Using Object Detection Methods to Detect Fashion Trends in University Students' Attire

Wei-Her Hsieh[1] and Second Author[2[1111-2222-3333-4444]]

[1] National Pingtung University of Science and Technology, Taiwan
weiher@mail.npust.edu.tw

[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. With the advancement of technology and the development of artificial intelligence, research methods in fashion design are continuously evolving. This project aims to use object detection techniques to analyze the attire of university students, exploring and uncovering fashion trends. Firstly, we selected YOLOv8 as object detection models, then trained the model and used the training set as well as adjusted parameters to improve accuracy. Secondly, we applied annotation tools to label collected photos, marking different types of clothing (e.g., tops, pants, shoes). To further enhance the quality and variety of the dataset, we leveraged a pre-trained model based on the DeepFashion2[10] dataset. By leveraging big data and machine learning technologies, this project will establish an automated system capable of analyzing university students' fashion styles in real-time and with accuracy, providing forward-looking insights for fashion design. In this research, we have chosen to focus on one of the most produced fashion items globally - t-shirts. Mass manufactured knit tops such as t-shirts, polo shirts, and sleeveless tops are chosen as they are one of the most ubiquitous items shown in our preliminary testing (explained further in the methodology section), but more importantly, as they are some of the most resource intensive apparel items to produce. In our Dataset creation we create 14 new categories of T-Shirt with annotation tool RoboFlow[13] and segmentation technique, which represent different shape, size and color patterns of T-Shirts. The students of the Department of Fashion Design and management were the main contributors to this Dataset. Fashion design is not only an art but also a discipline closely related to society and culture. Traditional methods of analyzing fashion trends often rely on surveys and street observations, which are time-consuming and subjective. With the development of artificial intelligence and big data technologies, applying object detection techniques to attire analysis can enhance efficiency and provide more objective and comprehensive data.

Keywords: Object Detection, Deep Learning, Fashion Trends

1 Introduction

Humans are social creatures. Their behavior and actions are influenced heavily through social interactions and cultural norms, yet each individual possesses a sense of uniqueness. They express their uniqueness and individuality through art and fashion. This creates volatility and uncertainty. So predicting fashion trends accurately through a single mathematical formula is really challenging. Still fashion prediction, in other words fashion forecasting poses huge benefits for the fashion industry and marketplaces. It is essential for low wastage and high margin of profits. It is easier to manage large inventory with fashion trends predicting methods. So the industry is constantly looking for better and efficient ways of predicting fashion trends.

The age-old traditional methods such as Delphi Methods [1] and surveys relies heavily on human intervention, which is inefficient for modern fast paced fashion trends prediction [2].

After the revolution of Machine Learning [3], Deep Learning [4], Object Detection [5] and Big Data [6] fashion industry also emerged into a new wave of opportunities. These technologies can be used to predict fashion trends far better and accurately as they rely on compute units and Datasets to produce results instead of human intervention. These are far cheaper and time efficient than humans. In the Object Detection method, Deep Learning models are trained on annotated Dataset, later they can be used to detect clothing items. Simply put, machines can be used as a substitute for human eyes. This reduces volatility and uncertainty significantly to further use the detected result to predict fashion trends.

In our work, the primary purpose is to use these highly effective technologies in real world scenarios, especially in student attire'. Study shows adolescents take higher risk and bold decisions than grown-ups [9]. This sector is untapped and more investigation needed. No significant study took place where large amounts of university surveillance videos are being analyzed for predicting students' attire accurately. Through our study, we have shed light to this profitable untapped sector. With our study researchers would be able to understand students' clothing behaviors and businesses would be able to make far more accurate decisions on inventory management for student attire.

During the initial period of our research we incorporate existing clothing Datasets. DeepFashion2 [10] and VITON-HD [11] are two prominent clothing datasets. DeepFashion2 is a comprehensive fashion dataset. It contains 491K diverse images of 13 popular clothing categories from both commercial shopping stores and consumers. We used a YOLOv8 [12] model trained on DeepFashion2 to predict on the Department of Fashion Design and Management of National Pingtung University of Science and Technology's surveillance videos(Closed-circuit television-CCTV). The result shows the prominent clothing item among the 13 clothing items is T-Shirt.

In the second phase of our research we created our own Dataset of T-shirts, as DeepFashion2 has only one category, which does not represent the different type and shape of T-Shirts. Also due to the nature of the surveillance video footage, as it has a certain angle and resolution, DeepFashion2 struggles to capture real world scenarios and perform poorly and produce a lot of false positive results. In our Dataset creation we create 14 new categories of T-Shirt with annotation tool RoboFlow [13] and segmentation technique, which represent different shape, size and color patterns of T-Shirts. The students of the Department of Fashion Design and management were the main contributors to this Dataset. They posed in different angles for the photoshoot. We used different cameras to mimic surveillance like images. This performs better due to its relatedness to the desired scenario [8-9].

At the final stage of this research we trained the YOLOv8 model with our Student T-Shirt Dataset. To predict fashion trends we use this model on surveillance videos and also videos taken from university cafeteria with hand held camera to have a wider range of distribution. In addition, we used Object Tracking and Counting algorithms to perform fashion trends prediction.

The rest of this paper is organized in the following order. Section 2 presents the Literature Review. Section 3 Methodology of the research. Section 4 presents Results. Section 5 and 6 will be the Discussion and Conclusion of the whole research.

The Dataset can be found here: https://github.com/R-sany/Ai4Fashion_NPUST

2 Literature Review

2.1 Fashion Trend

Fashion trend refers to the popular attributes of clothing items during a certain period. This might range from something as holistic as head-to-toe style, silhouette and proportion, to color, fabric, and down to finer details of a garment.Previously, fashion relied on a trickle-down economy theory where the upper echelon of society dictated the desirable way to dress [14]. As society evolves and social structure changes, so do fashion trends. On the contemporary fashion landscape, trends no longer exclusively work on a simple trickle down economy [15]. Out of the current involvement of various more complex moving parts in the making of a fashion trend and the growing intensity of competition amongst businesses, trend forecasting was born.

2.2 Trend Forecasting and its importance in fashion business

Fashion trend forecasting is a field that predicts the coming and going of fashion trends by assimilating observations of the present, past, and future elements that influence fashion. The field is essential for fashion businesses to thrive.Traditionally, fashion companies always stage shows and produce before an intended season comes. For example, a spring-summer collection would be showcased during the fall of the previous year and then produced months ahead of the intended season it arrives in store [16]. This requires designers to be able to foresee what is going to be on-trend or still relevant in the upcoming season. This is even more important with today’s fast fashion business model, where designs are delivered in “drops”, where new styles can come in store as often as weekly [16]. The importance of a correct trend forecast can’t be undermined as an inaccurate one can negatively impact sales and company’s financial health.In this research, we have chosen to focus on one of the most produced fashion items globally - t-shirts. Mass manufactured knit tops such as t-shirts, polo shirts, and sleeveless tops are chosen as they are one of the most ubiquitous items shown in our preliminary testing (explained further in the methodology section), but more importantly, as they are some of the most resource intensive apparel items to produce. A 0.5-lb cotton t-shirt requires 700 gallons of water during processing [17]. Compounding this with the fact that around 15 billion t-shirts are produced annually, among which approximately 2 billions are sold, focusing on predicting t-shirt trends can be beneficial not only for the shareholders of textile and fashion companies, but also the environment.

