

## Seminario de Tesis 2 – Descripción del avance semanal

**Código:** 20184017

**Nombre:** Paul Cristian Percca Julca

**Tema de Tesis:** Identificación Automática Del Comportamiento De Los Clientes De Una Tienda Retail Mediante Secuencias De Video Utilizando Aprendizaje Profundo

**Asesor:** Ivan Sipiran

N°	Fecha de envío del avance	Descripción del Avance
1 – Hasta el 08/09/2019	10/09/2019	Conversión de Tesis a Artículo, Antecedentes Teóricos (Redes Neuronales, Deep Learning, y YOLO) y Trabajos Relacionados

# Identificación Automática del Comportamiento de Clientes de una Tienda Retail mediante Secuencias de Video utilizando Aprendizaje Profundo

Paul Percca, Ivan Sipiran  
Pontificia Universidad Católica del Perú  
Lima, Perú  
cristian.percca@pucp.edu.pe, isipiran@pucp.edu.pe

**Resumen**—Las tiendas físicas usan múltiples técnicas para analizar el comportamiento del cliente, la mayoría de estas, se centra en la captura de datos en caja, al finalizar el proceso de compra [1]. Por lo tanto, se pierde información valiosa sobre el comportamiento del cliente, durante su estadía en la tienda. Conocer la satisfacción de los clientes, es extremadamente valioso para las tiendas [6]. Conocer las rutas de movimiento que tomaron los clientes dentro de una tienda y conocer el grado de interés de los clientes frente a algún producto, es importante para los gerentes de tienda, ya que, esos conocimientos puede ayudar a crear mejores campañas de marketing y publicidad dirigida, así como optimizar el almacenamiento y la operación de la tienda [15]. Este artículo se enfoca en análisis de video en una tienda retail con el fin de identificar las zonas de la tienda con mayor y menor tráfico así como identificar las zonas con mayor demanda de consumo utilizando técnicas del estado del arte en deep learning como son object detection (YOLOv3) y pose estimation (Deep Pose).

**Index Terms**—Object Detection, YOLOv3, Pose Estimation, Deep Pose

## I. INTRODUCCIÓN

Las tiendas retail, realizan una amplia variedad de actividades, muchas de las cuales pueden ser asistidas por análisis de video, como la planificación de diseños de tiendas basadas en estadísticas de ruta de clientes, el recuento de clientes histórico e instantáneo, entradas y salidas de tiendas, colas de pago, entre otras. Cada transacción comercial y paso durante el proceso de compra, genera una gran cantidad de datos [11], que pueden generar indicadores para la toma de decisiones.

Entre esas actividades se encuentra la prevención de pérdidas y la medición del nivel de satisfacción de sus clientes. La prevención de pérdidas se clasifica en cuatro grupos primarios: pérdidas internas (trabajadores de la tienda), externas (supuestos clientes, mediante cambio de etiquetas, fraude en la devolución de productos, entre otros), proveedores y administrativas [2].

Por otro lado, muchos clientes se sienten frustrados debido a que no encuentran su producto deseado, ya sea por haber sido colocados en góndolas inadecuadas o en zonas de la tienda que no son muy transitadas, esto generan una mala experiencia al cliente, lo cual se traduce en pérdidas de venta. Si bien se realizan estrategias comerciales y operativas para el posicionamiento de los productos, y estudios de

mercado para ello, esta información no se obtiene en tiempo real y muchas veces quedan obsoletas al ser aplicadas.

Hoy en día existen varias soluciones para análisis de datos de e-commerces, sin embargo, no se tiene mucha información de lo que sucede en la tienda física, por lo tanto, esto puede resultar un inconveniente para los administradores de la tienda; mediante el uso de análisis de video, aprendizaje profundo y procesamiento de imágenes, se tiene una recopilación más activa de la información [6], información como, cuántos clientes visitan la tienda, cuántos de las personas que entraron a la tienda compraron, cuánto tiempo permaneció una persona en la tienda, cuál fue la trayectoria de un cliente en la tienda, o cuáles son los productos donde la gente se detiene más o menos tiempo a observar, identificar el interés mediante los gestos que muestra un cliente al estar parado frente a un producto específico, es una información muy importante para la toma de decisiones, ya que se puede tener una visión general y tomar esta información para realizar campañas de marketing o reorganizar la ubicación de los productos de ser necesario.

El enfoque de este trabajo se centra en identificar las zonas de la tienda con mayor y menor tráfico y así como identificar las zonas con mayor demanda de consumo utilizando técnicas del estado del arte en deep learning como son object detection (YOLOv3) y pose estimation (Deep Pose) esta información va a ser mostrada en heatmaps de manera que pueda ser de mayor provecho para los gerentes de tienda y así tomar decisiones que ayuden a la recolocación de productos con el fin de aumentar las ventas, así mismo este análisis se podrá realizar a demanda pudiendo conocer el comportamiento de los clientes a cuando se requiera.

## II. ANTECEDENTES TEÓRICOS

### II-A. Redes Neuronales Artificiales (ANN)

Las redes neuronales artificiales son un paradigma de programación inspirado en la biología que permite a un algoritmo aprender a partir de un conjunto de datos observacionales [10]. La investigación en este campo ha tenido grandes avances, desde que McCulloch y Pitts propusieron su modelo de neurona que procesaba cada entrada multiplicada por un peso específico, para luego sumarlas y obtener una salida de 0 y 1 [9] como se aprecia en la Figura 1 tomado de [13] hasta la actualidad donde se cuenta con Deep Learning.



fotograma y el seguimiento tomaba entre 2 y 4 ms cuando había un primer plano que se debía seguir.

