Assisting clothing pattern recognition for the visually impaired: A two-stage classifier approach

Danica Fiteni
Institute of ICT
Malta College or Arts,
Science and Technology
Paola, Malta
danica.fiteni.b31354@mcast.edu.mt

Daren Scerri
Institute of ICT
Malta College or Arts,
Science and Technology
Paola, Malta
daren.scerri@mcast.edu.mt

***Abstract*—Finding clothes to wear can be difficult for someone who has difficulty seeing clearly or is completely blind. In this paper, we develop a new computer-vision technology that will tell the types of clothes and whether they are plain or unplain. Given that they are vision impaired people, this can be used on a device with a webcam and a speaker. The output will be a video with voice speech. In order to accomplish this, we propose a two-stage classifier system that recognizes clothing types such as top and bottom in the first stage and combines the clothing texture type such as plain and unplain in the second. Our project achieved an average accuracy of 81.2 percent.**

***Index Terms*—Computer Vision, YOLOv7, Blindness, Object Detection, vision impairment, Project, Paper, Clothes textures, clothes types**

I. Introduction

At least 2.2 billion people have a short or far vision impairment. Around 1 billion of these cases could have been prevented or have yet to be addressed. This 1 billion includes people with moderate or severe distance vision impairment or blindness due to uncorrected refractive error (88.4 million), cataracts (94 million), age-related macular degeneration (8 million), glaucoma (7.7 million), diabetic retinopathy (3.9 million) (1), and near vision impairment due to uncorrected presbyopia (826 million) (2) [1]. Age-related macular degeneration, cataract, diabetic retinopathy, glaucoma, and uncorrected refractive errors are the main causes of visual impairment.

People must find appropriate clothing to wear on a daily basis. Finding clothes that are the right texture can be difficult for impaired vision people. Some people depend on others for what they can wear while others do not have anyone to depend on.

The aim of this study is to develop a solution to assist persons who have impaired vision in selecting clothing. People who have impaired vision can be instructed on what they are wearing via a camera and text speech with this type of technology. The hypothesis behind the research is to use computer vision to help blind people detect clothes types if they are plain or unplain. An automated clothing annotation is suitable with different clothing images containing different types of clothing/patterns which can be simply re-trained. Given the above hypothesis 3 research questions were identified:

1) Which dataset is ideal to detect different types of clothes?
2) How the algorithm works?
3) What is the performance of a clothes detection method based on automated clothes detection annotation datasets when compared to other computer vision algorithms?

II. Literature Review

A. Computer Vision

Artificial intelligence (AI) is now an essential part of our daily life. Computer vision is a field of AI, and the aim of computer vision is to identify and understand things in videos, digital images, and other visual inputs. In this situation, computer vision attempts to recreate both how humans see and how humans process what they see.

One important task of computer vision is object detection. It has to do with locating a region of interest inside an image and classifying this region in the same way that a standard image classifier does. One popular model of object detection is YOLOV7(You Only Look Once)." Yolov7 is a real-time object detector currently revolutionizing the computer vision industry with its incredible features. The official YOLOv7 provides unbelievable speed and accuracy compared to its previous versions. Yolov7 weights are trained using Microsoft's COCO dataset, and no pre-trained weights are used" [2]. Ross Girshick and his team at Microsoft Research in 2014 developed one of the earliest successful attempts to address the object detection problem using deep learning was the R-CNN (Regions with CNN features) model [3].

YOLO evolved throughout time for better and faster performance. YOLOV1 was the first YOLO and YOLOV8 is the most recent. The YOLO algorithm works by starting with an image/video and attempting to understand what is in that image/video. The goal of the YOLO method is to appropriately output bonding boxes for that specific object in the image/video. The YOLO algorithm first divides the image into SxS grids. As a result, each cell will output the prediction along with the matching bounding box. After it finds the object midway, the cells must increase the x and y coordinates such that each output and label is relative to the cell, and each bounding box for each cell has: [x, y, w, h] coordinates.