2.3 Trend Forecasting Methods

Trend forecasting used to rely on methods that are subjective as they utilised qualitative research, surveys, street observation, and even intuitive judgement made by experts [18]. They manually note down the happenings in the street and elsewhere, and reflect them on their prediction. This, albeit a romantic idea, is heavily subjective and often inaccurate on top of being time consuming. Thus, it is no longer a reliable method to be used in today's fashion industrial landscape, where data is exchanged in real time through social media and micro trends happen in weeks time [19].Current face of trend forecasting has changed dramatically since its first inception. Nowadays, trend forecasting has become an entire industry on its own where many have assumed the form of global agencies and corporations such as WGSN, Peclers Paris, and many others. To cope with the ever increasing speed of fashion, trend forecasting companies have embraced the use of digital technologies. Methods to track trends through data analysis and AI are now common practice, as demonstrated in this research [20].

2.4 AI and its Role in the Future of Trend Forecasting

Artificial intelligence has various definitions, but a broad one that is often used is the technology that allows machines to perform complex human skills, notably in decision making [3]. The promises of AI in fashion trend forecasting are big. Some of them include further automation, speed, and accuracy in predicting trends. AI's ability to process and analyze vast amounts of data at scale enables designers, brands, and retailers to make data-driven decisions that were previously not possible.Ways to incorporate AI have been adopted by trend forecasting agencies. These include the use of machine learning, natural language processing, and deep learning. The latter two have been used extensively to predict trends based on consumer behavior and shopping patterns, social media usage, and historical data [20]. In this research, we aim to utilise deep learning in order to determine the trend of knit t-shirts worn in the stated demography as further explained in the methodology part of this research.
Deep learning is a more advanced subset of machine learning. It uses more than one layer of neural networks to analyse data and provide output. These layers process the data it is given differently and will learn from the previous layer's result. Like machine learning, it also has the ability to self-correct. This way, it is able to process more data (text and images), requires less human involvement, and is often more accurate in the output it produces compared to machine learning [3].

In trend forecasting, one instance where deep learning can be useful is to recognize an object. This is done by 'teaching' an algorithm or model what different pictures are through a massive amount of related pictures called dataset. The different layers will learn different aspects of these images until it can recognise or classify relevant images as the object the algorithm is supposed to detect. This object detection method is being utilized by trend forecasting agencies to, for example, process vast amounts of runway pictures or social media content, to street photography and online stores to determine whether a trend is coming up or going away [20]. In this research, we utilized instant segmentation, another subset of deep learning, to detect knit t-shirts and used the result to create graphs of what is the most popular type of t-shirts during a certain period of time in the designated setting and demography.

3 Methodology

3.1 Dataset Collection and Annotation

The dataset for this research was created by capturing images of university students wearing a variety of t-shirts in different real-world environments, including the university cafeteria and the General Education building, during peak hours. This allowed us to obtain a wide range of samples, including students of different body types, movements, and clothing styles. In some instances, students volunteered to be photographed, and additional images were provided by faculty members. To ensure the model could generalize well, we captured multiple images of the same t-shirt from various angles.

However, the data collection period coincided with the summer holidays of the students, which led to a reduction in the number of students available for image collection. To overcome this challenge, approximately half of the images in our dataset were sourced from the internet. These images were obtained through direct searches using keywords matching our labels, such as “short sleeve motif shirt” and “sleeveless motif/graphic t-shirt.” This approach was essential for capturing less common subcategories like sleeveless motif t-shirts, which were underrepresented in our initial collection efforts.

In addition to ensuring the diversity of clothing styles, we also focused on capturing images that would simulate real-world conditions for the model. Some images were taken from upfront angles, where students would approach the camera or handheld devices directly. This was done to replicate the type of interactions expected when the system tracks students in real-time scenarios.

To further enhance the quality and variety of the dataset, we leveraged a pre-trained model based on the DeepFashion2[10] dataset. As mentioned in the Introduction section, this model was trained on over 491,000 images with 13 different clothing categories. The use of this model allowed us to focus on specific outerwear categories relevant to our research. The chosen categories were “short sleeve top”, “Long sleeve top”, “Short sleeve outerwear”, “Long sleeve outerwear”, “Vest”, “Sling”, “Shorts”, “Trousers”, “Skirt” , “Short sleeve dress”, “Long sleeve dress”, “Vest dress”, “Sling dress”.

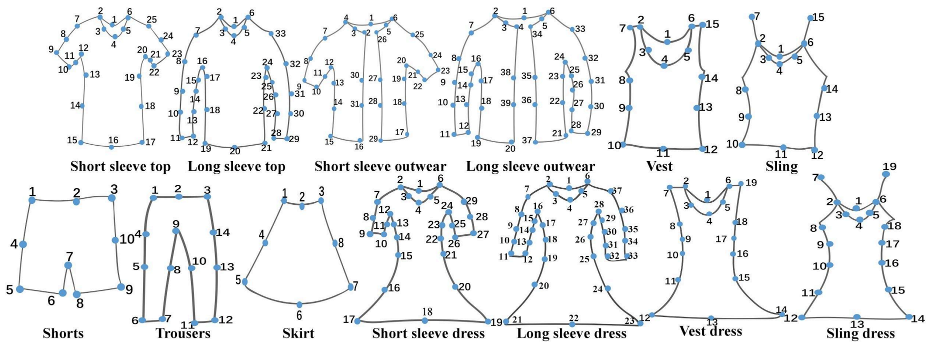

Figure 1: DeepFashion2 dataset and its labels

This curated selection of categories provided a solid foundation for the dataset, aligning with our research goals to predict fashion trends based on t-shirt and outerwear types.

Given that accurate annotation is critical for object detection, we employed the instance segmentation method to label the boundaries of the t-shirts in the collected images. The labels were assigned to the most appropriate class based on visual inspection. This manual labeling process, though time-consuming, was crucial to the success of the model. The dataset contains several classes of t-shirts, including categories such as "cropped motif short sleeve," "oversized solid color sleeveless," and "polo shirt short sleeve," among others. The base dataset contains approximately 1,500 annotated images with 14 different categories(listed below), which were subsequently uploaded to the Roboflow platform for further preprocessing and augmentation.

