

Research Article

Enhancing Image Content Analysis in B2C Online Marketplaces

Hilal Müleyke Yuksel¹, Arma Değer Mut², Alper Ozpinar^{3*}

¹ Pazarama, <https://orcid.org/0009-0008-7491-3693>, hilal.yuksel@pazarama.com,

² Pazarama, <https://orcid.org/0009-0001-1611-8666>, arma.mut@pazarama.com,

³ Istanbul Commerce University, <https://orcid.org/0000-0003-1250-5949>, alper.ozpinar@ticaret.edu.tr,

* Correspondence: alper.ozpinar@ticaret.edu.tr; +90 552 336 46 24

(First received September 22, 2023 and in final form December 18, 2023)

Reference: Yuksel, H., M., Mut, A., D., Ozpinari A. Enhancing Image Content Analysis in B2C Online Marketplaces. The European Journal of Research and Development, 3(4), 229-239.

Abstract

The automation of image analysis in Business-to-Consumer (B2C) online marketplaces is critical, especially when managing vast quantities of supplier-uploaded product images that may contain various forms of objectionable content. This study addresses the automated detection of diverse content types, including sexual, political, and disturbing content, as well as prohibited items like alcohol, tobacco, drugs, and weapons. Furthermore, the identification of competing brand logos and related imagery is examined for competition and ethical reasons. The research integrates custom transfer learning models with the established Microsoft and Google Vision APIs to enhance the precision of content analysis in e-commerce settings. The introduced transfer learning model, trained on a comprehensive dataset, exhibited a significant improvement in identifying and categorizing the specified content types, achieving a notable true positive rate that surpasses traditional API performances. The findings reveal that the "Pazarama Model", with its transfer learning framework, not only delivers a more accurate and cost-effective content moderation solution but also demonstrates enhanced efficiency by reducing the image processing time and associated costs. These results support a shift toward specialized transfer learning models for content moderation, advocating for their adoption to maintain content integrity and enhance user trust within e-commerce platforms. The study advocates for continued refinement of these models, suggesting the integration of multimodal data to further advance the content analysis capabilities in B2C environments.

Keywords: Computer Vision, Image Content Analysis, E-Commerce, Transfer Learning, Categorization

1. Introduction

In today's digital age, e-commerce platforms have become the lifeblood of global trade, transcending physical limitations and reshaping the landscape of product marketing and sales [1]. This burgeoning online ecosystem reflects not only the prowess of technological innovation but also a fundamental transformation in consumer behaviour, where convenience and variety reign supreme [2]. As these platforms burgeon, they amass a staggering inventory – a sea of products, each vying for attention, allure, and relevance within a diverse global audience [3].

However, this abundance presents a new set of challenges. Managing a vast and diverse product inventory demands meticulous organization and strategic planning. E-commerce platforms must grapple with [4]–[8]:

- Product categorization and searchability: Ensuring products are efficiently categorized and discoverable through intuitive search interfaces is crucial for customer satisfaction and conversion rates.
- Inventory optimization: Balancing stock levels, managing product lifecycles, and minimizing dead stock requires sophisticated inventory management systems and data-driven insights.
- Personalization and localization: Catering to the preferences and cultural nuances of a global audience necessitates personalized product recommendations, targeted marketing campaigns, and localized product descriptions.

Navigating this sea of products requires not just technological prowess but also a deep understanding of consumer behaviour, market trends, and the ever-evolving digital landscape. E-commerce platforms that successfully master this art will ride the wave of online retail dominance, ensuring their products find their perfect harbour in the hearts and minds of customers worldwide [9].

Managing such an extensive array of products presents a significant challenge. The complexity lies not only in the sheer volume of items but also in the diverse nature of these products, each with unique specifications, images, and descriptions [10], [11]. This necessitates a robust system that can handle the influx of data efficiently and accurately [12].

