Artificial Intelligence Review (2022) 55:5707–5729
https://doi.org/10.1007/s10462-022-10142-3

When AI meets store layout design: a review

Kien Nguyen[1] · **Minh Le**[2] · **Brett Martin**[1] · **Ibrahim Cil**[3] · **Clinton Fookes**[1]

Published online: 10 February 2022

Abstract
An efficient store layout presents merchandise to attract customer attention and encourages customers to walk down more aisles which exposes them to more merchandise, which has been shown to be positively correlated with the sales. It is one of the most effective in-store marketing tactics which can directly influence customer decisions to boost store sales and profitability. The recent development of Artificial Intelligence techniques, especially with its sub-fields in Computer Vision and Deep Learning, has enabled retail stores to take advantage of existing CCTV infrastructure to extract in-store customer and business insights. This research aims to conduct a comprehensive review on existing approaches in store layout design and modern AI techniques that can be utilized in the layout design task. Based on this review, we propose an AI-powered store layout design framework. This framework applies advanced AI and data analysis techniques on top of existing CCTV video surveillance infrastructure to understand, predict and suggest a better store layout.

Keywords Video analytic · CCTV visual intelligence · Business intelligence · Store layout · Retail layout

✉ Kien Nguyen
k.nguyenthanh@qut.edu.au

Minh Le
minhlth@ueh.edu.vn

Brett Martin
brett.martin@qut.edu.au

Ibrahim Cil
icil@sakarya.edu.tr

Clinton Fookes
c.fookes@qut.edu.au

[1] Queensland University of Technology, Brisbane, Australia

[2] University of Economics Ho Chi Minh, Ho Chi Minh, Vietnam

[3] Sakarya University, Serdivan, Turkey

1 Introduction

Annually, an average woman spends 140 h on 45 trips shopping for clothes and shoes, and 95 h on 84 trips shopping for food (Forbes 2015), it is the main source contributing to $24 trillion annual total retail sales (https://www.statista.com/statistics/443522/global-retail-sales/). Retail stores such as supermarkets and warehouses are allocating the highest percentage of total spend to sophisticated marketing strategies compared with other industries (21.9%) to attract more customers to the store, to keep customers browsing longer and being happier within the stores, which will consequently result in their more spending (https://www.brafton.com.au/blog/content-marketing/the-ultimate-list-of-marketing-spend-statistics-for-2019-infographic/). Store layout, the design of a store's floor area and the placement of items in the store, is one of the most effective in-store marketing tactics which can directly influence customer decisions to boost store sales and profitability (Larson et al. 2005). An efficient store layout presents merchandise to attract customer attention (Rhee and Bell 2002) and encourages customers to walk down more aisles which exposes them to more merchandise, which has been shown to be positively correlated with the sales (Martin 2011).

Currently, the conventional approach to designing store layout is based on a passive reaction to customer behavior (Cil 2012; Cil et al. 2009). For example, retailers examine sales data (e.g., by product category), amend the store layout and introduce in-store displays and promotions to increase or maintain sales. However, these conventional design approaches may not reflect the actual behavior of the customers visiting the stores, which may vary significantly from one residential population to another and from one store to another due to the differences in the customers visiting the store. Importantly, the conventional design process does not reflect (1) how customers actually navigate through store aisles, (2) how much time customers actually spend in each section, and (3) what visible emotion (e.g., happiness) customers exhibit in response to a product. With current in-store technology, such data can be obtained and analyzed by applying AI to the video surveillance of Closed-Circuit TeleVision (CCTV) infrastructure which has been long and widely employed in stores as a security measure to reduce shoplifting and employee theft (Fig. 1).

Recent advances in AI with its sub-fields in computer vision, machine learning and especially in deep learning, have led to breakthroughs in many tasks, with results that match or surpass human capacity (LeCun et al. 2015). The retail community has leveraged the power of AI for many tasks particularly related to the purchase transaction, such as pay-with-your-face, check-out free grocery stores by Amazon Go, visual and voice search by Walmart, Tesco, Kohl's, Costco, and customer satisfaction tracking and behaviour prediction (https://spd.group/artificial-intelligence/ai-for-retail/?fbclid=IwAR0HM8tP2vQ9MI6jE2lrkD7JnyBP1NMlEAgRWqWWKKlHoctFctHnPC60J9M#Route_Optimization). In 2019, the retail sector leads in global spending on AI systems, with $5.9 billion invested in automated customer service agents, shopping advisers, and product recommendation platforms (https://spd.group/artificial-intelligence/ai-for-retail/?fbclid=IwAR0HM8tP2vQ9MI6jE2lrkD7JnyBP1NMlEAgRWqWWKKlHoctFctHnPC60J9M#Route_Optimization). However, store layout design still lags behind in the AI era. Although research has highlighted the potential of CCTV to capture consumer movement in stores (Newman and Foxall 2003; Nguyen et al. 2017a, 2017b), a framework showing how AI-derived insights can be used for store design has been lacking. Indeed, current retailing research on AI emphasizes consumer perceptions of AI (Davenport et al. 2020; Grewal et al. 2020; Roggeveen and Sethuraman 2020) rather than how AI can be used to inform store layout design.

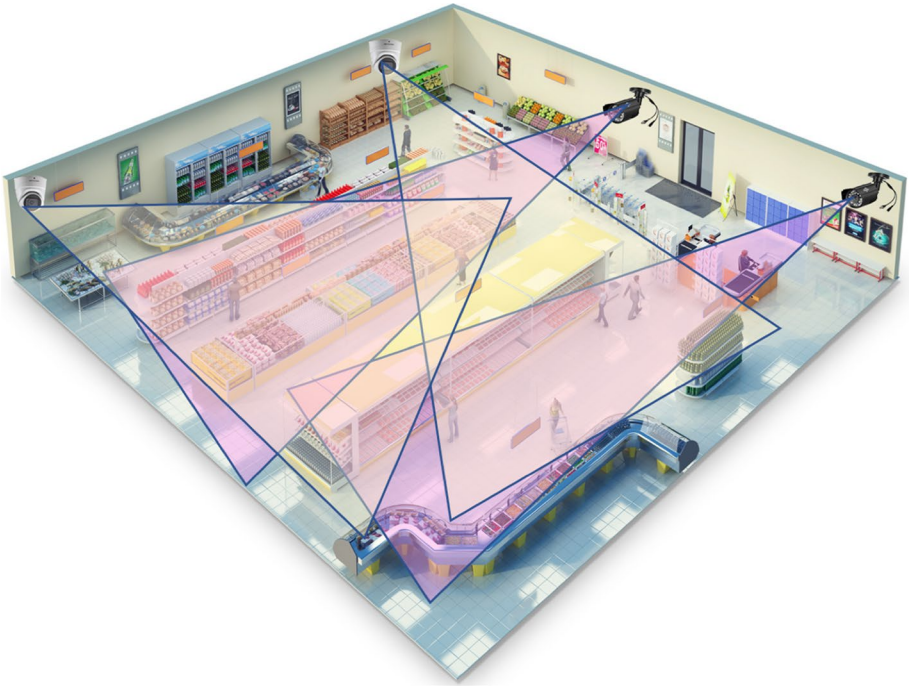


Fig. 1 Widely-employed security CCTV cameras in retail stores can provide in-store customer-behaviour insights to inform and improve store layout design

This research aims to conduct a comprehensive review on existing approaches in store layout design and modern AI techniques that can be utilized in the layout design task. Based on this review, we propose an AI-powered store layout design framework. This framework applies advanced AI and data analysis techniques on top of existing CCTV video surveillance infrastructure to understand, predict and suggest a better store layout. The framework facilitates customer-oriented store layout design by translating visual surveillance data into customer insights and business intelligence. It answers the following questions: How do shoppers really travel through the store? Do they go through every aisle, or do they change from one area to another in a more direct manner? Do they spend much of their time moving around the outer ring of the store, or do they spend most of their time in certain store sections? The big and rich visual data from the CCTV infrastructure allows us to answer such questions and optimize the store layout design towards both customers' convenience and satisfaction, and thereby increasing retailers' sales.

Scope: Retail is one of the biggest markets for AI. Currently, many forms of AI are used such as chat bots, price adjustment and predictions, visual search, virtual fitting rooms and supply chain management and logistics (https://spd.group/artificial-intelligence/ai-for-retail/?fbclid=IwAR0HM8tP2vQ9MI6jE2lrkD7JnyBP1NMlEAgRWqWWKKlHoctFctHnPC60J9M#Route_Optimization). This article focuses on the visual AI and analysis techniques which can be applied on top of the CCTV infrastructure of retails stores to improve customer-oriented store layout design. This article represents interdisciplinary research between AI and marketing, this article provides a foundation for academic researchers and practitioners from both fields to collaborate on this problem. In addition, this article offers

marketers and managers in retail insights to optimize store layout design, customer satisfaction and sales.