You Only Look Once (YOLO) is a state-of-the-art, real-time object detection algorithm introduced in 2015 by Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi in their famous research paper "You Only Look Once: Unified, Real-Time Object Detection" [4]. YOLO is popular because for its speed, detection, good generalization and its open-source. A study found out back in 2020 that YOLO was the fastest and beyond other object detectors with 91 FPS while the lowers was "Faster R-CNN" with a 5 FPS[4]. In figure 1 below there is a comparison of frames processed per second (FPS) implementing the Faster R-CNN, R-FCN, SSD, and YOLO models using input images with different resolutions

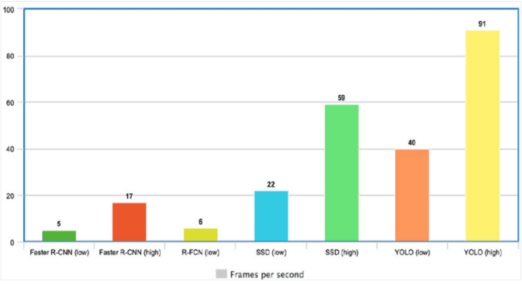


Fig. 1.

B. Clothes colors and pattern

Fashion has become an important part of our daily lives. Some people take it very seriously, while others do not. Mixing and matching patterns and colors is an art, but it must be done with caution because too many colors and patterns in the same clothing might seem unappealing. The ideal way to match patterns is to wear one solid color and one print, such as plain jeans and a dotted shirt.

There are different methods that can assist an individual in learning how to color-match clothes. They are the monochrome look, the contrasting combination, the soft combination, the neutral color costume, and the combination of neutral and complementary colors. The Monochrome Look is that in one outfit, one can wear 2-3 distinct shades of the same hue with varying levels of intensity and brightness. Another matching method is the contrasting combination is created by using colors that intensify each other, resulting in a strong contrast." Contrasting colors are those that are opposite each other on the color wheel. Some of the best contrasting colors are black and white, blue and orange, and green and purple. These color combinations are eye-catching and can make a statement in any room" [5]. Another one is the Soft Combination which uses soft colors like ochre yellow, burnt orange, sage, burgundy, and plum. Another method is the Neutral color outfit like shades of white, beige, camel, shades of grey, brown, and black. The last method is A mix of neutral and complementary colors, where one can combine neutral colors with complementary colors. "I know that there's this classic fashion rule that says you shouldn't't have more than 3 colors in one outfit. I don't think it's such great advice as it limits our self-expression. I believe you can use 1 color or 5 colors and still look amazing. Plus, a successful outfit does not only come from your ability to mix and match colors but also what style of clothes you match, how well everything fits you, and how confident you feel wearing it "(Wonder Wardrobe).

We used to be told that patterns and colors do not combine, but nowadays we see a wide range of colors and patterns. However, the smart dress rule is to wear two items of clothing with different patterns, such as plain and dotted or plain and striped, because wearing two types of clothing with the same pattern appears uninteresting and amateurish.

C. Help blind people recognize clothes using Computer Vision

[6] investigated how they could assist colorblind and blind people in recognizing their clothing. [6] used 3 datasets: 1) Brodatz album texture database; 2) color and texture matching (CTM) dataset for clothes; 3) texture detection (TD) dataset for clothes. "We have used it to compare our method to state-of-the-art methods. The CTM database contains 128 images of clothes with complex patterns, multiple colors, and lighting changes. The TD dataset contains 45 clothes images with different patterns and colors. Among these images, 23 of them have textures and 22 images without texture" [6].

TABLE I
COMPARISON WITH THE STATE-OF-THE-ART RESULTS OF TEXTURE MATCHING ON BRODATZ ALBUM DATASET AND CTM DATASET

Methods	Khouzani and Zadeh		Yingli Tian and Shuai Yuan	
Datasets	Brodatz	CTM	Brodatz	CTM
Accuracy	96.7%	67.7%	97%	82.7%

[6] developed a new approach for matching clothes with multiple colors and complex designs in order to assist visually impaired persons by discriminating texture and color information. To handle complex texture patterns and lighting fluctuations, they integrate the Radon transform, wavelet features, and co-occurrence matrix.

For their algorithm they developed a color classifier to detect multiple colors such as red, orange, yellow, green, cyan, blue, purple, pink, black, grey, and white and after they developed a simple edge-based texture detection method. Only images with texture patterns are put through to texture matching. Three databases assess the suggested clothes-matching method. Two of the databases contain images of clothing with various texture patterns, colors, and lighting changes. "The results demonstrate that our method is robust and accurate for clothes with complex patterns and multiple colours," they concluded.

In another study, [7] used MATLAB algorithm to implement pattern recognition and color identification. [7] had prepared two datasets: one for training and one for testing. The training dataset consisted of 627 sample data images with four different patterns: plaid, striped, pattern, and irregular. Each image has a resolution of 140 X 140 pixels. The testing datasets were expected to be recorded images from a camera-based system.

