SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I3P116

Volume 13 Issue 3, 210-229, March 2026

Original Article

Product Visibility in Advertising: A Deep Learning-based Analysis of Food and Cosmetic Advertisements

Madhura Joshi[1], Poojaa Krishnan[2], Utkarsha Savkare [3], Sheel Dongre[4], Shilpa Deshpande[5], Rashmi Apte[6], Mangesh Bedekar[7]

[1,2,3,4,5]*Computer Engineering Department, Cummins College of Engineering for Women, Pune, Maharashtra, India.*
[6]*Koushiki Innovision, Pune, India.*
[7]*School of Computer Science and Engineering, Dr. Vishwanath Karad, MIT World Peace University, Pune, India.*

[5]*Corresponding Author: shilpa.deshpande@cumminscollege.in*

Received: 12 December 2025 Revised: 14 January 2026 Accepted: 22 February 2026 Published: 31 March 2026

Abstract *- Advertising today relies a lot on visuals. When a product is shown clearly in the advertisement, it increases attention and helps people remember and purchase the brand. Product detection in advertisements is hence the crucial need of businesses in today's world. Most existing systems mainly look for logos, advertisement sections, or general categories. They do not explore how clearly and how often the product appears throughout the whole video. Manual review of an advertisement video takes a lot of time, and it depends on personal judgment. Older computer vision techniques fail to detect products in real-world conditions like blur, objects blocking the view, changing brightness, and fast scene changes. With the intent of addressing the earlier-mentioned issues, this paper offers a system called a Product Visibility Analysis for Advertisements (ProVis-Ad) to automatically detect the product and measure its visibility in the real advertisements. The proposed work utilizes Deep Learning models, namely, You Only Look Once version 8 (YOLOv8), version 5 (YOLOv5), Faster Region-based Convolutional Neural Network (R-CNN), and Single Shot Multibox Detector (SSD300) to detect the product in an advertisement. The research includes eight datasets made from 155 real food and cosmetics advertisements, containing 10,810 labelled frames made using Roboflow and the Computer Vision Annotating Tool (CVAT). The ProVis-Ad system calculates product visibility for each video frame to identify the total duration of the presence of the product on screen in the advertisement. The results show a clear difference in the performance across models and advertisement domains. Food advertisements show more consistent product visibility than cosmetics advertisements. The system also introduces a new brand-level price measure to check if the products with a higher price get more on-screen attention. Experimental results demonstrate that the performance of YOLOv8 is superior to that of other models with respect to accuracy and F1-score for the detection of products in advertisements.*

Keywords *- Advertisement analysis, Computer vision, Deep learning, Frame-level detection, Product visibility.*

1. Introduction

Today, advertising is dependent on visuals to a large extent, where how the product appears on screen strongly affects how consumers pay attention to it, remember the brand, and judge its value. As digital advertising [1] has grown rapidly, understanding how often and how clearly a product is shown in an advertisement has become very important for both marketers and researchers. Visibility of a product can influence purchasing decisions, make the brand identity stronger, and add to the overall convincing impact of an advertisement, particularly in domains that are heavily dependent on visual indications, such as food and cosmetics. Studies show that food advertising influences consumer behaviour and helps maintain brand loyalty [2], while cosmetic advertisements are strongly dependent on visuals, how the products are handled, and aesthetic elements to catch the attention of the audience [3]. Both the food and cosmetic sectors rank among the most advertised domains and are represented by how aesthetic and visually appealing they are. Many times, food advertisements are dependent on detailed product shots, textures, and packaging, whereas cosmetic advertisements put more focus on appearance, application, and freshness. Since food and cosmetics use different visual strategies while focusing on product presentation, they provide a strong contrast for examining product visibility patterns in advertisements. Even after having so much importance, analysing advertisements at scale is still difficult. Many organizations still get the advertisements manually reviewed or depend on heuristic-based techniques to measure the product presence. Traditional video-analysis methods, such as template matching and logo detection, fail frequently when they deal with real-world difficulties that include motion blur, occlusion, shiny surfaces, varying lighting conditions, and fast-changing scenes [4, 5].

Advertisements are highly expressive and multimodal; they use visuals, people, and context to send convincing messages, which makes automated analysis more difficult [6, 7]. These issues show a major research gap that, even though advertisements are becoming more visually rich and varied in style, the current analysis methods do not effectively show how and how often the product itself appears on screen throughout an advertisement. Most of the previous work focuses on logo detection, brand localization, or identifying advertisement segments [4], but does not provide systematic ways to measure continuous product visibility across full-length advertisement videos. Before the development of modern object detection methods, object classification could be seen to mainly depend on hand-crafted feature extraction methods [8]. Recent progress in computer vision now makes it possible to analyse real-world advertisement videos at scale, which in turn helps in accurate and reliable measurement of product presence [9]. Models like You Only Look Once (YOLO), Single Shot Multibox Detector (SSD), and Faster Region-based Convolutional Neural Network (R-CNN) can find objects correctly even if the light is poor, part of the object is covered, or the background is complicated [5, 10, 11]. Deep learning has remarkably advanced object detection. It has become central to modern advertisement analysis because it can automatically study large amounts of visual content that would otherwise take a lot of manual work [12]. One-stage detectors like YOLO and SSD are very fast and can detect objects in real time [10]. Whereas, two-stage detectors such as Faster R-CNN are slower but give more accurate results because they refine the regions they examine [5, 13]. These technologies have also been applied in advertising tasks, including logo detection and brand recognition [4]. However, even with improvements in automated advertisement detection, most earlier studies look at finding advertisement segments or logos, not at tracking how the actual product appears throughout the entire advertisement video. In order to solve these issues, this work introduces Product Visibility Analysis for Advertisements (ProVis-Ad), a system based on deep learning for automated calculation of how much time a product is visible in real-world video advertisements. The system aims to overcome the drawbacks of traditional advertisement-analysis techniques by making automatic frame extraction, product detection, and measurement of the visibility of the product in advertisements using well-established object detection models. ProVis-Ad not only identifies whether a product appears or not, but also measures how consistently it is visible across an advertisement, which is directly related to how much the consumers give attention to the advertisement and the marketing effectiveness of the advertisement. Product visibility measurement in the video advertisement is useful beyond studying consumer behaviour. It is easier for compliance teams to verify if the sponsored advertisements meet the required visibility guidelines. Media analytics teams can use this data to track the product placement trends and also get a better understanding of the marketing strategies. By keeping the focus on real-world advertising settings, the proposed approach contributes to the current research in multimedia analysis, computer vision, and marketing analytics, and shows how important it is to have scalable, data-driven methods for advertisement evaluation. Based on the findings of the proposed work, this study promotes the hypothesis that *Food advertisements exhibit stronger product presence and higher screen visibility than Cosmetics advertisements.*

1.1. Need for Solution

The need for the ProVis-Ad system stems from key research and industry gaps, which are described below.

1. Manual methods of advertisement-analysis are slow, individualized, and not scalable, which makes it difficult to examine large datasets of advertisements [8].
2. Existing automated methods mainly focus on logo detection or segmentation of advertisements, but not on tracing the actual product that is shown throughout an advertisement, even though product-centric visual analysis enables richer and more meaningful interpretation of advertising content [14].
3. Traditional computer vision methods fail when it comes to real-world conditions, such as fast scene changes, changing lighting conditions, or partially visible products [4, 5].
4. Moreover, earlier work shows that high-level indicators, such as how many views the advertisement gets or the appearance of a celebrity in the advertisement, do not show the advertisement's effectiveness in a reliable manner. This urges the need for content-based measurements like product visibility [15].
5. Modern object detectors exist, but how they perform in real-world advertisements is still not explored, especially by considering different categories of products.
6. These gaps show the need for a systematic, scalable, and model-independent system that can measure how much time the product is visible throughout the entire advertisement video.

1.2. Contributions

The proposed work directly marks the research gaps by giving a complete and practical solution. In this work, the authors present a novel system called a Product Visibility Analysis for Advertisements (ProVis-Ad) to automatically detect the product and measure its visibility in real advertisements. Thus, the specific contributions of this research work are given below.

- Eight domain-based datasets are created using 155 real advertisement videos from the food and cosmetics domains. To make the datasets larger and more diverse, annotated images from Roboflow Universe are also added to them, giving the complete dataset of 10,810 images across all the domains.
- Detection of the presence of products in advertisements

using four distinct deep learning models for object detection, which are trained and tested for each subdomain, allowing a strong and fair comparison under real advertisement conditions.

- Product visibility in terms of the percentage of time duration for which the product is visible on screen in the entire advertisement is computed.
- A new brand-level average price measure is introduced to check if product cost is associated with how particularly it appears, which is something earlier studies have not explored.
- Comparative analysis of product visibility for the food and cosmetics domains, as well as across different subdomains and brands within each domain.
- Comparative analysis of various deep learning models based on accuracy, precision, recall, and F1-score for detecting the presence of a product in advertisements.