A pivotal aspect of this product management is the self-service upload of product images by suppliers. This approach has been increasingly adopted by e-commerce platforms to streamline the process and empower suppliers. However, with this autonomy comes the responsibility of maintaining quality and consistency in the product images. Poor quality

or irrelevant images can significantly impact the user experience, potentially leading to a loss in sales and customer trust.

In this context, several e-commerce platforms have emerged as successful examples of implementing effective image upload and management systems. These platforms have leveraged advanced technologies to facilitate a smooth, efficient process while ensuring the quality of the uploaded images [13], [14].

Central to this technological advancement is the role of autonomous machine learning, particularly in image analysis and categorization. Machine learning algorithms have the capability to automatically process and categorize vast amounts of image data, reducing the need for manual intervention. This not only streamlines the process but also enhances accuracy and consistency in product representation.

This surge in the application of machine learning has led to the exploration of specialized tools like the Microsoft Vision API [15]–[17] and Google Vision API [18], [19]. These APIs provide powerful image recognition capabilities, enabling platforms to automatically analyze and categorize images at scale. The Microsoft Vision API, for example, offers features like object detection, brand recognition, and even sentiment analysis within images. Similarly, the Google Vision API provides robust image labeling, landmark detection, and facial recognition features. While these tools have significantly contributed to the advancement of image analysis in e-commerce, they are not without limitations. Their generalized models, while effective in broad applications, often lack the nuanced understanding required for specific e-commerce contexts.

In the evolving landscape of e-commerce, a significant transformation has been observed in the approach to product image analysis. The traditional, labor-intensive manual processes have been largely supplanted by automated solutions driven by artificial intelligence (AI) [20]. Central to this transformation is the integration of machine learning (ML) and deep learning (DL) techniques, particularly deep neural networks, which have shown remarkable capability in interpreting complex visual data [21].

Convolutional Neural Networks (CNNs), inspired by the neural architecture of the human brain, have been extensively utilized for pattern recognition in visual data. These networks are adept at analyzing various attributes of product images, including object detection, color and texture analysis, and understanding contextual nuances [22], [23]. Despite their proficiency, the direct application of these generalized AI models to the unique context of e-commerce has encountered notable challenges. E-commerce images, characterized by variability in quality and presentation, often defy the standardized parameters of these models [24].

The introduction of transfer learning has been pivotal in addressing these challenges. This approach involves adapting pre-trained neural networks to the specific context of e-commerce imagery. By leveraging the foundational knowledge acquired from general datasets, these networks are fine-tuned to cater to the nuanced requirements of product categorization in online marketplaces.

The adoption of transfer learning has several advantages [25]. Primarily, it reduces the necessity for extensive training data and computational resources, building upon the pre-existing knowledge base of the models. Moreover, it enhances the accuracy and context sensitivity of the models, enabling them to distinguish between subtle differences in products and accurately classify them according to e-commerce standards [26].

Recent advancements in this domain extend beyond transfer learning. Research into generative adversarial networks (GANs) for the generation of synthetic training data has shown potential, especially in addressing the scarcity of labeled data in certain e-commerce categories [27]–[29]. Additionally, the convergence of image analysis with natural language processing (NLP) is being explored. This interdisciplinary approach allows for a more comprehensive understanding of products, combining insights from visual data and textual descriptions [30], [31].

As the e-commerce sector continues its rapid expansion, the demand for sophisticated image analysis systems is expected to escalate. The integration of deep learning and transfer learning, along with ongoing technological advancements, is poised to bring significant benefits to e-commerce platforms. These benefits encompass a range of applications, from enhancing personalized shopping experiences to optimizing product categorization and streamlining supplier processes.

2. Materials and Methods

The project's objective was to analyse images uploaded by sellers to the e-commerce platform, Pazarama.com, for the detection of potentially objectionable content. A total of 55 million images were scanned during this study.

