https://doi.org/10.31891/csit-2023-3-3
UDC 004.89: 004.3

Olga PAVLOVA, Andrii BASHTA, Andrii KUZMIN
Khmelnytskyi National University

APPROACHES OF BUILDING A REAL-WORLD OBJECT DETECTOR DATA SOURCE

In our constantly developing world virtual, augmented, and mixed reality technologies are becoming integral parts of our daily lives. In the current stage of Information Technology field development, technologies of virtual, augmented and mixed reality can be seen in almost all areas of human life. Nowadays AR is used in Marketing and Advertising, Education, Medicine, Automotive, Game Development, Navigation and other areas of our everyday life. Therefore, object detection is a crucial task in computer vision and AI applications, enabling machines to identify and locate objects within images or video frames. The accuracy and performance of an object detector heavily rely on the quality and diversity of the training data. This paper is aimed at finding the approaches of building a real-world object detector data source to be able to create a model for detecting a sport games surfaces using the Action & Vision App. During this research several structured approaches of building an object detector data source have been built, drawing inspiration from Apple's Create ML documentation on the topic. Additionally, real-world applications available on both the App Store and Google Play that leverage object detection technology were showcased and analyzed. In the course of study a dataset of objects has been collected and then utilized to build a robust detection model, tailored to function seamlessly with Vision and Core ML frameworks on iOS devices. The trained object detection model, informed by the diverse dataset and robust training process, is employed to identify and outline tables and rectangles in each frame of the video stream. The model and the proposed approaches will be further applied to develop the method of object detection in the real world and create a mobile application for sport games simulation, that would help players to practice their skills out of the training field.

Keywords: object detection, computer vision, AI applications, training data, Create ML, structured data source, real-world applications, App Store, Google Play.

Ольга ПАВЛОВА, Андрій БАШТА, Андрій КУЗЬМІН
Хмельницький національний університет

ПІДХОДИ ДО ПОБУДОВИ ДЖЕРЕЛ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ ОБ'ЄКТІВ У РЕАЛЬНОМУ СВІТІ

На сучасному етапі розвитку сфери інформаційних технологій технології віртуальної, доповненої та змішаної реальності можна побачити практично у всіх сферах життя людини. Сьогодні AR використовується в маркетингу та рекламі, освіті, медицині, автомобілебудуванні, розробці ігор, навігації та інших сферах нашого повсякденного життя. Тому виявлення об'єктів є ключовим завданням у програмах комп'ютерного зору та штучного інтелекту, що дозволяє машинам ідентифікувати та знаходити об'єкти в зображеннях або відеокадрах. Точність і продуктивність детектора об'єктів значною мірою залежать від якості та різноманітності навчальних даних. Ця стаття спрямована на пошук підходів до побудови реального джерела даних детектора об'єктів, щоб мати можливість створити модель для виявлення поверхонь для спортивних ігор за допомогою програми Action & Vision. Під час цього дослідження було створено кілька структурованих підходів до створення джерела даних детектора об'єктів, черпаючи натхнення з документації Apple Create ML на цю тему. Крім того, були продемонстровані та проаналізовані реальні програми, доступні як в App Store, так і в Google Play, які використовують технологію виявлення об'єктів. У ході дослідження було зібрано набір даних об'єктів, який потім використано для побудови надійної моделі виявлення, адаптованої для бездоганної роботи з фреймворками Vision і Core ML на пристроях iOS. Навчена модель виявлення об'єктів, яка базується на різноманітних наборах даних і надійному процесі навчання, використовується для ідентифікації та окреслення таблиць і прямокутників у кожному кадрі відеопотоку. Модель і запропоновані підходи будуть надалі застосовані для розробки методу виявлення об'єктів у реальному світі та створення мобільного додатку для симуляції спортивних ігор, який допоможе гравцям відпрацьовувати свої навички поза тренувальним полем.

Ключові слова: виявлення об'єктів, комп'ютерний зір, програми ШІ, навчальні дані, Create ML, джерело структурованих даних, реальні програми, App Store, Google Play.

Introduction

Currently, virtual, augmented and mixed reality technologies can be increasingly found in various areas of human life and even in everyday use. These technologies are already actively used in such fields as education, medicine, marketing and advertising, automotive industry, and navigation. Therefore, involving these technologies in application development is a relevant task. Object detection has a crucial part in various applications, from computer vision to autonomous vehicles. One of the key factors that influence the accuracy and robustness of an object detector is the quality of the training data. In our previous works we proposed the application of augmented reality for navigation purposes [1] and for Objects 3D Models visualization using Augmented Reality [2].