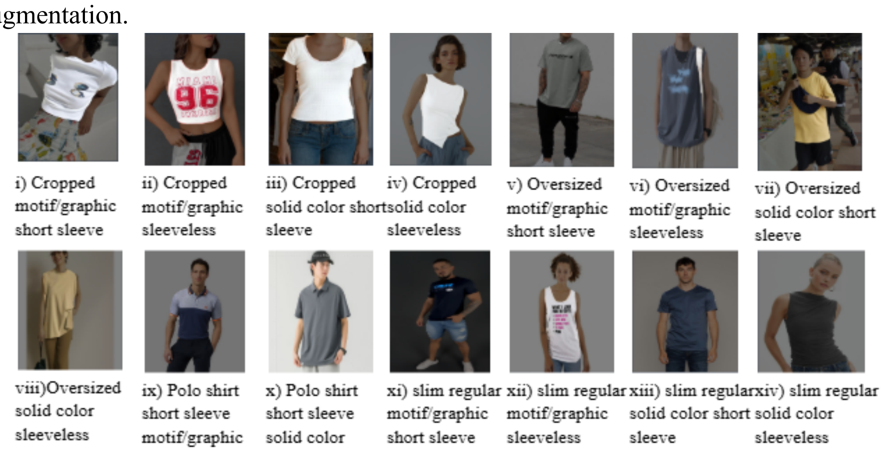

Figure 2: Sample of all fourteen newly created T-Shirt items.

3.1.2 Dataset Annotation and Labeling Process

To accurately label the t-shirts in our dataset, we used Roboflow's Smart Labeling Tool, which uses AI to help automate the annotation process. This tool was especially helpful for quickly outlining the boundaries of clothing, saving a lot of time during the preparation phase. However, there were some instances—like overlapping clothes, intricate designs, or unusual shapes—where the automated tool struggled. In such cases, we switched to manual labeling using the polygon tool, which allowed us to trace the exact boundaries of the clothing more precisely.

We put extra care into this annotation process, as it was critical for teaching the model to distinguish between different types of t-shirts. These labels were carefully assigned based on features like sleeve length, patterns, and overall style, ensuring the dataset was comprehensive and well-structured for training.

Here are a few examples from our dataset that highlight its diversity, including different t-shirt styles, angles, and environments.

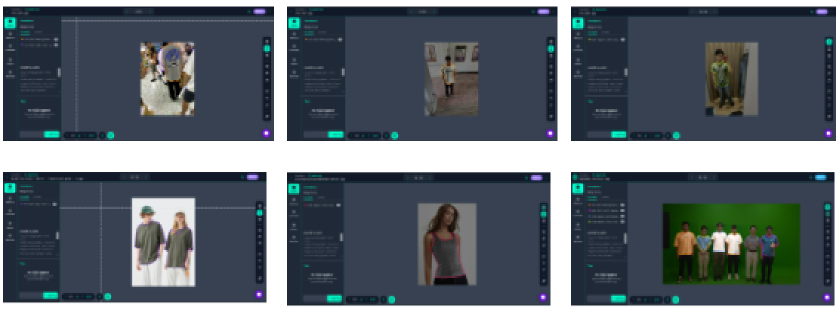

Figure 3: Annotated images from the base dataset used for the training of the base model

3.1.3 Dataset Generation

The final dataset for training the base model comprised 3,474 images in the training set, 226 in the validation set, and 151 in the test set, following extensive preprocessing and augmentation. This dataset represents a significant improvement over the previously generated dataset, which contained 3,186 images in the training set, 199 in the validation set, and 139 in the test set, with limited augmentations applied.

The preprocessing phase involved ensuring consistency in the dataset by:

i) Auto-Orienting all images to standardize their alignment.

ii) Resizing images to a uniform resolution of 640 x 640 pixels using a stretch-based method.

To enhance the diversity and robustness of the dataset, multiple augmentation techniques were applied to each training example, generating three outputs per input image. These augmentations include:

Geometric Transformations:

i) 90° Rotations: Applied both clockwise and counter-clockwise.

ii) Random Rotations: Within a range of -45° to +45°.

iii) Shear Transformations: Horizontal shear of ±0° and vertical shear of ±45°.

Color Adjustments:

i) Grayscale Conversion: Applied to 15% of images to simulate monochrome environments.

ii) Saturation Variations: Randomized between -9% and +9%.

Noise and Blur:

Noise Addition: Up to 2.59% of pixels to simulate image artifacts.

Bounding Box Blur: Introduced blur effects of up to 3.3px for bounding box edges.

Bounding Box Augmentations:

Zoom Crops: Between 0% minimum and 40% maximum zoom, ensuring the subject remains detectable.

Bounding Box Rotations: Randomized within -45° to +45°.

In contrast, the previously generated dataset had limited augmentations, consisting of:

90° Rotations: Clockwise and counter-clockwise.

Shear Transformations: Horizontal ±0° and vertical ±45°.

Noise Addition: Up to 0.1% of pixels.

Impact of Augmentations: The enhanced augmentation strategy significantly enriched the dataset by introducing greater variability, which is crucial for improving the model's ability to generalize across diverse conditions[24]. The augmented dataset now includes a wider range of visual features, such as varied orientations, lighting conditions, and noise levels, closely mimicking real-world scenarios.

3.2 Model Selection and Training

In this research, we employed two key techniques for different tasks: instance segmentation and object detection. Instance segmentation was initially chosen for predicting clothing categories because it not only classifies clothing but also provides pixel-level delineation of each item. This is particularly crucial in cases where clothing items have similar shapes or overlap, allowing us to achieve more precise boundaries compared to traditional object detection, which only offers bounding boxes.

On the other hand, object detection was used for real-time tracking of the clothing items. For this task, we leveraged the track function of YOLOv8, as it is designed to track moving objects across video frames. While instance segmentation allows us to identify and segment out each clothing item accurately, object detection with YOLOv8 helps us track these items over time, which is essential for counting and analyzing fashion trends in dynamic environments.

After establishing this distinction, we proceeded with YOLOv8 for the detection and tracking aspects of our model. YOLOv8 has a proven track record of delivering high performance in real-time tasks, making it a natural fit for our use case of tracking clothing in a busy, dynamic setting.

The training process was performed using the YOLOv8 framework. The following code snippet illustrates the model training procedure:

```python
from ultralytics import YOLO


def train_model():
    model = YOLO('yolov8x-seg.pt')  # Load the YOLOv8 model
    data_yaml = "path_to_data.yaml"  # Path to dataset configuration
    model.train(data=data_yaml, epochs=50, batch=16, imgsz=640, device='cuda')
```

Figure 4: Code snippet of the illustrated training process

In this setup, the model was trained for 50 epochs using a batch size of 16 and an image size of 640 pixels. To address class imbalances within the dataset, we applied a class-weighting strategy by adding more images from underrepresented classes and retraining the model with a lower epoch count. The use of a robust dataset and iterative fine-tuning enabled the model to achieve progressively higher accuracy in detecting t-shirt types.