Our key contributions are twofold:

- Comprehensively reviewing existing approaches in store layout design and modern AI techniques that can be utilized in the layout design task. Section 2 reviews conventional methodology for supermarket layout design, discussing the design aims, layout types and conventional design approaches. Section 3 reviews how modern visual AI techniques are being used in retails, to analyze customer emotion and behaviors while shopping.
- Proposing a novel AI-powered store layout design framework. Section 4 provides details how the proposed framework applies advanced AI and data analysis techniques on top of existing CCTV video surveillance infrastructure to understand, predict and suggest a better store layout.

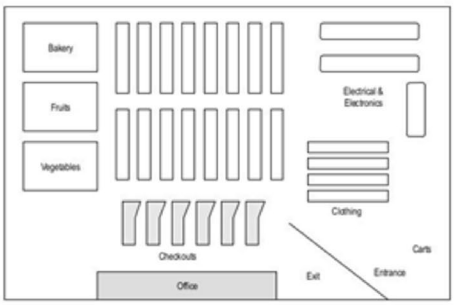
2 Conventional methodology for supermarket layout design

2.1 Layout design aims

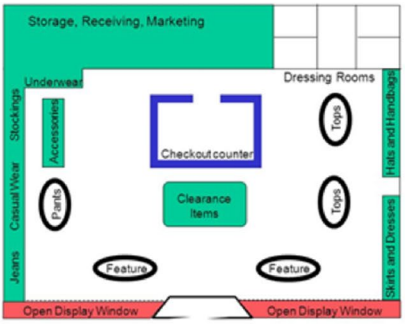
The ultimate goal of layout design is to increase the sale of stores by navigating consumer behavior (Vrechopoulos et al. 2004). Store layout aims to provide four factors such as perceived usefulness, ease of use, entertainment, and time-consuming (Hansen et al. 2010). The layout of a retail store is a key element in its success, and can increase store sales and profitability (Larson et al. 2005). The goal can be compiled into the following factors.

Expose customers to more products: to attract more purchasing decisions. To optimize the picking up more products, retailers may arrange the essential products at the end of aisles (Tan et al. 2018). Endcaps effectiveness provides the prominent location and attracts the higher shopper traffic (Page et al. 2019). In addition, during a special occasion, such as Easter, stores can put Easter eggs at the end of the aisle of the biscuits area. Customers need to walk through a variety of products to approach daily products. For example, during the travelling time, the essential products should be put at the end of the aisle, consumers need these products, then will be likely go through the entire aisles to pick them. It increases the opportunities to sell other products on the way they go (Page et al. 2019). Designing the shop floor layout for the sales magnets to display products of interest enables customers to circulate around many sections and allows significant contact with a variety of products (Ohta and Higuchi 2013).

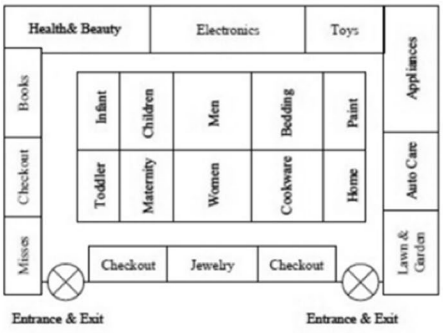
Increase browsing time: leading to different cluster configurations for short, medium, and long trips (Larson et al. 2005). The time spent around the supermarket is more or less to demonstrate whether customers are interested in shopping at supermarkets or not (Kim and Kim 2008; Sorensen 2016). Shopping can be a way to reduce stress and enjoy free time (Guiry et al. 2006; Hart et al. 2007). By planning the store layout, retailers encourage customers to flow through more aisles and see a wider variety of merchandise, which can increases the potential for more sales (Cil 2012), even lead to compulsive buying (Geetha et al. 2013). For example, the freeform layout shown in Fig. 2 is a free-flowing and asymmetric arrangement of displays and aisles, employing a variety of different sizes, shapes, and styles of display. This layout increases the recreation time consumers spend in the store



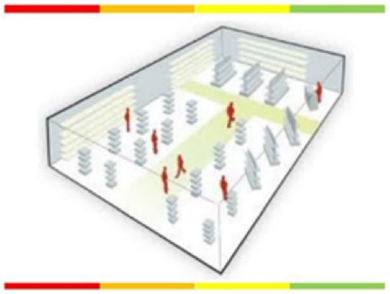
(a) The grid store layout



(b) The free-form store layout



(c) The racetrack store layout



(d) The circulation spine layout

Fig. 2 Four retail store layout types: **a** grid, **b** free form, **c** racetrack and **d** circulation

(Larsen et al. 2020; Lindberg et al. 2018). The trip duration increases store performance metrics by understanding consumers' needs and retailers' profits.

Easily find related products: to increase customer satisfaction. A store layout arranged in sets of products intended for purposeful buying can increase sales (Hansen et al. 2010). Thus, stores can arrange substitute goods and complementary goods in the same area. For example, tea and coffee are substitute goods. Grouping products in this way replaces consumers finding tea in the beverage section, cheese in fresh cheese, and cereal in the cereal section. This layout leads the customer along specific paths to visit as many store sections as possible (Kim and Kim 2008; Vrechopoulos et al. 2004). In addition, the industrial store layout normally arranges related products together such as in the bakery area (with bread, cakes, biscuits, and so on), and the vegetable area (with carrots, beans, and so on). Further, exposing consumers to well-presented merchandise in such areas can result in higher sales (Kiran et al. 2012).

Cost-control and Calculate stock inventory: setting up a suitable layout and using AI to track image and traffic flows allow retailers to control cost, calculate in-time stocks, and fill in products in the shelves (Yang and Chen 1999). Frontoni et al. indicated that machine learning can predict the available display area in the store and notice the lack of goods on the shelves to the center (Frontoni et al. 2017). Employees receive the notice and quickly fill out the product, not to miss consumers. For example, Walmart is testing by using robots

to scan shelves for missing items, and products that need to be restocked with offering the changing price tags and weights (https://spd.group/artificial-intelligence/ai-for-retail/?fbclid=IwAR0HM8tP2vQ9MI6jE2lrkD7JnyBP1NMlEAgRWqWWKKlHoctFctHnPC60J9M#Route_Optimization).

2.2 Layout types

There are four major store layout types: the grid layout, the freedom layout, the racetrack store, and the circulation spine layout (Cil 2012) as shown in Fig. 2.

The grid layout is a rectangular arrangement of displays and long aisles that generally run parallel to one another (Worse to come 2020). Fixtures and displays are laid parallel to the walls (Barghash et al. 2017). This type of layout is a popular choice for supermarkets, grocery stores and chain pharmacies (Vrechopoulos et al. 2004). The grid layout is known as end caps, and staple items such as milk and eggs are put at the back of the store. Thus, when consumers seek out essential products, they walk past a series of products to get them. This product exposure increases the potential for sales.

The free-form layout is a free-flowing and asymmetric arrangement of displays and aisles, employing a variety of different sizes, shapes, and styles of display (Cil 2012). This layout can increase the time consumers spend in a store. It also provides a wider view of products to consumers than the grid layout format. This increases the chance of consumers engaging in exploratory behavior where they look at new products.

The racetrack layout: offers a major aisle to control customer traffic to the store's multiple entrances, known as a loop layout, "usually in the shape of a circle, square, or rectangle-and then returns the customer to the front of the store" (American Marketing Association, 2020). This layout leads the customer along a specific path to ensure that they are exposed to as many store sections as possible (Vrechopoulos et al. 2004).

Circulation spine layout: is one where there is a traffic loop around the entire store, but the layout also includes a customer path right through the middle of the store (Cil 2012; Langevin et al. 1994).

2.3 Layout design approaches

Conventional retail store layout design approaches mainly rely on three criteria: product categories, cross-elasticity and consumption universes.

Product categories: This layout mainly displays the manufacture products, or categories products and bases on the industry implication (Cil 2012). Here, supermarkets cluster product groups to assist consumers in picking them faster and easier. For example, supermarkets frequently group products in a bakery area (e.g., grouping products such as, bread, cakes, and biscuits), vegetable area (e.g., beans, cabbages), and fruit area (e.g., apples and oranges). The shopper is used to finding products on the same shelf or in the same aisle (Mowrey et al. 2018). Jones et al. indicate that arrangement by product category can enhance consumer impulse buying (Jones et al. 2003).

Cross-elasticity: This layout emphasizes changing sales of one product through price changes in another product (Cil 2012; Hansen et al. 2010; Kamakura and Kang 2007), capturing cross-product interactions in demand via prices (Hwang et al. 2005; Murray et al. 2010). According to Walters and Jamil (2003), product categories are placed side by side following cognitively logical pairs, considered as cross-elasticities. Thus, it captures the use association among categories. With this type of layout, consumers can buy products

easily by comparing their strengths and weaknesses on the same store visit (DrÃšze et al. 1994; Murray et al. 2010).

