Computer Vision: When Machines See What Humans Don’t

Carla Gabriela Rădulescu
Informatics
IMC Krems University of Applied Sciences
Krems an Der Donau, Austria
23IMC10633@fh-krems.ac.at

***Abstract*—This paper aims to introduce computer vision with both positive and negative aspects. The ”Introduction” provides an overview of the field, offering definitions and examples related to artificial intelligence and machine learning. Along the way, everyday applications such as self-driving cars and medical imaging stand out. These topics are then examined in the “Related Work” section, which combines a literature review and my own exploration of four data sets. The “Background” section provides historical context for the development of the field. Finally, the paper justifies the chosen approach, answers the research questions and presents the conclusions drawn from the analysis.**

***Index Terms*—Computer Vision (CV), Artificial Intelligence (AI), Machine Learning (ML), Supervised learning, Unsupervised learning, Deep learning, Convolutional Neural Networks (CNNs), Autonomous driving, Algorithm, Dataset, Medical imaging, learning, object detection, image classification, image processing, edge cases, data augmentation, bias, LIDAR**

I. Introduction

One of the most important senses that we possess as humans is vision. From the moment we are born, the ability to see the world is one of our first developed capabilities. Through our eyes, we gather an immense amount of information effortlessly, without the need to learn how. It simply comes naturally as we explore and observe our surroundings.

The same is true for animals. However, when it comes to computer vision (CV), the process of perceiving what is around us is far more complex and entirely dependent on a vast amount of data.

A. *What is Computer Vision?*

Computer vision is a field of study within **artificial intelligence (AI)** that uses **machine learning (ML)** and **neural networks** to allow computers to get a better understanding of reality. The main goal is to extract meaningful information from visual data to support tasks such as decision making. If AI allows computers to process and reason, computer vision helps them perceive, interpret, and comprehend images and videos. [1]

Visual perception technology bridges the gap between **raw visual input** and **actionable insights**, all inspired by human visual perception. [2] This domain is of great importance and appears in numerous real-world applications ranging from facial recognition to medical diagnostics.

Before exploring these applications in detail, it is essential to explain the behind-the-scenes process and introduce how machine learning and neural networks serve as the foundation for modern computer vision.

B. *About Machine Learning*

Machine learning is a rapidly advancing subset of AI that enables systems to **learn patterns** from data rather then **relying on hard-coded rules**. Both **supervised learning** and **unsupervised learning** are essential in computer vision. The key difference is that supervised learning involves training models with labeled data to predict or classify new instances (e.g. using techniques like regression and classification), whereas unsupervised learning discovers hidden structures in data without explicit labels (e.g. clustering or dimensionality reduction). ML plays a critical role in computer vision tasks such as **image classification**, **object detection**, and **feature extraction**. [2]

Figure 1 illustrates the workflow of supervised learning, showing how labeled inputs and training labels are used to train a model to predict or classify new data.

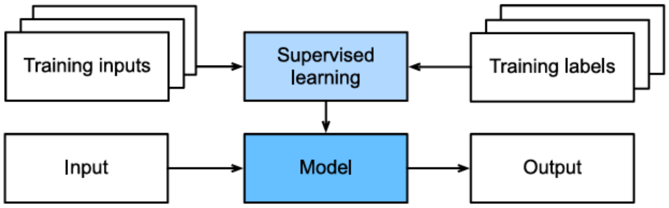

Fig. 1. Supervised learning workflow, showing the relationship between inputs, labels, the model, and outputs. Adapted from [2] and [5].

In this paper, the focus will be more on supervised learning methods, especially classification, as these align with the planned exploration of future datasets.

The first question would be: how does a machine ”learn”? Unlike a child learning from experience, ML models **identify patterns** and **adjust based on feedback**.

Imagine a situation where a child sees a cat for the first time. They may not immediately understand what it is. But after seeing several cats - each with different colors, fur patterns, and sizes - they begin to recognize common features. The child may begin to associate fur, whiskers, and pointed ears with the term “cat.” Over time, their brain refines their understanding, allowing them to identify a cat even in a new situation.