Comparing YOLOv8 to its predecessors (YOLOv4, YOLOv5, and YOLOv7), YOLOv8 brings significant improvements in performance, especially in the context of real-time detection. YOLOv8 benefits from advancements in both architecture and training techniques, offering faster inference times, more accurate bounding boxes, and improved ability to handle small and overlapping objects. This makes it an ideal model for our task, where precise and fast tracking of clothing items in crowded university spaces is required.

Here is a brief comparison of its predecessors:

YOLOv4: While YOLOv4 was considered highly effective for object detection tasks, its speed was compromised in some scenarios when working with small or densely packed objects, such as students wearing similar clothing.	**YOLOv5**: A substantial improvement over YOLOv4 in terms of performance and speed. However, it still struggled with classifying fine-grained objects, such as different types of t-shirts, especially in video-based environments with fast-moving subjects.	**YOLOv7**: YOLOv7 introduced **efficiency improvements**, particularly for detecting objects in complex, cluttered scenes. However, it did not address pixel-level segmentation, which was a critical need for our clothing prediction task.

Figure 5: Model comparison of previous YOLO architectures.

YOLOv8, in contrast, optimizes both accuracy and speed, with enhancements to its model architecture(figure below), including support for more precise detections in challenging conditions such as occlusion and motion blur, which are common when tracking students in a busy environment.

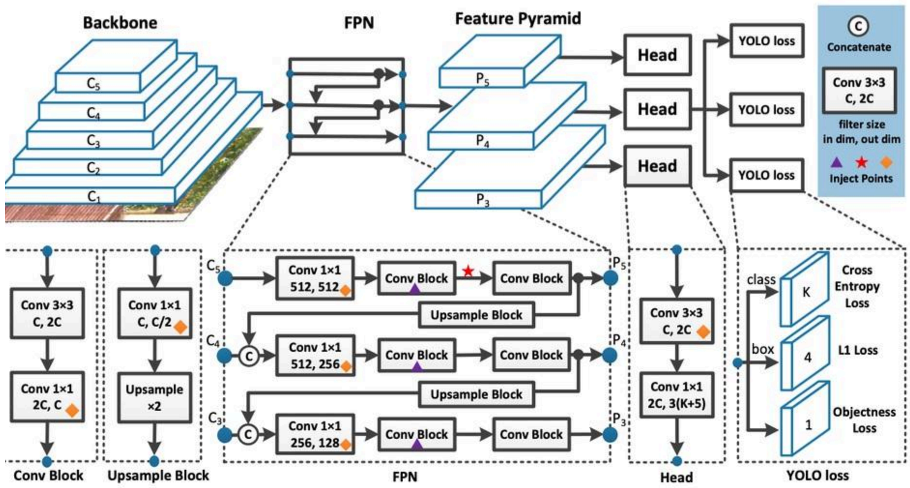

Figure 6: YOLOV8 architecture diagram[23]

While instance segmentation is paramount for accurate classification, YOLOv8's real-time tracking feature proved beneficial for our live detection and counting of different clothing pieces. The track function of YOLOv8(discussed in the next section) enabled the model to follow individual clothing items across video frames, counting them as distinct entities and helping us ensure accurate trend analysis.

3.3 Real-time Detection and Tracking

After training the model on static images, we tested its performance in real-time scenarios using video feeds from various sources, including handheld cameras, CCTV, and web cameras. To replicate the conditions of the dataset and ensure the model's robustness, we set up a web camera in the university, capturing live video during busy hours when student traffic was high. This setup allowed us to track students wearing different t-shirts in real time.

Object tracking, a key feature of the YOLOv8 model, was utilized to monitor the movement of t-shirts within the camera frame. This is analogous to tracking vehicles on highways, a widely studied application in the field of computer vision. By employing object tracking, we were able to avoid counting the same person multiple times, ensuring that each t-shirt class was accurately tracked and counted. The following code snippet demonstrates how the model processes video frames and tracks t-shirt types:

```python
import cv2
from ultralytics import YOLO

def process_video(model_path, video_path, output_video_path):
    model = YOLO(model_path)
    cap = cv2.VideoCapture(video_path)  # Open the video
    success, frame = cap.read()

    while success:
        tracks = model.track(frame, persist=True, show=False, classes=classes_to_count, conf=0.8)
        # Track and count objects in the frame
        success, frame = cap.read()

    cap.release()
```

Figure 7: Code snippet demonstrates how the model processes video frames and tracks t-shirt types

In this process, each frame is analyzed for t-shirt classes, and the movement of objects is tracked across multiple frames. The system then labels the tracked t-shirts as "in" or "out" of the frame, ensuring that each t-shirt is counted only once. This methodology enhances the reliability and accuracy of the real-time analysis.

3.4 Iterative Evaluation and Improvement

The model's performance is evaluated iteratively using video feeds captured in the university's common areas. After running the model on the video, frames containing t-shirts are extracted and saved into their respective folders. Any misclassified images are reviewed, and their labels are corrected. These corrected images, along with newly collected data, are then fed back into the model in bulk. This iterative process of collecting new images, retraining the model, and improving predictions has proven effective in enhancing the model's accuracy over time[15].

The iterative evaluation approach follows the cycle of collecting more data, correcting errors, and refining the model's predictions. This continuous feedback loop ensures that the model gradually adapts to varying conditions, such as changes in lighting, different camera angles, and diverse student appearances.

The following code demonstrates the error correction process, where misclassified frames are saved and relabeled:.

```python
import cv2
import os

def save_misclassified_frames(model, video_path, output_folder):
    cap = cv2.VideoCapture(video_path)

    while cap.isOpened():
        success, frame = cap.read()
        if not success:
            break

        # Process the frame and get predictions
        results = model(frame)
        # Save misclassified frames
        if some_condition(results):
            save_frame_to_folder(frame, output_folder)

    cap.release()
```

Figure 8: Code demonstrates the error correction process, where misclassified frames are saved and relabeled

This process helps improve the model's generalization capabilities as new data is continuously integrated into the training pipeline. The evaluation is ongoing, with the model being updated and refined regularly based on the latest data and feedback

4 Results

In this section, results of this entire research will be presented in sequential manner. 4.1 Clothing Detection in Surveillance Videos with DeepFashion2 and YOLOv8, discussing the initial phase of the research.In 4.2 Fine tuned YOLOv8 Deepfashion2 model with our Dataset, second phase of the research will be discussed. At 4.3 Iterative evaluation results of our trained YOLOv8 Model and Dataset and 4.4. Fashion Trend result, rest of the result will be presented.

4.1 Clothing Detection in Surveillance Videos with DeepFashion2 and YOLOv8

As mentioned earlier, during the initial phase of this research, we have used YOLOv8 [12] trained on DeepFashion2 [10] Dataset to detect clothing items on Department of Fashion Design and Managements surveillance videos.

Figure 9: Sample prediction on Surveillance Videos using DeepFashion2 and YOLOv8. It is detecting thirteen clothing items.