In the preliminary phase of this study, a critical evaluation of the efficacy of two prominent computer vision services—Microsoft Computer Vision and Google Vision—was conducted. The objective was to rigorously assess their capabilities in accurately identifying “racy” and “adult” content within a diverse dataset of images, given their relevance and sensitivity in e-commerce platforms.

The employed methodology adopted a binary classification approach for Microsoft's API, wherein images were distinctly labelled as 'true' if they contained the specified content, signifying the presence of either racy or adult material, and 'false' if such content was not

detected. In juxtaposition, Google's API was subjected to a more granular analysis, providing a likelihood classification spectrum. This spectrum ranged from 'very unlikely' to 'very likely', denoting the varying degrees of probability that an image contained the content.

A meticulously curated sample of images underwent analysis through both APIs. The detection rate, defined as the proportion of images classified within each category by the respective service, was meticulously calculated.

2.1. Racy Content Analysis

For Microsoft Computer Vision, out of the 1,418 images analysed, 368 were identified as 'true' for containing racy content, yielding a detection rate of approximately 25.95%. The remaining 74.05% were classified as 'false,' indicating no racy content was detected.

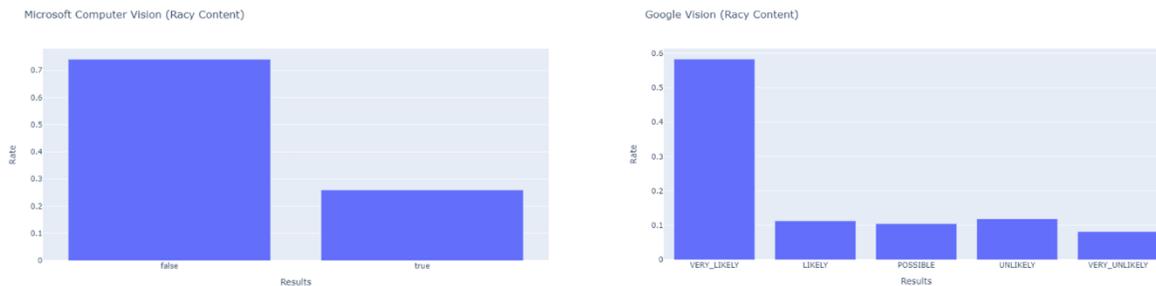


Figure 1: Racy Content Analysis

Google Vision's results were categorized into five likelihood levels. The service tagged 758 images as 'very unlikely' to contain racy content, constituting the majority at a rate of approximately 58.30%. The 'unlikely' and 'possible' categories followed, with rates of 11.84% and 10.46% respectively. Notably, 'likely' and 'very likely' categories had smaller representations in the dataset as can be seen from Figure 1.

2.2. Adult Content Analysis

The Microsoft Computer Vision API's results for detecting adult content in the dataset show that out of 1,418 images, 368 were classified as containing adult content ('true'), resulting in a detection rate of approximately 25.95%. The majority of the images, 1,050 of them, were categorized as 'false,' with no adult content detected, leading to a rate of 74.05%.