2. Related Work

2.1. Literature Review

A review of earlier studies shows that advertisement analysis has been surveyed in areas like classifying advertisements, recognizing logos and brands, and detecting objects in general. Early work by Li et al. [16] found that advertisement images include unique visual and text patterns that make them different from regular images. Studies like the one by Krishnan and Sitaraman [1] show that where visuals are placed in an advertisement, and how clearly they are seen, strongly affect how people react to the advertisement. According to Li and Lo [17], clear and important brand visuals help customers remember the brand more easily, and make visible product imagery very important. Deep learning has helped this field to progress notably. Hoi et al. [4] have introduced LOGO-Net, a large-scale system for deep logo recognition, while Ratner et al. [18] have explored advertising analytics based on images using classical image features. Major object detection models such as Faster R-CNN [5], SSD [10], and YOLO family [11, 19, 20] have become strong foundations for advertising-related visibility tasks, including the detection of outdoor advertising structures like billboards and estimation of their visual prominence [21]. Multiple types of approaches, such as Ad-Net [22], ActiveAd [23], and AdVision [24], show how deep learning can process advertisements on a large scale for advertisement analysis. Challenges in video detection, such as motion blur and rapid scene changes, are discussed by Zhu et al. [25]. In addition to this, domain-specific works such as Rodrigues et al. [2] and Li and Zhang [12] show that machine learning plays a very important role in understanding how advertisements influence customers. These works help in telling and explaining the research direction. Work on viewer attention and brand fame [1, 17] shows why measuring product visibility is important, which is a main goal of the proposed system. Logo-detection studies [4, 18] demonstrate that object-level detection is possible, but since they focus only on logos rather than full product packaging, this work aims to build a more detailed detection system. Modern object detectors [5, 10, 11, 19, 20] show a good performance for broad datasets, which adds to the decision to evaluate different types of architectures, such as YOLOv5, YOLOv8, SSD300, and Faster R-CNN on real advertisement videos. Previous works on advertisement classification and multimodal detection [22, 23] show that deep learning can handle advertisement analysis on a large scale. However, these works do not focus on frame-by-frame detection of product presence, which is the main goal. Even after great progress, object detection models like YOLOv8 [26] still face difficulties in handling occlusion, dissimilarities in scale, and backgrounds with complexity, which negatively affect the performance in real-world conditions. The quality and variety of datasets, inconsistent annotation, and showing limited variations in the real world, are a major limitation to achieving reliable and generalized performance in detection [26].

The challenges that come up in video detection research [25] show how important it is to create datasets that are carefully annotated and are domain-based, such as the one proposed in this system. Domain-based studies [2, 12] show that it is important to study what appears in the advertisements, not just detect the advertisements themselves. This supports the goal of ProVis-Ad to calculate the product visibility across food and cosmetic advertisements. Table 1 shows the comparison of earlier studies with the proposed ProVis-Ad system by considering how advertisement content is used, supported functionality, methodologies used, datasets, and limitations noticed. The table highlights the gaps in existing approaches and shows how the proposed system uniquely addresses these limitations using a product-focused and frame-by-frame analysis of the visibility of the product.

2.2. Research Gap

After looking at the past research such as advertisement classification [1, 16, 22, 23], logo detection [4, 18], and general object detection [5, 10, 11, 19, 20], it is observed that each research gives useful ideas but are not fully successful in solving the detected limitations as noted from the literature review. Advertisement-classification methods can tell what kind of advertisement it is, but they do not detect individual objects or products inside the advertisement.

Logo detection methods work well when the logos are visible, but they fail in situations where the product is shown without a clear logo, or when the logo is blocked by any other object, or when the advertisement uses creative visuals, which are the most common situations that happen in real advertisements. Studies based on detection in videos identify the challenges of moving imagery [25], but they do not focus on real customer products or advertisement environments that include multiple domains [28]. The research in digital advertising until today has majorly focused on comparing which ad formats, like video vs image, are more

effective based on various metrics of engagement, such as reach, impressions, and click-through rates, instead of analyzing the actual visibility of products within the advertisements [29].

Studies mainly focus on analyzing the effects of visual appeals and recommendations by celebrities on the purchase behaviour of consumers throughexperimental or campaign-based approaches [30].

Table 1. Qualitative comparison of prior works in advertisement analysis and object detection

Reference	Use of Advertisement Content	Supported functionality	Methodology	Dataset Used	Remarks
Rodrigues et al. 2023 [2]	High-level categorization of all the advertisements	Food vs non-food classification	CNN video features -Traditional ML techniques	YouTube food advertisements	- Does not perform object detection - Does not compute the visibility of the product in the advertisement - Limited variety in advertisements - Does not analyse product regions
Hussain et al. 2017 [7]	Uses visual indicators to match advertisements with catalog images	Advertisement-product linking	Uses deep features with an approach based on data retrieval	Private matched advertisement dataset	- No detection of product presence in advertisements - Does not compute the visibility of the product in the advertisement and depends on metadata
Li and Zhang 2024 [12]	Detects products and faces in images from social media	Engagement pattern analysis	Pretrained CNN feature extraction	Instagram advertisement posts	- Requires manual filtering of data - No frame-wise calculation of visibility performed.
Zhang et al. 2016 [27]	Focuses on object and scene context	Placement of advertisements in videos (not detection)	Optimization with Object-level matching	Small private dataset	- Does not analyse product presence in advertisements - Does not compute the visibility of the product in advertisements - Traditional approaches used in place of deep learning - Scalable to a limited extent

Proposed system: ProVis-Ad	Fine-grained product detection inside real advertisement videos	Detection of the presence of products in advertisements On-screen product visibility duration computation	Best performing deep-learning model selected for product detection out of - YOLOv8, YOLOv5, SSD300, and Faster R-CNN	155 real-world advertisement videos across 8 subdomains	- Continuous visibility measurement of the product - Analysis across various domains - Benchmarking of multiple models - Insights aligned with marketing

Similarly, marketing studies [2, 12] explore how advertisements influence customers, but they do not use computer vision to measure how often the products appear in the advertisement videos or to analyse how the visibility of products changes across multiple domains. Collectively, the existing work shows a noticeable gap in the research: Even though different approaches have focused on various aspects of advertisement understanding, none of them provide a centred, product-level, frame-by-frame way of detection across multiple domains. To overcome the limitations observed in earlier studies, the proposed ProVis-Ad system uses an approach that completely focuses on the product and its frames for advertisement analysis. Previous methods focus on broad classification of advertisements, visibility of logo, or detection of objects in single-frames. On the contrary, ProVis-Ad continuously tracks the presence of the actual advertised product throughout the entire duration of the real-world advertisement video. The system uses object detection models based on deep learning, which are tested in varied and challenging advertising conditions, allowing robust detection even if the products appear partially, briefly, or without clear logos. A visibility metric specific to domains in advertisements is used in ProVis-Ad, which directly supports the goals of advertising by measuring how long a product remains on screen. Furthermore, the system is designed in such a way that it can support multiple product domains and brands, which makes it scalable and consistent in real advertisement environments. In this way, ProVis-Ad is a cohesive system that overcomes the gap between computer vision methods and practical visibility measurement of products in advertisements.

2.3. Novelty of the Proposed ProVis-Ad System

As shown in Table 1, the proposed ProVis-Ad system is qualitatively compared with other approaches dealing with advertisement analysis and object detection. From Table 1, it is evident that the existing research on advertisement analysis fails to give a product-level, frame-by-frame analysis of visibility across multiple domains. As compared to earlier works, the ProVis-Ad system offers various distinguishing features as follows. The system detects full product packaging rather than only logos or any advertising parts, which helps better understand how visible the product is throughout the advertisement. The system employs the best-performing deep learning model for product detection that allows a fair and consistent measurement across various advertisements. By preparing domain-wise datasets and providing brand-level price-visibility analysis, the proposed system brings a new marketing insight.

3. Methodology

This section explains all the steps that are used in the work. The process begins by collecting and annotating the data, then moving on to training the model, testing it on different advertisements, and finally doing the comparison analysis. The proposed workflow makes sure that the comparisons remain fair across all four models. Also, consistency is maintained across food and cosmetics advertisements. In addition to this, the system measures product visibility reliably using both automated and manual checks.

3.1. Dataset Creation

The datasets are created for this research to analyse how much time the product appears in the advertisement in real-world video advertisements across two commercially significant domains: Food and Cosmetics. The food domain is selected because food advertisements are designed in such a way that they instantly grab attention with the use of rich colours, tempting visuals, and reveal products quickly, which can trigger hunger and buying decisions. These advertisements commonly focus on taste, texture, freshness, and immediate craving, and hence make the product visibility extremely dynamic and fast. On the other hand, the cosmetics domain is selected because cosmetic advertisements focus on beauty, self-image, and transformation. They focus more on close-up shots, elegant packaging, brand identity, and gradual visual change in appearance. Both domains are dependent more on storytelling using visual methods, but in very different ways. Hence, they provide an appropriate comparison to study how product presence changes in different advertising styles. Further, each domain is divided into four subdomains, which gives eight subdomains in total:

1) Food: Biscuits, Chocolates, Noodles, Soft Drinks
2) Cosmetics: Facewash, Foundation, Perfume, Soap

For every subdomain, five popular brands are selected based on their significance in the market and how often they are advertised. From each brand, four different advertisement videos are collected, which leads to a total of around 20 advertisements per subdomain. After combining, the datasets result in an overall total of 155 video advertisements. Structured and well-balanced datasets are generated. Figure 1 depicts the overall layout of these datasets. These datasets establish a constant representation of different domains, subdomains, and product types. That is why the datasets make it possible to compare product visibility fairly between food and cosmetics advertisements. The datasets are created using standard methods in video object detection. This makes sure that there is a required variety and balanced classes so that the models can be trained and tested properly [1, 16].