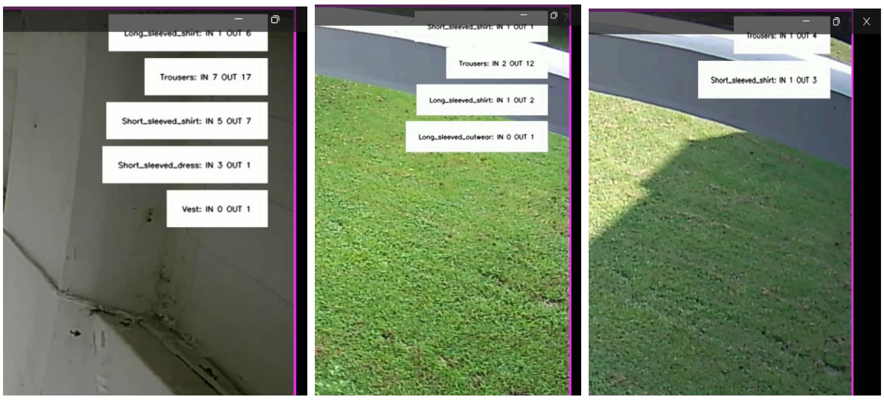

Figure 10: Sample outputs of “In” and “Out” counting methods for Surveillance or CCTV videos. In the upper corner of each frame of prediction could be found. Once an item of desired class or label enters the boundary box, it increases the counts for that specific label.

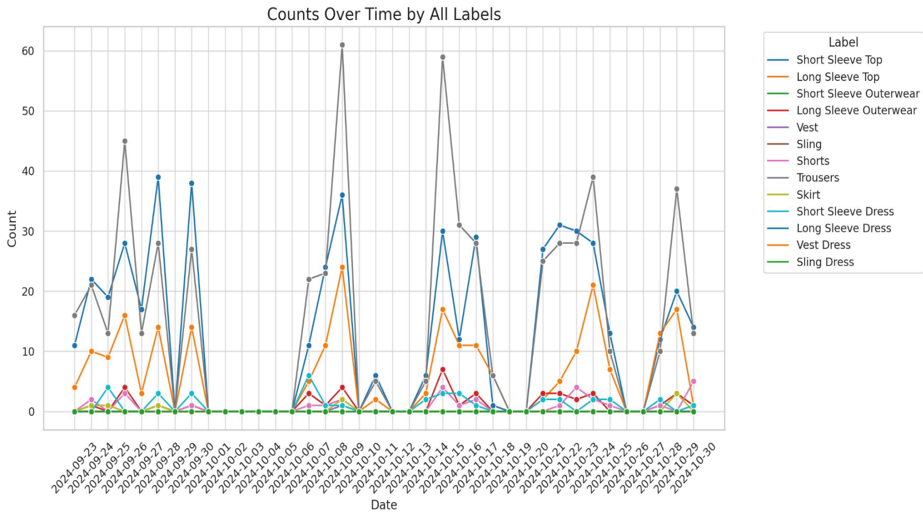

Figure 11: Appearance of each clothing item of DeepFashion2 in surveillance videos in the month of September and October. The Short Sleeve Top (Blue) or T-Shirt has less outliers. The Trousers have two spiking outliers

Table 1. Total appearance of each clothing item of DeepFashion2

.

Label	Total Count
Trousers	593
Short Sleeve Top	504
Long Sleeve Top	236
Long Sleeve Outerwear	42
Short Sleeve Dress	38
Shorts	31
Skirt	8
Vest	3
Long Sleeve Dress	0
Short Sleeve Outerwear	0
Sling	0
Sling Dress	0
Vest Dress	0

Although the total count of Trousers (Table 1) is higher than the Short Sleeve Top or in other word T-Shirts, the Short Sleeve has constant values. On the contrary Trousers have changed variably throughout the months. It has two spiking outliers in the middle (Figure 9). This is the reason we choose T-Shirt or Short Sleeve Top instead of Trousers to conduct our research further.

4.2 Fine tuned YOLOv8 Deepfashion2 model with our Dataset

After Dataset creation, we fine tuned the existing YOLOv8 DeepFashion2 model with our Dataset. In our result we found that this model performed poorly on surveillance videos. It produces a significant amount of false positive results.

Figure 12: From left, the fine tuned YOLOv8 DeepFashion2 with our dataset detects “Pants” as “Oversized motif-graphic short sleeve”. It also falsely detects “Sidebar” as “Oversized motif-graphic short sleeve” and “Oversized motif-graphic sleeveless”

We assume(Figure 12) this could happen due to the sheer size of the DeepFashion2 dataset. To solve this problem we have trained a YOLOv8 model iteratively with our dataset.

The results of this process could be found in 4.3 Iterative evaluation results of our trained YOLOv8 model and dataset.

4.3 Iterative evaluation results of our trained YOLOv8 Model and Dataset

At the second stage of our research,, we have trained a YOLOv8 model with our newly created T-Shirt Dataset iteratively. To illustrate the trained model performance and metrics, we have presented our final training and evaluations in the next few pages.

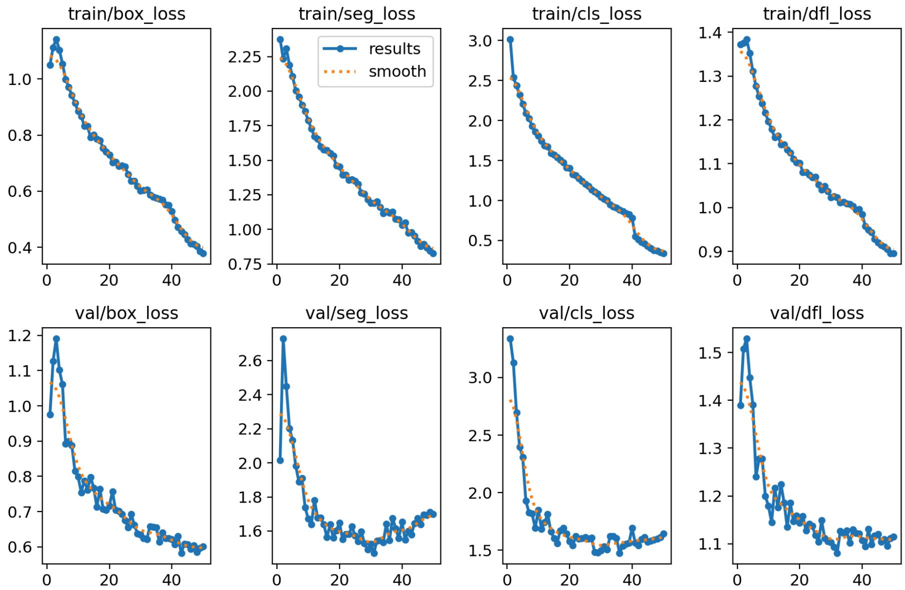

Figure 13: These diagrams show the training losses of YOLOv8 on our Dataset.

The X axis of this diagram represents Epoch and the Y axis represents the Loss values. From the diagrams we could observe that the losses decays over each Epoch, making a downward curve and ultimately reaching a value nears to zero.