Consumption universes: This refers to a "consumer-oriented store layout approach through a data mining approach" (Cil 2012). According to this layout, breakfast products including tea, bread, cheese, and cereal are presented in the same place (Cil 2012). This approach replaces finding tea in the beverage section, cheese in fresh cheese, and cereal in the cereal section. Other types of universe include the baby universe or tableware universe which can be clustered as different product categories (Borges 2003). Clustering products around consumer buying habits can have significant appeal to busy consumers. In such situations, consumers can feel satisfaction because the retailer appears to understand their needs and life circumstances, which in turn can positively affect consumer purchasing decisions. Indeed, store layout design can not only satisfy consumer requirements, but can also let consumers more easily accept the price, increases loyalty (Bill and Dale 2001; Koo and Kim 2013), and repeat purchasing (Soars 2003). In addition, store layout affects consumer behavior in terms of in-store traffic patterns, shopping atmosphere, shopping behavior, and operational efficiency (Bill and Dale 2001; Donovan et al. 1994). Store atmosphere increases the positive mood, and then consumers satisfaction with shopping (Anic et al. 2010; Hussain and Ali 2015), and then they will be willing to buy again (Jalil et al. 2016).

3 How visual AI techniques are being used in retailing

This section discusses how modern visual AI techniques are being used in retailing to understand consumers and their shopping behavior. The overarching aim of this research is understanding consumers and their behaviors while shopping with the current supermarket layout. The analysis can lead to better understanding of consumer shopping satisfaction, which can be used to optimize store layout design thereby making the shopping experience more convenient, more productive and more satisfying for consumers. AI and its sub-fields in computer vision, machine learning and deep learning function like the brain to process and interpret the footage coming from the eyes (CCTV cameras). Modern AI video analytic can be applied to footage coming from a network of CCTV cameras for insights into customer shopping activity. This includes customer-centric tasks (Gupta and Ramachandran 2021) such as customer detection, shopping cart detection, customer identification, customer tracking, customer emotion recognition and customer action recognition.

3.1 Customer and shopping cart detection

Object detection

Object detection involves AI techniques locating and classifying an object based on a range of predefined categories. In the store design setting, objects to located include humans (i.e., shoppers), shopping items, and the shopping cart. After object detection has been performed, bounding boxes are returned, where each box presents the spatial location and extent of each object instance in the image or video. For example, Fig. 3 shows an example of how an object detection algorithm detects the customer and the items which have been picked up in the shopping cart.

Object detection has been an active area of research (Zou et al. 2019) since it is the foundation to solve complex and high-level vision tasks such as identification and tracking (to be discussed later). Object detection is not simple as objects can vary significantly due

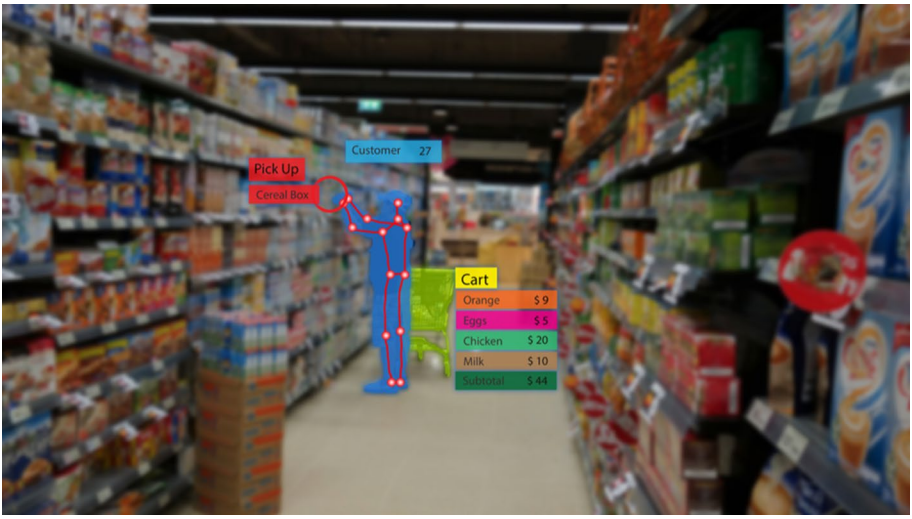


Fig. 3 Object detection in a supermarket setting

to variations in their pose, scale, resolution and lighting. For example, a person can wear any type of clothing from working uniforms to pajamas to go shopping. In addition, the person can stand anywhere in the recorded scene, at any distance from the camera and from any angle to the camera. Recent advances in deep learning, after the breakthrough results proposed by Krizhevsky et al. (2012), have significantly improved object detection performance to a level sufficient for real life deployment. The modern object detectors currently being deployed are divided in two categories: two-stage detectors and one-stage detectors.

- Two-stage detectors first extracts category-independent regions, then apply classification on the deep features of each region. State of the art examples of this category are Faster RCNN (Ren et al. 2015) and Mask RCNN (He et al. 2017).
- One-stage detectors unify the two stages into one by directly predicting class probabilities and bounding box offsets from full images with a single feed-forward CNN in a monolithic setting, that does not involve region proposal generation or post classification. State of the art examples of this category are YOLO (Bochkovskiy 2020) and EfficientNet (Tan and Le 2019).

In the retail store setting, humans and projects are the key objects of interest for detection. Detecting these objects exhibit unique and/or extreme challenges compared to general object detection. These challenges include complex backgrounds, uneven lighting, unusual viewing angle, specularity (Santra and Mukherjee 2019) and severe occlusions among groups of instances of the same categories (Cai et al. 2021; Karlinsky et al. 2017). Several works have employed HOG features and SVM for human detection in a retail store (kuang et al. 2015; Marder et al. 2015; Ahmed et al. 2017). However, these handcrafted based approaches are vulnerable to imaging conditions. Recently, many deep learning based approaches have been proposed. Cai et al. (2021) proposed a cascaded localization and counting network which simultaneously regresses bounding boxes of objects and counts the number of instances. Nogueira et al. (2019) proposed to preprocess surveillance videos with foreground-background classification before

feeding the RGB data and the foreground mask to their RetailNet for detection. Kim et al. (2019) investigated the performance of a range of popular deep learning based detectors including YOLO, SSD, RCNN, R-FCN and SqueezeDet in the retail setting. They found that Faster RCNN and R-FCN are the most accurate detectors for retails.

Pose estimation

Pose estimation aims to obtain posture of the human body from given images or footage. While object detection techniques infer the presence and location of customers, pose estimation techniques infer the body posture through body keypoints such as head, shoulders, elbows, wrists, hips, knees, and ankles (Chen et al. 2020). In the store design setting, pose estimation provides deep details to understand the action and interaction of a customer with surrounding objects. There are two categories of pose estimation: 2-dimension (2D) and 3-dimension (3D). While the 2D pose estimation task is predicting the location of body joints in the image (in terms of pixel values), the 3D pose estimation task predicts a three-dimensional spatial arrangement of all the body joints as its final output (Chen et al. 2020). In the 2D pose task, there are two main categories:

- Regression-based methods: attempt to learn a mapping from image to kinematic body joint coordinates by an end-to-end framework and generally directly produce joint coordinates. Typical examples of this category are DeepPose (Toshev and Szegedy 2014; Luvizon et al. 2019).
- Detection-based methods: predict approximate locations of body parts or joints, usually are supervised by a sequence of rectangular windows or heatmaps (Chen et al. 2020). Typical examples of this category are PartPoseNet (Tang and Wu 2019) and HRNet (Wang et al. 2020).

In the 3D pose task, there are two main categories:

- Model-based methods: employ a parametric body model or template to estimate human pose and shape from images. Two most popular parametric models in the literature are SMPL (Loper et al. 2015) and kinematic model (Mehta et al. 2017). The pose estimation task is then interpret ted as a model parameter estimation task, which requires much less parameters.
- Model-free methods: either directly map an image to 3D pose or estimate depth following intermediately predicted 2D pose from 2D pose estimation methods. Typical examples of this category are MocapNET (Qammaz and Argyros 2019) and 3DMDN (Li and Lee 2019).

Heatmap analytics

The output of human detection can be used for constructing business heatmaps. A heatmap can provide a visual summary of information by the two-dimensional representation of data, in which values are represented by colours as illustrated in Fig. 4. There can be a number of ways to display heatmaps, but they all share one thing in common—they use colours to draw the relationships between data values that would be much harder to understand if presented in a sheet of numerical values. For example, the following heatmap employs warmer colours for locations where all customers spend more time there. Heatmaps are being used to understand sales and marketing.

