An Automated Tool for Creating Clothing Catalog Databases: MyEyes–Fashion

Mariana Carvalho, Daniel Rocha, and Vítor Carvalho(✉)

2Ai - School of Technology, IPCA, Barcelos, Portugal
vcarvalho@ipca.pt

Abstract. In the field of object identification and segmentation the development of automated dataset creation tools holds significant importance. Recognizing the necessity for abundant data, these tools serve to lighten the manual load of data organization, amplifying efficiency and enabling researchers to focus on the essence of their analysis. Integrated into a project that aims to improve the quality of life and well-being of blind people, this paper presents the steps of creating a tool capable of facilitating the data augmentation of databases to feed artificial intelligence algorithms in the identification and extraction of its features. Using a computer vision detection system, a tool was developed to identify and extract garments and catalog them in a database. For user interaction it was created an interface that allows the insertion of elements in the database and the extraction of necessary resources for easy dataset creation. In the segmentation and classification algorithm of the garments, it was obtained an accuracy between 75% and 100% in its recognition.

Keywords: Blind People · Garments · Image Segmentation · Neural Networks · Deep Learning

1 Introduction

In the dynamic landscape of machine learning and artificial intelligence, the pivotal role of training data quality and quantity in determining model success cannot be overstated. The performance of these models is intrinsically linked to the caliber of the data they are trained upon. A promising solution that has emerged is automated data augmentation – a process that not only addresses data scarcity but also amplifies a model's capacity to generalize across diverse scenarios. As industries enthusiastically adopt these cutting-edge technologies, the demand for precise and resilient models has witnessed an exponential surge. However, this heightened demand is met with a significant impediment – the insufficiency of diverse and comprehensive datasets. The acquisition, curation, and annotation of such datasets pose substantial challenges, demanding significant financial resources and time commitments. This predicament is particularly accentuated in fields where data collection is financially demanding, time-intensive, or ethically intricate. Consequently, the scarcity of meticulously labeled, high-caliber data persists as a stumbling

E. Brooks et al. (Eds.): DLI 2023, LNICST 589, pp. 49–61, 2024.
https://doi.org/10.1007/978-3-031-67307-8_5

block in the progress of machine learning applications. This is precisely where the inception of an automated data augmentation tool takes center stage, holding the promise of a transformative solution. This paper outlines the development of a tool designed to extract information about clothing from images and systematically catalog them within a database. This advanced system is a pivotal component of an ongoing project aimed at enhancing the quality of life for blind individuals, particularly in terms of improving their ability to manage their clothing [1–9]. It addresses a critical gap in the project - the lack of suitable datasets that meet its specific requirements. The tool's primary purpose is to create larger and more comprehensive datasets, which serve as a critical resource by providing valuable inputs that contribute to the enhancement of algorithms utilized in the ongoing research initiative [8, 9]. This endeavor is motivated by the long-standing recognition of a pressing need within the blind community, a need that extends to the broader project, in which this tool plays a crucial role, as blind individuals have experienced persistent challenges in accessing clothing-related information autonomously [7]. As a result, the primary overarching goal is to significantly elevate the quality of life for blind users by addressing this need, providing them valuable information, and enabling confident and independent clothing choices.

The paper is structured into five sections. Section 2 covers the related work, Sect. 3 provides an overview of the project, Sect. 4 explains the methodology, Sect. 5 presents the experimental results, and Sect. 6 concludes with the final remarks.

2 Related Work

This project is associated with the My Eyes initiative [1–9], which is focused on nurturing self-sufficiency in the everyday activities of individuals with visual impairments, particularly in the context of selecting their attire.