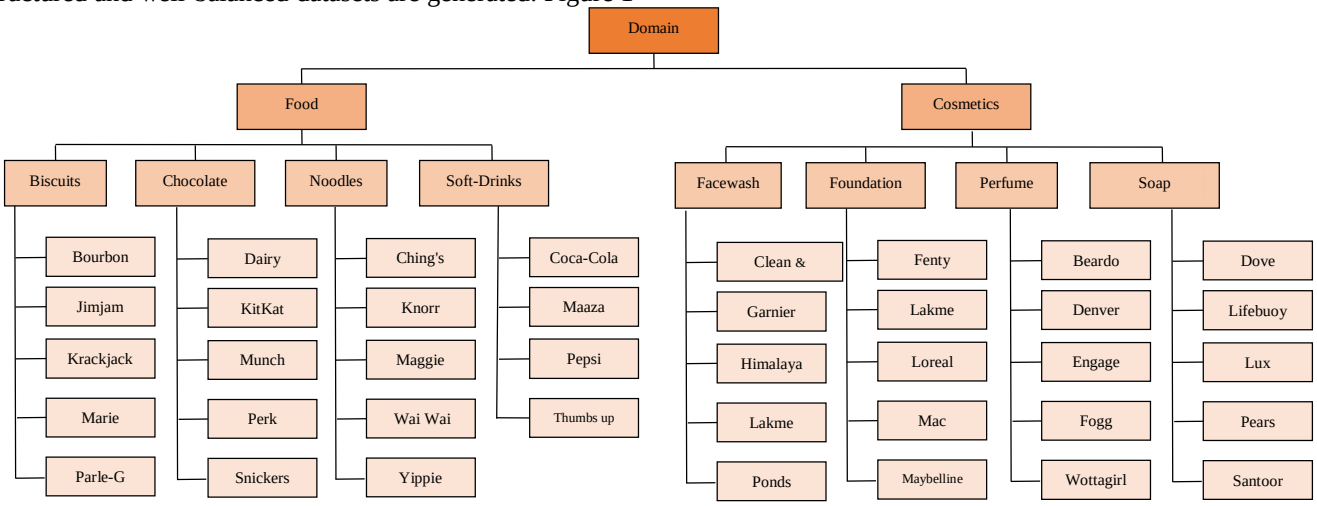


Fig. 1 Hierarchical structure of the food and cosmetics advertisement dataset

3.1.1. Data Collection

All advertisement videos are collected from publicly available sources, mainly YouTube and official brand marketing channels. This makes sure that the datasets contain only original and real-world advertising content rather than any controlled or artificial content. This helps to improve the real-world significance of the proposed analysis. At the time of the process of collecting the data, careful attention is given to ensure visual and corresponding diversity across the datasets. The selected advertisements show variations in:

1) Different lighting effects, such as indoor, outdoor, and studio lights.
2) Different camera movements, such as static shots, panning, and fast cuts.
3) Types of shots taken, such as close-ups, medium shots, and wide angles.
4) Different types of packaging aspects, like size, colour, visibility, and branding elements.
5) Presence of actors or models (single actors, families, celebrities).
6) Showing the usage of the product, such as product handling, application, and consumption.
7) Different types of backgrounds, such as clean backgrounds or cluttered real-world scenes.

This variety is needed in order to train deep learning models that can work well under uncontrolled visual settings. Previous research shows that if models are trained on varied and complex scenes, it helps to improve their ability to perform in real-world environments [9]. Earlier work also shows that detection models that are trained on datasets that are visually similar usually fail under real-world changes. This happens mainly in video advertisements where motion blur, blocking due to objects, and fluctuations in lighting are common [1, 16]. A wide variety of visuals and advertisement styles are covered in these datasets. Because of this, the datasets help to build a robust product detection system. The models that are proposed in this work perform consistently across various domains. They can also capture real product visibility patterns in both food and cosmetic advertisements.

3.1.2. Data Annotation

After the collection of data, all advertisement videos are processed for frame-by-frame annotation to perform supervised training of the object detection models. First, each video is uploaded to the Roboflow platform, where it is automatically converted into different frames of images at a rate of 1 Frame Per Second (1 FPS). This rate is selected so that the ProVis-Ad system does not become too heavy while recording the product throughout the entire advertisement.

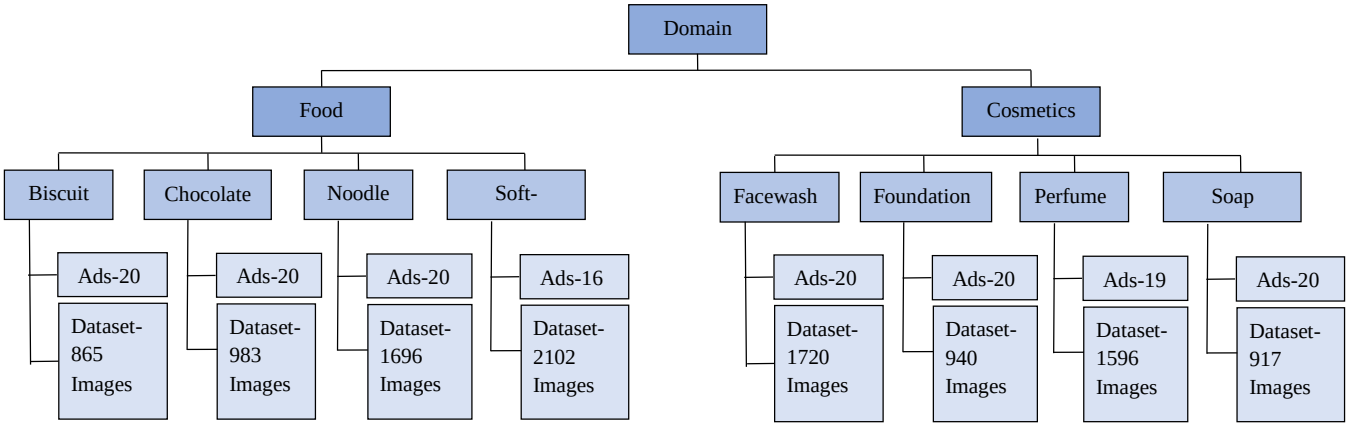
The frames that are extracted earlier are manually labelled by drawing boxes around the visible products. Annotation is done using both Roboflow and the Computer Vision Annotation Tool (CVAT), which gives flexibility and good accuracy while labelling. All products are labelled as just one class called "Product". This is because the goal is to study how often the product appears in advertisements, not to identify different brands or types. Using one class also prevents learning between domains, and it helps the model work better across different domains and subdomains. Bounding-box annotation is used because it is the most commonly used and effective method for object detection. Earlier approaches show that the use of consistent annotation using bounding boxes notably improves how well the model learns and is able to locate the objects [4, 17]. In addition to this, supervision using bounding-boxes is well-suited for frames that are derived from videos where the scale, orientation, and occlusion of objects often change over time [18]. To make the datasets distinct, open-source images of products are added from the Roboflow Universe to each subdomain. This stops the model from leaning towards only a few advertisement styles. These images are chosen carefully so that they look similar to real advertisement frames. They should match the correct brands and product types. The images that are repeated or not clear have been deleted.

After the selection of these images, the Roboflow images are combined with the manually annotated advertisement frames. This combined set of images is the final dataset for each subdomain. The extra images help the model to perform better. They also make the model more diverse and better at generalizing. The model gets better as it learns from many different real-world visual examples [5, 27].

3.1.3. Data Analysis

After the work is done with the annotations and addition of extra data, a detailed check of the datasets is performed. The quality, balance, and structure of the data in all eight subdomains are observed. The goal of this step is to make sure that the datasets are good for training deep learning models. A check is made to ensure that the datasets are not biased and that they are diverse enough. This helps the models to simplify during testing. The following aspects for each subdomain are checked:

1) Total number of images
2) Frame-wise product distribution
3) How much does the product size vary
4) How often the product is partly covered
5) Variety in lights and backgrounds
6) How evenly are different brands represented



Note: Ads = Advertisements

Fig. 2 Categorization of food and cosmetic advertisement data used for model training and evaluation

The sizes of the final dataset range from 800 to over 2100 images per subdomain, which ensures that a sufficient amount of data is available to train deep convolutional models. Additionally, the combination of both frames derived from advertisements and the Roboflow Universe dataset images improves the variety in datasets. This is important to refrain from overfitting and improve the strength of detection in real-world settings [5, 27]. Special care is taken to incorporate the datasets with complex samples, as follows:

1) Products that are partially covered (for example, soap and facewash covered by foam),
2) Small-scale objects (for example, perfume and foundation containers),
3) High reflectivity packaging, and
4) Frames that are affected by motion blur are caused by the fast movement of the camera.

The above-described challenging visual patterns are intentionally kept in the datasets because earlier work shows that the use of complex samples during training is important to improve the strength and real-world performance of detectors [10, 25]. In addition to this, different brands are distributed equally within each subdomain to make sure that

no single brand dominates the datasets. This ensures that the model is not biased towards specific packaging styles or visual indicators, and fair brand-level comparative analysis takes place during testing. The domain-wise and subdomain-wise distribution of advertisements used for model training and evaluation is illustrated in Figure 2. This structured data analysis step makes sure that the datasets are balanced, diverse, and comprise real advertising conditions. This, in turn, provides a reliable foundation for further training, evaluation, and comparative analysis.

3.1.4. Data Cleaning and Pre-Processing

Standard cleaning and pre-processing steps are then applied to the final datasets before training the detection models.

This makes the data the same, visually in line, and fair among all the models. Because the created datasets are derived from various sources, such as frames of advertisements and images from Roboflow Universe, this is an important step.

This method removes noise, fixes irregularity, and develops the data properly for deep-learning object detection. Steps of pre-processing are applied to all datasets in an equal manner as follows:

1) Auto-orientation: All images are automatically re-oriented using metadata so that they are properly aligned. This makes sure that errors in models that are caused by rotated or flipped images do not occur and helps to keep the alignment of images consistent during training.
2) Image resizing: Each image in the dataset is resized to a fixed size of 640 × 640 pixels to ensure fast processing, along with accurate detection of the product. This input size is suitable for all four detection models used in this work.
3) Class label standardization: All annotations are merged into a single class called "Product". This removes unnecessary dissimilarities in labels. Also, it shifts the focus from detailed classification to strong detection of product presence, which is the main goal of advertisement visibility analysis. Roboflow is used to do all the pre-processing steps. It also helps to keep a track of different dataset versions.

To ensure smooth training of all the models, each subdomain dataset has been exported in YOLO, COCO, and Pascal-VOC format.

These pre-processing steps are important for more stable training, reducing noise, and for the fair evaluation of the datasets from both domains. Earlier works also show that standard resizing, fixing image orientation, and keeping labels consistent help improve detection accuracy and make results easier to compare between models [5, 18].