In this paper, we will explore several structured ways to build an object detector data source, taking inspiration from Apple's Create ML documentation on the topic. Additionally, we will showcase real-world applications available on both the App Store and Google Play that leverage object detection technology.

Domain analysis

Object detection is a fundamental task within the domain of computer vision and artificial intelligence. It finds application in various industries and domains, driving innovations and advancements in safety, efficiency, and user experiences [3-5].

Autonomous Vehicles: In the domain of autonomous vehicles, object detection plays a pivotal role in enabling self-driving cars to perceive their surroundings and make real-time decisions. The ability to detect pedestrians, vehicles, traffic signs, and obstacles is critical for ensuring safe and reliable autonomous navigation.

Surveillance and Security: Surveillance systems heavily rely on object detection to monitor and identify suspicious activities, intruders, or potential threats in real-time. This technology has significant implications in enhancing security measures in public spaces, airports, and critical infrastructure.

Augmented Reality: Object detection is a core technology in augmented reality (AR) applications. By recognizing and tracking objects in the real world, AR systems can overlay virtual objects or information seamlessly, enriching user experiences in gaming, education, and marketing.

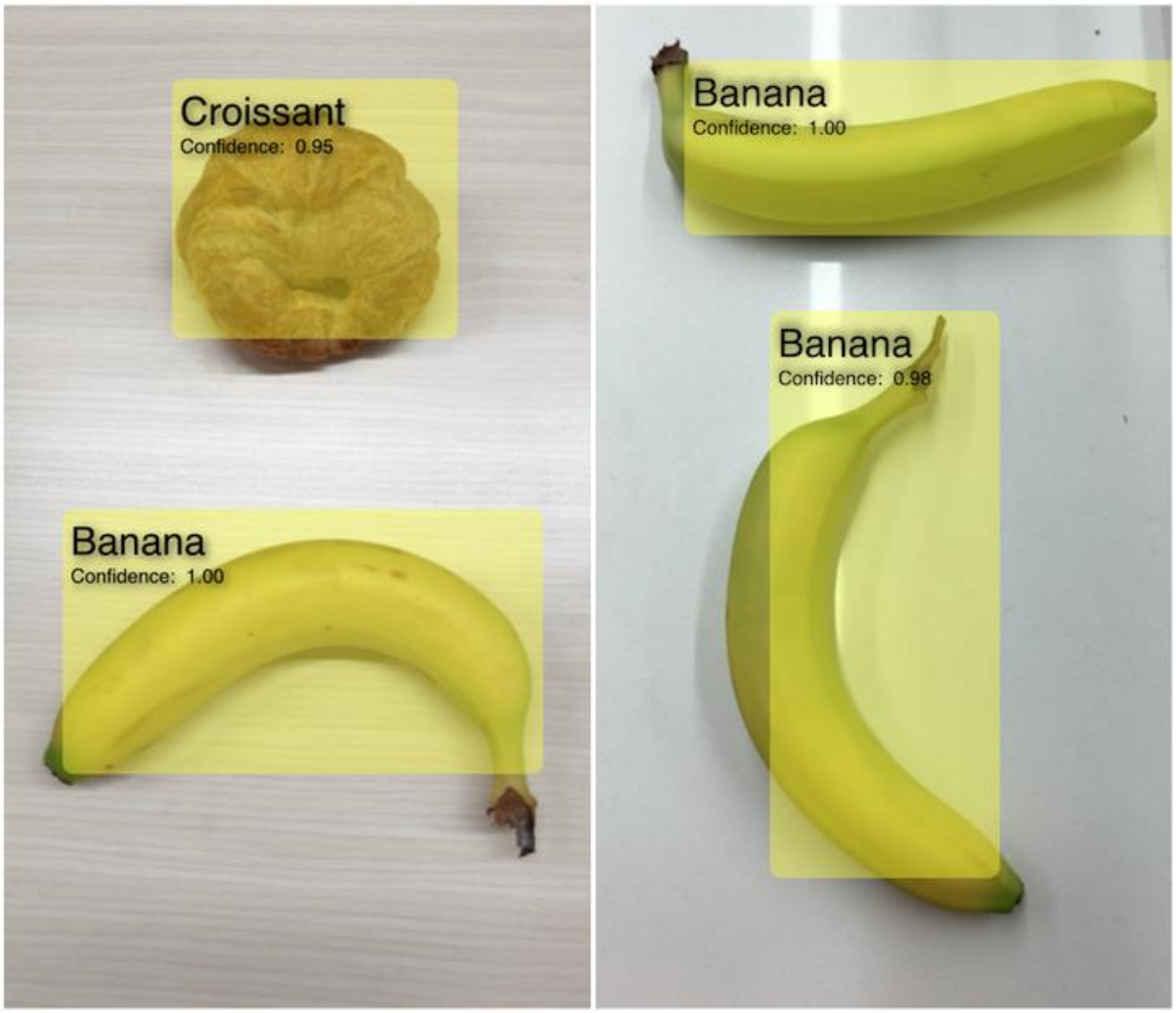

Fig. 1. Bananas detection

Medical Imaging: In medical imaging, object detection is utilized to identify and locate anomalies or specific structures within medical images, assisting in diagnosis and treatment planning. Applications range from detecting tumors in MRI scans to identifying cells in microscopy images.