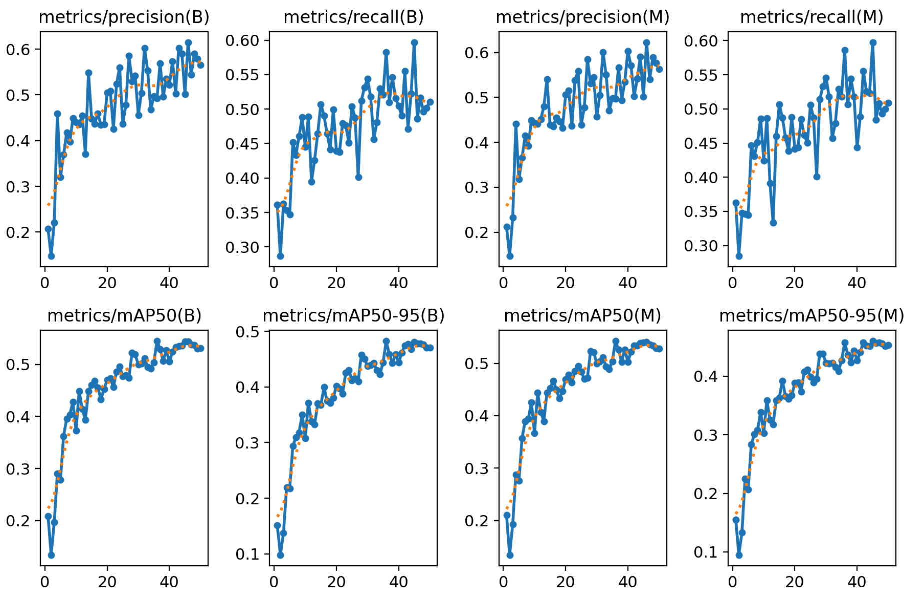

Figure 14: The diagrams shown above are the accuracy metrics recorded over each Epoch during training of the YOLOv8 model with our Dataset. The accuracy goes high to each individual metric as Epoch increases.

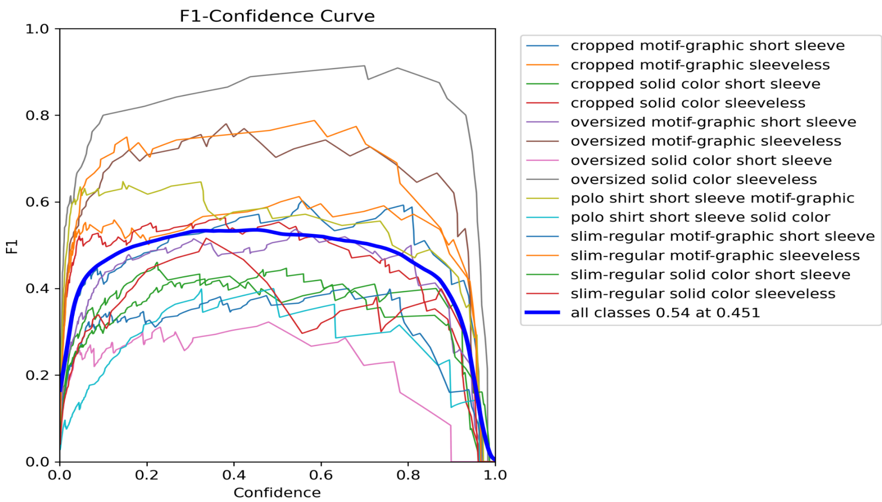

Figure 15: This plot shows the F1-score at different confidence thresholds for each class. Each line represents a different class, and the thicker blue lines represent the overall performance for all classes combined.

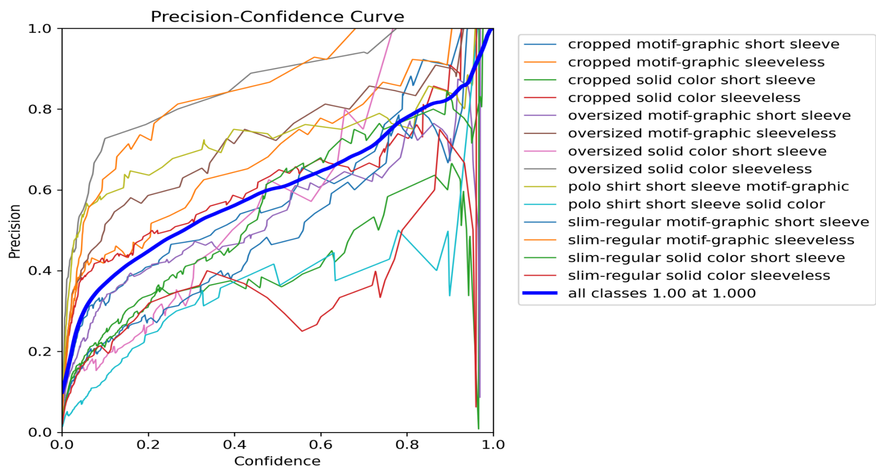

Figure 16: This plot shows the Precision-Recall -score at different confidence thresholds for each class. Each line represents a different class, and the thicker blue lines represent the overall performance for all classes combined.

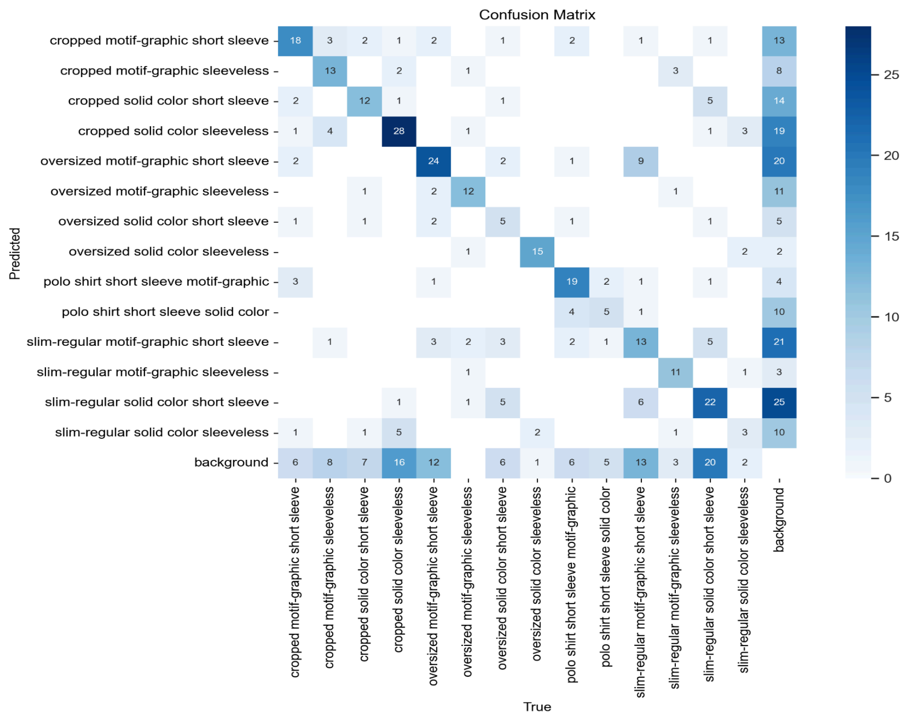

Figure 17: Confusion matrix shows the true labels in X axis and predicted labels Y axis across multiple classes. In our case the 14 newly created labels . The scale values represent sample counts, with deeper blue indicating more frequent predictions Correct predictions can be observed along the diagonal, while off-diagonal cells reflect wrong predictions.