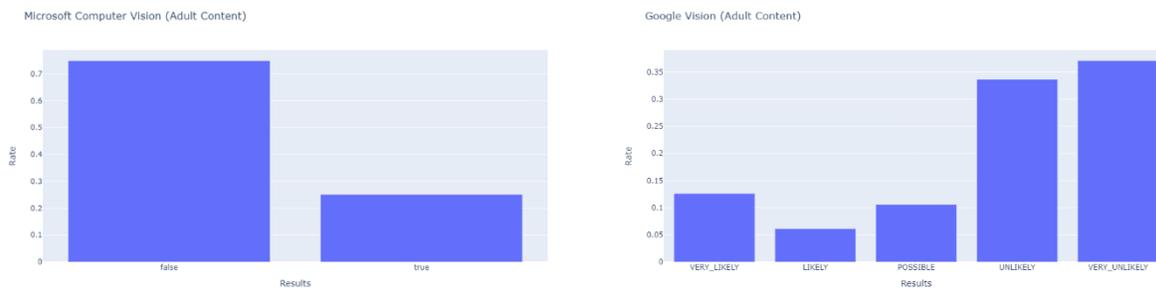


Figure 2: Adult Content Analysis

Google Vision API categorized the images with varying degrees of likelihood for containing adult content. Out of the total images analyzed, 175 were marked as 'very likely' to contain adult content, which translates to a rate of 12.58%. The 'likely' category comprised 85 images with a rate of 6.11%, and 'possible' included 147 images at a rate of 10.56%. Notably, the majority of images fell into the 'unlikely' and 'very unlikely' categories, with rates of 33.64% and 37.09% respectively, indicating a lower probability of containing adult content according to Google's assessment, as can be seen from Figure 2.

2.3. Pazarama Improved Model

The research approach within the paper employed EfficientNet V2, a transfer learning algorithm from the PyTorch library, known for its pre-trained models optimized for efficiency. To customize this model for a specialized dataset, a unique dataset was constructed. The dataset was balanced using the `compute_class_weight` method from the Scikit-learn library, aiming to mitigate any imbalance within the data and foster a homogeneous learning environment.

Images of varying dimensions were standardized to a single resolution and converted into tensors. This standardization was critical for establishing consistency during the training process and improving the model's performance outcomes.

Fine-tuning of the model was conducted through iterative research. Specific layers of the model were sequentially unlocked and trained, resulting in the model achieving the desired performance. Notably, an ensemble learning approach was not adopted. Instead, the specific layers were adjusted at certain epochs. Additionally, the Vision Transformers (ViT) model from Hugging Face was evaluated.

On average, 100,000 images were screened daily, with approximately 1-2% of these images being flagged as potentially objectionable and reported to the Category team.

In terms of performance, Microsoft's API demonstrated a 25% true positive rate (TPR), Google's API achieved a 78% TPR, while the customized model for Pazarama reached a 96% TPR.

The major categories screened for objectionable content included:

- Sexual content
- Political content
- Alcohol, tobacco, and drugs
- Weapons
- Disturbing content
- Competing brand logos and related imagery

3. Result

The custom model's performance is notably exceptional, particularly when considering the challenging nature of content classification within e-commerce. The F1-scores, a harmonic mean of precision and recall, are indicative of the model's ability to balance the identification of objectionable content (true positives) with the exclusion of non-objectionable content (true negatives).

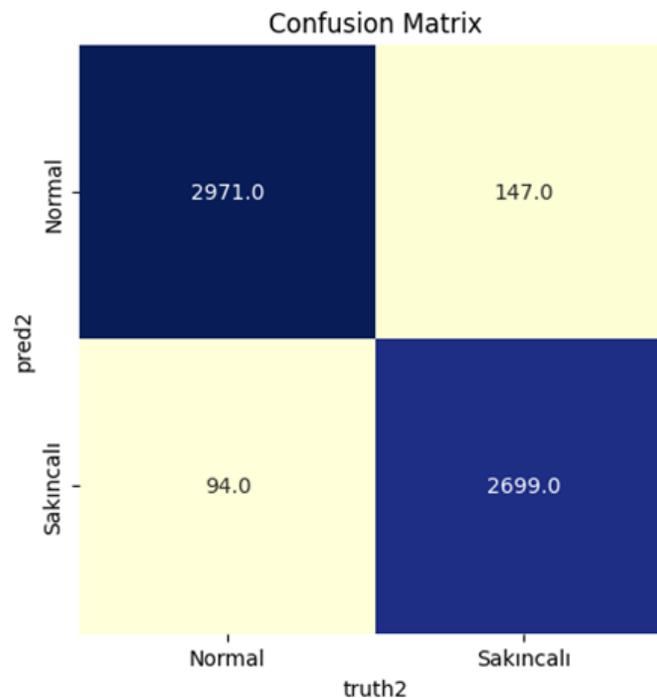


Figure 3: Confusion Matrix for Pazarama Model

Figure 3 and Table 1 shows the details of the improved model.