E-commerce and Retail: In e-commerce and retail, object detection facilitates various applications such as visual search, product recommendation, and inventory management. By recognizing objects in images or videos, retailers can offer personalized shopping experiences and optimize supply chain operations.

Robotics: In robotics, object detection is crucial for enabling robots to interact with their environment and perform tasks autonomously. From industrial robots identifying objects on assembly lines to service robots navigating in dynamic environments, object detection enhances the capabilities of robotic systems.

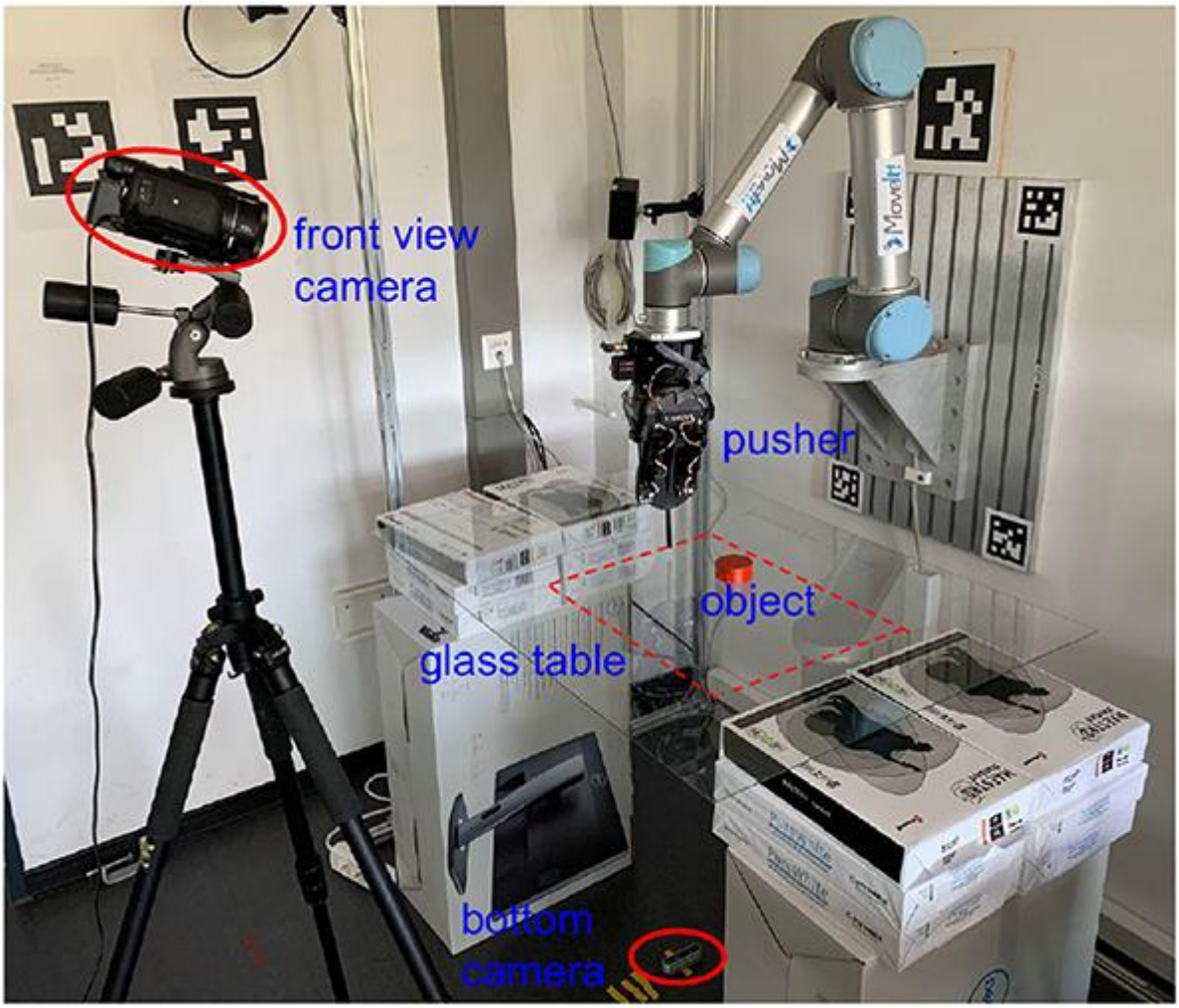

Fig. 2. Robot detects the objects and works with them

Accessibility: Object detection technologies have also been leveraged to enhance accessibility for individuals with visual impairments. By identifying and describing objects in real-time, these applications empower users to navigate their surroundings more effectively.