3.2. Functional Overview

This subsection describes the core functional architecture of the proposed ProVis-Ad system, which is developed to get an automated measure of how much time the product is visible in video advertisements. The functional overview focuses specifically on the input-processing-output pipeline of the system and explains how raw video advertisements are converted into numerical data of the product visibility score. Figure 3 shows the functional workflow for the ProVis-Ad system. The proposed system goes through the following step-by-step functional modules:

1) Advertisement Video Input: The ProVis-Ad system takes raw video advertisements as input. These videos are collected from platforms that are publicly available and are either from the Food or the Cosmetics subdomains. Each input video acts as the primary data unit for calculating the visibility of the product.
2) Frames Extraction from the video advertisement: The advertisement videos are taken as input, and each one of them is converted into a set of image frames by using a fixed rate of 1 Frame Per Second (1 FPS). This helps to ensure that:

a. Frames are extracted consistently with the same time intervals for all the advertisements.
b. Computational load is decreased.
c. Product appearance is shown in a balanced manner throughout the video duration.

3) Pre-Processing and Standardization: A pre-processing step is applied to the extracted image frames to make sure that they are suitable for all detection models. This includes the following steps.

a. Auto-orientation to fix the alignment of the image.
b. Resized to 640 × 640 pixels to keep the input size of the image constant for all models.

These steps make sure that all frames are in the same format before they are given to the detection models.

4) Product Detection using Deep Learning models: The pre-processed image frames are given to four different trained models, viz. YOLOv8, YOLOv5, Faster R-CNN, and SSD300. Each model carries out frame-by-frame detection of the product and gives the raw outputs as:

a. Bounding box coordinates
b. Detection confidence score for each bounding box

A single class "Product" is used in the process of detection, as the objective of this work is to calculate the on-screen duration of product presence in the advertisement and not to identify specific brands. Ultimately, out of the above-mentioned four models, in the ProVis-Ad system, the deep learning model with the best performance is employed for the

detection of products in an advertisement. The details of the models used in the product detection are described in the subsection that follows.

5) Frame-Level Visibility Metrics Calculation: The raw detection outputs generated in Module 4 are further processed by the ProVis-Ad system to give the following key detection metrics:

a. Presence or absence of product.
b. Number of detected bounding boxes.
c. Detection confidence value per frame (average of the confidence scores of total bounding boxes in a frame).

These results at the frame level become the basic units to calculate product visibility.

6) Product Visibility Computation (ProVis-Ad Output): The final output of the ProVis-Ad system is the Product Visibility Score (%), which is the measure of the amount of time a product appears on screen. It is defined as:

$$Product\ Visibility = \frac{n}{N} \times 100 \quad (1)$$

Where, Product Visibility Signifies product visibility score, n denotes the number of frames in which the product is detected, and N indicates the total number of frames in the advertisement. This visibility percentage is the main quantitative output of ProVis-Ad, which is directly derived from the frame-by-frame detections of the products. It acts as a key indicator of how much the product is visible on screen in video advertisements.

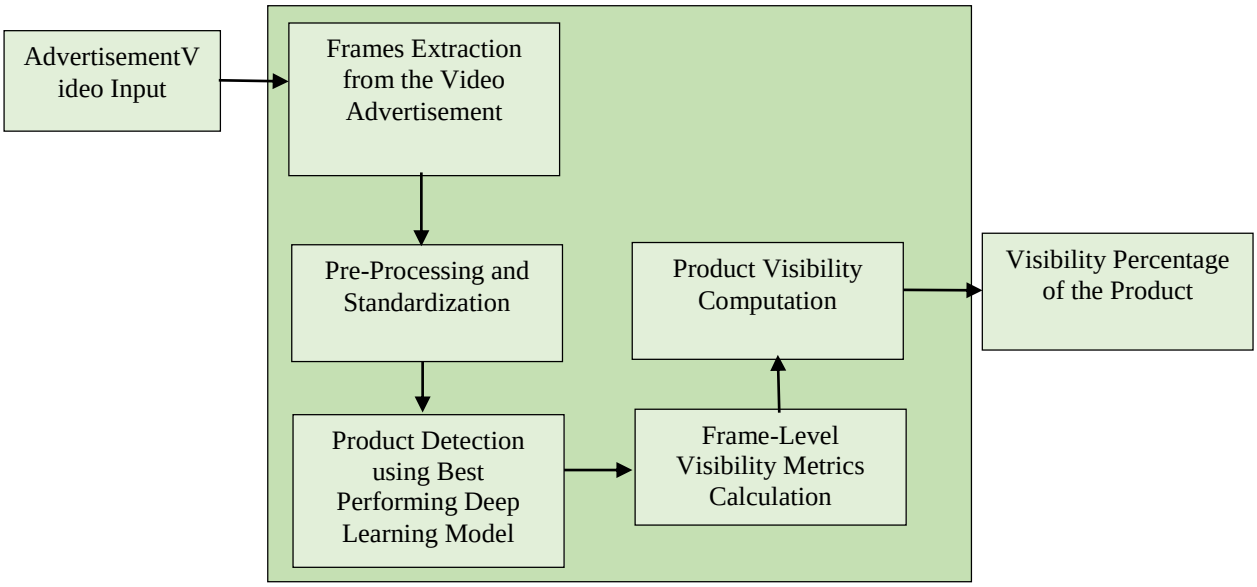


Fig. 3 Functional overview of the provis-ad system

3.3. Product Detection Algorithms

Recent progress in object detection and semantic segmentation has presented strong methods that possess the capability to classify objects and locate them accurately within visual scenes. Traditional advertisement analysis requires manual frame-by-frame review most of the time, which makes the process slow and complex, especially for long video durations [31]. In this work, four well-known object detection models are explored to effectively detect the product and evaluate its visibility in the real video advertisements. By using different types of models, both one-stage and two-stage detectors can be tested. It gives a balanced and fair comparison analysis. These object detection models are described below.

3.3.1. YOLOv8

YOLOv8 is the latest version in the YOLO family, which uses a one-stage object detection approach. It brings in important enhancements in how the network is built and how features are combined. It also enhances the way it detects objects without anchors. Because of these updates, YOLOv8 performs better in difficult and complex scenes [26]. YOLOv8 is very good at:

1) Detecting big and small objects.
2) Detecting the objects even when they are covered by something.
3) Handling the backgrounds that change continuously or are cluttered.

The YOLOv8 model is chosen as a baseline because it is accurate and faster than others. Studies also say that YOLOv8 is able to achieve higher precision, recall, and mean Average Precision (mAP) as compared to earlier YOLO versions, which makes it more reliable for real-world detection tasks, mainly in cluttered scenes and videos with movements [19, 32, 33].

3.3.2. YOLOv5

YOLOv5 is a one-stage detector that is lightweight and performs fast computations. It is widely used for real-time video object detection. It is known for:

1) Fast inference speed
2) Ease of deployment
3) Good performance on objects that are of medium siz

Even after being a bit weaker than YOLOv8 in feature learning, YOLOv5 gives a proper guideline in order to compare how new upgrades in models affect their ability to detect the visibility of a product. Its use in real-time commercial vision systems has been noted worldwide [20]. YOLO-based models work well for video frame object detection and comparing different detection models in video analysis tasks [34].

3.3.3. Faster R-CNN

Faster R-CNN is based on earlier models such as R-CNN and Fast R-CNN, and provides major improvements in speed and accuracy [13]. It is a two-stage detector in which the first step is to detect the areas where the object might be present.

After that, it finds the object in each region and makes the boxes around it more accurate. This method leads to higher precision, especially when the environment is controlled and clear. The Faster R-CNN, which is included in this work, analyzes the following.

1) How does the performance change on objects that are small and partially covered
2) How reactive it is to complex textures and shiny packaging of products
3) Whether it focuses more on precision as compared to real-time models

However, since it is slower and heavier for computation, it is less suitable for real-time use, especially for video-based analysis [2].

3.3.4. SSD300

SSD300 is another one-stage detector which is designed for fast object detection. It detects objects directly from multiple feature maps at different scales, which makes it a good model for quick and effective processing. Prior studies show that SSD architectures are able to provide high speed along with handling objects at multiple scales, which makes them suitable for real-time applications [35]. However, even though SSD300 provides high speed, it poses challenges concerning:

1) Small object detection
2) Dense scenes
3) Shiny and low-contrast objects

It is mainly included to find out how older real-time detectors compare with modern YOLO-based models under real-world advertisement conditions [36]. Reasons based on which multiple types of models are selected for exploration are as follows.

1) Real-time and high-accuracy models can be reasonably compared.
2) How well the models behave for the Food and Cosmetics domains, respectively, can be tested.
3) How the visibility changes across different models can be analysed.
4) Results can be confirmed with the use of one-stage and two-stage detectors together.

Such evaluation of multiple models is highly recommended in object detection research to reduce algorithmic bias and to make sure that the results do not depend on any single model [37, 38].

4. Performance Evaluation

4.1. Experimental Setup

All experiments in this research are carried out using GPU-based cloud environments, mainly Google Colab and Kaggle, to ensure that enough resources for computation are available for deep learning video analysis. Python (v3.10) is used to build the models for product detection along with PyTorch, which is a deep learning framework.

More supporting libraries, such as OpenCV, are used for the extraction of frames, preprocessing images, and visualization. Annotation and management of datasets are carried out using the Roboflow and CVAT tools, and results visualization is done using Matplotlib and Seaborn.

4.1.1. Hardware Configuration

The experiments are run on NVIDIA Tesla T4 GPU with 12 GB RAM and cloud-based session storage for training of models and deducing the results. This setup gives sufficient memory and processing capacity in order to handle large datasets of video frames effectively.

4.1.2. Input Configuration

All image frames are resized to a fixed resolution of 640 × 640 pixels before training the model and deduction of results to make sure that the consistency of image frames is maintained across all models. Video advertisements are processed at a sampling rate of 1 Frame Per Second (1 FPS) to keep the analysis of frames at equal time intervals and also reduce the load on computation.