Regarding the support of blind people in their selection of garments, we can find projects [10, 11] that can, overall, identify 11 colors (red, orange, yellow, green, blue, cyan, purple, pink, black, grey, and white) and 4 patterns (plaid, striped, pattern less and irregular). The results are given in the form of sound. In assistance to online shopping, it was developed, in 2018, the BrowseWithMe tool [12], which aids blind people gain independence and progress from passive listeners to active seekers of information. Through artificial intelligence, computer vision, and natural language processing, a product's web page is converted to a structured representation that allows the user to ask the assistant for the information he/she is seeking (e.g., price, color, description, etc.). Regarding clothing combinations, the development of a web page platform is described in [12]. This project utilizes Near Field Communication for garments identification. The clothing features are inserted manually, and it is possible to automatically identify the color of the garment. For future work, optimizations such as the automatic insertion of garments and characteristics extraction, resorting to a camera, image processing, and automatic learning algorithms, are taking place.

In the identification of garments, Yamaguchi et al. [13] propose a method of clothes analysis able to classify 53 different categories and segment each piece of clothing by exploiting super-pixels and model of flexible pieces for human pose estimation. To achieve this, the process involves retrieving similar images from the analyzed image database. Subsequently, the extracted images and tags are used to analyze the image under examination.

Using Feature Pyramid Networks, a neural network model developed by Martinsson and Mogren [14] which performs the semantic segmentation of fashion images, that can be used in conjunction with Conditional Random Fields. The model learns which body parts of the human body in which different clothing items are typically placed on and, although demonstrating good results, it shows some difficulties in distinguishing between garments used on the same body part, for example, a t-shirt and a blouse.

A model of Human-centric Clothing Segmentation via a Deformable Semantic Locality-preserving Network [15] puts together the original image and one semantically similar in the dataset, being that by having the same pose or appearance. Using a Convolutional Neural Network model based on the deformable convolutions this model can extract the non-rigid features of the clothing.

A system for garment segmentation and color classification [16] was created, given an image taken from an online fashion store - background removal is performed to obtain a binary mask. The photo type detection module first classifies the masks according to the shooting type. Consequently, according to the shooting type, both skin and additional garments and accessories are removed to obtain a clear picture of the object of interest. Finally, a garment color descriptor or HOG-based descriptors of garment shapes and textures are computed on the selected object and used for color retrieval or classification.

Using CNNs, Cychnerski et al. [17], developed a computer vision system for accurate detection and classification of clothes for e-commerce images. Utilizing the DeepFashion dataset, which contains box annotations for locations of clothes. Automatic gathering of labels resulted in an average of 83% rate of correct labels.

Kayed et al. [18], used a CNN based LeNet-5 architecture to train parameters of the CNN on the Fashion MNIST dataset, showing results of accuracy of over 98%. Outperforming both the classical CNN model and the other existing state-of-the-art models in literature.

A survey paper on Image Data Augmentation for Deep Learning [19] explains that Deep convolutional neural networks are heavily reliant on big data to avoid overfitting. Overfitting refers to the phenomenon when a network learns a function with very high variance such as to perfectly model the training data. Data Augmentation encompasses a suite of techniques that enhance the size and quality of training datasets such that better Deep Learning models can be built using them.

An article on Data Augmentation in Classification and Segmentation [20], recognizes that deep learning models may require a large amount of data to achieve satisfying results. Unfortunately, the availability of sufficient amounts of data for real-world problems is not always possible, and it is well recognized that a shortage of data easily results in overfitting. This issue may be addressed through several approaches, one of which is data augmentation. The paper examines classification and augmentation techniques in computer vision tasks, including segmentation and classification, and suggests new strategies.

From the research undertaken it becomes evident that the development of tools aimed at assisting the blind community is still below the anticipated and necessary level.

Although garments detection systems have shown promising advancements, they continue to encounter challenges, particularly with confusion between different categories of clothing. One major contributing factor to this issue is the scarcity of high-quality garments datasets with accurate annotations. In response to this challenge came the development of this study.

3 Project Overview

The main objective of this work is the creation of a tool that allows the extraction of garments from images and their cataloging in a database to feed artificial intelligence algorithms in the identification and extraction of its features. For that reason, the tool can create JSON files and save them in a folder, along with the original image. This can be done in two ways, the inputting of images or the retrieving of data from the existing database.

