engineering proceedings

MDPI

Proceeding Paper

Shoe Recommendation System Integrating Generative Artificial Intelligence and Convolutional Neural Networks for Image Recognition †

Chin-Chih Chang [1,*], Chi-Hung Wei [2], Ray-Nan Liao [1], Sean Hsiao [3] and Chyuan-Huei Thomas Yang [4]

[1] Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu 30012, Taiwan; m11002020@chu.edu.tw
[2] Ph.D. Program in Engineering Science, Chung Hua University, Hsinchu 30012, Taiwan; urderrick@gmail.com
[3] Department of Computer Science and Information Engineering, Ming Chuan University, Taoyuan 33348, Taiwan; sean.hsiao@mail.mcu.edu.tw
[4] School of Information Engineering, Shandong Vocational and Technical University of International Studies, Rizhao 276826, China; chyang@swut.edu.cn
* Correspondence: changc@chu.edu.tw
† Presented at the 2024 IEEE 6th Eurasia Conference on IoT, Communication and Engineering, Yunlin, Taiwan, 15–17 November 2024.

Abstract: We developed a shoe recommendation system that integrates generative artificial intelligence (AI) and convolutional neural networks (CNNs) to enhance image recognition and personalize recommendations. The system utilizes CNNs to accurately identify shoe types from user-uploaded images. Utilizing the capabilities of generative AI, the system generates custom shoe suggestions based on weather and location. The proposed system minimizes the need for manual searching but enhances user experience by providing an efficient, automated, and visually driven solution for selecting shoes. The effectiveness of integrating image recognition and generative techniques paves the way for advancements in AI-driven fashion recommendation systems. The developed method offers a powerful tool for increasing customer engagement and satisfaction by delivering personalized and fashion-forward shoe recommendations.

Keywords: convolutional neural network (CNN); generative AI; image recognition; recommendation system; personalized shopping

Academic Editors: Teen-Hang Meen, Chi-Ting Ho and Cheng-Fu Yang

Published: 8 May 2025

Citation: Chang, C.-C.; Wei, C.-H.; Liao, R.-N.; Hsiao, S.; Yang, C.-H.T. Shoe Recommendation System Integrating Generative Artificial Intelligence and Convolutional Neural Networks for Image Recognition. *Eng. Proc.* **2025**, *92*, 62. https://doi.org/10.3390/engproc2025092062

1. Introduction

The advancement of technology has made the Internet become an integral part of daily life. The widespread use of mobile devices has made accessing online information convenient and ubiquitous. Shoes are essential in everyday life. On the street, we encounter a variety of shoes in different styles and types, and numerous options confuse consumers. A photo of favorable shoes can be taken for future purchase. Current recommendation systems on shopping platforms struggle to meet users' needs due to the growing diversity of shoes.

Using deep learning, we trained models to extract shoe features and combined them with generative artificial intelligence (AI) and recommendation systems to enhance recommendation accuracy and efficiency. We created a fast and personalized shopping experience by utilizing generative AI's natural language processing capabilities to offer precise and thoughtful suggestions.

In this article, Section 1 presents the introduction. Section 2 discusses related work. Section 3 explains the research methodology. Section 4 presents the experimental results and analysis. Section 5 concludes this study and propose recommendations for future work.

2. Related Works

2.1. Deep Learning and Convolutional Neural Network (CNN)

Deep learning is a machine learning approach that uses artificial neural networks to simulate the human brain and automatically learn features from data. Large datasets are processed through multiple layers of neural networks to extract useful information [1]. This method is widely used in image recognition, speech recognition, autonomous driving, and gaming. Well-known deep learning methods include AlphaGo, facial recognition, and traffic sign detection. Deep learning, supported by big data and powerful computational resources, enables the resolution of complex problems without the need for manual feature design.

CNN is a deep learning model architecture, particularly effective for handling image data. It utilizes convolution operations to automatically learn and extract features from images by capturing local patterns while reducing the number of parameters to enhance both efficiency and performance [2,3]. The basic CNN structure includes convolutional layers, pooling layers, activation functions, and fully connected layers. Each layer serves a specific function in the network's operation. Convolutional layers mimic the human visual system by detecting features such as edges, while pooling layers reduce the dimensionality of feature maps, optimizing computation. Non-linear activation functions, especially ReLU (Rectified Linear Unit), address vanishing gradients to enhance network efficiency. Fully connected layers are responsible for the final output, classifying or predicting based on the extracted features.