For the process of pattern and color identification, the author used a Graphical user interface screen that was developed in MATLAB where the images where imported from testing

datasets. MATLAB also enables direct audio output after processing, which is essential for blind persons to recognize the colour and pattern of their clothes.

Radon transforms provide image orientation in two forms: isotropic and anisotropic. Smooth principal orientations are used in pattern less images, plaid has two principal orientations, and stripped has one. Principle orientations are not exclusive to irregular patterns. (see Fig 1)

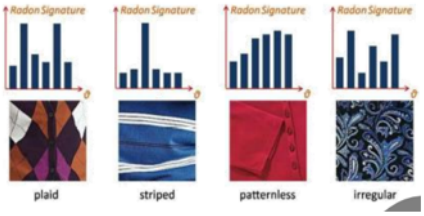


Fig. 2. Pattern and its Radon Signature

[7] stated that for cloth colour identification, " is carried out by normalized histogram of colour of each cloth image in to the HSI colour space. HIS stands for hue, intensity and saturation. As per the concept, colour space is quantized which is depending on the relations between hue, saturation, and intensity. Pixels in the image for each image are quantized by the system into 15 distinct colours as follows: white, grey, black, pink, magenta, purple, cyan, sky blue, blue, orange, yellow, yellowish green, green, light green and red. In case of cloth image having two or more than two colours, the dominant colours (the one having pixels larger than 5 percentage in the whole image) will be used as output". Finally, the results of sample photos with pattern and colour are shown in table 2.

TABLE II
SAMPLES IMAGES WITH PATTERN AND COLOR RESULT

Pattern	Plain	Striped	Patternless	Irregular
Colour	yellow(49%) orange(36%) black(9%)	blue(75%) white—(19%)	red(98%)	black(41%) red(26%) blue(6%) green(5%)

III. Research Methodology

The aim of this project is to create a working system that will automatically understand and label different types of clothes and patterns using a webcam. This study's research aims are as follows:

1) Two datasets including a variety of clothing images. One dataset will indicate if the clothing item is a top or a bottom, while the other will indicate whether the texture is plain or unplain.
2) Use a camera to recognize different styles of clothing by using a computer vision algorithm.

The Research Pipeline below indicates how it is supposed to work(Fig. 3)

A. Datasets

Makesense.ai was used to annotate the images. In this application, various image types were uploaded with different labels such as "plain," "unplain," and "top," "bottom." Object detection was the method that was used. After a bounding box was made with the correct label. Two-stage classifier projects with different data were created separately, one project for the texture of the clothing while the other for the type of clothing. Two folders named 'data-set' was created in the two projects, one for the images and the other for their labels. Each folder was then divided again into train and val. One project named 'plain and unplain' contains 196 images in the train folder and 47 images in the val folder while the other project named 'top and bottom' had 299 in the train folder and 80 in the val folder.

B. Model training

YOLOv7 was used to train the system to identify different clothes types and patterns. YOLO is trained on the COCO dataset. A YOLO model's architecture is shown in Figure 4. in contrast to others. Unlike other techniques that require two scans of an image, YOLO only needs to look at the input image once. As shown in Figure 5, YOLO models operate by dividing the image into a grid of S x S and then assigning probabilities for a bounding box to appear in a given grid cell. YOLO is a single-shot algorithm, which makes it very quick at identifying objects in an image. YOLOv7, the most recent iteration of the concept, is based on the bag of freebies model [8].

Since YOLOv7 is the fastest object detector, it was chosen for this project (Figure 1). The two projects simply created a new pre-trained weights for the two existing projects after being trained with the YOLO pre-trained weights. Actually, the new "best.pt" and "topbottom.pt" are the pre-trained model that were set up instead of the YOLO pre-trained weights. However, since the classifier path did not work, the two projects had to be combined into a single project with two different pre-trained weights. Since it was in one project, the parameters had to be doubled with different variable names.

C. Clothes detection

Following training,the YOLOv7 began identifying various types of clothing and patterns with labeled bounding boxes. Successfully, a label showing whether the clothing is top, bottom, plain, or unplain was printed in the left-hand corner of each image(Fig. 6).