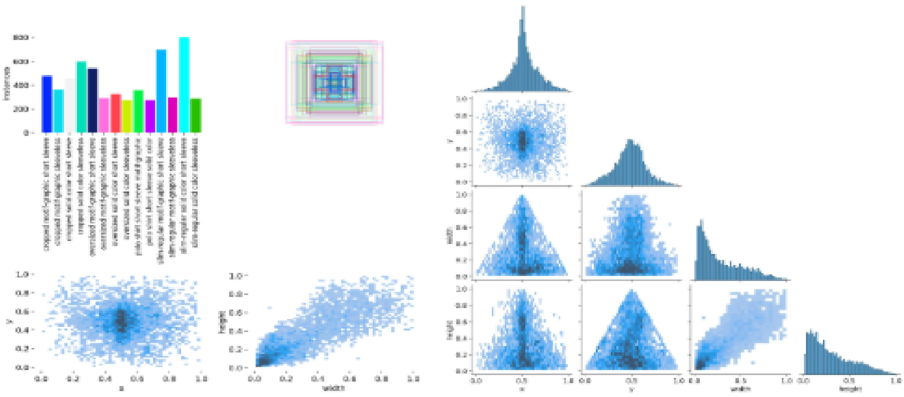

Figure 18: From left image sizes of each label and on the right labels correlogram.

4.4. Fashion Trend results

In 4.4 Fashion Trend results, the final trend prediction will be provided, which is an automated process to find the final prediction outcome.

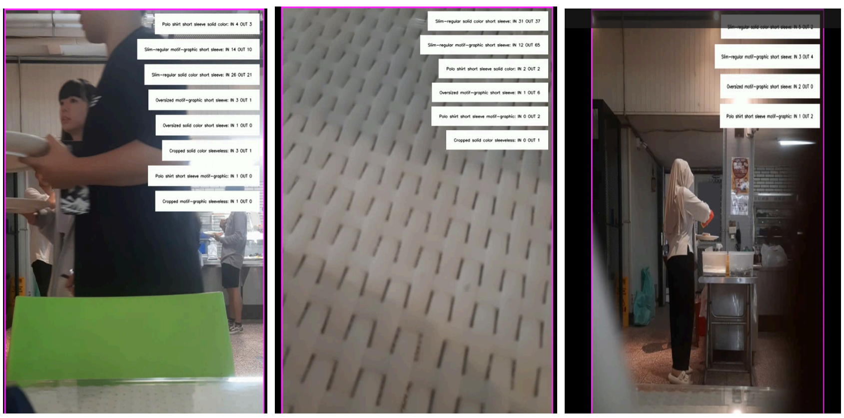

Figure 19: Sample outputs of “In” and “Out” counting methods for Cafeteria videos.

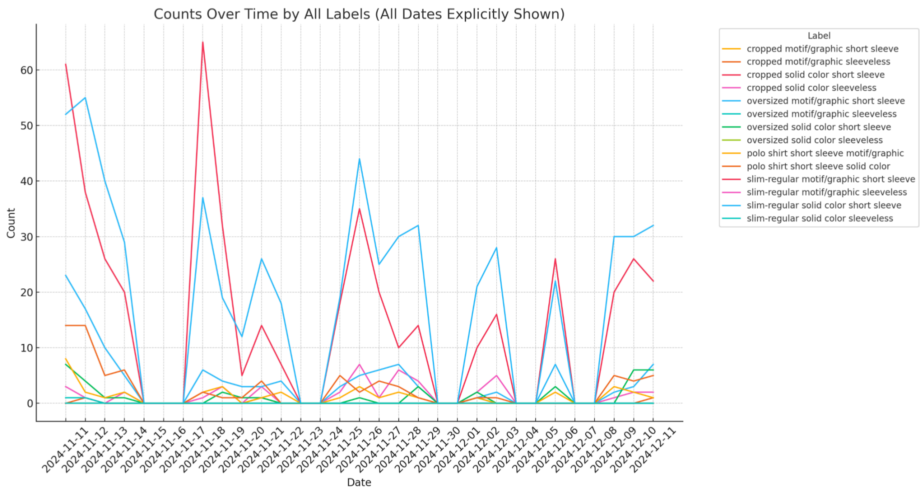

Figure 20: Appearance of each T-Shirt. Each color represents individual labels. “Slim-regular solid color short sleeve”(blue) has the highest and most consistent appearance. Followed by “slim-regular Motif/Graphic Short Sleeve", “Oversized Motif/Graphic Short Sleeve”, and so forth.

Table 2. Total appearance of each T-Shirt item.

Label	Total Count
Slim-Regular Solid Color Short Sleeve	601
Slim-Regular Motif/Graphic Short Sleeve	485
Oversized Motif/Graphic Short Sleeve	121
Short Sleeve Solid Color Cropped	78
Solid Color Sleeveless Oversized	45
Solid Color Short Sleeve Polo Shirt	38
Short Sleeve Motif/Graphic Cropped	38
Motif/Graphic Sleeveless Oversized	3
Motif/Graphic Sleeveless Cropped	3
Solid Color Short Sleeve Oversized	0
Solid Color Sleeveless Slim-Regular	0
Motif/Graphic Sleeveless Slim-Regular	0

In Table 2 the final counting result of the predicted labels can be found. It shows the total appearance of each label throughout November and December of 2024. Top three most frequent items are “slim-regular Motif/Graphic Short Sleeve", and “Oversized Motif/Graphic Short Sleeve”. From figure 20 we can observe these items decay over time. The most likely reason is the weather. Further analysis is presented in next chapter 5. Discussion.

5 Discussion

The results shown in section 4.1 and 4.2 were the evidence of poor performance of previous dataset such as DeepFashion2. Although this model can be used for generic fashion trends prediction, the labels are too broad to create meaningful, industrial trend prediction. For example, the most prominent T-Shirt item according to the final result found in section 4.4 is “Slim-regular solid color short sleeve” (blue) (Table 2). This label is more useful from an industry standpoint as it is more informative to the parties reading the trend forecast, in comparison to DeepFashion2’s general labels that were used in the preliminary testing of this research (skirts, vest, etc.).