A residual network (ResNet) is used to solve the problem of vanishing gradients in deep networks with residual connections [4,5]. These connections allow inputs to bypass certain layers, improving gradient flow and accelerating convergence. ResNet variants, including ResNet-50 and ResNet-152, have been widely used in image classification, such as object detection and natural language processing (NLP).

DenseNet (dense convolutional network) solves the vanishing gradient and improves parameter efficiency by introducing dense connections [6]. In DenseNet, every layer is connected to all subsequent layers, maximizing feature reuse and reducing redundant computations, thereby enhancing model performance. You Only Look Once (YOLO) is a real-time object detection algorithm [7]. Unlike traditional methods, YOLO outputs multiple bounding boxes and classifications simultaneously, drastically reducing computation and boosting detection speed. With continuous improvements, YOLO versions such as YOLOv2 through YOLOv8 have enhanced accuracy and speed, making it used most in real-time applications including autonomous driving and surveillance systems [8]. YOLO excels in fast detection but struggles with detecting small objects and generating precise bounding boxes. The latest version is available on the Ultralytics website [9].

2.2. Recommendation System

A recommendation or recommender system analyzes users' historical behavior and preferences to suggest personalized content or products to provide the most relevant information and enhance platform engagement [10,11]. Widely used in e-commerce, social media, and streaming services, recommendation systems include content-based, collaborative filtering, and hybrid recommendations.

Content-based recommendation suggests items similar to those a user has liked before, without relying on other users' data. Collaborative filtering, on the other hand, recommends

items based on similarities between users or items, either by suggesting what similar users have liked or by finding items similar to those the user has shown interest in. Hybrid recommendation systems combine the strengths of both approaches to improve accuracy and diversity. For example, they recommend similar items based on content and then use collaborative filtering to broaden the selection.

The integration of deep learning with traditional recommendation methods advances recommendation systems. Traditional methods often struggle with sparse data and limited feature extraction, but deep learning enhances them by uncovering complex patterns and relationships within large datasets. Neural networks analyze user–item interactions deeply and capture trends and attributes that simpler models may overlook. This hybrid approach leads to accurate and personalized recommendations, especially in cases of sparse data or changing user preferences, enabling smart and adaptive recommendation systems.

2.3. Generative AI

Generative AI is designed to create text, images, music, and videos [12,13]. Unlike traditional AI systems that focus on classification or prediction, generative AI enables new content based on existing data patterns. By learning from large datasets, it captures statistical patterns to generate new data. Applications include text generation, image creation, music composition, video production, and deepfake technology. Generative adversarial networks (GANs), variational autoencoders (VAEs), and transformer architectures are pivotal for the creation of high-quality and diverse outputs [14,15]. The generative pre-trained transformer (GPT), a popular transformer architecture, has gained widespread recognition through products such as ChatGPT and Copilot.

Generative AI platforms have gained immense popularity recently, particularly ChatGPT, Gemini, Copilot, and Sora, which generate diverse content in the form of text, images, and videos. As businesses and individuals increasingly adopt these platforms, the potential for enhanced creativity and productivity continues to grow, transforming how we interact with digital content [16].

In summary, generative AI shows immense creative potential across various fields from texts and images to music and medical imaging. As the technology advances, its applications are increasing, driving innovation in AI. However, addressing the ethical and bias issues posed by generative AI is a challenge.

3. Methodology

3.1. Research Methodology

The shoe recommendation system collects and preprocesses a diverse dataset of shoe images and metadata (e.g., brand, size, and user preferences). Data augmentation and normalization techniques are applied to enhance the dataset, while CNNs are used for shoe image classification and feature extraction. CNN identifies visual patterns in shoes, such as shape and texture, and extracts high-level features that serve as inputs for the recommendation model.

A generative AI platform, such as ChatGPT, is utilized to create new shoe designs based on user preferences, weather conditions, and location. By synthesizing novel shoe images that align with visual and user behavior patterns, these models complement content-based filtering and recommend visually similar shoes. The recommendation system integrates both content-based and collaborative filtering, combining visual features with user behavior data to provide personalized recommendations.