Similarly, ML models train on vast datasets, gradually improving their ability to distinguish between objects, such as identifying an animal regardless of changes in lighting, angle, or background.

C. *Deep Learning and Neural Networks*

Deep learning is an essential technology in computer vision and a subset of machine learning. It uses **multilayered neural networks** to mimic human brain processes such as making decisions. [3]

Neural networks are computational models inspired by the brain, consisting of interconnected nodes, which plays the role of neurons, that **process information in layers**. Deep learning extends this by stacking multiple layers, often comprising hundreds or thousands, to enable increasingly complex pattern recognition. [3]

Consider again the example of a child learning to recognize a cat. Initially, the child might associate a cat with the first one they see, perhaps a small, white cat in a park. As they encounter more cats of different sizes, colors, and fur patterns, their brain forms deeper, more refined connections, allowing them to identify a cat even if they see an entirely new breed.

Similarly, deep learning models improve their accuracy by training on large and diverse datasets. Unlike traditional machine learning models, which rely on predefined rules, deep neural networks adjust their internal representations across multiple layers, enabling the recognition of complex features such as shapes, textures, and spatial relationships. [3]

Deep learning is pivotal in two applications that this scientific paper explores: **autonomous driving**, where it allows vehicles to detect objects, lane markings, and road signs; and **medical imaging**, where it aids in diagnosing diseases by identifying tumors and abnormalities. [2], [3]

D. *Convolutional neural networks*

Convolutional neural networks (CNNs) are a subset of deep learning designed to **process and analyze visual data**. Unlike traditional neural networks, CNNs use convolutional layers to automatically extract spatial features from images. Their architecture includes **convolutional layers**, **pooling layers**, **activation functions**, and **fully connected layers**, which allow them to recognize complex visual patterns efficiently. [4]

Due to their hierarchical feature-learning capability, CNNs **outperform traditional machine learning models** in complex visual tasks: **image classification**, **object detection**, and **video analysis**. Despite their success, challenges such as **computational efficiency** and **data requirements** remain areas of ongoing research. [4]

Just as a child gradually learns to recognize a cat by observing various features, a CNN helps machines learn to “see” by analyzing images in layers. Instead of viewing an image as a whole, a **CNN breaks it down into small parts**, identifying patterns such as edges, shapes, and textures. Through repeated training, it refines its ability to recognize objects, much like how human perception improves with experience.

E. *Applications and Challenges*

Now that the process behind computer vision is clearer, I want to mention its wide range of real-world applications. Everyday examples include facial recognition, self-driving cars, medical diagnostics, surveillance, augmented reality, robotic automation, sports performance analysis, and agricultural monitoring. The list continues to expand as technology advances.

In this paper, I will focus on two key applications mentioned earlier: autonomous cars and medical imaging. However, first, I would like to highlight two additional interesting examples.

Facial recognition is a fascinating feature commonly used in daily life, such as unlocking our smartphones. In some cases, for example in China, it is widely used for public surveillance, identity verification, and even financial transactions. While the widespread implementation of this technology demonstrates its potential, it also raises significant ethical and privacy concerns.

Understanding visual complexity in **urban streetscapes** is another growing research area. A recent study examined how humans and computer vision systems perceive complexity in urban environments. By analyzing geo-referenced photographs through both human surveys and computer vision models, researchers found that visual complexity is influenced by factors such as contrasting colors and sharp edges. These insights have practical applications in urban design, helping to reduce visual pollution caused by road signage, advertising, and telecommunication systems. [6]

Although computer vision has advanced significantly, one of the most intriguing challenges remains: **What makes vision such a difficult task for machines?** Unlike computer graphics, where we start with a predefined model and render an image, computer vision works in reverse: starting from raw data and attempting to infer meaningful information. The inverse problem introduces uncertainty and complexity. As an example, “Consider a robot trying to estimate the distance to an obstacle: It is usually safer to underestimate than to overestimate.” [2] Reconstructing the world - its object shapes, colors, lighting, and even beyond - makes vision a profoundly challenging task for machines, despite its intuitive ease for humans. [2]

Consider one more time the example with a child learning to recognize a cat. Over time, after encountering diverse cats, they refine their understanding and can generalize across variations. In contrast, a machine must be trained on thousands of labeled images to develop similar recognition capability. Even then, minor changes in lighting, angle, or background can confuse the model, demonstrating the challenge of replicating human perception in artificial systems.