Table 1: Classification Confusion Matrix

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Normal</i>	0.95	0.97	0.96	3065
<i>Objectionable</i>	0.97	0.95	0.96	2846
<i>Accuracy</i>			0.96	5911
<i>Macro Avg</i>	0.96	0.96	0.96	5911
<i>Weighted Avg</i>	0.96	0.96	0.96	5911

The overall high performance across most categories suggests that the custom model is a substantial improvement over the existing solutions, providing a tailored approach to content moderation. This is supported by the comprehensive nature of the categories, which encompass a wide array of potential e-commerce items and the types of objectionable content that might be associated with them.

4. Discussion and Conclusion

The comparative study evaluated three distinct content moderation systems: Microsoft's content filter, Google's Vision API, and Pazarama's custom model employing transfer learning. The evaluation metrics focused on the detection of racy and adult content, other disturbing content, and restricted items such as tobacco, alcohol, and firearms. Furthermore, the study assessed the cost-effectiveness and efficiency of each system by analyzing model training costs, processing time, and price per 100,000 images screened. Microsoft's system exhibited a true positive rate (TPR) of 25% for racy and adult content, indicating a relatively low detection capability in this category. For other types of disturbing content, the system's performance was even lower, with a TPR of 8%. Additionally, the Microsoft system was deemed insufficient for detecting restricted items, highlighting significant limitations in its content filtering capabilities.

Google's Vision API showed an improved performance with a 78% TPR for racy and adult content and a 12% TPR for other disturbing content. While this marks a substantial improvement over Microsoft's system, it still indicates room for improvement in identifying non-explicit but potentially objectionable materials. Google's system faced limitations similar to Microsoft's regarding the detection of restricted items, being classified as restricted but not adequately effective.

Pazarama's transfer learning-based model significantly outperformed both Microsoft and Google, with a TPR of over 85% for both racy and adult content, as well as other disturbing content. This model also demonstrated a TPR of over 90% for detecting restricted items, showcasing a superior capability to identify a broader range of objectionable materials effectively.

These results indicate that the transfer learning approach adopted by Pazarama provides a more efficient and cost-effective solution for content moderation in e-commerce platforms. The enhanced performance in detecting a wide range of objectionable content, combined with lower costs and faster processing times, positions transfer learning as a superior method for real-time content filtering.

The implications of these findings are significant for the e-commerce sector, where the rapid and accurate screening of large volumes of images is essential. The adoption of efficient AI-driven content moderation systems can help platforms maintain community standards, comply with regulatory requirements, and enhance user trust and safety.

In conclusion, the custom model represents a significant advancement in automated image classification for e-commerce platforms, offering a nuanced and highly accurate tool for content moderation that could significantly reduce the need for manual review and oversight.

5. Acknowledge

The authors acknowledge that specific cost comparisons of the models have not been disclosed within this document due to proprietary commercial considerations and the variability of pricing structures. Detailed pricing information is available and can be furnished upon direct inquiry to the corresponding author or responsible entity.

References

- [1] J. Canton, *The extreme future: The top trends that will reshape the world in the next 20 years*. Penguin, 2006.
- [2] L. T. T. Tran, "Managing the effectiveness of e-commerce platforms in a pandemic," *Journal of Retailing and Consumer Services*, vol. 58, p. 102287, 2021.
- [3] S. Vaidhyathan, *The Googlization of everything:(and why we should worry)*. Univ of California Press, 2012.
- [4] B. Yoo and N. Donthu, "Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL)," *Quarterly journal of electronic commerce*, vol. 2, no. 1, pp. 31–45, 2001.
- [5] S. S. Srinivasan, R. Anderson, and K. Ponnavaolu, "Customer loyalty in e-commerce: an exploration of its antecedents and consequences," *Journal of retailing*, vol. 78, no. 1, pp. 41–50, 2002.
- [6] M. Wolfinbarger and M. C. Gilly, "Shopping online for freedom, control, and fun," *Calif Manage Rev*, vol. 43, no. 2, pp. 34–55, 2001.