Understanding the specific domains and applications of object detection helps researchers, developers, and practitioners tailor their approaches and data collection methods to address the unique challenges and requirements of each use case. The quality and relevance of the training data are paramount in achieving accurate and efficient object detection systems across diverse applications.

Analysis of existing solutions and technologies

Object detection technology has found its way into numerous real-world applications, revolutionizing various industries and enhancing user experiences. Leveraging computer vision and AI algorithms, these applications have demonstrated the practicality and effectiveness of object detection in addressing real-life challenges. In this section, we will explore three prominent real-world applications that utilize object detection to provide valuable insights, improve accessibility, and transform the way we interact with our surroundings.

Application "Google Lens" (Available on Google Play). Google Lens is an innovative application that leverages object detection to provide users with instant information about the world around them. By simply pointing the smartphone camera at objects or scenes, Google Lens can identify landmarks, plants, animals, and a wide range of everyday objects. The app then delivers relevant search results, detailed information, and even language translation based on the recognized objects. Whether users are exploring a new city, trying to identify a particular species of flora or fauna, or need to understand foreign language text, Google Lens offers an intuitive and interactive experience. By incorporating object detection technology, Google Lens has revolutionized how we interact with the environment, making information readily available at our fingertips.

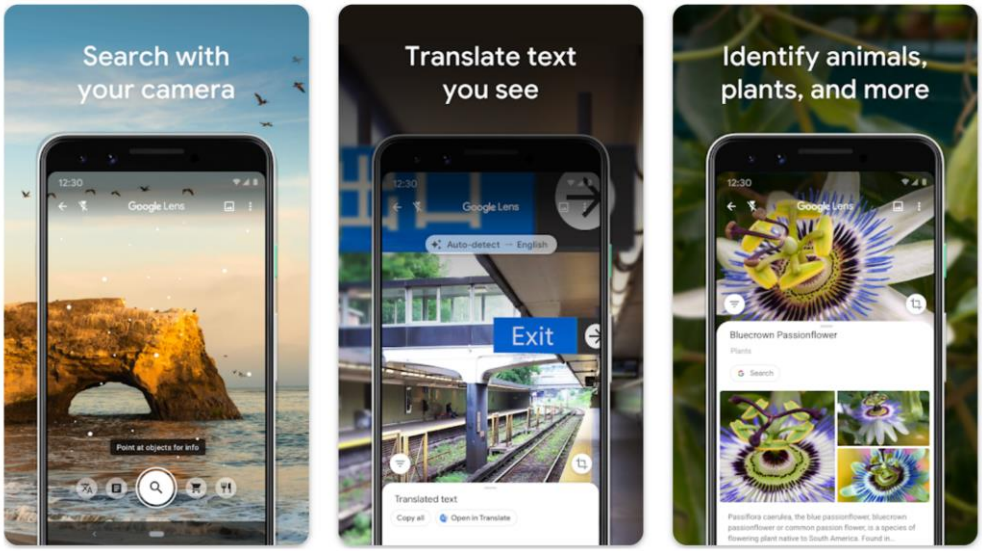

Fig. 3. Google lens in Google Play

Application "CamFind" (Available on App Store and Google Play). CamFind is a powerful image recognition app that utilizes object detection to identify objects captured in photos. Whether it's household items, fashion accessories, or artworks, CamFind can accurately recognize and classify objects, allowing users to learn more about them. The app provides detailed information, shopping links, and related content based on the identified objects. CamFind's seamless integration of object detection technology has bridged the gap between the physical and digital worlds, enabling users to explore and discover information effortlessly. From identifying unfamiliar objects to finding similar products online, CamFind exemplifies the practical applications of object detection in everyday life.