Using the JSON file, we can effortlessly obtain the segmented image through annotation tools, avoiding the need for manual point-by-point annotation as the alternative method.

Thus, the system is divided into two main stages the extraction of garments from images and their cataloging in a database and the retrieval of information from the database (Fig. 1).

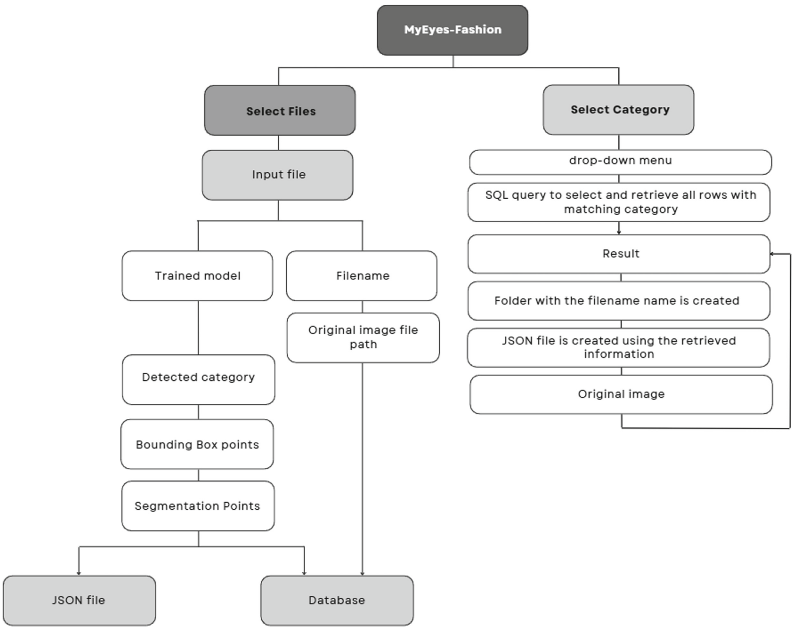

Fig. 1. Project Overview

4 Methodology

To be able to extract garments from images it became necessary to develop a dataset and a model capable of semantic segmentation.

Our dataset consists of 2000 images, distributed equally amongst 8 categories of garments, those being: dresses; jackets; pants; polos; shirts; shoes; shorts; and t-shirts. The 250 images per class are split into training and validation images with ratios of 0.8 and 0.2, respectively. A mix of images from the public dataset Fashion Product Images [21], as well as pictures taken in a home environment, were used. With those collected it was necessary to annotate every image, those were made resorting to the Roboflow tool [22]. The dataset was exported under the data format COCO JSON.

On the Google Colab platform [23], utilizing the Detectron2 [24] library, as well as others necessary for image manipulation, such as OpenCV, our model was trained under various conditions. A pre-trained Instance Segmentation model was utilized, "COCO-InstanceSegementation/mask_rcnn_R_50FPN_3x.yaml" through transfer learning.

Having as an objective the ability to catalog the garments extraction data, an algorithm in Python was written, and for that we used the *psycopg2* library, to connect to the PostgreSQL [25] database. The table includes the following attributes:

- "Id" – An auto-incrementing primary key column;
- "Category Id" – A integer column that stores the detected category in integer form;
- "Category" – A non-nullable VARCHAR column with a length of 45 characters to store the category name;
- "Filename" – A non-nullable VARCHAR column with a length of 45 characters column to store the file name;
- "BoundingBox" – A TEXT column to store the bounding box coordinates under array form;
- "OriginalImageFilePath" – A VARCHAR column with a length of 300 characters to store the file path of the original image;
- "SegmentationPoints" – A TEXT column to store the segmentation points.

To allow the upload of images, the performance of object detection and segmentation, and retrieval of data from a database based on the detected categories, a GUI was developed "MyEyes-Fashion", using the Tkinter library. The application includes a main menu with buttons to upload images for processing "Select Files", to select a category to retrieve data from the database "Select Category" and to delete incorrect information from the database "Delete Incorrect Data from Database".