IV. Findings & Discussion of Results

It is ideal to have the camera footage capture different impaired people wearing different clothes in order to successfully fulfill the goals of this study and directly address the key issue of developing clothes detection for impaired people. Since they cannot see, these results will also have speech output. As seen in Fig 7, the text in the left corner is what the program will need to generate speech; for example, in Fig 7 it did not

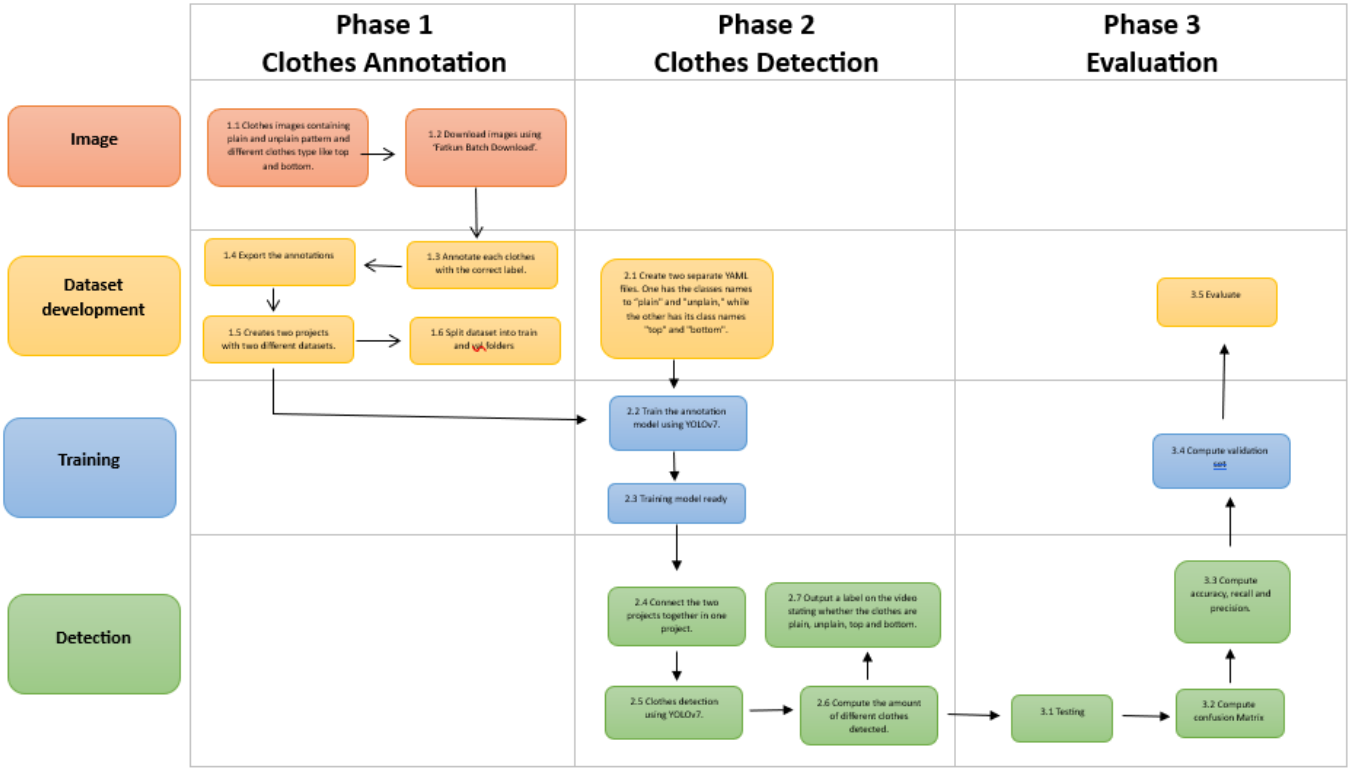


Fig. 3. Pipeline

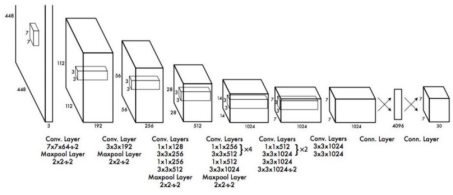


Fig. 4. YOLO model's architecture

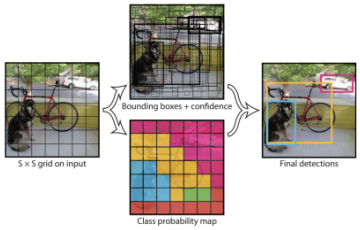


Fig. 5. Bounding Boxes

Fig. 6. Detecting the top and the texture

identify the texture of the top but did detect the plain bottom, therefore the result will be: The Top is and the Bottom is Plain

Two different impaired vision people were attempting to use the prototype in the following four scenarios. The textured top failed, but the unplain bottom was a success in footage 1. It only detected the plain top in footage 2. Footage 3 failed to recognize the unplain top and mistakenly identified it as a plain top, while the bottom failed to figure out whether it is plain or not. In footage 4, the textured top was successful while it missed the plain bottom. The algorithm only detects three annotations, as shown in Fig 7. The top and bottom are always detected while selecting one annotation for the texture, whether for the top or bottom.