Training the system involves using CNN for image recognition and implementing the hybrid recommendation model. The system is evaluated using metrics including accuracy and precision for image classification and user satisfaction metrics for recommendations.

The system includes a user interface for uploading shoe images and viewing suggestions to ensure ethical considerations related to data privacy and fairness.

3.2. System Flowchart

The system flowchart consists of model training, website setup, and outcomes, as shown in Figure 1.

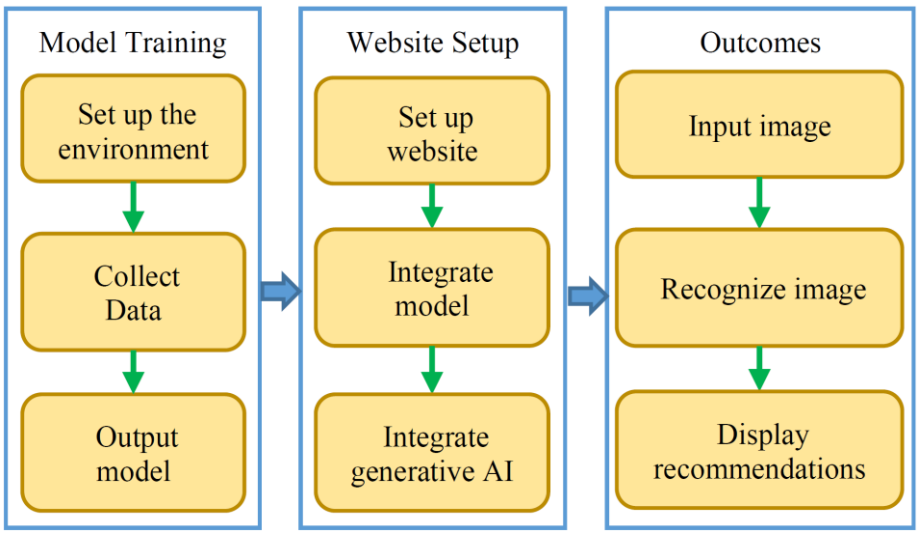

Figure 1. System flowchart.

3.2.1. Model Training

The environment is set up, and the dataset is prepared to generate the output model.

3.2.2. Website Setup

The environment and webpage are constructed by integrating the model and generative AI for recommendations. The website is built on the Flask framework [17].

3.2.3. Outcomes

This phase encompasses inputting an image, recognizing the image, and displaying both the recommendations and the reasoning behind the generative AI recommendations.

3.3. Dataset

The shoe dataset used in this research was provided by Roboflow and can be accessed from the following websites:

- https://universe.roboflow.com/visart-looan/vision_artificial-f7qwx (accessed on 2 October 2024)
- https://universe.roboflow.com/ronnie-paguia/ut-zap50k_heels/dataset/1 (accessed on 2 October 2024)

The dataset includes six categories, with each category containing 3000 training images, 1000 validation images, and 1000 test images. Data preprocessing is critical in deep learning to ensure that the data meet the requirements for model training. Data preprocessing includes data labeling, data collection and cleaning, data partitioning, data augmentation, scaling, and normalization. Data preprocessing is conducted to enhance the diversity and richness of the data and reduce overfitting.

3.4. Training

DenseNet excels in efficiently learning feature hierarchies for classification tasks. ResNet is used for deep architectures, enabling accurate classification even with deep networks. YOLOv8 is optimal for object detection and recognition tasks, enabling real-time shoe identification in images. Each model offers advantages depending on whether the task focuses on classification, feature extraction, or real-time detection in the shoe recommendation system. After training the model using DenseNet, ResNet, or YOLOv8 on the shoe dataset, optimized weights for different shoe types are predicted to detect shoes in images. Along with the model, evaluation metrics such as accuracy, loss, precision, recall, and a confusion matrix are estimated to assess performance. The YOLOv8 model produces bounding boxes around detected objects, along with class labels. The trained model is saved in a serialized format (e.g., .h5 or .pt) for further use, and logs/plots are created to track the model's training history and present metrics over time. This information is used to ensure the model is ready for deployment and capable of making predictions on new data.