4.1.3. Training Configuration

The datasets are split into 70% for training, 20% for validation, and 10% for testing to make sure that the model learning is balanced and evaluation is not biased towards any

product. AdamW is used as the optimizer for YOLO-based models and SSD300. Whereas the Stochastic Gradient Descent (SGD) optimizer is used for Faster R-CNN. All models are trained using a batch size of 16 for 250 epochs, and an early stopping patience is given as 60 to make sure the model does not overfit. A single-class ("Product") is used for all the experiments to ensure consistent learning of the models.

4.1.4. Testing Configuration

During testing, confidence thresholds are used to remove predictions that have low confidence. Each trained model is used independently to detect products in the advertisements. Frame-by-frame detections are then combined to calculate the product visibility (%), average confidence, and statistics at the brand-level and subdomain-level.

4.1.5. Generation of Ground-Truth

For accurate testing and correct calculation of visibility, a separate ground-truth generation process is carried out. The testing advertisements are again converted into sets of image frames using a custom Python script. The frames are manually annotated in the CVAT tool and exported in YOLO, COCO, and Pascal VOC formats. This export process in multiple formats makes sure that they are suitable for all desired models for the work and allows an accurate one-to-one correspondence between test frames and ground-truth labels. This experimental setup guarantees that the datasets are reproducible, no partiality is observed towards any model, and the comparative analysis of the performance for the Food and Cosmetics advertisement domains is accurate.

4.2. Evaluation Metrics

Every model gives predictions for all test frames and calculates various measures, which include accuracy, precision, recall, F1-score, and average confidence.

4.2.1. Results of Ground Truth and Detection

Evaluation in ProVis-Ad is performed at the frame level. For each frame, the prediction of the model is compared with the ground truth extracted from manually annotated image frames and is divided into the following categories:

1) True Positive (TP): The model is able to correctly detect the product that is advertised in a frame where it is present.
2) False Positive (FP): The model detects the product in a frame in which it is not actually present, which happens mostly due to objects that look similar or due to a mess in the background.
3) False Negative (FN): The model fails to detect the product in a frame where it is actually visible, which usually happens if the objects are partially covered, affected due to motion blur, small in size, or appear for a shorter period of time.
4) True Negative (TN): The model is able to correctly detect the frames in which the product is not present. These outcomes are used to compute all evaluation metrics and ensure that visibility calculations of products in the advertisements are based on accurate and reliable manual annotations.

4.2.2. Accuracy

Accuracy measures how correctly the product is detected in all the frames, including the frames with or without the product present in them. It is the sum of the number of true positive frames and true negative frames divided by the total number of frames. In ProVis-Ad, accuracy shows the reliability of a model. However, most advertisement videos have many frames without the product. Because of this, accuracy alone does not fully show product visibility performance and should be considered together with the other metrics [26].

4.2.3. Precision

Precision shows how many of the detected product frames are actually correct [13]. It is the division of the number of true positive frames by the sum of true positive and false positive frames. In advertisement analysis, high precision is important because wrong detections can falsely increase product visibility, which can give a wrong idea about how much a product is actually shown [26].

4.2.4. Recall

Recall calculates how many frames in which the product is actually visible are successfully detected by the model [13]. It is the division of the number of true positive frames by the sum of true positive and false negative frames. In ProVis-Ad, recall shows how much the system is able to capture all instances of the appearance of the product, which includes brief, partial, or partially covered appearances. Low recall results in lower measurement of product visibility than the actual value, mostly in advertisements that move quickly or are highly edited [26].

4.2.5. F1-score

The F1-score combines precision and recall into a single formula and shows a balance between correct and missed detections of the product. It is calculated as twice the product of precision and recall, divided by the sum of precision and recall. In the product visibility analysis, the F1-score shows how well a model balances correct product detection and avoiding false detections. This makes it very useful for comparing different detection models in the ProVis-Ad system.

4.2.6. Average Confidence

Average confidence is the mean confidence score, which is given by the detection model for all detected product instances across frames. It is the sum of the confidence scores that are given by the model for detected products, divided by the total number of product detections. In the ProVis-Ad system, higher average confidence shows that the visibility of

a product is clearer and consistent, whereas lower confidence values usually arise due to complex conditions such as motion blur, partial covering of objects, shine, or small product size [13, 19].

4.3. Results and Analysis

In this section, the experimental results of the selected models are presented across 155 real video advertisements. The comparative analysis of the product visibility and average pricing is done based on brands and subdomains for the food and cosmetics domains. The performances of various deep learning models used in the product detection are compared on the basis of accuracy and F1-score.

4.3.1. Brand-Level Comparative Analysis of Product Visibility and Average Pricing

At the brand level, the output of the ProVis-Ad system obtained in the form of Product Visibility Scores is compared with the corresponding average prices of the products.

Figure 4 graphically demonstrates the results of this comparison for the Food domain.

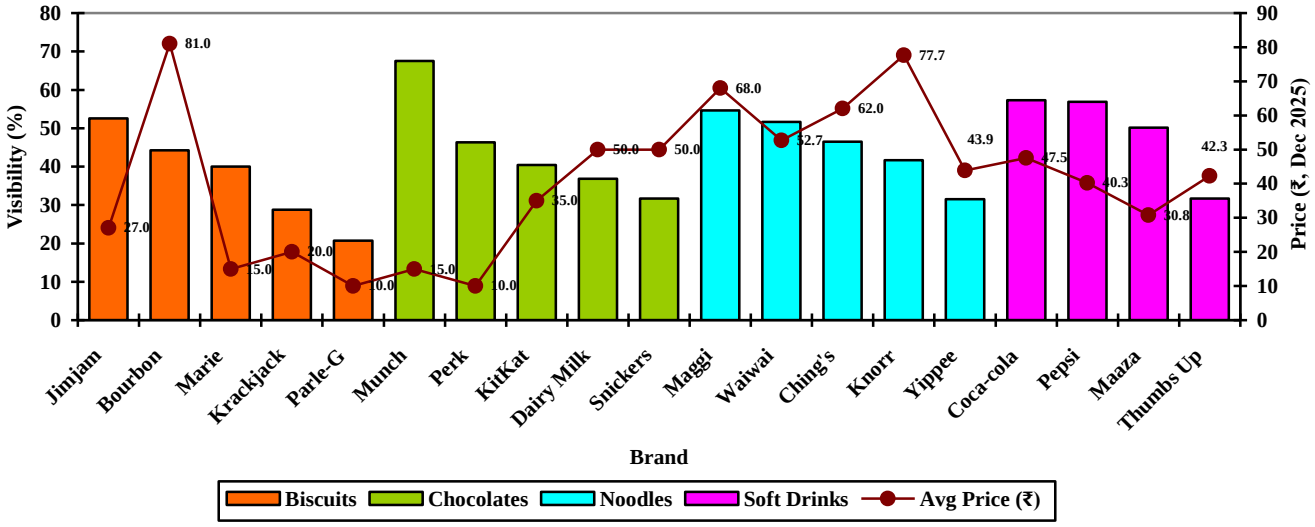


Fig. 4 Food visibility vs average pricing (₹, as of Dec. 2025)

The results from Figure 4 show that in the Food domain, the visibility percentage maintains a consistency in high visibility of about 40-60% frame-presence range.

A weak to moderate positive association is observed between average pricing and visibility for the food subdomain.

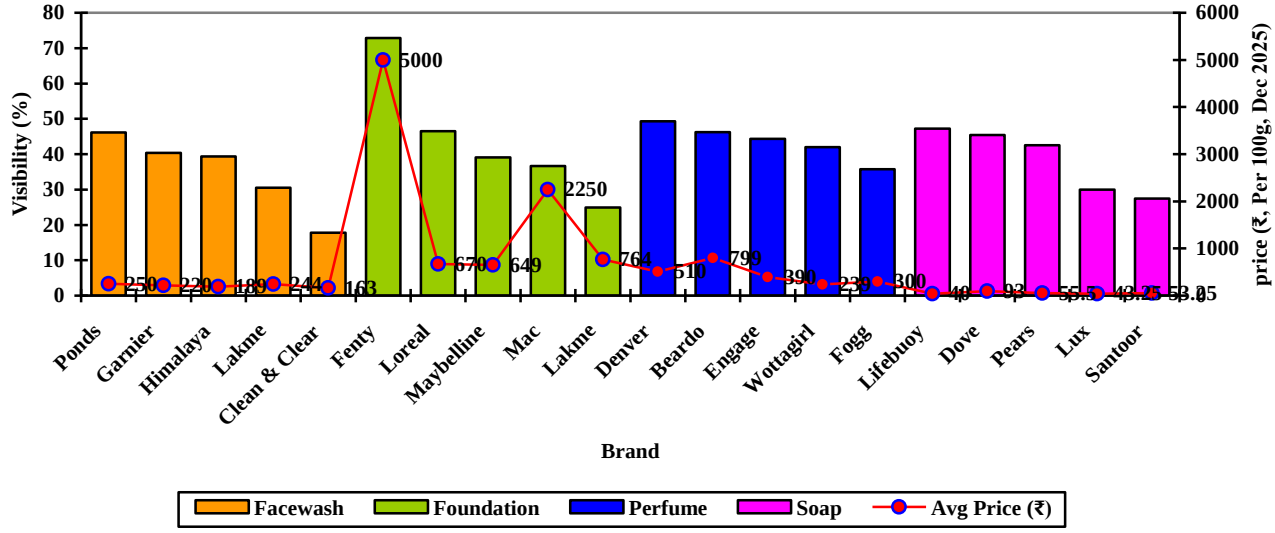


Fig. 5 Cosmetics visibility vs average pricing (₹, as of dec. 2025, per 100 g)

This pattern exposes food advertising strategies and shows that they highly rely on direct product display, pouring shots, packaging close-ups, and visuals of product consumption to strengthen brand recall [2], along with price acting as a co-influencing strategic variable instead of a strictly dependent variable.