Once the user uploads the images, our pre-trained model processes them. This allows us to acquire the detected class, the image with a bounding box and object segmentation, the predicted segmentation mask, along with the bounding box coordinates from the model's outputs. Using this information, we can perform further processing to obtain an image that contains only the detected object, along with the contour points necessary for image segmentation. Subsequently, a JSON file is generated, and all the previously mentioned information is stored in the database.

To access the information from the database, a drop-down menu was implemented using the 'Menu' widget in the Tkinter GUI. The menu contains the following items:

"dress," "jacket," "pants," "polo," "shirt," "shoes," "short," and "t-shirt." Upon connecting to the database, we retrieve all items that match the selected category and create a corresponding JSON file, which is then saved alongside the original image.

Additionally, to ensure that the database only contains accurate information, an option has been implemented in the user interface. This option enables the user to delete image information based on its file name. By doing so, any incorrect detection can be easily identified and removed from the database. This functionality empowers the user to maintain a clean and reliable database by selectively removing data that may not meet the desired criteria. This way, the database remains accurate and relevant, allowing for improved performance and meaningful results in subsequent processes.

5 Experimental Results

Using the created dataset, without data augmentation, our model was trained, using the training parameters as follows: "ims_per_batch" set to 1, "base_lr" to 0.00025, "batch_size_per_image" set as 128. We aimed to explore the impact of different training iteration rounds on the model's performance. For each experiment, we varied the number of training iterations while keeping other settings constant. The results are presented in Table 1 and analyzed to understand the effect of training duration on the model's accuracy and convergence.

Table 1. Training model results (%) for different training iterations (7500, 10000, 30000).

Number of training iterations	7 500	10 000	30 000
mAP	77.40	80.76	85.10
mAP50	82.05	85.29	89.80
mAP75	81.26	84.25	88.56
Dress AP	76.64	76.13	76.52
Jacket AP	80.06	80.37	75.71
Pants AP	89.03	86.14	86.85
Polo AP	87.64	91.66	92.06
Shirt AP	47.20	67.97	80.00
Shoes AP	98.70	99.05	99.63
Short AP	76.86	78.20	86.70
T-Shirt AP	63.14	66.56	83.39
Total Loss	15.48	09.88	08.88
Accuracy	98.55	98.61	98.88

Observing Table 1 one can see that the model underwent training using different iteration rounds, and as the number of iterations increases, noticeable improvements were observed. Among the various iterations, the model trained with 30000 iterations stood out as the chosen one for algorithm development due to its consistently good results. It achieved a mean average precision (mAP) of 85.1%, with precision levels ranging from 75.71% for jackets to 99.63% for shoes.

Using this model we can obtain the category of the detected garment, the bounding box, and object only images; some examples of the obtained results are shown in Fig. 2.

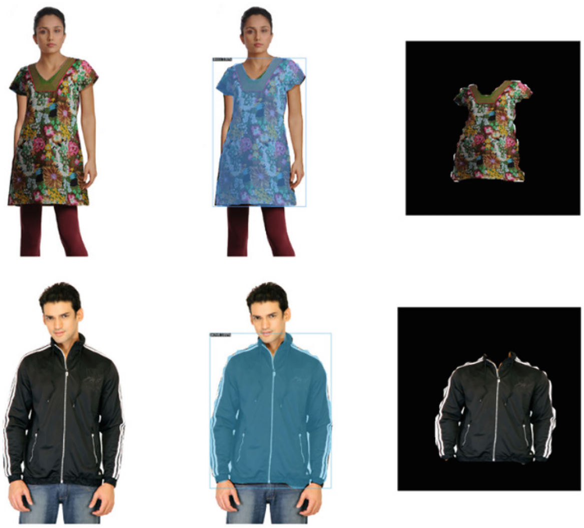

Fig. 2. Results images. Left to right: original, bounding box, and object only images. Examples for dresses and jackets.

While not common, some confusion occurs when the input image features a woman wearing a t-shirt or a shirt, leading the model to misidentify them as a dress. Additionally, the model occasionally confuses long shorts with pants, contributing to further instances of confusion.