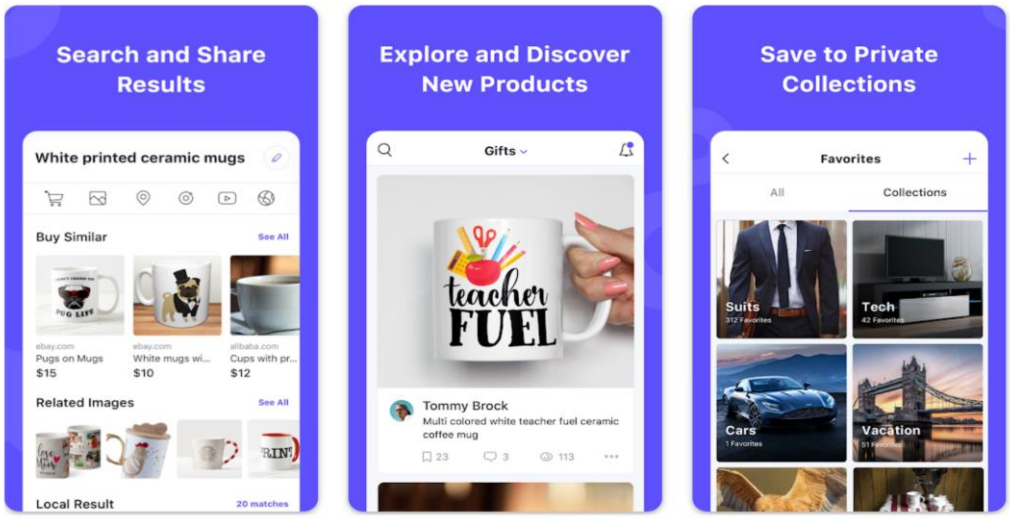

Fig. 4. CamFind in Google Play

Application "TapTapSee" (Available on App Store and Google Play). TapTapSee is an exceptional accessibility application that utilizes object detection to empower visually impaired users. By capturing images through their smartphone cameras, users can rely on TapTapSee to identify objects and receive auditory descriptions in real-time. The app is capable of recognizing a wide range of objects, including everyday items, products, and even scenes. By incorporating object detection, TapTapSee enhances the independence and confidence of visually impaired individuals, enabling them to navigate their surroundings more effectively. This application exemplifies the transformative impact of object detection in promoting inclusivity and accessibility.

Fig. 5. TapTapSee in AppStore

Now let’s compare these three apps by platform availability, functionality and benefits for users.

Application	Platform	Main Functionality	Key Benefit
Google Lens	Google Play	Instant object identification	Provides detailed information and search results based on recognized objects
CamFind	App Store,	Image recognition and object classification	Offers detailed information and shopping links for identified objects
TapTapSee	App Store,	Object recognition for visually impaired	Provides auditory descriptions of objects in real-time to enhance accessibility for users with visual impairments

Real-world applications using object detection technology have showcased its versatility and significance in various domains. From the interactive and informative experience offered by "Google Lens" to the seamless shopping and discovery facilitated by "CamFind," object detection has revolutionized how we interact with the world around us. Moreover, applications like "TapTapSee" have demonstrated the profound impact of object detection in enhancing accessibility for visually impaired individuals. As the technology continues to advance, we can expect even more innovative and transformative applications leveraging object detection to address real-life challenges and improve user experiences across diverse industries.

Structured Ways of Building an Object Detector Data Source

Object detection models heavily rely on high-quality training data for accurate and reliable performance. Building a structured object detector data source involves various approaches and techniques to ensure the dataset's diversity, precision, and scalability. One of the primary methods is hand-annotated datasets, where human experts meticulously label objects in images. Although this process can be time-consuming, it offers precise annotations, making it suitable for specialized use cases. To streamline the manual annotation process, developers can utilize tools like Labelbox, RectLabel, and VoTT.

Another effective strategy for building a robust dataset is leveraging pre-trained datasets such as COCO, ImageNet, and PASCAL VOC for transfer learning. By starting with pre-trained models and fine-tuning them on specific datasets, developers can significantly speed up the training process and improve object detection performance. Create ML, a popular tool in the Apple ecosystem, supports this technique, making it accessible to developers.

Data augmentation techniques also play a crucial role in enhancing the diversity of the dataset. By applying various transformations like rotation, flipping, scaling, and color changes, developers can create multiple variations of the original data, thereby improving the detector's generalization and reducing overfitting.

In some cases, real-world data may not be sufficient to cover all possible scenarios. Here, synthetic data generation comes into play. Developers can create synthetic data using 3D rendering engines or generative models,

allowing them to control object placements, backgrounds, and lighting conditions. This approach helps address the limitations of real-world data collection and further augments the dataset's diversity.

To optimize the training dataset and reduce annotation efforts, active learning strategies can be implemented. Active learning involves selecting the most informative samples for manual annotation, effectively focusing the annotation efforts where it matters the most. By doing so, developers can achieve higher detection accuracy while minimizing the annotation workload.