The consistent visibility of products across brands implies that clear and repeated visuals positively influence the memory and buying intention of consumers [12, 17]. Figure 5 graphically depicts the results of the comparison for the Cosmetics domain. The results from Figure 5 show that in the Cosmetics domain, the visibility percentage maintains a consistency in high visibility of about 30-50% frame-presence range.

Figure 5 graphically depicts that cosmetic visibility does not follow a uniform relationship across brands. Facewash and soaps achieve higher visibility at low to moderate prices due to wider customer reach, while premium foundations maintain strong visibility despite high prices because brands focus more on attracting customers through their strong image and appeal. Perfumes show a balanced pattern, where mid-range pricing aligns with consistently high visibility.

Overall, cosmetic visibility is affected more by additional factors like cinematic focus on face, emotional cues, and application aesthetics than the product itself [3].

4.3.2. Subdomain-Level Comparative Analysis of Product Visibility

To understand how brands generalize over subdomain categories, the results of Product Visibility Scores are shown in Figures 6 and 7. Figure 6 graphically represents the results in the form of average values of the visibility scores in the subdomains of the Food domain. The results in Figure 6 demonstrate that, as a result of product detection, for the subdomain of Soft Drinks, the visibility score (49.01%) is the highest as compared to the other subdomains.

Whereas, the visibility score is the lowest for the Biscuits subdomain (37.25%). The average values of the visibility scores in the subdomains of the Cosmetics domain are graphically demonstrated in Figure 7. The results in Figure 7 demonstrate that, as a result of product detection, for the subdomain of Foundation, the visibility score (43.99%) is the highest as compared to the other subdomains. At the same time, the visibility score is the lowest for the Facewash subdomain (34.86%).

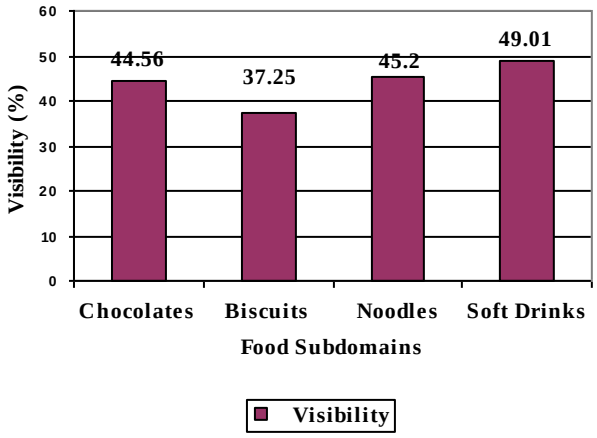


Fig. 6 Subdomain-level visibility percentage (food)

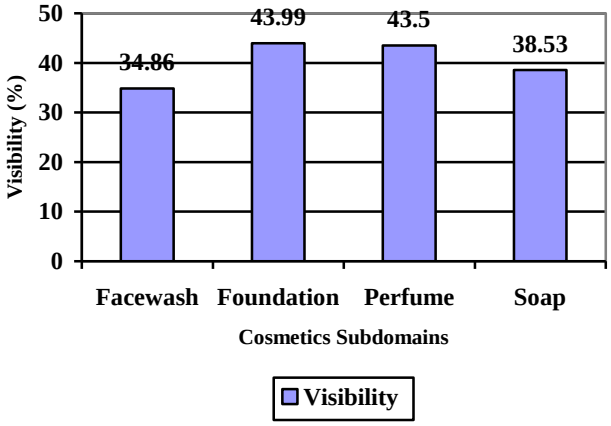


Fig. 7 Subdomain-level visibility percentage (cosmetics)

4.3.3. Subdomain-Level Comparative Analysis of Accuracy, Precision, Recall, and F1-score

To analyse the various brands over subdomain categories, the results of accuracy, precision, recall, and F1-score obtained when all four deep learning models discussed in Section 3.3 are applied for product detection are averaged out. Table 2 shows these results for the subdomains of the Food domain. For the subdomain of Soft Drinks, the values of accuracy and F1-score in the detection of products are the highest in comparison with the other subdomains. Moreover, in terms of accuracy, the highest performance of the models is observed for the Soft Drinks subdomain, and models show a lower performance in the Chocolates subdomain.

The results from Table 2 depict that a notable variation in recall is observed across the subdomains. The range of recall is observed between 0.60 in Biscuits, and 0.76 in Soft Drinks, and accuracy is maintained relatively stable at about 0.81 on average. It can be derived that detection performance differs by the model's capability to recover ground-truth product-present frames instead of considering classification correctness. With increasing recall values, the downstream analysis is supported by a large proportion of ground-truth frames. Thus, from the results in Table 2 and Figure 6, it is justified that product visibility is directly proportional to recall in the Food subdomains.

Table 2. Average performance: subdomains of the food category

Metrics	Subdomains			
	Chocolates	**Biscuits**	**Noodles**	**Soft Drinks**
Accuracy	0.79	0.81	0.81	0.83
Precision	0.71	0.64	0.73	0.75
Recall	0.66	0.6	0.7	0.76
F1-score	0.66	0.59	0.68	0.74

Table 3 shows the results in the form of average values of the performance metrics in the subdomains of the Cosmetics domain. From the results in Table 3, the values of accuracy and F1-score in the detection of products are the highest for the subdomain of Soap in comparison with the other subdomains.

Moreover, in terms of accuracy, high performance of the models is observed for the Soap subdomain, and models show low performance in the Foundation subdomain. Similar to the Food domain, recall variation, although in a smaller range (0.68-0.70), is observed in the cosmetics domain while maintaining a range of 0.79-0.82 accuracy values. From the results in Table 3 and Figure 7, Foundation shows the lowest recall (0.67) with visibility (43.99%) and hence shows fewer recoveries of ground-truth frames. Highest recall is observed for Facewash (0.70) with the lowest visibility (34.86%).

This contradicts the dependency as seen for the Food subdomains. Thus, it shows a non-proportional relationship between product visibility and recall in the cosmetics domain, demonstrating that visibility is governed by the advertisement presentation style rather than detection recovery.

Table 3. Average performance: subdomains of the cosmetics domain

Metrics	Subdomains			
	Facewash	**Foundation**	**Perfume**	**Soap**
Accuracy	0.81	0.79	0.79	0.82
Precision	0.71	0.67	0.66	0.69
Recall	0.7	0.67	0.68	0.69
F1-score	0.66	0.65	0.64	0.68

Visibility analysis is performed with the help of Ground-truth-validated recall, influenced by recall-driven visibility in the food subdomain and presentation-driven visibility trend in the cosmetics subdomain.

4.3.4. Domain-Level Comparative Analysis of Product Visibility

On evaluating the visibility (%) of the products in two main domains: Food and Cosmetics, the average visibility (%) of the overall domains is calculated. Figure 8 graphically shows that overall, the products in the Food domain possess 44.01% visibility in the advertisements, whereas products in the Cosmetics domain have 40.22% visibility in the advertisements.

4.3.5. Assessment of Model-Level Performance in Product Detection

A meaningful output of the product presence is drawn out by comparing model performance based on metrics that are aggregates of accuracy, precision, recall, F1-score, and average confidence for every model described in Section 3.3 to choose the best one.

Table 4 shows the actual values of performance metrics for selected models and is used to present the results in the form of a radar chart, as shown in Figure 9. The radar chart shows a clear and consistent ranking across the models. YOLOv8 forms the largest and most balanced polygon and achieves the highest values for accuracy (0.92), recall (0.81), and F1-score (0.81).

The results support the inferences that state YOLOv8's multi-scale feature extraction, improved backbone, and optimized detection head [19, 20]. YOLOv5 is close to the performance of YOLOv8 but slightly lacks recall (0.64) and F1-score. Faster R-CNN and SSD300 have smaller radar ranges, implying weaker generalization and lower consistency.

Faster-R-CNN has a lower recall (0.57) and precision, while SSD300 lacks specifically in precision and F1-score (0.63). Behaviours of these two models resemble the limitations of classical detectors while dealing with reflective or small, irregularly visible objects as observed in cosmetic advertisements [10, 11].

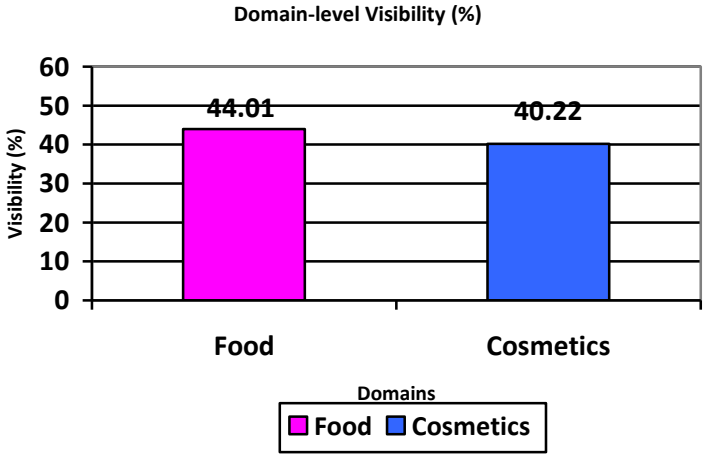


Fig. 8 Domain-level visibility percentage (food vs cosmetics)