As the developed database serves as an efficient repository, effectively storing all the essential data required for the project, we can interact with the database and retrieve vital information. Such as:

- Detected category id;
- Detected category;
- Filename;
- Bounding Box points;
- Original Image File Path;
- Segmentation points.

By having this stored data, accessing all the necessary information to create a JSON file becomes effortless.

Regarding our interface, on the main page, our GUI incorporates three buttons: "Select Category," "Select Files," and "Delete Incorrect Information from Database", Fig. 3. These buttons offer users distinct functionalities to choose garment categories for image retrieval as select input images for processing, and delete inaccurate data from the database, respectively.

Fig. 3. GUI main page.

When the user clicks the "Select Files" button, a pop-up window appears, Fig. 4, providing the user with the option to select images. This feature enables users to choose one or multiple images from their local storage for further processing and analysis within the application.

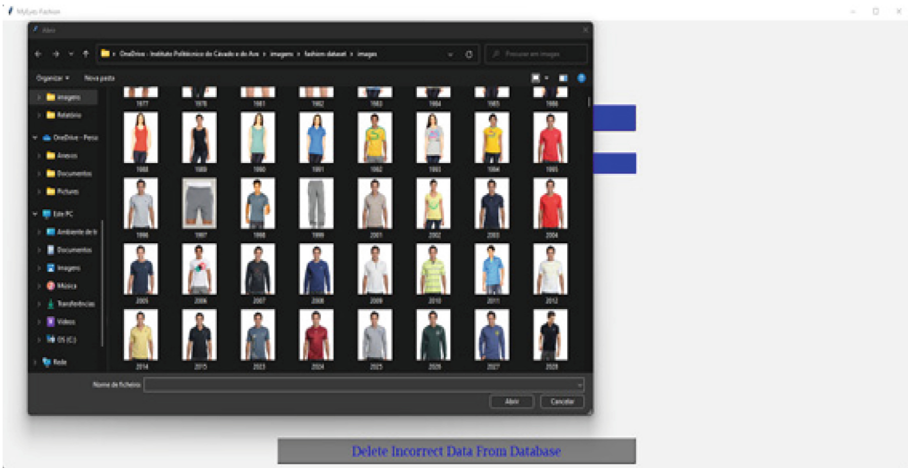

Fig. 4. Pop-up window after using the "Select Files" button.

Once the user selects files, the application processes them by identifying the category of garment detected and extracting the bounding box coordinates for each image.

Furthermore, a JSON file is generated. The information is saved on the database. The original, bounding box, and object only images and class are displayed on the interface, as can be visualized in Fig. 5.

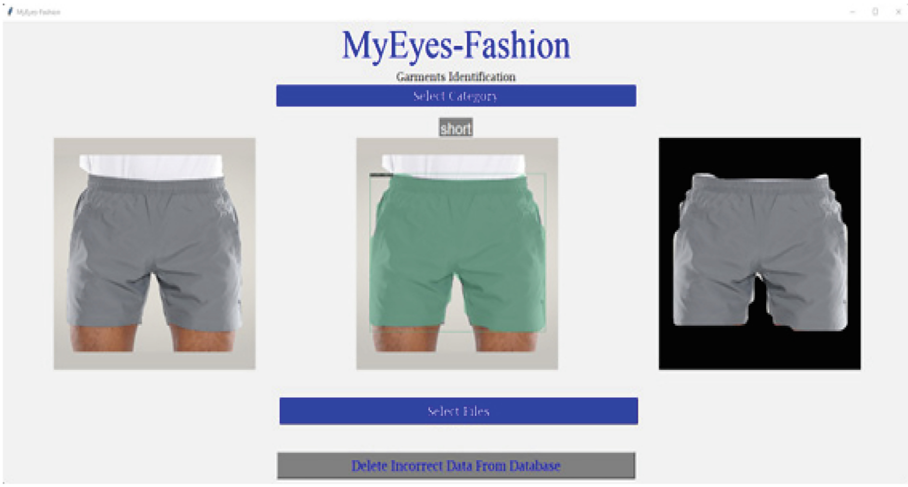

Fig. 5. Interface displaying the detected class and, from left to right, the original, bounding box, and object only images

When the user clicks the "Select Category" button, a drop-down menu appears, presenting all the available garment categories for selection. From this menu, the user can choose a specific category of interest, Fig. 6. Upon selection, the application creates the corresponding folders containing the original images and JSON files associated with that category.