3.5. Integrating Generative AI for Enhanced Shoe Recommendations

Using generative AI through an application programming interface (API) such as OpenAI's API, the system generates synthetic shoe images to augment the dataset, increase the diversity of the training data, and reduce overfitting. This helps the model accurately predict new shoe designs based on learned patterns, contributing to product innovation. By using OpenAI's API, developers can seamlessly integrate such capabilities into the system, enhancing its ability to adapt and innovate.

After recognizing a shoe image, the system uses OpenAI's API to generate personalized recommendations based on user preferences, weather, and location (Figure 2). For instance, the system suggests shoes that match the user's style and provides AI-generated explanations, enhancing satisfaction and trust. By connecting to OpenAI's API, these personalized suggestions are generated to enhance interaction and engagement.

```
def get_gpt_recommendation(prompt):
    openai.api_key =
    response = openai.ChatCompletion.create (
        model="gpt-4",
        messages=[
            {"role": "system", "content": "You are a helpful assistant."},
            {"role": "user", "content": prompt}
        ]
    )
    return response.choices[0].message['content'].strip()
```

Figure 2. GPT-generated recommendation.

GPT models generate shoe descriptions via the API and explain how they fit with various outfits and styles to enhance user understanding and shopping experience. Integrated with OpenAI's API, the system surpasses basic image recognition and provides personalized suggestions, synthetic designs, and descriptive insights, making recommendations smart, tailored, and user-centric.

4. Results and Discussions

4.1. Implementation

The system was built using the hardware and software in this study, as presented in Table 1.

Table 1. Implementation environment.

Hardware	
Processor	Intel® Core™ i7-12700 2.10 GHz
Memory	DDR4 32G
Graphics Card	NVIDIA GeForce RTX 4070Ti Super (16G)
Software Development Environment	
Programming Language	Python 3.10.12
Operating System	Windows 11
Development Tools	Anaconda3
Programming Language	Python 3.10.12
Tensorflow	2.10.1
Pytorch	2.4.1
CUDA	11.2
Web Framework	Flask 3.0.0

4.2. Results

We compared the accuracy and performance of different models. Using diagrams and tables, the training effectiveness of each model was visualized. The accuracy of ResNet50 is presented in Figure 3, while Figure 4 illustrates the loss of DenseNet201.

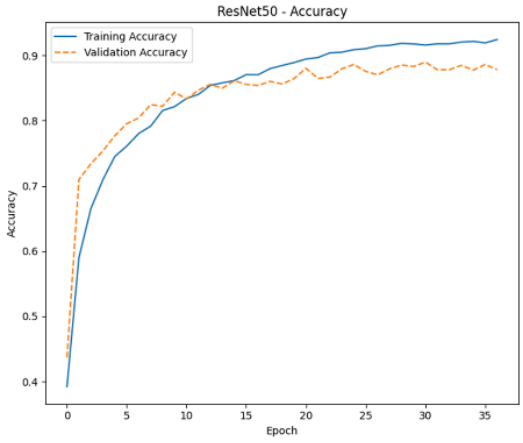

Figure 3. ResNet50 accuracy.

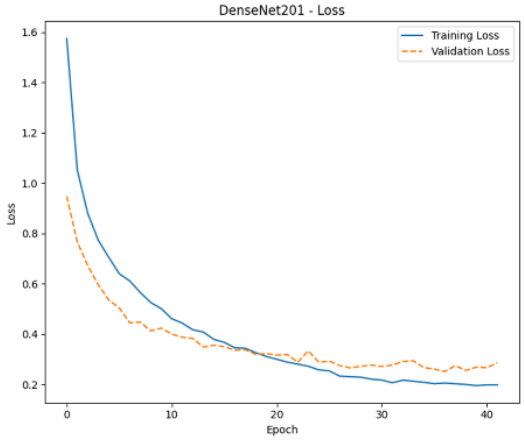

Figure 4. DenseNet201 loss.

The comparison of model performance on the shoe dataset is shown in Table 2. DenseNet201 achieved high accuracy on the training and validation sets, but its performance decreased significantly on the test set, indicating possible overfitting. It had the longest training time, suggesting that its deeper architecture required more time to converge, yet it did not provide superior generalization. ResNet50 provided a good balance between accuracy and training time but suffered from a drop in generalization performance. However, it had a much shorter training time than DenseNet201. YOLOv8 outperformed DenseNet201 and ResNet50, achieving the highest accuracy and time efficiency.

Table 2. Comparison of model performance on shoe dataset.

