each image in 30-degree increments and cropping the center at multiple scales (1–4 depending on the resolution of the original image), resulting in 8232–17,304 samples per class, or 77,052 images total[2].

ALGORITHMS AND VALIDATION

To identify textures, we repurposed a state-of-the-art convolutional neural network model (ResNet-101 [6]) that was pre-trained on the *ImageNet* object dataset [9]. Using a standard transfer learning approach to avoid overfitting when insufficient data is available [4], we fixed all layers except for the final densely connected classification layer, and trained the weights for that layer using our dataset. To ensure that each class was equally likely when training, we randomly sampled 6400 images per class for the training set and 1600 images for a test set, discarding the rest.

Classification accuracy on the test set was 91.7%, suggesting that this approach should work well in general. On our smaller finger-mounted camera dataset [7], which contained 400 images across the six classes and included variations in distance, rotation, perspective, and fabric tension, accuracy was 72.8%. Most errors were caused by confusion due to insufficient context or coarse threads (*e.g.*, Figure 3). For example, *zigzag* was the worst performing class, likely because the camera's proximity to the fabric obscured much of the pattern. Roughly 14.5% of images were also misclassified as *checkered* or *floral*, most likely due to confusion from the coarse threads. Half of the images were captured with the finger-worn camera held 5cm from the fabric, while the other half were captured from a distance of 12cm; if only the latter images are considered (N=200), accuracy rises to 78.0%. Finally, continuing the training process to fine-tune the classifier using roughly half of the finger-camera images (N=36 per class) increases accuracy to 96.5%, suggesting additional images from the target domain can boost performance.

ONGOING WORK

In addition to identifying patterns, our real-time system identifies the dominant color in each image. To do so, we

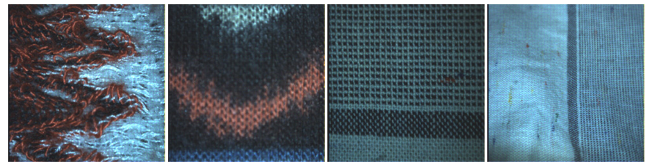

Figure 3. Example misclassifications (actual→predicted)

use k-means clustering with a dynamic number of clusters selected using the "jump method" [11], which searches for the point at which additional clusters stop greatly reducing error. We name each cluster center using the XKCD color survey results [8] to provide commonly accepted names for 48 RGB color values. The level of detail is user-configurable, including the number and complexity of the colors named in each image (*e.g.*, "green, purple, brown", or "lime green, lilac, beige"). Users can also identify multiple colors by moving their finger across the fabric. We did not evaluate color accuracy but if properly calibrated, our finger-worn approach should mitigate issues with lighting and distance that impact existing color identifiers.

The system runs at approximately four frames per second on a desktop computer.[3] The current implementation tracks classification results for the most recent two seconds. To reduce noise from misclassifications, patterns are reported by majority vote, with unclear results labeled "unknown". Colors are only reported if they are named consistently across frames—for example, results of "blue and light blue" and "blue and gray" would be reported simply as "blue". Users can press a button to hear the most recent result via text to speech, or hold for continuous updates.

DISCUSSION AND FUTURE WORK

Our pattern classification accuracy when fine-tuned with images from a finger-mounted camera is similar to our prior work (97% vs 99%), but our end-to-end deep learning approach reduces training overhead and should be much more generalizable and robust. Our use of a single online source does introduce some bias in our dataset, which future work should investigate. And to mitigate errors caused by lack of context or distracting details (*e.g.*, coarse threads), the camera should likely be positioned farther back on the user's finger or wrist. This change would still allow users to easily query multiple locations and combine automated feedback with their own sense of touch.

Future work will need to investigate performance and usability with visually impaired users, and assess the potential benefits of our approach compared to existing aids. Additionally, how best to convey complex color and pattern information to users is still an open question.

ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

[1] https://github.com/hardikvasa/google-images-download

[2] https://github.com/lstearns86/clothing-pattern-dataset

[3] Dell Precision Workstation, dual Intel Xeon CPU @ 2.1 GHz, NVIDIA GeForce GTX 1080

REFERENCES

1. Brytech. Brytech Color Teller. Retrieved from http://www.brytech.com/colorteller/
2. Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P Bigham, and Amy Hurst. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. In *Proceedings of ASSETS 2012*, 135–142. https://doi.org/10.1145/2384916.2384941
3. Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. 2014. Describing Textures in the Wild. In *Proceedings of CVPR 2014*, 3606–3613.
4. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: a deep convolutional activation feature for generic visual recognition. In *Proceedings of ICML 2014*, I-647. Retrieved June 13, 2018 from https://dl.acm.org/citation.cfm?id=3044879
5. GreenGar Studios. Color Identifier. Retrieved from https://itunes.apple.com/us/app/color-identifier/id363346987?mt=8
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of CVPR 2016*, 770–778. https://doi.org/10.1109/CVPR.2016.90
7. Alexander J. Medeiros, Lee Stearns, Leah Findlater, Chuan Chen, and Jon E. Froehlich. 2017. Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation. In *Proceedings of ASSETS 2017*, 393–394. https://doi.org/10.1145/3132525.3134805
8. Randall Munroe. 2010. Color Survey Results. Retrieved from https://blog.xkcd.com/2010/05/03/color-survey-results/
9. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3: 211–252. https://doi.org/10.1007/s11263-015-0816-y
10. Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved June 12, 2018 from http://arxiv.org/abs/1409.1556
11. Catherine A Sugar and Gareth M James. 2003. Finding the Number of Clusters in a Dataset. *Journal of the American Statistical Association* 98, 463: 750–763. https://doi.org/10.1198/016214503000000666
12. Yingli Tian and Shuai Yuan. 2010. Clothes Matching for Blind and Color Blind People. In *Proceedings of ICCHP 2010*, 324–331. https://doi.org/10.1007/978-3-642-14100-3_48
13. Xiaodong Yang, Shuai Yuan, and YingLi Tian. 2014. Assistive Clothing Pattern Recognition for Visually Impaired People. *IEEE Transactions on Human-Machine Systems* 44, 2: 234–243. https://doi.org/10.1109/THMS.2014.2302814
14. Shuai Yuan, YingLi Tian, and Aries Arditi. 2011. Clothing matching for visually impaired persons. *Technology and Disability* 23, 2: 75–85. https://doi.org/10.3233/TAD-2011-0313