Fig. 6. Interface displaying the drop-down menu for category selection.

For each element with a matching category on the database, a new folder is created with the filename, and a copy of the original image and the corresponding created JSON file are stored inside this folder.

When the user clicks the “Delete Incorrect Data from Database” button, an entry widget is generated, accompanied by a “Confirm Delete” button (as shown in Fig. 7). Within this entry widget, the user can input a specific filename associated with the data they wish to remove. Upon clicking the “Confirm Delete” button, the application initiates the deletion process, removing the data corresponding to the provided filename from the database.

Fig. 7. Interface modifications when the “Delete Incorrect Data from Database” button is used

The JSON files created through both the “Select Files” and “Select Category” processes follow the same approach. The resulting JSON file serves as a valuable input for annotation tools. By using this JSON file in conjunction with the original image, users can efficiently annotate images, simplifying the data augmentation process. The result of applying this can be observed in Fig. 8.

Leveraging the JSON file and employing annotation tools presents a more efficient and streamlined solution. The JSON file conveniently contains all the necessary information, such as object segmentation points, enabling annotation tools to automatically generate the data annotation.

In contrast, opting for the alternative approach would entail resorting to manual point-by-point annotation, a meticulous and time-consuming process. This method would require identifying and marking each point or region within the image, which can prove to be highly laborious. Moreover, as the complexity and scale of the dataset grow, the effort and time invested in this manual annotation process would increase exponentially, potentially becoming impractical or unfeasible.

Fig. 8. Result of the use of JSON file.

6 Final Remarks

Part of a larger project dedicated to assisting visually impaired individuals, the primary objective of this initiative was to create a tool that simplifies the augmentation of image datasets. This was achieved by isolating clothing items within images and organizing them into a comprehensive database. Subsequently, this enhanced dataset formed the basis for training new algorithms crucial to the development of a tool tailored to the specific needs of the visually impaired community. These algorithms are instrumental in the identification and extraction of vital features. They are an integral part of an application designed to improve the daily lives of blind individuals by providing them with independent clothing management. This, in turn, significantly impacts their self-esteem, allowing them to make confident and autonomous choices about their attire. Ensuring that the blind community can access detailed information about their clothing, ultimately empowering them to take greater control of their wardrobe and, by extension, boosting their self-confidence. This was achieved by extracting garments from images and cataloging them in a database. Subsequently, this enriched dataset served as the foundation for training new algorithms integral to the development of a tool catered to the needs of the blind community. These algorithms facilitate the identification and extraction of essential features to enhance the user experience. The results obtained show that our model obtained precision levels on a spectrum from 75.7% for jackets to 99.6% for shoes. The system can store and retrieving all the information regarding an object's segmentation on a database and creating JSON files that will facilitate the augmentation of datasets.

In the scope of future work, one of our primary goals is to develop an autonomous learning system that continuously improves itself by recognizing new categories and learning from its mistakes. Additionally, we plan to adapt the current user interface into a web-based platform. This transition to a web page will enable seamless access and interaction with the tool across various devices and platforms, promoting greater accessibility and convenience for users.

Acknowledgements. This work was funded by the project NORTE-01–0145-FEDER-000042, supported by Northern Portugal Regional Operational Program (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). It was also funded by national funds, through the FCT – Foundation for Science and Technology and FCT/MCTES in the scope of the project UIDB/05549/2020 and UIDP/05549/2020.