II. Related Work

This section details two notable applications of computer vision: **autonomous vehicles** and **medical imaging**. I will begin by examining autonomous cars, highlighting their benefits and challenges, with a focus on LIDAR technology and its underlying mechanisms. Following this, I will explore the role of computer vision in the medical field, analyzing its significance and addressing its challenges. Finally, I will conduct an analysis of publicly available datasets from Kaggle [1] related to these innovations, providing insights and visualizations along the way to support the discussion.

A. *Autonomous vehicles*

Whether a pleasure or a necessity, cars are integral to our daily lives. With advancements in technology, they have transformed dramatically, offering faster, safer, and more convenient transportation. Yet, for some, driving can be stressful or inaccessible due to disabilities or other constraints. The advent of self-driving cars has promised to address these challenges, sparking both excitement and skepticism.

The following studies provide valuable insights into the advancements and challenges of autonomous vehicles, focusing on traffic management, inference frameworks, and LIDAR integration.

1) **Intelligent Vehicle Violation Detection System**
Traffic violations remain one of the leading causes of accidents worldwide. Addressing this issue, the study ”Intelligent Vehicle Violation Detection System Under Human-Computer Interaction and Computer Vision” [9] proposes a robust system that detects violations in real-time using computer vision.
The system employs the **YOLOv4 algorithm** for vehicle detection, specializing in small objects like license plates. Trajectory tracking is enhanced using the **Kalman filter**, which predicts motion accurately. A key innovation is the **human-computer interaction interface**, designed to simplify usability for law enforcement agencies.
Datasets such as BIT-Vehicle and AppoloScape were used, containing over 49,000 annotated images of traffic scenarios. Preprocessing techniques, including **mean filtering** and **histogram equalization**, improved image quality, boosting detection accuracy to 96.86%. This system showcases the potential of computer vision to improve road safety and traffic management.

2) **Mastering Computer Vision Inference Frameworks**
Autonomous vehicles rely on efficient inference frameworks to process real-time data, making optimization a critical area. In their study, Pochelu and Castro-Lopez [10] conducted an empirical evaluation of multiple inference frameworks across hardware setups like **Tesla V100** and **Ampere A100 GPUs**.
Their analysis highlights the trade-offs between speed, memory consumption, and power efficiency. For instance, frameworks like TensorRT outperformed others in prediction speed, while ONNX-RT was more memory efficient. These findings highlight the importance of adapting inference frameworks to specific hardware for autonomous driving applications, where real-time processing is essential.
The study emphasizes the importance of **post-training optimization**, which can tailor models to specific hardware requirements. Results from their benchmarks underline the need for optimized inference frameworks to ensure the seamless operation of self-driving systems.

3) **LIDAR and Camera Fusion for 3D Terrain Reconstruction**
To navigate complex environments, autonomous vehicles must perceive their surroundings accurately. The study ”Global Terrain Registration of LiDAR and Camera Fusion Using Multiple Calibrators” [11] addresses this challenge by combining **LIDAR** and camera data for **3D terrain reconstruction**.
LIDAR sensors offer precise distance measurements, while cameras capture rich texture and color information. By combining these modalities, the system achieves high-accuracy perception. Techniques such as **RANSAC** and loop closure detection enhance calibration and optimize point cloud registration.
Applications of this research extended beyond autonomous vehicles to fields like cultural heritage preservation and gaming. The findings demonstrate that integrating LIDAR and camera technologies significantly improves the accuracy of 3D perception, enabling safer navigation in autonomous systems.