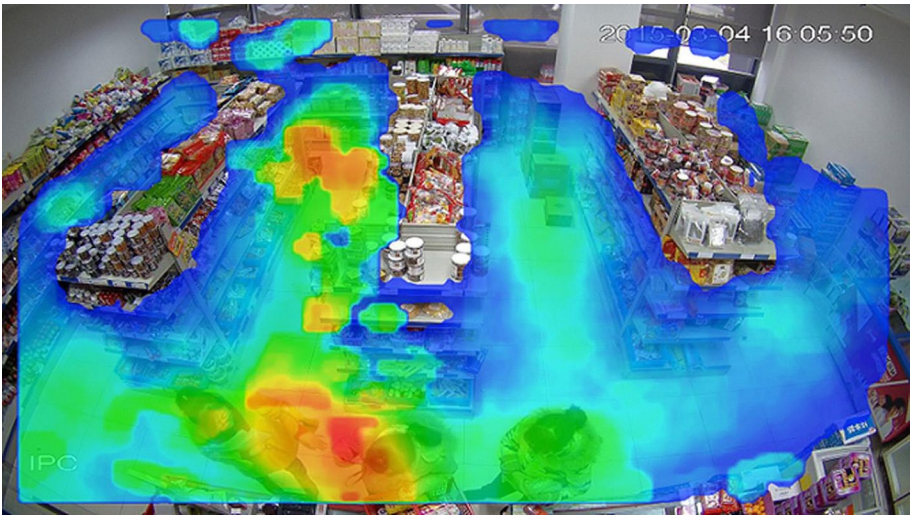


Fig. 4 Heatmaps in a supermarket setting

Fig. 5 Face recognition in a supermarket setting

3.2 Customer identification

Facial recognition

The facial region is one of the most natural human characteristics for identification due to the way we recognize humans. Facial recognition relies on the science of biometrics in detecting and measuring various characteristics, or feature points, of human faces for identification. When a new face comes into the scene, it is compared with a gallery of faces, which have been previously collected to infer the identity of the new face. In the supermarket setting, facial recognition can be used as a customer identification system that does not require physical ID cards to track either identity or shopping history as shown in Fig. 5.

Facial recognition has a long history and is actively researched within both the research community and industry. Facial recognition is not simple because of facial expressions—which can distort a person's face, poses—in which non-frontal angles could occlude parts of the face, and lighting—which can illuminate the face non-uniformly. Recent advances in deep learning, after the breakthrough results proposed by Krizhevsky et al. (2012), have brought facial recognition performance to a level equal or even surpassing that of humans. The modern face recognizers are based on deep learning techniques. Compared with conventional feature engineering approaches, deep learning face recognition approaches employ deep networks to train end-to-end without the need for human feature engineering involvement. The prominent approach in the deep learning category is FaceNet proposed in 2015 by researchers from Google Inc. Schroff et al. (2015). FaceNet achieved the state-of-the-art accuracy on the famous LFW benchmark, approaching human performance on the unconstrained condition for the first time (FaceNet: 97.35% vs. Human: 97.53%) by training a 9-layer model on 4 million facial images. Recent research has even managed to dramatically boost the performance to above 99.80% in ArcFace (Deng et al. 2019) in 2019.

While the use of facial recognition in the retail setting is still controversial, especially due to the privacy concern, it is still a possible application of AI in the retail setting. However, performing facial recognition in the retail CCTV cameras could exhibit challenges that have to be further investigated. The first challenge is the small resolution of the faces presented in images due to the subject-camera distance or the actual resolution of a CCTV camera (Nguyen et al. 2012; Lin et al. 2007). The second challenge is the pose of a face or non-frontal faces. The unconstrained dynamics of customers while travelling a shop may make it challenging to capture an ideal frontal face for recognition. Super-resolution can be used to improve the resolution of facial images, either in the pixel domain or feature domain (Nguyen et al. 2018; Jiang et al. 2021). Generative models can be used to synthesize a face from a multitude of angles to deal with the angle view challenge (Wang and Deng 2021).

Customer characterization

A person's face contains not only information about identity, but also important demographic attributes, including a person's age, gender and ethnicity. As shown in Fig. 5, the age, gender and ethnicity of both customers are being extracted from the facial image. In the supermarket setting, this information could be used in many ways for customer-tailored targeting.

Modern face attributes estimation algorithms also rely heavily on deep learning to extract deep features from the facial image (Zheng et al. 2020). The prominent approaches in facial attribute estimation are Multi-task Convolutional Neural Network (MT-CNN) with local constraint for face attribute learning (Cao et al. 2018) and Deep Multi-Task Learning (DMTL) approach for Heterogeneous Face Attribute Estimation proposed in 2018 in Han et al. (2018). Experiments on large scale facial datasets showed promising 86%, 96% and 94% accuracy for age, gender and race estimation respectively.

3.3 Customer tracking

Human trajectory tracking

While each camera can detect the location of a human, tracking customers over a whole CCTV camera network is of great interest to business analytic. Human tracking over camera networks is composed of two functional modules: human tracking within a camera and human tracking across non-overlapping cameras. The human tracking within a camera

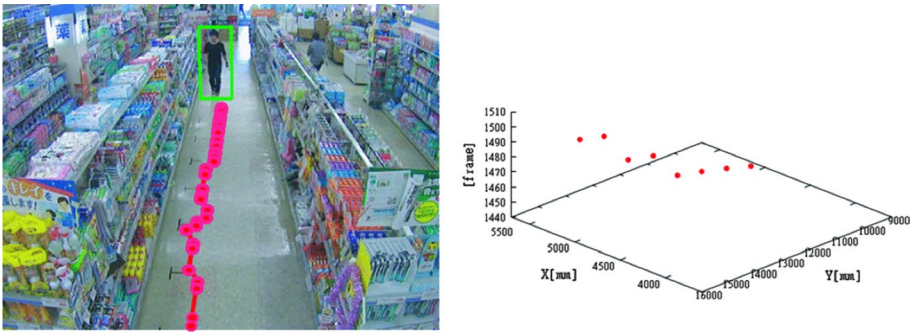


Fig. 6 Object tracking in a supermarket setting

focuses on locating human objects in each frame of a given video sequence from a camera, while the human tracking across multiple cameras concentrates on associating one tracked human object from the Field Of View (FOV) of a camera with that from the FOV of another camera (Fernando et al. 2018) (Fig. 6).

Human tracking within a camera

Human tracking within a camera generates the moving trajectories of human objects over time by locating their positions in each frame of a given video sequence. Based on the correlation among the human objects, the human tracking within a camera can be categorized as two types, the generative trackers and the discriminative trackers.

- Generative trackers: each target location and correspondence are estimated by iteratively updating the respective location obtained from the previous frame. During the iterative search process for human objects, in order to avoid exhaustive search of the new target location to reduce the cost of computation, the most widely used tracking methods include Kalman filtering (KF) (Kalman 1960), Particle filtering (PF), and kernel-based tracking (KT).
- Discriminative trackers: all the human locations in each video frame are first obtained through a human detection algorithm, and then the tracker jointly establishes these human objects' correspondences across frames through a target association technique. The most widely used target association techniques include joint probability data association filtering (JPDAF), multiple-hypothesis tracking (MHT), and flow network framework (FNF).

There are two major problems in tracking: single-object tracking (SOT) and multiple-object tracking (MOT) (Wu et al. 2021). In a crowded supermarket, MOT is preferred considering it can track simultaneously multiple customers. The current state of the art MOT is based on deep learning such as ByteTrack (Zhang et al. 2021), which achieved 80.3 MOT accuracy on the test set of MOT17. New deep learning architectures such as transformers have also been shown achieving top performance in tracking benchmarks such as TransMOT (Chu et al. 2021).

In the specific setting of a retail store, a number of work have attempted to employ modern trackers for the customer tracking task. Nguyen and Tran (2020) employed Siamese-family trackers for tracking customers in crowded retail scenes. Leykin and Tuceryan (2005) designed a Bayesian jump-diffusion filter to track customers in a store and performs

a number of activity analysis tasks. While most modern trackers can be applied to the customer tracking task in a retail store setting, the tradeoff between speed and accuracy has to be considered (Wojke et al. 2017).

Human tracking across cameras

Human tracking across non-overlapping cameras establishes detected/tracked human objects' correspondence between two non-overlapping cameras to successfully perform label handoff. Based on the approaches used for target matching, human tracking across cameras can be divided into three main categories, human re-ID, CLM-based tracking, and GM-based tracking.

In addition, the use of multiple cameras allows us to obtain the 3D position of subjects using triangulation, but there are works capable of obtaining the 3D position of humans using a single camera (Neves et al. 2015), which reduces the system cost.

3.4 Customer emotion recognition

Emotion recognition

Humans express emotion through observable facial expressions such as raising an eyebrow, eyes opening or changing the expression of their mouth (e.g., smiling). Understanding customer emotion while they are browsing through the store could provide marketers and managers a valuable tool to understand customer reactions to the products they sell (Martin 2003; Martin and Lawson 1998). Given this potential, emotion recognition has grown significantly to a $20 billion industry. Many big names like Amazon, Microsoft and IBM now advertise "emotion analysis" as one of their facial recognition products.