To ensure the dataset remains up-to-date and relevant, it is crucial to follow best practices for maintaining a structured object detector data source. Regularly updating the dataset based on new scenarios, edge cases, and emerging trends is vital for adapting the object detector to changing real-world conditions.

In conclusion, building a structured object detector data source involves a combination of various techniques and strategies. Hand-annotated datasets, transfer learning with pre-trained models, data augmentation, synthetic data generation, active learning, and semi-supervised learning all contribute to creating a comprehensive and diverse training dataset. Following best practices for dataset maintenance ensures the object detector remains accurate and performs effectively in real-world applications.

Creating a model for detecting a sport games surfaces using the Action & Vision App

In the realm of modern technology, the Action & Vision app stands as a remarkable testament to the synergistic capabilities of Vision and Core ML technologies. By seamlessly integrating these powerful tools, the app ventures into the intricate world of sports environments, deciphering player movements and interactions with objects with unprecedented precision.

Central to this innovative endeavor is the development of an app that not only captures sports scenes but also comprehends them. The app's underlying framework involves the meticulous curation of a diverse image dataset, meticulously annotated to facilitate the training of a robust model. This pivotal phase is elevated with the assistance of Roboflow, a specialized service that streamlines data collection and annotation, ensuring the model's accuracy and efficacy.

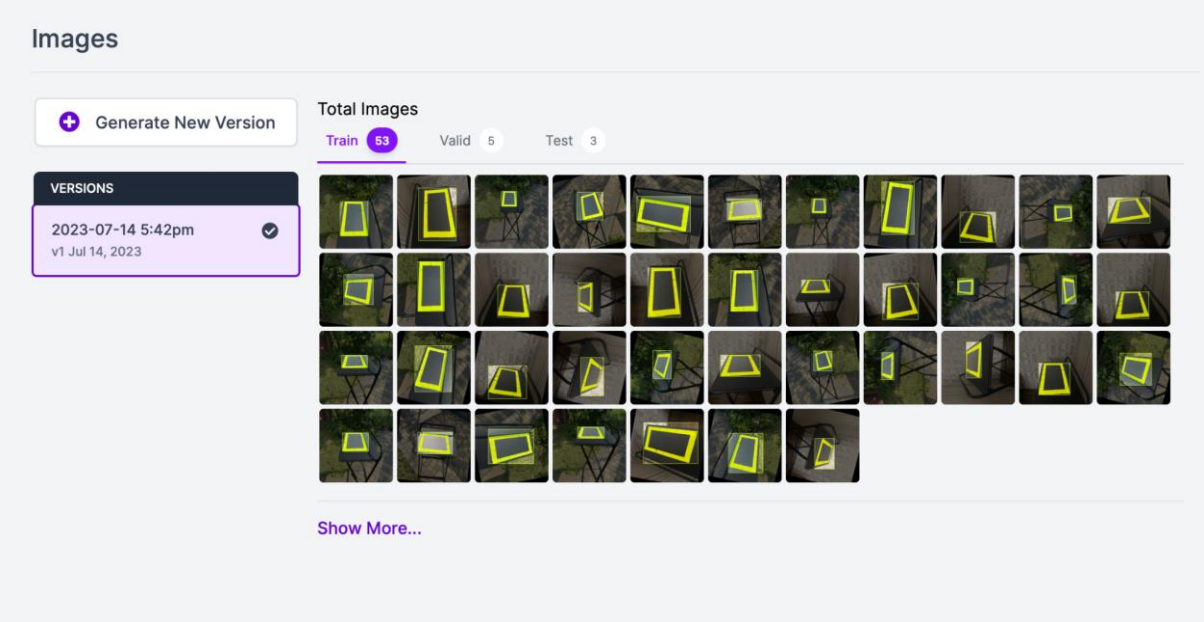

Fig 6. Annotating Images on Roboflow

Once the dataset is curated and annotated via Roboflow, it is transformed into a comprehensive training dataset. This dataset is then utilized to build a robust detection model, tailored to function seamlessly with Vision and Core ML frameworks on iOS devices.

The core functionality of the app hinges upon its ability to perform real-time scene analysis and accurately identify tables and rectangles within the video frames. This is achieved through the intelligent application of VNCoreMLRequest, a Vision framework component. The trained object detection model, informed by the diverse dataset and robust training process, is employed to identify and outline tables and rectangles in each frame of the video stream.