Model	Training Set Accuracy	Validation Set Accuracy	Test Set Accuracy	Training Time (s)	Epochs
ResNet50	92.41%	88.96%	73.04%	6970.95	37
DenseNet201	94.92%	91.9%	76.28%	17,801.18	42
YOLOv8	99.24%	98.20%	98.20%	5348.10	50

4.3. Website with Generative AI

The website interface was constructed with shoe recognition models and generative AI for personalized user experience. After uploading an image of a shoe, the system recognizes and classifies the shoe using models such as YOLOv8 or ResNet50. Once the shoe is identified, the generative AI enhances the process by offering tailored recommendations.

User Interface and Image Upload

Upon visiting the website, users can upload a shoe image through the interface. The system processes the image and displays the detected shoe type, along with relevant information such as the brand and style, as shown in Figure 5.

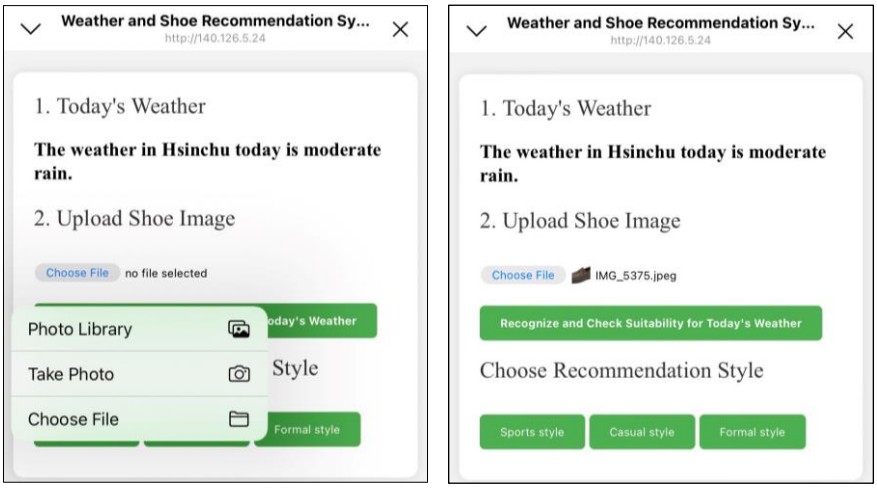

Figure 5. Uploading image interface.

5. Generative AI for Personalized Recommendations

After recognition, generative AI generates personalized shoe recommendations based on weather and location, while also providing AI-generated explanations and visuals to support each suggestion (Figure 6). The user experience is enhanced by offering practical, tailored recommendations along with clear reasons and images for each choice. For example, if the weather is rainy, the system might suggest waterproof or durable shoes, while sunny conditions could prompt recommendations for lighter, breathable footwear. By considering the user's geographic location, the AI tailors suggestions to fit regional

styles or preferences, offering both personalized and practical recommendations in suited to the user's environment.