Emotion recognition algorithms work by employing computer vision techniques to locate the face, and identify key landmarks on the face, such as corners of the eyebrows, tip of the nose, and corners of the mouth (Nguyen et al. 2017a, 2017b). Most approaches in the literature classify emotion into one of seven expression classes: fear, anger, joy, sadness, acceptance or disgust, expectancy, surprise (Li and Deng 2020). Beyond market research, emotion detection technology is now being used to monitor and detect driver impairment, test user experience for video games and to help medical professionals assess the well-being of patients (Don't look now 2020).

Multimodal emotion recognition

Humans also express emotion in many other ways such as raising their voices and pointing their fingers. Studies by Albert Mehrabian in the 80s (Mehrabian 1981) established the 7–38–55% rule, also known as the "3V rule": 7% of communication is verbal, 38% of communication is vocal and 55% of communication is visual. Multimodal Emotion Recognition approaches rely on a combination of facial, body and verbal signals to infer the emotion of a subject (Sharma and Dhall 2021). One state of the art example of multimodal emotion recognition is End-to-End Multimodal Emotion Recognition Using Deep Neural Networks (Tzirakis et al. 2017), in which the network comprises of two parts: the multimodal feature extraction part and the RNN part. The multimodal part extracts features from raw speech and visual signals. The extracted features are concatenated and used to feed 2 LSTM layers. These are used to capture the contextual information in the data. Other than LSTM, a Deep Belief Network can also be employed to further capture the interaction between multiple modalities (Nguyen et al. 2017a, 2017b). Contextual information can be further captured and modelled to improve the accuracy (Mittal et al. 2020).

In the supermarket setting, three main modalities of interest to multimodal emotion recognition are facial expression, speech and body gesture. Due to the distance from the

Fig. 7 Action recognition in a supermarket setting

cameras to the customers in the supermarket setting, facial expression and body gesture are more popular than speech.

3.5 Customer action recognition

Understanding customer behaviors is the ultimate goal for business intelligence. How customers behave or act reveals their interest in the product. Obvious actions like picking up products, putting products into the trolley, and returning products back to the shelf have attracted great interest for the smart retails. Other behaviors like staring at a product and reading the box of a product are a gold mine for marketing to understand the interest of customers in a product. From a marketing point of view, these behaviors provide evidence to investigate the elements of the decision-making process of purchase that determines a particular choice of consumers and how marketing tactics can influence consumers. Empirical researches on consumer behaviour are primarily based on the cognitive approach, which allows to predict and define possible actions that lead to the conclusion and to suggest implications for communication strategies and marketing (Le 2019; Martin and Nuttall 2017; Martin and Strong 2016).

Understanding and recognizing behaviors of a customer are based on the body gestures of the customer in relation to the shelves, the products and the trolley/basket he/she is using (Gammulle et al. 2020). The analysis is based on recognizing the head orientation, eye gazing, and 2-dimensional (2D) and 3-dimensional (3D) pose estimation as illustrated in Fig. 7. Compared with the traditional pose estimation approaches which required motion capture systems or depth cameras, modern approaches only need a monocular RGB video to infer 2D and 3D skeleton in real time.

Modern action recognition algorithms are based on the temporal movement of the skeleton estimated from the pose to extract motion patterns form a certain skeleton sequence and classify the action into a number of action categories of interest (Herath et al. 2017;

Zhang et al. 2019). There are three main approaches to process the sequence of a 2D and 3D skeleton:

- RNN-based: Recurrent Neural Networks are networks that process a sequence of time-series data recursively. The skeleton sequence estimated in an action video is fed into a RNN, which will model the temporal relationship among the sequence to classify the action observed into one of the pre-defined action categories. One state of the art example of this category is HBRNN-L (Yong et al. 2015).
- CNN-based: Convolutional Neural Networks are networks that process images (2D CNNs) and videos (3D CNNs) using a hierarchy of convolutional layers which gradually learn high level semantic cues with its natural equipped excellent ability to extract high level information. The skeleton sequence estimated from in an action video is usually transformed into a pseudo-image to be processed by a CNN. One state of the art example of this category is Caetano et al. (2019).
- GCN-based: Graph Convolutional Networks are networks that perform convolution operations on a graph, instead of on an image composed of pixels. Since skeleton data is naturally a topological graph instead of a sequence vector or pseudo-image, GCNs have been adopted for this task. One state of the art example of this category is ST-GCN (Yan et al. 2018).

Many researchers have applied these modern action recognisers to the customer action recognition task in a retail store setting. Liu et al. proposed combined hand feature (CHF), which includes hand trajectory, tracking status and the relative position between hand and shopping basket, classify arm motions into several classes of arm actions (Liu et al. 2015). Hoang et al. proposed a hierarchical finite state machine to detect human activities in retail surveillance (Trinh et al. 2011). Wang (2020) proposed a hierarchy-based system for recognizing customer activity in retail environments. Behaviors of customers can also be analysed offline when there is no need for real time instant decision. Researchers have also considered combining conventional RGB cameras with Depth sensors, which provide richer information about the 3D location and subject shapes for customer interaction analysis (Frontoni et al. 2013),

4 How AI techniques can be employed to improve layout design

Based on the analysis and review in Sects. 2 and 3, we propose a comprehensive framework to apply visual AI and data analytic techniques to the store layout design task. This framework will be referred to as AI-powered Store Layout Design.

Conceptual design

The conceptual diagram of the Sense-Think-Act-Learn (STAL) framework is presented in Fig. 8. The highest-level architecture consists of Sense—Think—Act—Learn modules. Firstly, "Sense" is to collect raw data, i.e., video footage from a store's CCTV cameras for subsequent processing and analysis, similar to how humans use their senses. Secondly, "Think" is to process the data collected through advanced AI and data analytic, i.e., intelligent video analytic and AI algorithms, similar to how humans use their brains to process the incoming data. Thirdly, "Act" is to use the knowledge and insights from the second phase to improve and optimize the supermarket layout. The process operates as a continuous learning cycle. An advantage of this framework is that it allows

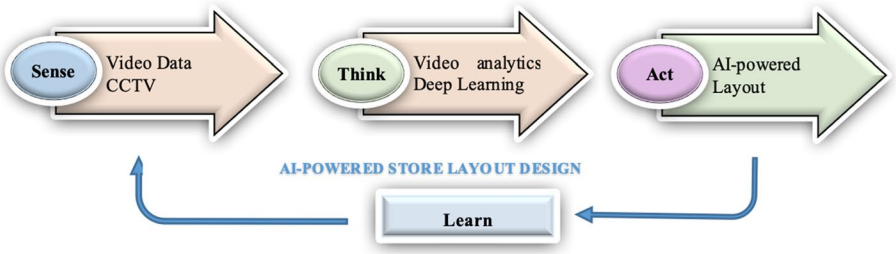


Fig. 8 STAL: cyclic conceptual diagram sense—think—act—learn model for store layout design

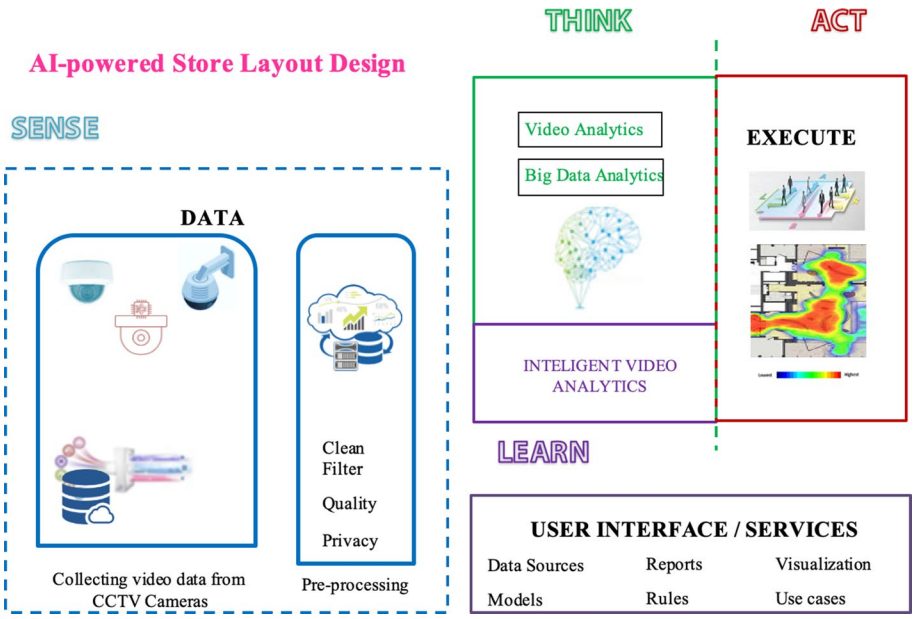


Fig. 9 AI-powered store layout ecosystem architecture

retailers to test store design predictions such as the traffic flow behavior when customers enter a store or the popularity of store displays placed in different areas of the store (Ferracuti et al. 2019; Underhill).

System architecture

The proposed framework has a multiple-layer architecture corresponding to multiple phases of the conceptual diagram. The proposed framework is presented in Fig. 9.