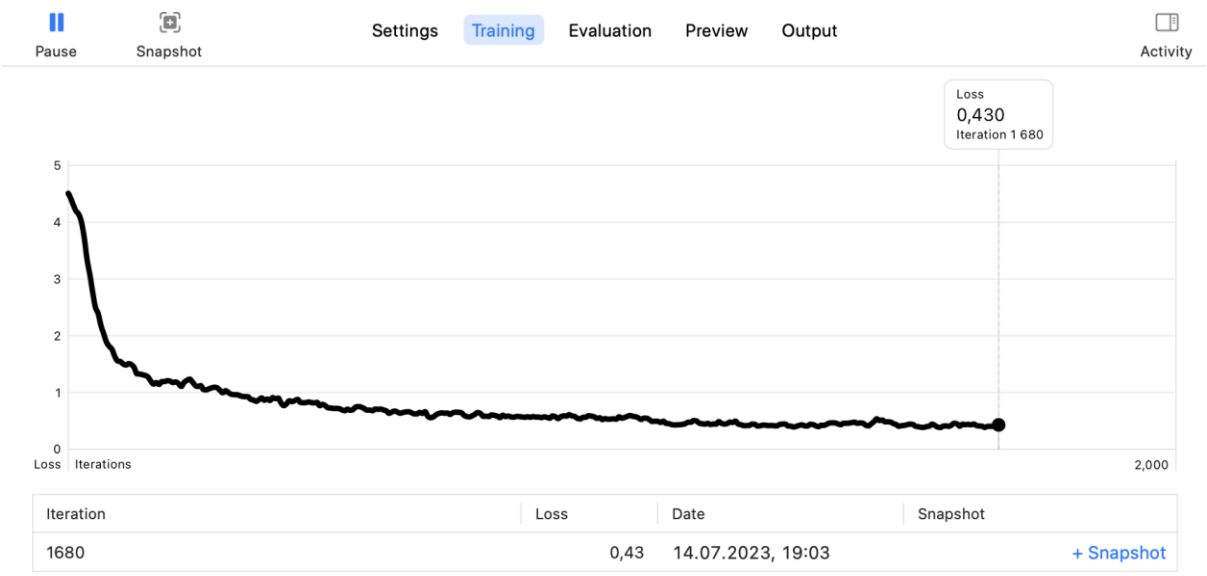

Fig 7. Generating Model with Create ML

Subsequent to the successful generation of the detection model, the app harnesses the capabilities of Core ML Preview mode. This feature enables the integration of the model into the application and allows for real-time testing and validation of the model's performance on sample video frames. This iterative testing process ensures the model's accuracy and reliability in identifying tables and rectangles across various scenarios.

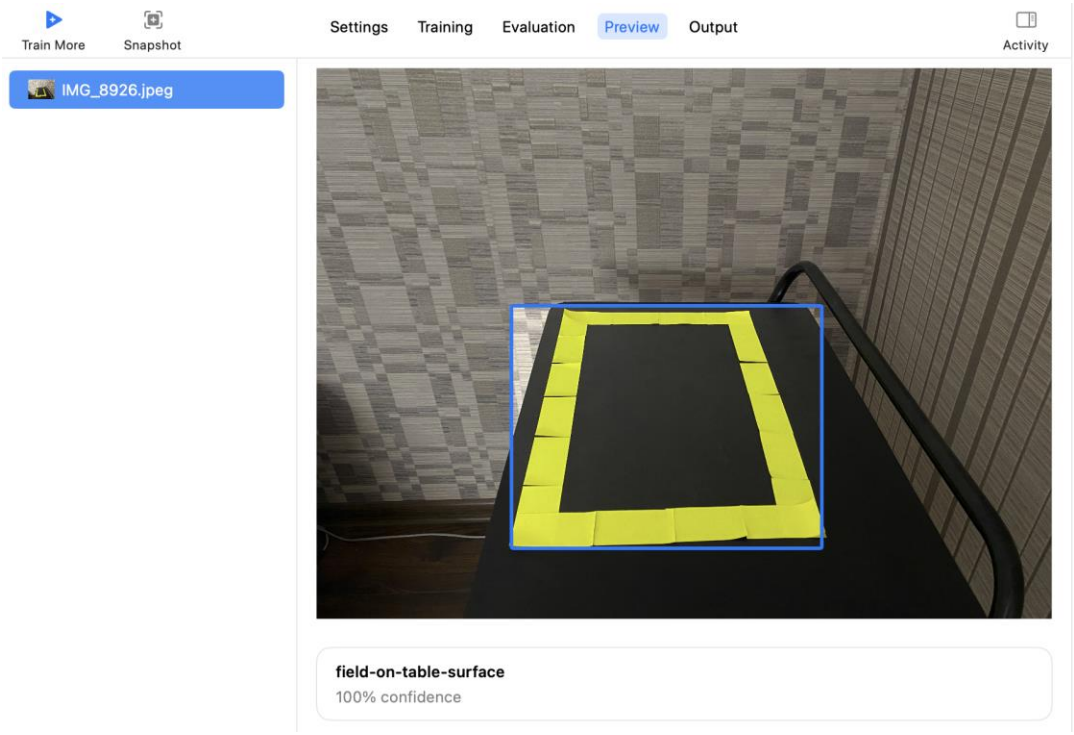

Fig 8. Using Core ML Preview Mode

Fig 9. Using the model in the iOS app

In essence, the integration of machine learning models via Core ML within the Action & Vision app showcases the profound impact of Vision and Core ML on advancing object detection capabilities. This marriage of technologies not only augments the functionality of the app but also serves as a testament to the transformative potential of machine learning and computer vision in the modern technological landscape.