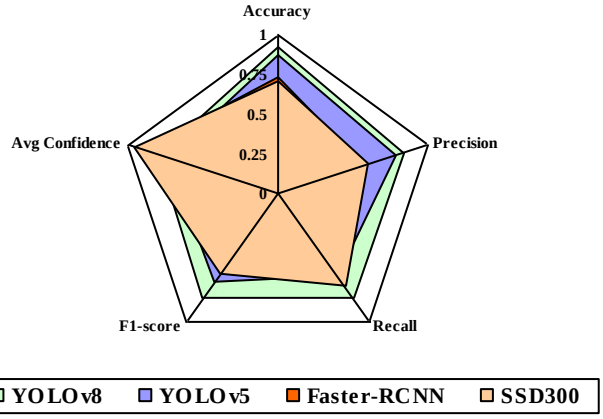


Fig. 9 Model performance radar chart: YOLOv8 vs YOLOv5 vs faster-RCNN vs SSD300

Table 4. Comparative results of models in product detection

Metrics	Models			
	YOLOv8	YOLOv5	Faster R-CNN	SSD300
Average Confidence	0.8	0.74	0.9	0.97
Accuracy	0.92	0.88	0.74	0.7
Precision	0.84	0.78	0.57	0.59
Recall	0.81	0.64	0.57	0.72
F1-score	0.81	0.69	0.53	0.63

Figure 10 graphically demonstrates the values of the performance metrics for the explored models. Average confidence determines the surety with which the model detects the product. However, high confidence is not reliable for detection. SSD300 and Faster R-CNN produce greater confidence scores (0.97 and 0.90, respectively), but with low precision and F1-scores demonstrating over-confidence and inaccurate predictions. Whereas, YOLOv8 shows a balanced average confidence (0.80), having strong precision, recall, and F1-score, proving that correctness associated with confidence has reliable predictions, as in the case of YOLOv8. Based on the averaged metrics, the following trend is observed with respect to the best performance of models. YOLOv8 > YOLOv5 > Faster R-CNN > SSD300

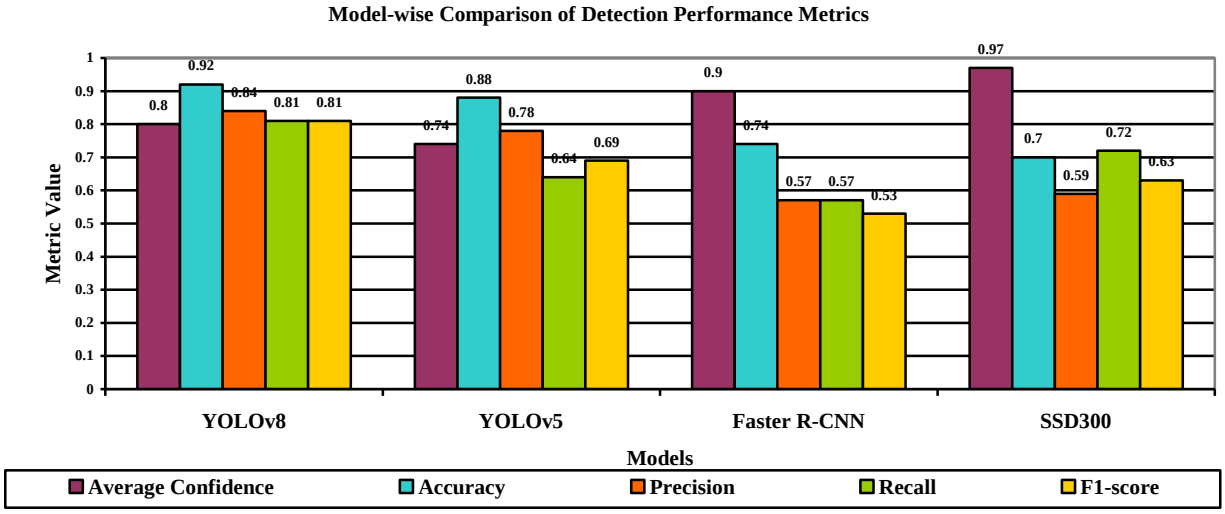


Fig. 10 Model-wise performance in product detection

Model-wise evaluation is important as it signifies visibility measurement is not entirely dependent on the advertisement content, but also the capabilities of the models to detect products under dynamic filming conditions. Figures 11-14 show representative frames obtained during the testing phase of the proposed system. In these frames, the product is automatically detected and marked with the bounding boxes using the YOLOv8 model. The results show that Food advertisements (Figures 11 and 12) usually display the product clearly and for a longer time. The packaging is also often visible and is detected correctly in multiple frames. On the other hand, cosmetic advertisements (Figures 13 and 14) show the product less often. The focus in these advertisements is more on people, facial expressions, and emotions. So, the product appears for shorter periods. These frames clearly match and support the domain-level analysis results. They confirm that the food advertisements have higher and more consistent product visibility than cosmetic advertisements.

4.3.6. Visual Evidence for Product Visibility across Advertisements

Fig. 11 Product detection by the system - food advertisement (frame 1)
source: youtube, "kitkat, crispy, delicious fingers" https://youtu.be/2Lkc0Yb6i-w?si=_jYcGtDfHFKAvj5k

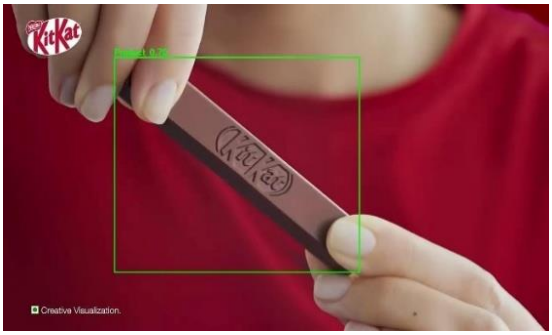


Fig. 12 Product detection by the system - food advertisement (frame 2)
source: youtube, "kitkat, crispy, delicious fingers" https://youtu.be/2Lkc0Yb6i-w?si=_jYcGtDfHFKAvj5k

Fig. 13 Product detection by the system - cosmetics advertisement (frame 1) source: youtube, "lux flaw-less glow anushka sharma & virat kohli" https://www.youtube.com/watch?v=uRpj5hyqN_c

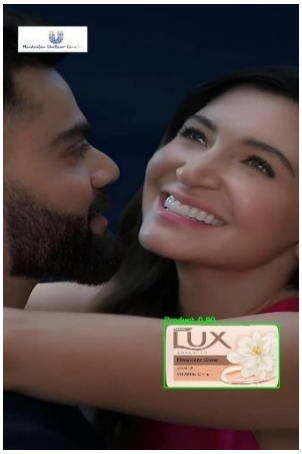


Fig. 14 Product detection by thesystem - cosmetics advertisement (frame 2) source: youtube, "lux flaw-less glow anushka sharma & virat kohli" https://www.youtube.com/watch?v=uRpj5hyqN_c

5. Observations and Findings

The comparative analysis using ProVis-Ad reveals systematic performance differences across domains, brands, and models that directly support the proposed hypothesis.

5.1. Why Food Advertisements Have Greater Product Visibility

1) Food brands usually show the package and the food being served or eaten, which helps make the product visible throughout the advertisement.
2) Food products generally have well-defined shapes like packets, bottles, and boxes, which makes it easier for detectors such as YOLO and SSD to localize them [5, 10].
3) In contrast, products like perfume bottles, foundation, and facewash tubes cover a very small part of the frame, which creates challenges for detecting small objects [10, 25].
4) Cosmetic advertisements commonly emphasize faces, skin texture, and transformation, which reduces the total number of frames where the actual product is visible.
5) In soap and facewash advertisements, the visible product often becomes the foam or lather rather than the actual container or pack of product, which leads to inconsistent percentages of visibility within the cosmetic subdomain.
6) As per the analysis in this work, the Food product visibility is 40-60%, and the Cosmetics product visibility is 30-50% of the total duration of the advertisement.

5.2. Role of Product Pricing in Visibility

1) Food products use pricing and on-screen visibility as supporting factors of the marketing strategy.

a. When the average price is plotted against visibility, food brands group together with moderate prices and high visibility.

b. This supports the claim that food products focus on showing the packaging often to highlight value and help people remember the brand [12, 17].

2) Cosmetic products show very little or no link between visibility and price. High-priced cosmetic products, such as perfumes and foundations, show lower visibility because:

a. Product shots are brief and visually appealing.

b. The goal is that the product should be emotionally associated with the customer rather than the product display.

3) This aligns with findings that premium cosmetic brands focus more on emotional appeal than on showing visuals related to functionality of the product [3].
4) Higher-priced food brands often maintain a longer screen presence of the product in the advertisements to strengthen the perceived value.
5) However, in cosmetics, premium brands frequently give priority to emotional storytelling, which in turn reduces the actual product's screen time.

5.3. Manual vs Model Observations

Manual observations bring in visual checking of advertisement frames to study the amount of time the product is visible on screen without using automated model outputs. These observations confirm the same trend:

1) Food advertisements visually display the product more frequently, with longer continuous shots.
2) Cosmetic advertisements, however, show irregular visibility of products, which is often disturbed by model focus, hands covering the product, or transitions due to the application of the product.

As illustrated in Section 4.3, the analysis of product detection models supports and quantifies these observations, and shows that YOLOv8 is very close to human observations.

5.4. Hypothesis Validation

The proposed hypothesis is validated through an analysis that goes through multiple levels by considering different aspects of product visibility instead of depending on just one metric.

1) First, the domain-level visibility analysis compares the average product visibility in food and cosmetic advertisements. This analysis shows that the product visibility patterns differ between domains in how often and how continuously the product appears on screen.
2) Second, brand-level visibility analysis focuses on how individual brands show their products in advertisements. This helps to check if the overall trends observed in each domain apply to most brands or are driven only by a few advertisers.
3) Third, qualitative inspection is carried out using manual observations. Each frame is carefully examined, which shows that food advertisements give a longer screen time to their products. On the other hand, the cosmetic advertisements display the products in brief moments, often drawing attention away from the product by focusing on faces, expressions, or story-driven demonstrations.
4) Finally, the analysis of pricing correlation checks whether the cost of a product affects the visibility strategies across different domains. This analysis helps to determine if a higher visibility of a product is due to its pricing or to its advertising practices specific to the domain.

Together, these aspects validate the following hypothesis: Food advertisements exhibit stronger and more consistent product visibility than cosmetic advertisements.