The data layer in the SENSE phase includes data streams from the CCTV cameras and video recordings, metadata, customer data, market layout data, etc. In addition to dynamically formed information sources (flowing data), there may also be static information sources such as floor data. The collectors in the data layer collect this data continuously or when demand occurs at certain intervals and a resource pool is created. Data will be filtered and cleaned to improve the quality and privacy. At this stage, it will also be possible to transform the data into a structural form. Given privacy is a key

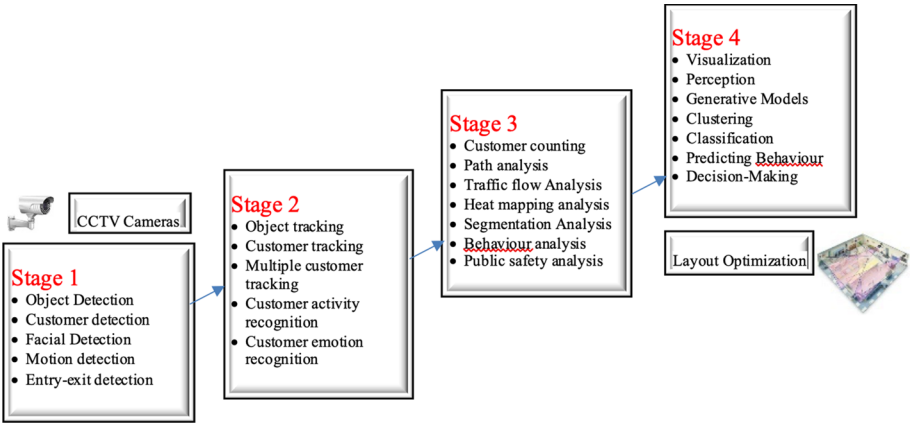


Fig. 10 A processes of intelligent video analytic for supermarket layout optimization

concern for customers, data can be de-identified or made anonymous, for example, by examining customers at an aggregate level (Lewinski et al. 2016).

The pre-processing layer in the SENSE phase applies several techniques to improve the quality of captured images and videos such as cleaning, de-noising, de-blurring and transformation correcting. Since there is an intense data flow from the CCTV cameras, a cloud-based system can be considered as a suitable approach for supermarket layout analysis in processing and storing video data. Large video data can be stored successfully and steadily on cloud storage servers for extended periods. The proposed architecture is intelligent and fully scalable. Local client-server architectures are not a suitable solution for supermarkets, considering the need for system maintenance and qualified personnel and the financial costs.

The intelligent video analytic layer in the THINK phase plays the key role in interpreting the content of images and videos. Intelligent video analysis involves a variety of AI, computer vision, machine learning, and deep learning techniques for detection, identification, tracking, analyzing, and extracting meaningful information from images and video streams. Details on these techniques have been discussed in Sect. 3. A general video analysis process for supermarket layout optimization based on deep learning is summarized in Fig. 10.

The execute layer in the ACT phase employs the analytical results and insights from the analytic layer to take actions by improving layout, measuring the success of the improved layout, evaluating the results obtained and continuous revision of the created layout. Two examples of using the STAL framework would be studying maps of customer density or time spent in stores (see Ferracuti et al. 2019) to generate optimal layouts. Layout variables managers can consider include store design variables (e.g., space design, point-of-purchase displays, product placement, placement of cashiers), employees (e.g., number, placement), and customers (e.g., crowding, visit duration, impulse purchases, use of furniture, waiting queue formation, receptivity to product displays).

Data flow

The information flows between the components that make up the architectural structure designed for the layout to be designed with AI support is shown in Fig. 11. In the Sense stage, the data flows out of the Data Generation and Pre-processing layers in the form of

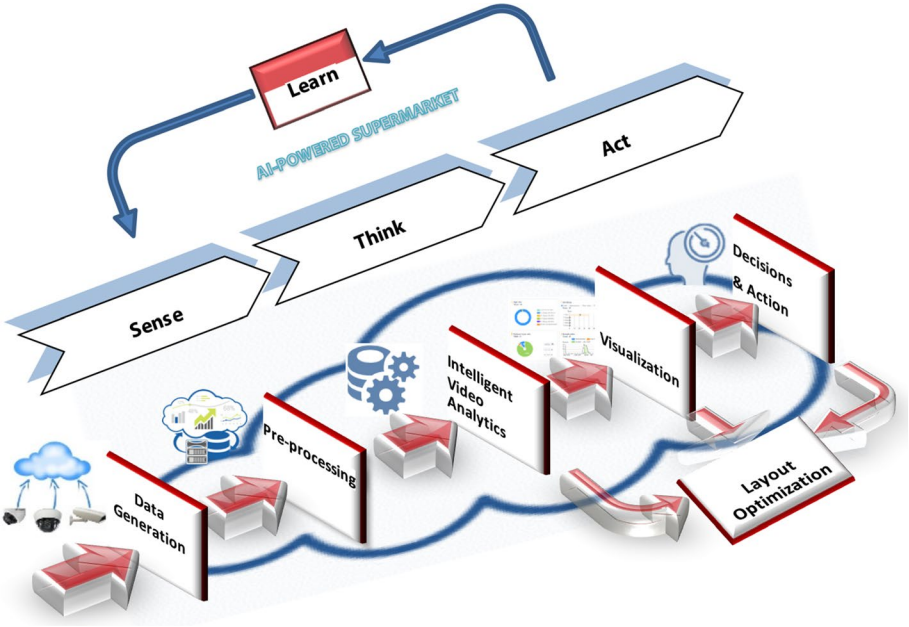


Fig. 11 AI-powered supermarket layout design model and data flow

raw videos and processed videos, respectively. In the Think stage, the data flows out of the Intelligent Video Analytic in the form of analysis and interpretation such as measurements, statistics, and analytic. The analysis will be fed to the Visualization layer to output graphs and diagrams. Ultimately, intelligent decisions are made, suggestions are presented, and inferences are made to improve the layout. In this framework, a cyclical process (sense-think-act) is repeated as a result of the machine learning which is taking place. This iterative process results in changes in store layout which should improve customer satisfaction. For example, research shows how there is often a mismatch between customer expectations and what retailers do regarding shelf placement. Further, this mismatch has the potential to make customers frustrated (Valenzuela et al. 2013). Following the STAL model would reduce such discrepancies between consumer expectations and store design (Valenzuela et al. 2013). In so doing, it provides an example of how AI can be used to enhance business productivity (Kumar et al. 2021). Feedback from the STAL model can also be used to aid managers in the market segmentation of different groups of consumers with different behavior patterns (e.g., time in store, size and frequency of purchase, price and store display promotion sensitivity).

5 Discussion and conclusions

Improving supermarket layout design is one important tactic to improve customer satisfaction and increase sales. This paper reviews existing approaches in the layout design task, AI and big data techniques that can be applied in the layout design problem, and most importantly proposes a comprehensive and novel framework to apply AI techniques on top

of the existing CCTV cameras to interpret and understand customers and their behavior in store. There are a number of research directions to be further explored:

- Research performance of the proposed framework in real stores at different scales and types of merchandises;
- Research the impact of unique retail store settings on the performance of AI components (e.g. detection, tracking, identification and behaviour analysis);
- Research the impact of the change in marketing strategies caused by the insights generated from the proposed framework.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

References

Ahmed AH, Kpalma K, Guedi AO (2017) Human detection using hog-svm, mixture of gaussian and background contours subtraction. In: 2017 13th International conference on signal-image technology internet-based systems (SITIS), pp 334–338.https://doi.org/10.1109/SITIS.2017.62

Anic ID, Radas S, Lim LK (2010) Relative effects of store traffic and customer traffic flow on shopper spending. Int Rev Retail Distrib Consum Res 20(2):237–250

Artificial intelligence for retail in 2020: 12 real-world use cases. https://spd.group/artificial-intelligence/ai-for-retail/?fbclid=IwAR0HM8tP2vQ9MI6jE2lrkD7JnyBP1NMlEAgRWqWWKKlHoctFctHnPC60J9M#Route_Optimization. Accessed: 2020-10-15

Barghash MA, Al-Qatawneh L, Ramadan S, Dababneh A (2017) Analytical hierarchy process applied to supermarket layout selection. J Appl Res Ind Eng 4(4):215–226

Bill M, Dale M (2001) Superstore interactivity: a new self-service paradigm of retail service? Int J Retail Distrib Manag 29(8):379–389

Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: optimal speed and accuracy of object detection. CoRR arxiv: abs/2004.10934 (2020)

Borges A (2003) Toward a new supermarket layout: from industrial categories to one stop shopping organization through a data mining approach. In Proceedings of the 2003 society for marketing advances annual symposium on retail patronage and strategy, Montreal, November 4–5

Caetano C, Sena J, Brémond F, Dos Santos JA, Schwartz WR (2019) Skelemotion: a new representation of skeleton joint sequences based on motion information for 3d action recognition. In: 2019 16th IEEE international conference on advanced video and signal based surveillance (AVSS), pp 1–8