Experiments and directions of further work

As object detection technology continues to advance, there are several directions for future work to enhance the capabilities and applications of object detectors. First, exploring more advanced data augmentation techniques and generative models can further improve dataset diversity and expand the training data pool.

Additionally, the integration of domain adaptation methods can help improve the generalization of object detectors to new environments and domains. This is particularly relevant in scenarios where the training data may differ significantly from the test data.

Furthermore, research on active learning strategies tailored specifically for object detection can lead to more efficient annotation processes and better utilization of limited labeled data.

Lastly, investigating methods to address potential biases in the training data and ensuring the fairness and ethics of object detection models is essential, especially in applications with high social impact.

By continuously exploring and advancing these areas, object detection models can continue to evolve and provide valuable insights and solutions in a wide range of real-world applications.

Conclusions

In our constantly developing world virtual, augmented, and mixed reality technologies are becoming integral parts of our daily lives. The study has highlighted the significance of object detection in the domains of computer vision and AI, as it empowers machines to identify and locate objects in images or video frames. The accuracy and performance of object detectors are intricately linked to the quality and diversity of the training data, making the establishment of a reliable data source a critical endeavor.

Through a systematic exploration of methodologies and insights derived from Apple's Create ML documentation, this research has successfully constructed structured approaches for building a real-world object detector data source. Additionally, the paper has shed light on various real-world applications that leverage object detection technology, available on platforms like the App Store and Google Play. The culmination of this work is the development of a robust object detection model, tailored to seamlessly integrate with Vision and Core ML frameworks on iOS devices. This model excels in identifying and outlining tables and rectangles within video streams, particularly in the context of sports game surfaces. The practical implications of this research extend beyond the academic realm and pave the way for the creation of a mobile application, the Action & Vision App, which will simulate sports games. This application holds the potential to revolutionize the way players practice and enhance their skills, extending the boundaries of traditional training fields.

References

1. Pavlova, O., Bashta, A., Kravchuk, S., Hnatchuk, Y., Bouhissi, H.E. Augmented Reality Based Technology and Scenarios For Route Planning and Visualization. CEUR Workshop Proceedings, 2022, 3156, pp. 613–623
2. Pavlova O., Bashta A., Kostiuk M. and El Bouhissi H. Technology and Scenarios for Objects 3D Models visualization using Augmented Reality CEUR Workshop Proceedings, 2023, 3373, (IntelITSIS, Khmelnytskyi, March 23-25) pp. 343-353
3. Kang, J., Tariq, S., Oh, H., & Woo, S. S. (2022). A survey of deep learning-based object detection methods and datasets for overhead imagery. IEEE Access, 10, 20118-20134.
4. Markov, T., Zhang, C., Agarwal, S., Nekoul, F. E., Lee, T., Adler, S., ... & Weng, L. (2023, June). A holistic approach to undesired content detection in the real world. In Proceedings of the AAAI Conference on Artificial Intelligence .Vol. 37, No. 12, pp. 15009-15018
5. VS, V., Oza, P., & Patel, V. M. (2023). Towards online domain adaptive object detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 478-488.

Olga Pavlova **Ольга Павлова**	PhD, Associated professor of Computer Engineering & Information Systems Department, Khmelnytskyi National University, Khmelnytskyi, Ukraine, e-mail: olya1607pavlova@gmail.com https://orcid.org/0000-0001-7019-0354	доктор філософії, доцент кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, Хмельницький, Україна.
Andriy Bashta **Андрій Башта**	PhD Student of Computer Engineering & Information Systems, Khmelnytskyi National University, Khmelnytskyi, Ukraine, e-mail: andreybashta@gmail.com https://orcid.org/0000-0002-0775-1347	аспірант кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, Хмельницький, Україна.
Andrii Kuzmin **Андрій Кузьмін**	Master Student of Computer Engineering & Information Systems, Khmelnytskyi National University, Khmelnytskyi, Ukraine, e-mail: andriy1731@gmail.com	студент-магістр кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, Хмельницький, Україна.