References

1. Rocha, D., Carvalho, V., Oliveira, E., Gonçalves, J., Azevedo, F.: MyEyes-automatic combination system of clothing parts to blind people: first insights. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1−5. IEEE (2017)
2. Rocha, D., Carvalho, V., Gonçalves, J., Azevedo, F., Oliveira, E.: Development of an automatic combination system of clothing parts for blind people: myeyes. Sensors Transducers (2018)
3. Rocha, D., Carvalho, V., Oliveira, E., Gonçalves, J.: MyEyes - automatic combination system of clothing parts for blind people: prototype validation. In IEEE (2017)
4. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: Extracting clothing features for blind people using image processing and machine learning techniques: first insights. In: Tavares, J.M.R.S., Natal Jorge, R.M. (eds.) VipIMAGE 2019. LNCVB, vol. 34, pp. 411–418. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32040-9_42
5. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: Design of a smart mechatronic system to combine garments for blind people: first insights. In: HealthyIoT 2019 - 6th EAI International Conference on IoT Technologies for HealthCare (2019)
6. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: A model approach for an automatic clothing combination system for blind people. In: EAI DLI 2020 - 5th EAI International Conference on Design, Learning & Innovation (2020)
7. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: Understand the importance of garments identification and combination to blind people. In: IHIET (2021)
8. Rocha, D., Soares, F., Oliveira, E., Carvalho, V.: Blind people: clothing category classification and stain detection using transfer learning. Appl. Sci. **13**(3), 1925 (2023)
9. Rocha, D., Pinto, L., Machado, J., Soares, F., Carvalho, V.: Using object detection technology to identify defects in clothing for blind people. Sensors (2023)
10. Nagaraj, B., Francis, A.: Recognizing clothes patterns and colors for blind people using neural network. IEEE (2015)
11. Anuradha, S.G., Ashwini, T.: Clothing color and pattern recognition for impaired people. Int. J. Eng. Comput. Sci. (2016)
12. Stangl, A.J., Kothari, E., Jain, S.D., Yeh, T., Grauman, K., Gurari, D.: Browse with me: an online clothes shopping assistant for people with visual impairments. In: ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 107–118 (2018)
13. Yamaguchi, K., Kiapour, M., Ortiz, L., Berg, T.: Parsing clothing in fashion photographs. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2012)
14. Martinsson, J., Mogren, O.: Semantic segmentation of fashion images using feature pyramid networks. In: Proceedings - 2019 International Conference on Computer Vision Workshop, pp. 3133–3136 (2019)
15. Ji, W., Li, X., Wu, F., Pan, Z., Zhuang, Y.: Human-centric clothing segmentation via deformable semantic locality-preserving network. Trans. Circ. Syst. Video Technol. 4837–4848. IEEE (2020)

16. Manfredi, M., Grana, C., Calderara, S., Cucchiara, R.: A complete system for garment segmentation and color classification. Mach. Vis. Appl. **25**(4), 955–969 (2013). https://doi.org/10.1007/s00138-013-0580-3
17. Cychnerski, J., Brzeski, A., Boguszewski, A., Marmolowski, M., Trojanowicz, M.: Clothes detection and classification using convolutional neural networks. In: 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE (2017)
18. Kayed, M., Anter, A., Mohamed, H.: Classification of garments from fashion MNIST dataset using CNN LeNet-5 architecture. In: Proceedings of 2020 International Conference on Innovative Trends in Communication and Computer Engineering, pp. 238–243 (2020)
19. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 1–48 (2019). https://doi.org/10.1186/s40537-019-0197-0
20. Alomar, K., Aysel, H.I., Cai, X.: Data augmentation in classification and segmentation: a survey and new strategies. J. Imaging (2023)
21. FashionProductImages. https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset. Accessed 10 July 2023
22. Roboflow. https://roboflow.com/. Accessed 07 July 2023
23. Google Colab. https://colab.research.google.com/. Accessed 07 July 2023
24. Detectron2. https://ai.meta.com/tools/detectron2/. Accessed 10 July 2023
25. PostgreSQL. https://www.postgresql.org/. Accessed 10 July 2023