Cai Y, Wen L, Zhang L, Du D, Wang W (2021) Rethinking object detection in retail stores. In: The 35th AAAI conference on artificial intelligence (AAAI 2021)

Cao J, Li Y, Zhang Z (2018) Partially shared multi-task convolutional neural network with local constraint for face attribute learning. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, pp 4290–4299

Chen Y, Tian Y, He M (2020) Monocular human pose estimation: a survey of deep learning-based methods. Comput Vis Image Underst 192:102897

Chu P, Wang J, You Q, Ling H, Liu Z (2021) Transmot: Spatial-temporal graph transformer for multiple object tracking. CoRR arxiv: abs/2104.00194
Cil I (2012) Consumption universes based supermarket layout through association rule mining and multidimensional scaling. Expert Syst Appl 39(10):8611–8625
Cil I, Ay D, Turkan YS (2009) Data driven decision support to supermarket layout. In: Proceedings of the 8th WSEAS international conference on artificial intelligence, knowledge engineering and data bases, AIKED'09, pp 465–470. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin
Davenport T, Guha A, Grewal D, Bressgott T (2020) How artificial intelligence will change the future of marketing. J Acad Mark Sci 48(1):24–42
Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 4685–4694
Donovan RJ, Rossiter JR, Marcoolyn G, Nesdale A (1994) Store atmosphere and purchasing behavior. J Retail 70(3):283–294
Don't look now: why you should be worried about machines reading your emotions. https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science. Accessed 16 Oct, 2020
DrÃšze X, Hoch SJ, Purk ME (1994) Shelf management and space elasticity. J Retail 70(4):301–326
Fernando T, Denman S, Sridharan S, Fookes C (2018) Tracking by prediction: a deep generative model for mutli-person localisation and tracking. In: 2018 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1122–1132
Ferracuti N, Norscini C, Frontoni E, Gabellini P, Paolanti M, Placidi V (2019) A business application of rtls technology in intelligent retail environment: defining the shopper's preferred path and its segmentation. J Retail Consum Serv 47:184–194
Frontoni E, Raspa P, Mancini A, Zingaretti P, Placidi V (2013) Customers' activity recognition in intelligent retail environments. In: New trends in image analysis and processing—ICIAP 2013. Springer, Berlin, Heidelberg, pp 509–516
Frontoni E, Marinelli F, Rosetti R, Zingaretti P (2017) Shelf space re-allocation for out of stock reduction. Compu Ind Eng 106:32–40
Gammulle H, Denman S, Sridharan S, Fookes C (2020) Fine-grained action segmentation using the semi-supervised action gan. Pattern Recogn 98:107039
Geetha M, Bharadhwaj S, Piyush S (2013) Impact of store environment on impulse buying behavior. Eur J Mark 47(10):1711–1732
Grewal D, Noble SM, Roggeveen AL, Nordfalt J (2020) The future of in-store technology. J Acad Mark Sci 48(1):96–113
Guiry M, Mägi AW, Lutz RJ (2006) Defining and measuring recreational shopper identity. J Acad Mark Sci 34(1):74–83
Gupta S, Ramachandran D (2021) Emerging market retail: transitioning from a product-centric to a customer-centric approach. J Retail. https://doi.org/10.1016/j.jretai.2021.01.008
Han H, Jain AK, Wang F, Shan S, Chen X (2018) Heterogeneous face attribute estimation: a deep multitask learning approach. IEEE Trans Pattern Anal Mach Intell 40(11):2597–2609
Hansen JM, Raut S, Swami S (2010) Retail shelf allocation: a comparative analysis of heuristic and meta-heuristic approaches. J Retail 86(1):94–105
Hart C, Farrell AM, Stachow G, Reed G, Cadogan JW (2007) Enjoyment of the shopping experience: impact on customers' repatronage intentions and gender influence. Serv Ind J 27(5):583–604
He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: 2017 IEEE international conference on computer vision (ICCV), pp 2980–2988
Herath S, Harandi M, Porikli F (2017) Going deeper into action recognition: a survey. Image Vis Comput 60:4–21
Hussain R, Ali M (2015) Effect of store atmosphere on consumer purchase intention. IDEAS working paper series from RePEc
Hwang H, Choi B, Lee MJ (2005) A model for shelf space allocation and inventory control considering location and inventory level effects on demand. Int J Prod Econ 97(2):185–195
Jalil NAA, Fikry A, Zainuddin A (2016) The impact of store atmospherics, perceived value, and customer satisfaction on behavioural intention. Procedia Economics and Finance **37**, 538 – 544. The Fifth international conference on marketing and retailing (5th INCOMaR) 2015
Jiang J, Wang C, Liu X, Ma J (2021) Deep learning-based face super-resolution: a survey. ACM Comput Surv. arxiv: abs/2101.03749
Johnson E The Real Cost of Your Shopping Habits, Forbes 2015

Jones MA, Reynolds KE, Weun S, Beatty SE (2003) The product-specific nature of impulse buying tendency. J Bus Res 56(7):505–511 (**Retailing Research**)
Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45
Kamakura WA, Kang W (2007) Chain-wide and store-level analysis for cross-category management. J Retail 83(2):159–170
Karlinsky L, Shtok J, Tzur Y, Tzadok A (2017) Fine-grained recognition of thousands of object categories with single-example training. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 965–974. https://doi.org/10.1109/CVPR.2017.109
Kim HY, Kim YK (2008) Shopping enjoyment and store shopping modes: the moderating influence of chronic time pressure. J Retail Consum Serv 15(5):410–419
Kim CE, Dar Oghaz MM, Fajtl J, Argyriou V, Remagnino P (2019) A comparison of embedded deep learning methods for person detection. Prague, pp 459–465
Kiran V, Majumdar M, Kishore KK (2012) Innovation in in-store promotions: effects on consumer purchase decision. Eur J Bus Manag 4:36–44
Koo W, Kim YK (2013) Impacts of store environmental cues on store love and loyalty: single-brand apparel retailers. J Int Consum Mark 25(2):94–106
Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, pp 1097–1105
Kuang Wu Y, Wang HC, Chang LC, Chou SC (2015) Customer's flow analysis in physical retail store. Procedia Manufacturing **3**, 3506–3513 (2015). 6th International conference on applied human factors and ergonomics (AHFE 2015) and the affiliated conferences, AHFE 2015
Kumar V, Ramachandran D, Kumar B (2021) Influence of new-age technologies on marketing: a research agenda. J Bus Res 125:864–877
Langevin A, Montreuil B, Riopel D (1994) Spine layout design. Int J Prod Res 32(2):429–442
Larsen NM, Sigurdsson V, Breivik J, Orquin JL (2020) The heterogeneity of shoppersâ supermarket behaviors based on the use of carrying equipment. J Bus Res 108:390–400
Larson JS, Bradlow ET, Fader PS (2005) An exploratory look at supermarket shopping paths. Int J Res Mark 22(4):395–414
Le MTH (2019) Brand fanaticism: scale development
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–44
Lewinski P, Trzaskowski J, Luzak J (2016) Face and emotion recognition on commercial property under eu data protection law. Psychol Mark 33(9):729–746. https://doi.org/10.1002/mar.20913
Leykin A, Tuceryan M (2005) Tracking and activity analysis in retail environments. https://www.researchgate.net/publication/228907903_Tracking_and_Activity_Analysis_in_Retail_Environments_Technical_Report_620
Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE Trans Affect Comput 1–1
Li C, Lee GH (2019) Generating multiple hypotheses for 3d human pose estimation with mixture density network. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 9879–9887
Lin F, Fookes C, Chandran V, Sridharan S (2007) Super-resolved faces for improved face recognition from surveillance video. In: Lee SW, Li SZ (eds) Adv Biom. Springer, Berlin, Heidelberg, pp 1–10
Lindberg U, Salomonson N, Sundstrom M, Wendin K (2018) Consumer perception and behavior in the retail foodscapeâ a study of chilled groceries. J Retail Consum Serv 40:1–7
Liu J, Gu Y, Kamijo S (2015) Customer behavior recognition in retail store from surveillance camera. In: 2015 IEEE international symposium on multimedia (ISM), pp 154–159. https://doi.org/10.1109/ISM.2015.52
Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ (2015) SMPL: a skinned multi-person linear model. ACM Trans Graph 34(6):248:1-248:16 (**Proc. SIGGRAPH Asia**)
Luvizon DC, Tabia H, Picard D (2019) Human pose regression by combining indirect part detection and contextual information. Comput Graph 85:15–22
Marder M, Harary S, Ribak A, Tzur Y, Alpert S, Tzadok A (2015) Using image analytics to monitor retail store shelves. IBM J Res Dev 59(2/3):31–311. https://doi.org/10.1147/JRD.2015.2394513
Martin BAS (2003) The influence of gender on mood effects in advertising. Psychol Mark 20(3):249–273
Martin BAS (2011) A stranger's touch: effects of accidental interpersonal touch on consumer evaluations and shopping time. J Consum Res 39(1):174–184
Martin B, Lawson R (1998) Mood and framing effects in advertising. Austral Mark J (AMJ) 6(1):35–50. https://doi.org/10.1016/S1441-3582(98)70238-1
Martin BAS, Nuttall P (2017) Tense from touch: examining accidental interpersonal touch between consumers. Psychol Mark 34(10):946–955