Since the dataset can only detect dotted texture, the texture pattern is not being recognized, which is why video 1(fig. 7) did not detect an unplain top since it is not dotted, but video 4(fig. 7) did because it is dotted.

TABLE III
CONFUSION MATRIX

	Top	Bottom	Plain	Unplain
Top	4	0	0	0
Bottom	0	4	0	0
Plain	0	0	3	1
Unplain	1	1	0	2

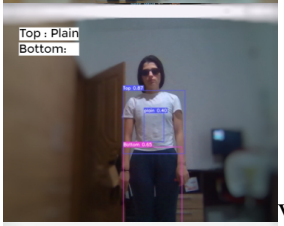
TABLE IV
RESULTS

	Dataset	Percent
Khouzani and Zadeh	CTM	67.7%
Yingli Tian and Shuai Yuan	CTM	82.7%
My dataset	YOLOv7	81.2%

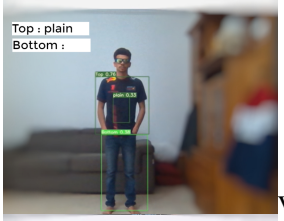
Ground-truth test results: In our test, accuracy was assessed on four footage to detect clothing styles and patterns. The methods of Khouzani and Zadeh [7] obtains 67.7 percent clothing matching rate with CTM dataset, whereas Tian and Yuan[6] achieves 82.7 percent texture matching with CTM dataset. The accuracy displayed in Table III was determined using a confusion matrix in the four footage. Our YOLOv7 dataset accuracy is 81.2 percent, as shown in Table IV.

This prototype is supposed to be 100 percent successful since it needs to be very accurate since they are impaired people that need to depend on this prototype.

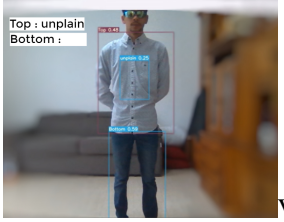
Video 1



Video 2



Video 3



Video 4

Fig. 7. Detected Persons in different footage

V. CONCLUSION

In this study, an effective technique for automatically annotating a variety of clothing and patterns from images and footage was provided. Question 3 of the hypothesis was confirmed correct since the accuracy score was 81.2 percent, which is slightly higher or on the same level as comparable to the state-of-the-art systems. To solve research question one, the system trains the model using a customized dataset including various types of clothing and pattern annotation. The training model is then processed into the model YOLOv7, which generates detections for research question 2. These detections show the analysis, location, and computation of different types and patterns of clothing. A problem was discovered because the method only detected one annotation for the top and bottom textures. These instances can be improved if the model is taught with more texture annotations, such as stripes and plaid. Another problem was that the dotted texture has to be quite stable in order to be identified. It was also recognized as a bottom type to a background object.

As an improvement, the program would have offered additional color detection so that they could have the clothing types, textures, and colors. It might be more efficient with these individuals.

The aim of the project is to use this prototype to enable visually impaired persons to know what they are wearing. As a result, as a suggestion, this system will concentrate on categorizing more patterns and transferring the method to a mobile phone application; photos may also recognize the types and patterns with audio output.

REFERENCES

[1] "Vision impairment and blindness." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
[2] A. Joy, "Introduction to YOLOv7." [Online]. Available: https://pythonistaplanet.com/yolov7/
[3] "YOLO Algorithm for Object Detection Explained [+Examples]." [Online]. Available: https://www.v7labs.com/blog/yolo-object-detection, https://www.v7labs.com/blog/yolo-object-detection
[4] "YOLO Object Detection Explained: A Beginner's Guide." [Online]. Available: https://www.datacamp.com/blog/yolo-object-detection-explained
[5] "List Of Contrasting Colors," Jun. 2022, section: Blog. [Online]. Available: https://www.colorsidea.com/list-of-contrasting-colors/
[6] Y. Tian and S. Yuan, "Clothes Matching for Blind and Color Blind People," in *Computers Helping People with Special Needs*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 6180, pp. 324–331, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-642-14100-3_48
[7] T. Walvekar and P. Reddy, "CLOTHING PATTERN RECOGNITION FOR VISUALLY IMPAIRED PEOPLE," vol. 3, no. 7.
[8] S. Vaidya, S. Kavthekar, and A. Joshi, "Leveraging YOLOv7 for Plant Disease Detection," in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Feb. 2023, pp. 1–6.