6. Conclusion

In this paper, a Product Visibility Analysis for Advertisements (ProVis-Ad) - a novel system for product detection and its visibility measurement in real advertisement videos has been presented. The ProVis-Ad system focuses on the comprehensive range of food and cosmetics product advertisements for analysis. Thus, for the evaluation, eight datasets from 155 real-world food and cosmetics video advertisements are created. The comparative analysis of product visibility shows that at the domain level, food advertisements consistently demonstrate about 40-60% product visibility, and cosmetics advertisements demonstrate about 30-50% product visibility. At the brand and subdomain level, it is further observed that "soft drinks" and "noodles" show the highest visibility in the food domain. In the cosmetics domain, however, "foundation" and "perfume" display a relatively higher visibility as compared to "facewash" and "soap". Experimental results have shown that product visibility is directly proportional to recall in the food subdomains. Whereas, results have depicted that there is a non-proportional relationship of product visibility and recall for the cosmetics domain, demonstrating that visibility is governed by advertisement presentation style rather than detection recovery.

Experimental results have further illustrated that YOLOv8 delivers the outstanding and most stable performance, achieving the highest accuracy (0.92), precision (0.84), recall (0.81), and F1-score (0.81). Results have also shown that YOLOv8 depicts a balanced average confidence (0.80), signifying its applicability in reliable predictions. Thus, as the YOLOv8 depicts the optimal performance, it is the best-suited model for product detection and visibility analysis in the ProVis-Ad system. Analysis of how product pricing is related to the visibility of the product shows domain-specific trends. The food domain analysis suggests that pricing and on-screen visibility jointly act as influential strategic factors, and hence, indicates that visibility of the product is used to strengthen recall, along with reflecting product cost. However, the cosmetic domain does not show the above trend considering price and visibility. The priority is given to emotional storytelling and aesthetics over the actual display of the product. To conclude, price-visibility reveals a stronger marketing strategy in Food and a weaker and more inconsistent one in Cosmetics. In summary, ProVis-Ad shows a strong practical value beyond research experiments. The system can help brands to measure how much their product appears in the advertisements and can also be used as media analytics through large-scale analysis of product placement trends. In addition, it can also help in compliance verification by validating contractual visibility requirements in advertisement videos. These applications boost the usefulness

of ProVis-Ad as a data-based system for advertisement analysis. Overall, ProVis-Ad offers a scalable and reliable way that helps advertisers, marketing analysts, and researchers measure product visibility. The proposed system supports automated analysis of advertisements and helps in advanced research in both computer vision and advertising analytics. Looking ahead, product detection can be enhanced by fine-tuning models for specific product categories and using new vision transformers in advertisements involving fast motions or style changes. Overall, these future enhancements would not only refine the system's accuracy but also bring a greater impact of the ProVis-Ad system as a tool for multimedia analysis, marketing science, and automated assessment for advertisements.

References

[1] S. Shunmuga Krishnan, and Ramesh K. Sitaraman, "Understanding the Effectiveness of Video Ads: A Measurement Study," *IMC '13: Proceedings of the 2013 Conference on Internet Measurement Conference*, Barcelona Spain, pp. 149-162, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[2] Michele Bittencourt Rodrigues et al., "Revolutionising Food Advertising Monitoring: A Machine Learning-based Method for Automated Classification of Food Videos," *Public Health Nutrition*, vol. 26, no. 12, pp. 2717-2727, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nuri Paşa Özer, and Ali Erkam Yarar, "An Analysis of Cosmetic Advertisements on Instagram," *Journal of Civilization and Society*, vol. 9, no. 1, pp. 40-56, 2025. [Google Scholar] [Publisher Link]

[4] Steven C.H. Hoi et al., "Logo-Net: Large-Scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks," *arXiv Preprint*, pp. 1-15, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[5] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 28, pp. 1-9, 2015. [Google Scholar] [Publisher Link]

[6] Rahma Dania, and Rosi Kumala Sari, "A Multimodal Analysis of Food Advertisement," *iNELTAL Conference Proceedings: The International English Language Teachers and Lecturers Conference*, pp. 86-92, 2020. [Google Scholar]

[7] Zaeem Hussain et al., "Automatic Understanding of Image and Video Advertisements," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1100-1110, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Areeg Fahad Rasheed, and M. Zarkoosh, "Optimized YOLOv8 for Multi-Scale Object Detection," *Journal of Real-Time Image Processing*, vol. 22, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Hieu Duong-Trung, and Nghia Duong-Trung, "Integrating YOLOv8-agri and DeepSORT for Advanced Motion Detection in Agriculture and Fisheries," *EAI Endorsed Transactions: on Industrial Networks and Intelligent Systems*, vol. 11, no. 1, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Wei Liu et al., "SSD: Single Shot MultiBox Detector," *Computer Vision - ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, vol. 9905, pp. 21-37, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ángel Morera et al., "SSD vs. YOLO for Detection of Outdoor Urban Advertising Panels Under Multiple Variabilities," *Sensors*, vol. 20, no. 16, pp. 1-23, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Hairong Li, and Nan Zhang, "Computer Vision Models for Image Analysis in Advertising Research," *Journal of Advertising*, vol. 53, no. 5, pp. 771-790, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Patrick Jonsson, "*A Deep Learning Approach to Advertisement Detection in Newspapers*," Master's Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2022. [Google Scholar]

[14] Petar Ristoski et al., "A Machine Learning Approach for Product Matching and Categorization: Use Case: Enriching Product Ads with Semantic Structured Data," *Semantic Web: - Interoperability, Usability, Applicability*, vol. 9, no. 5, pp. 707-728, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] James Hahn, and Adriana Kovashka, "Measuring Effectiveness of Video Advertisements," *arxiv Preprint*, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Dongfang Li et al., "On Detection of Advertising Images," *2007 IEEE International Conference on Multimedia and Expo*, Beijing, China, pp. 1758-1761, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hao Li, and Hui-Yi Lo, "Do you Recognize its Brand? The Effectiveness of Online In-Stream Video Advertisements," *Journal of Advertising*, vol. 44, no. 3, pp. 208-218, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[18] Edward Ratner, Schuyler Cullen, and James Quigley, "Object Recognition in Complex Video Scenes for Advertising Applications," *2015 49th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, pp. 1387-1392, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[19] Oluwaseyi Ezekiel Olorunshola, Martins Ekata Irhebhude, and Abraham Eseoghene Evwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Dillon Reis et al., "Real-Time Flying Object Detection with YOLOv8," *arXiv Preprint*, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Zuzana Berger Haladova, Michal Zrubec, and Zuzana Cernekova, "A Method for Estimating Roadway Billboard Salience," *SAP '24: ACM Symposium on Applied Perception 2024*, Dublin, Ireland, pp. 1-5, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Shervin Minaee et al., "Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection in Videos," *arXiv Preprint*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Jinqiao Wang et al., "ActiveAd: A Novel Framework of Linking ad Videos to Online Products," *Neurocomputing*, vol. 185, pp. 82-92, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[24] Faeze Zakaryapour Sayyad et al., "AdVision: An Efficient and Effective Deep Learning-based Advertisement Detector for Printed Media," *Machine Learning with Applications*, vol. 21, pp. 1-12, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Haidi Zhu et al., "A Review of Video Object Detection: Datasets, Metrics and Methods," *Applied Sciences*, vol. 10, no. 21, pp. 1-24, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Vivi Afifah, and Surni Erniwati, "Yolov8 for Object Detection: A Comprehensive Review of Advances, Techniques, and Applications," *IJACI: International Journal of Advanced Computing and Informatics*, vol. 2, no. 1, pp. 53-61, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[27] Haijun Zhang et al., "Object-Level Video Advertising: An Optimization Framework," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 520-531, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Tomohiko Takahashi et al., "Arbitrary Product Detection From Advertisement Video by using Object Independent Features," *2011 IEEE International Conference on Multimedia and Expo*, Barcelona, Spain, pp. 1-6, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[29] Khinsa Fairuz Zahirah et al., "Evaluating the Effectiveness of Digital Product Advertisement Type using Machine Learning and Shapley Additive Explanations Analysis," *Journal of Information and Communication Technology (JICT)*, vol. 25, no. 1, pp. 79-106, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[30] Rafaela Cajado Magalhães de Alencar et al., "Impact of Visual Appeals and Brand Ambassador in Online Food Advertising on Consumer Purchase Behaviour," *International Journal of Social Economics*, vol. 53, no. 1, pp. 149-162, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[31] Sukriti Dhang, Mimi Zhang, and Soumyabrata Dev, "AdSegNet: A Deep Network to Localize Billboard in Outdoor Scenes," *Signal, Image and Video Processing*, vol. 18, no. 10, pp. 7221-7235, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Wonkyung Kim et al., "A Deep Learning Approach for Identifying User Interest from Targeted Advertising," *Journal of Information Processing Systems*, vol. 18, no. 2, pp. 245-257, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Divya Nimma et al., "Object Detection in Real-Time Video Surveillance using Attention based Transformer-YOLOv8 Model," *Alexandria Engineering Journal*, vol. 118, pp. 482-495, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[34] Martin Magdin, and Zoltán Balogh, "Comparison Classification Algorithms and the YOLO Method for Video Analysis and Object Detection," *Scientific Reports*, vol. 15, no. 1, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Marco A. Moreno-Armendáriz et al., "Deep-Learning-based Adaptive Advertising with Augmented Reality," *Sensors*, vol. 22, no. 1, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, pp. 7464-7475, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[37] Hangyue Zhao, Hongpu Zhang, and Yanyun Zhao, "YOLOv7-sea: Object Detection of Maritime UAV Images based on Improved YOLOv7," *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Waikoloa, HI, USA, pp. 233-238, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[38] Frouke Hermens, "Automatic Object Detection for Behavioural Research using YOLOv8," *Behavior Research Methods*, vol. 56, no. 7, pp. 7307-7330, 2024. [CrossRef] [Google Scholar] [Publisher Link]