Martin BAS, Strong CA (2016) The trustworthy brand: effects of conclusion explicitness and persuasion awareness on consumer judgments. Mark Lett 27(3):473–485
Mehrabian A (1981) Silent messages: implicit communication of emotions and attitudes, 2nd edn. Wadsworth Pub. Co., Belmont
Mehta D, Rhodin H, Casas D, Fua P, Sotnychenko O, Xu W, Theobalt C (2017) Monocular 3d human pose estimation in the wild using improved cnn supervision. In: 2017 international conference on 3D vision (3DV), pp 506–516
Mittal T, Guhan P, Bhattacharya U, Chandra R, Bera A, Manocha D (2020) Emoticon: context-aware multimodal emotion recognition using frege's principle. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE Computer Society, pp 14222–14231
Mowrey CH, Parikh PJ, Gue KR (2018) A model to optimize rack layout in a retail store. Eur J Oper Res 271(3):1100–1112
Murray CC, Talukdar D, Gosavi A (2010) Joint optimization of product price, display orientation and shelf-space allocation in retail category management. J Retail 86(2):125–136 (**Special Issue: Modeling Retail Phenomena**)
Neves JC, Moreno JC, Barra S, Proença H (2015) Acquiring high-resolution face images in outdoor environments: a master-slave calibration algorithm. In: 2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS), pp 1–8. https://doi.org/10.1109/BTAS.2015.7358744
Newman A, Foxall G (2003) In-store customer behaviour in the fashion sector: some emerging methodological and theoretical directions. Int J Retail Distrib Manag 31:591–600
Newman AJ, Yu DK, Oulton DP (2002) New insights into retail space and format planning from customer-tracking data. J Retail Consum Serv 9(5):253–258
Nguyen PA, Tran ST (2020) Tracking customers in crowded retail scenes with siamese tracker. In: 2020 RIVF international conference on computing and communication technologies (RIVF), pp 1–6. https://doi.org/10.1109/RIVF48685.2020.9140794
Nguyen K, Sridharan S, Denman S, Fookes C (2012) Feature-domain super-resolution framework for gabor-based face and iris recognition. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2642–2649
Nguyen D, Nguyen K, Sridharan S, Ghasemi A, Dean D, Fookes C (2017a) Deep spatio-temporal features for multimodal emotion recognition. In: 2017 IEEE winter conference on applications of computer vision (WACV), pp 1215–1223
Nguyen D, Nguyen K, Sridharan S, Ghasemi A, Dean D, Fookes C (2017b) Deep spatio-temporal features for multimodal emotion recognition. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1215–1223
Nguyen K, Fookes C, Sridharan S, Tistarelli M, Nixon M (2018) Super-resolution for biometrics: a comprehensive survey. Pattern Recogn 78:23–42
Nogueira V, Oliveira H, Augusto Silva J, Vieira T, Oliveira K (2019) Retailnet: a deep learning approach for people counting and hot spots detection in retail stores. In: 2019 32nd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), pp 155–162. https://doi.org/10.1109/SIBGRAPI.2019.00029
Ohta M, Higuchi Y (2013) Study on the design of supermarket store layouts: the principle of sales magnet. Int J Soc Behav Educ Bus Ind Eng 7:209–212
Page B, Trinh G, Bogomolova S (2019) Comparing two supermarket layouts: the effect of a middle aisle on basket size, spend, trip duration and endcap use. J Retail Consum Serv 47:49–56
Qammaz A, Argyros AA (2019) Mocapnet: ensemble of snn encoders for 3d human pose estimation in rgb images. In: British machine vision conference (BMVC 2019). BMVA, Cardiff. http://users.ics.forth.gr/argyros/res_mocapnet.html
Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks, pp 91 – 99
Rhee H, Bell DR (2002) The inter-store mobility of supermarket shoppers. J Retail 78(4):225–237
Roggeveen AL, Sethuraman R (2020) Customer-interfacing retail technologies in 2020 and beyond: an integrative framework and research directions. J Retail 96(3):299–309
Santra B, Mukherjee DP (2019) A comprehensive survey on computer vision based approaches for automatic identification of products in retail store. Image Vis Comput 86:45–63
Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 815–823
Sharma G, Dhall A (2021) A survey on automatic multimodal emotion recognition in the wild. Springer, Cham, pp 35–64
Soars B (2003) What every retailer should know about the way into the shopper's head. Int J Retail Distrib Manag 31(12):628–637

Sorensen H (2016) Inside the mind of the shopper: the science of retailing, 2nd edn. Pearson, London
Tan M, Le QV (2019) Efficientnet: rethinking model scaling for convolutional neural networks, pp 10691–10700
Tan PJ, Corsi A, Cohen J, Sharp A, Lockshin L, Caruso W, Bogomolova S (2018) Assessing the sales effectiveness of differently located endcaps in a supermarket. J Retail Consum Serv 43:200–208
Tang W, Wu Y (2019) Does learning specific features for related parts help human pose estimation? In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 1107–1116
The ultimate list of marketing spend statistics for 2019 (infographic). https://www.brafton.com.au/blog/content-marketing/the-ultimate-list-of-marketing-spend-statistics-for-2019-infographic/. Accessed 15 Oct, 2020
Toshev A, Szegedy C (2014) Deeppose: human pose estimation via deep neural networks. In: 2014 IEEE conference on computer vision and pattern recognition, pp 1653–1660
Total retail sales worldwide from 2018 to 2022. https://www.statista.com/statistics/443522/global-retail-sales/. Accessed 15 Oct, 2020
Trinh H, Fan Q, Jiyan P, Gabbur P, Miyazawa S, Pankanti S (2011) Detecting human activities in retail surveillance using hierarchical finite state machine. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1337–1340. https://doi.org/10.1109/ICASSP.2011.5946659
Tzirakis P, Trigeorgis G, Nicolaou MA, Schuller BW, Zafeiriou S (2017) End-to-end multimodal emotion recognition using deep neural networks. IEEE J Sel Top Signal Process 11(8):1301–1309
Underhill P Why we buy : the science of shopping, updated and rev. edn. Simon & Schuster, New York
Valenzuela A, Raghubir P, Mitakakis C (2013) Shelf space schemas: Myth or reality? J Bus Res 66(7):881–888
Vrechopoulos AP, Oâ Keefe RM, Doukidis GI, Siomkos GJ (2004) Virtual store layout: an experimental comparison in the context of grocery retail. J Retail 80(1):13–22
Walters RG, Jamil M (2003) Exploring the relationships between shopping trip type, purchases of products on promotion, and shopping basket profit. J Bus Res 56(1):17–29
Wang M (2020) Consumer behavior analysis in the offline retail stores based on convolutional neural network. Suzhou, China
Wang M, Deng W (2021) Deep face recognition: a survey. Neurocomputing 429:215–244
Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W, Xiao B (2020) Deep high-resolution representation learning for visual recognition. In: IEEE transactions on pattern analysis and machine intelligence, pp 1–1
Wojke N, Bewley A, Paulus D (2017) Simple online and realtime tracking with a deep association metric. In: 2017 IEEE international conference on image processing (ICIP), pp 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962
Worse to come: February retail trade figures a preview of Coronavirus hit. https://www.miragenews.com/worse-to-come-february-retail-trade-figures-a-preview-of-coronavirus-hit/. Accessed 15 Oct, 2020
Wu J, Cao J, Song L, Wang Y, Yang M, Yuan J (2021) Track to detect and segment: an online multi-object tracker. In: IEEE conference on computer vision and pattern recognition (CVPR)
Yang MH, Chen WC (1999) A study on shelf space allocation and management. Int J Prod Econ 60–61:309–317
Yan S, Xiong Y, Lin D (2018) Spatial temporal graph convolutional networks for skeleton-based action recognition, pp 7444–7452
Yong Du, Wang W, Wang L (2015) Hierarchical recurrent neural network for skeleton based action recognition. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 1110–1118
Zhang HB, Zhang YX, Zhong B, Lei Q, Yang L, Du JX, Chen DS (2019) A comprehensive survey of vision-based human action recognition methods. Sensors 19:1005
Zhang Y, Sun P, Jiang Y, Yu D, Yuan Z, Luo P, Liu W, Wang X (2021) Bytetrack: multi-object tracking by associating every detection box. arXiv:2110.06864
Zheng X, Guo Y, Huang H, Li Y, He R (2020) A survey of deep facial attribute analysis. Int J Comput Vision 128(8):2002–2034
Zou Z, Shi Z, Guo Y, Ye J (2019) Object detection in 20 years: a survey. CoRR arxiv: abs/1905.05055