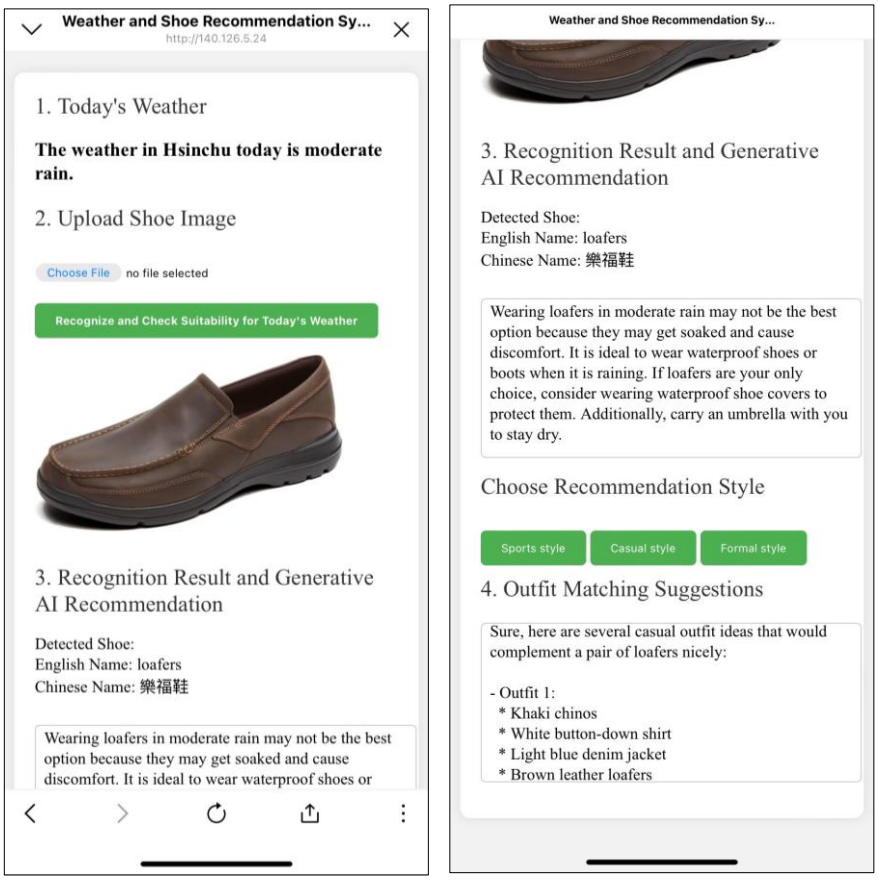

Figure 6. Recognition results and recommendations.

The integration of generative AI enhances the accuracy of shoe recognition and the overall shopping experience, offering users personalized, creative, and interactive suggestions tailored to their preferences.

6. Conclusions and Recommendations

The integration of generative AI and CNNs in the shoe recommendation system demonstrates significant advancements in personalization and image recognition. By utilizing CNNs for precise classification and generative AI for tailored recommendations, the system offers a more efficient, automated solution than traditional methods. The YOLOv8 model outperformed other models in terms of accuracy and training speed, highlighting its suitability for real-time applications. However, DenseNet201, while accurate, presented challenges with longer training times, emphasizing the need for model efficiency in practical scenarios. The system successfully streamlined the shoe selection process and improved user experience by offering visually driven, innovative recommendations.

Future developments are necessary to incorporate additional features such as shoe color, style, and patterns to enhance recommendation quality. By improving system scalability and integrating real-time feedback mechanisms, a more dynamic user interaction is enabled. Optimizing the development environment, perhaps through the use of Docker or virtual environments, improves compatibility and system stability.

Author Contributions: Conceptualization, C.-C.C.; Methodology, C.-C.C., C.-H.W., R.-N.L. and C.-H.T.Y.; writing—original draft preparation, C.-C.C.; writing—review and editing, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. John, D. *Kelleher, Deep Learning*; The MIT Press Essential Knowledge Series; The MIT Press: Cambridge, MA, USA, 2019; pp. 1–295.
2. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
3. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
4. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
5. Koonce, B. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: Berkeley, CA, USA, 2021; pp. 63–72.
6. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
7. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
8. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
9. Ultralytics YOLO11. Available online: https://docs.ultralytics.com/models/yolo11/ (accessed on 14 October 2024).
10. Aggarwal, C.C. *Recommender Systems*; Springer International Publishing: Cham, Switzerland, 2016.
11. Zhang, S.; Yao, L.; Sun, A.; Tay, Y. Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.* **2019**, *52*, 1–38. [CrossRef]
12. Dhamani, N.; Engler, M. Introduction to Generative AI. Manning Publications Co.: Shelter Island, NY, USA, 2024.
13. Feuerriegel, S.; Hartmann, J.; Janiesch, C.; Zschech, P. Generative AI. *Bus. Inf. Syst. Eng.* **2024**, *66*, 111–126. [CrossRef]
14. Aggarwal, A.; Mittal, M.; Battineni, G. Generative adversarial network: An overview of theory and applications. *Int. J. Inf. Manag. Data Insights* **2024**, *1*, 100004. [CrossRef]
15. Bengesi, S.; El-Sayed, H.; Sarker, M.K.; Houkpati, Y.; Irungu, J.; Oladunni, T. Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access* **2024**, *12*, 69812–69837. [CrossRef]
16. Alhur, A. Redefining healthcare with artificial intelligence (AI): The contributions of ChatGPT, Gemini, and Co-pilot. *Cureus* **2024**, *16*, e57795. [CrossRef] [PubMed]
17. Grinberg, M. *Flask Web Development*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2018.