- [7] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet Comput*, vol. 7, no. 1, pp. 76–80, 2003.
- [8] K. C. Laudon and C. G. Traver, *E-commerce 2019: Business, technology, society*. Pearson, 2020.
- [9] C. Christensen and M. Raynor, *The innovator's solution: Creating and sustaining successful growth*. Harvard Business Review Press, 2013.
- [10] Z. Huang and M. Benyoucef, "From e-commerce to social commerce: A close look at design features," *Electron Commer Res Appl*, vol. 12, no. 4, pp. 246–259, 2013.
- [11] C. Naumzik and S. Feuerriegel, "One picture is worth a thousand words? The pricing power of images in e-Commerce," in *Proceedings of The Web Conference 2020*, 2020, pp. 3119–3125.
- [12] S. Q. Lau, "Domain analysis of e-commerce systems using feature-based model templates," Master of Applied Science Thesis, Electrical and computer Engineering, University of Waterloo. Link: <http://gp.uwaterloo.ca/files/2006-lau-masc-thesis.pdf>, 2006.
- [13] S. Painuly, S. Sharma, and P. Matta, "Big Data Driven E-Commerce Application Management System," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2021, pp. 1–5.
- [14] G. Singh, H. Kaur, and A. Singh, "Dropshipping in e-commerce: A perspective," in *Proceedings of the 2018 9th International Conference on E-business, Management and Economics*, 2018, pp. 7–14.
- [15] A. Del Sole and A. Del Sole, "Introducing microsoft cognitive services," *Microsoft Computer Vision APIs Distilled: Getting Started with Cognitive Services*, pp. 1–4, 2018.
- [16] A. Del Sole and A. Del Sole, "Getting Started with the Computer Vision API," *Microsoft Computer Vision APIs Distilled: Getting Started with Cognitive Services*, pp. 5–15, 2018.
- [17] A. Del Sole, *Microsoft computer vision APIs distilled: Getting started with cognitive services*. Apress, 2017.
- [18] J. J. Omena, E. Pilipets, B. Gobbo, and J. Chao, "The potentials of Google Vision API-based networks to study natively digital images," *Diseña*, vol. 19, pp. 1–25, 2021.
- [19] D. Mulfari, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, "Using Google Cloud Vision in assistive technology scenarios," in *2016 IEEE symposium on computers and communication (ISCC)*, IEEE, 2016, pp. 214–219.
- [20] J. Bughin, E. Hazan, P. Sree Ramaswamy, W. DC, and M. Chu, "Artificial intelligence the next digital frontier," 2017.
- [21] T. H. Davenport, *The AI advantage: How to put the artificial intelligence revolution to work*. mit Press, 2018.
- [22] G. W. Lindsay, "Convolutional neural networks as a model of the visual system: Past, present, and future," *J Cogn Neurosci*, vol. 33, no. 10, pp. 2017–2031, 2021.
- [23] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [24] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *ArtifIntell Rev*, vol. 53, pp. 5455–5516, 2020.
- [25] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [26] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [27] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE access*, vol. 7, pp. 36322–36333, 2019.
- [28] H. Alqahtani, M. Kavakli-Thorne, and G. Kumar, "Applications of generative adversarial networks (gans): An updated review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 525–552, 2021.

- [29] I. Goodfellow et al., "Generative adversarial networks," *Commun ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [30] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, "Natural language processing advancements by deep learning: A survey," arXiv preprint arXiv:2003.01200, 2020.
- [31] E. Pons, L. M. M. Braun, M. G. M. Hunink, and J. A. Kors, "Natural language processing in radiology: a systematic review," *Radiology*, vol. 279, no. 2, pp. 329–343, 2016.