En la detección de personas también se han realizado investigaciones como: En [4] donde se realiza una investigación en la detección y seguimiento de humanos para videovigilancia con los datos de la Universidad Estatal de Ohio y se incorpora histogramas de gradientes orientados (HOG), la teoría de la prominencia visual y el modelo de predicción de prominencia Deep Multi-Level Network, se usó las funciones HGO para obtener una máquina de vectores de soporte (SVM) para detectar humanos en cualquier frame, para el seguimiento de personas se usó del algoritmo k - means que permitió encontrar los patrones de movimiento en los frames, lo cual es relevante para la reidentificación de la persona. además de ello se usó el algoritmo k - means para comparar y agrupar las características HOG de cuadros consecutivos para identificar el conjunto de puntos en la imagen que se asemejan a una persona en particular que se mueve en el video.

#### IV. DISEÑO DEL EXPERIMENTO

A continuación se explicará de forma detallada el diseño de la solución presentada en este trabajo de investigación:

Se han usado n secuencias de video de 720 px por 1280 px con una duración total de m horas para el presente trabajo de investigación tomadas de .....

Debido a que se está trabajando con videos, se ha realizado un preprocesamiento, que consiste básicamente en:

- Remover el audio de los videos.
- Dividir el video en frames y etiquetarlos.

Debido a que la detección de personas es un problema conocido se optó por un modelo de YOLO v3 (Darknet) preentrenado con el dataset PASCAL que inicialmente detectaba 40 clases entre las cuales se encontraba la clase Persona, se realizó Transfer Learning con el dataset Y mejorando el accuracy al detectar solo una clase.

#### V. EXPERIMENTACIÓN Y RESULTADOS

#### VI. DISCUSIÓN

#### VII. CONCLUSIONES Y TRABAJOS FUTUROS

Como trabajos futuros, se plantea aumentar más indicadores con información relevante para la toma de decisiones, así como la implementación de otros modelos de red neuronal, cuyo foco sea la seguridad y la detección de fraudes en las tiendas retail.

#### AGRADECIMIENTOS

#### REFERENCIAS

- [1] Anthony C Boucouvalas, Constantine J Aivalis, and Kleanthis Gatzio-  
lis. Integrating retail and e-commerce using web analytics and  
intelligent sensors. In *International Conference on E-Business and  
Telecommunications*, pages 3–35. Springer, 2015.
- [2] Ernie Deyle. The global retail theft barometer, 2015.
- [3] Tim Ellis. Performance metrics and methods for tracking in sur-  
veillance. In *Proceedings of the 3rd IEEE International Workshop  
on Performance Evaluation of Tracking and Surveillance (PETS'02)*,  
pages 26–31, 2002.
- [4] Vandit Gajjar, Ayesha Gurnani, and Yash Khandhediya. Human  
detection and tracking for video surveillance: A cognitive science  
approach. In *Proceedings of the IEEE International Conference on  
Computer Vision*, pages 2805–2809, 2017.
- [5] Ismail Haritaoglu and Myron Flickner. Detection and tracking of  
shopping groups in stores. In *Proceedings of the 2001 IEEE Computer  
Society Conference on Computer Vision and Pattern Recognition.  
CVPR 2001*, volume 1, pages I–I. IEEE, 2001.
- [6] Nabil Tahmidul Karim, Sanjana Jain, Jednipat Moonrinta, Matthew N  
Dailey, and Mongkol Ekpanyapong. Customer and target individual  
face analysis for retail analytics. In *2018 International Workshop on  
Advanced Image Technology (IWAIT)*, pages 1–4. IEEE, 2018.
- [7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning.  
*nature*, 521(7553):436, 2015.
- [8] Alex Leykin and Mihran Tuceryan. Detecting shopper groups in video  
sequences. In *2007 IEEE Conference on Advanced Video and Signal  
Based Surveillance*, pages 417–422. IEEE, 2007.
- [9] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas  
immanent in nervous activity. *The bulletin of mathematical biophysics*,  
5(4):115–133, 1943.
- [10] Michael Nielsen. Neural networks and deep learning.[online] neural-  
networksanddeeplearning. com, 2018.
- [11] Foster Provost and Tom Fawcett. *Data Science for Business: What you  
need to know about data mining and data-analytic thinking*. "O'Reilly  
Media, Inc.", 2013.
- [12] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improve-  
ment. *arXiv preprint arXiv:1804.02767*, 2018.
- [13] Raúl Rojas. *Neural networks: a systematic introduction*. Springer  
Science & Business Media, 2013.
- [14] Andrew W Senior, L Brown, Arun Hampapur, C-F Shu, Yun Zhai,  
Rogério Schmidt Feris, Y-L Tian, Sergio Borger, and C Carlson. Video  
analytics for retail. In *2007 IEEE Conference on Advanced Video and  
Signal Based Surveillance*, pages 423–428. IEEE, 2007.
- [15] Avi Yaeli, Peter Bak, Guy Feigenblat, Sima Nadler, Haggai Roitman,  
Gilad Saadoun, Harold J Ship, Doron Cohen, Omri Fuchs, Shila  
Ofek-Koifman, et al. Understanding customer behavior using indoor  
location analysis and visualization. *IBM Journal of Research and  
Development*, 58(5/6):3–1, 2014.