In section 4.3 the detail of the model evaluation is shown. Higher F1 scores indicates the model is performing well. Except for some labels, our dataset produced a steady curve. The same trade can be observed in the rest of the evaluation matrices such as Precision-Confidence Curve, Recall Confidence Curve and Confusion Matrix.

In future research, areas of improvement from our findings could include higher accuracy in predicting clothing forms that are obstructed by another object and the use of more detailed labels. The former is an observation made when the researchers tried to process testing results. Some t-shirts that were visually obstructed by objects such as the samples’ arms or jackets might yield inaccurate results. This was improved after further improvement in the dataset and more training, but was not eliminated completely. The latter is in regards to how our labels are still not as detailed as what a traditional trend forecasting report would require. For example, our labels do not inform what kind of pattern is relevant to the motif label or what color and texture are relevant to the solid color label. This was due to the limited function of instant segmentation and could be improved in future research.

6 Conclusion

In conclusion, we used instant segmentation and annotation techniques to determine the trend of t-shirts among students in National Pingtung University of Science and Technology in a specific period of time. In the process we have created a student specific T-Shirt dataset and trained a YOLOv8 model. We have also used tracking and counting methods to determine the student fashion trends through running those models on surveillance footage and videos taken with handheld cameras.

We found that previous clothing item datasets are not capable of producing advanced fashion trends analysis as its labels are too general and lack informative details that a trend report requires. Also most of the dataset is not reflecting real world scenarios. To mitigate this problem we created fourteen new t-shirt categories that include details relevant for fashion businesses such as length, presence of motifs and collars, and sleeve lengths . We tried to mimic the final testing environment in our dataset to increase accuracy and took photos in different locations, angles, height and shapes.

We can observe through the results that our model showed promising results on identifying clothing items. Our results suggested that “T-Shirt” is the most worn item and among t-shirts “Slim-regular solid color short sleeve” is the most prominent. About six hundred students have worn “Slim-regular solid color short sleeve” t-shirts in the months of November and December in 2024 at National Pingtung University of Science and Technology, which is the prominent T-Shirt item.

This result is produced after running our models on about 500GB of Surveillance and handheld camera videos, which haven't been done before. This study will provide a comprehensive method to businesses for conducting better fashion trends analysis on students attire in real world scenarios. As these methods purely rely on compute units, tracking, counting and sensors, businesses now can perform this analysis with less human intervention. This will reduce costs and time to perform fast fashion trends prediction, which will produce more profits with less inventory.

The future direction of this research could be adding more versatile clothing items to the datasets alongside adding more accurate colours. Gender specific research could also benefit the fashion industry.

Acknowledgments. This project is funded by Taiwan Educational Experience Program (TEEP) 2024. We want to appreciate the active participation of the students of the Department of Fashion Design and Management, National Pingtung University of Science and Technology, Taiwan.

Disclosure of Interests. The authors have no competing interests.

References

1. Noh, M., & Ulrich, P. (2013). Querying fashion professionals’ forecasting practices: the Delphi method. International Journal of Fashion Design, Technology and Education, 6(1), 63–70. https://doi.org/10.1080/17543266.2013.765510
2. Fink-Hafner, Danica & Dagen, Tamara & Doušak, May & Novak, Meta & Hafner-Fink, Mitja. (2019). Delphi Method: Strengths and Weaknesses. Advances in Methodology and Statistics. 2. 1-19. 10.51936/fcfm6982.
3. S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451
4. LeCun, Yann & Bengio, Y. & Hinton, Geoffrey. (2015). Deep Learning. Nature. 521. 436-44. 10.1038/nature14539.
5. A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," in IEEE Access, vol. 11, pp. 35479-35516, 2023, doi: 10.1109/ACCESS.2023.3266093.
6. S. Sagiroglu and D. Sinanc, "Big data: A review," 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, pp. 42-47, doi: 10.1109/CTS.2013.6567202.
7. Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., & Roy, P. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Computer Methods and Programs in Biomedicine, 213, 106504. https://doi.org/10.1016/j.cmpb.2021.106504

8. C. Luo, X. Li, L. Wang, J. He, D. Li and J. Zhou, "How Does the Data set Affect CNN-based Image Classification Performance?," 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 2018, pp. 361-366, doi: 10.1109/ICSAI.2018.8599448.
9. Steinberg, L. (2006). Risk taking in adolescence: What changes, and why? Annals of the New York Academy of Sciences, 1021(1), 51-58. https://doi.org/10.1196/annals.1308.005
10. Ge, Y., Zhang, R., Wu, L., Wang, X., Tang, X., & Luo, P. (2019). A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. CVPR. Retrieved from https://arxiv.org/pdf/1901.07973.pdf
11. Choi, S.; Park, S.; Lee, M.; Choo, J. VITON-HD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021
12. G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics.” https://github.com/ultralytics/ultralytics, 2023. Accessed: February 30, 2023.
13. Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.
14. Cartner-Morley, J. (2023, April 13). Equal parts practical and daring: How mary quant created look for a new way of living. The Guardian. https://www.theguardian.com/fashion/2023/apr/13/equal-parts-practical-and-daring-how-mary-quant-created-look-for-a-new-way-of-living
15. Atik, D., Cavusoglu, L., Ozdamar Ertekin, Z., & Fırat, A. F. (2022). Fashion, consumer markets, and democratization. Journal of consumer behaviour, 21(5), 1135-1148.
16. Kuipers, G., Brans, L., & Carbone, L. (2023). The myth of trickle-down: How fashions do (not) spread in European fashion magazines, and what this tells us about power and status in the global fashion system. In Fashion’s Transnational Inequalities(pp. 114-142). Routledge.
17. Crofton, S., & Dopico, L. (2007). Zara-Inditex and the growth of fast fashion. Essays in Economic & Business History, 25, 41-54.
18. Khan, M. M. R., & Islam, M. M. (2015). Materials and manufacturing environmental sustainability evaluation of apparel product: knitted T-shirt case study. Textiles and Clothing Sustainability, 1, 1-12.
19. DuBreuil, M., & Lu, S. (2020). Traditional vs. big-data fashion trend forecasting: an examination using WGSN and EDITED. International Journal of Fashion Design, Technology and Education, 13(1), 68-77.
20. Reilly, A., & Hawley, J. (2019). Attention deficit fashion. Fashion, Style & Popular Culture, 6(1), 85-98.
21. Shi, M., Chussid, C., Yang, P., Jia, M., Dyk Lewis, V., & Cao, W. (2021). The exploration of artificial intelligence application in fashion trend forecasting. Textile Research Journal, 91(19-20), 2357-2386.
22. Manuel Zahna , Douglas P. Perrin, Ph.D.a,b,*(2023, April 3) Efficient human-in-loop deep learning model training with iterative refinement and statistical result validation
23. Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., Ding, E., & Wen, S. (2020). PP-YOLO: An Effective and Efficient Implementation of Object Detector. arXiv. https://doi.org/10.48550/arXiv.2007.12099
24. Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0