TABLE I
BENEFITS AND CHALLENGES OF AUTONOMOUS VEHICLES

Source	Benefits	Challenges
Source 1 [9]	Real-time traffic violation detection; improved road safety; automation reduces human intervention.	Requires high-quality datasets; extensive preprocessing needed for accuracy.
Source 2 [10]	Efficient data processing; optimized speed, memory, and power consumption.	Hardware-specific trade-offs; post-training optimization is critical.
Source 3 [11]	Enhanced 3D perception; precise navigation in complex environments.	Requires advanced calibration and high computational power.

B. *Medical imaging*

Medicine has advanced rapidly alongside technology. Health remains one of the most important aspects of life, and we are fortunate to live in an era where technological advancements allow for a better understanding of our bodies and improved treatment of diseases. Computer vision has become a powerful tool in this field, with applications spanning neuroscience, skin disease diagnosis, and deep learning techniques. Below, I explore some key contributions from recent research:

1) **Neuroscientific Insights into Computer Vision**
Seba Susan’s study [12] investigates how neuroscience has influenced the development of computer vision. Early work on **artificial neurons** laid the foundation for architectures like CNNs, and biologically inspired

[1] Available: https://www.kaggle.com/

networks like VisNet and Beta-VAE aim to mimic the brain's visual pathways. These models are designed to be adaptive and efficient, inspired by the way human vision works.

Insights from neuroscience continue to shape the creation of biologically inspired neural networks, helping to make models more efficient and robust.

2) **Computer Vision in Skin Disease Diagnosis**

The survey by Gupta et al. [13] delves into computer vision's role in diagnosing various skin conditions, such as melanoma, psoriasis, and fungal infections. It explores image preprocessing, segmentation, feature extraction, and classification techniques, while also addressing challenges like variability in skin images, data annotation, and imbalanced datasets.

By proving diagnostic accuracy and accessibility, computer vision holds great promise for dermatology, though challenges like image variability and limited data still need to be addressed.

3) **Edge Deep Learning In Medical Diagnostics**

Xu et al.'s research [14] explores how edge computing and deep learning can enhance medical diagnostics. Lightweight models and compression techniques allow real-time processing on edge devices. The study also emphasizes the benefits of **edge deep learning**, such as reduced latency, lower bandwidth usage, and better data privacy.

Edge deep learning has the potential to transform medical diagnostics by enabling efficient real-time analysis, even in resource-limited environments.

TABLE II
BENEFITS AND CHALLENGES OF COMPUTER VISION IN MEDICAL IMAGING

Source	Benefits	Challenges
Source 1 [12]	Inspired neural networks with adaptive and efficient designs based on biological principles.	Limited success in fully replicating the complexity of biological vision.
Source 2 [13]	Improved accuracy and accessibility in diagnosing skin diseases using computer vision techniques.	Issues with skin image variability, annotation, and class imbalance.
Source 3 [14]	Real-time processing on edge devices with benefits like reduced latency and data privacy.	Resource constraints and the need for efficient lightweight models.

C. Datasets analysis

Datasets play a critical role in training supervised learning models, especially in the areas that have been previously analyzed. In this section, I examine four datasets to highlight their diverse applications, potential benefits and challenges.

The analysis begins with two datasets related to autonomous driving, focusing on traffic analysis and sign detection, showcasing their importance in enhancing transportation systems and autonomous vehicle technology. The discussion then shifts to two medical datasets, one focused on skin condition classification and another on detecting bone fractures, emphasizing their potential for advancing healthcare diagnostics and treatment.

The first dataset, **"The City Intersection dataset"** created by Mohamadreza Momeni[2], is a valuable resource for analyzing traffic scenarios at urban intersections. It contains 1321 training images, 392 validation images, and 189 test images, organized into three categories. The dataset is designed for applications like traffic flow analysis, city transportation planning, autonomous vehicle development, and vehicle-based advertising.

A possible use of this dataset is to train a machine learning model to classify vehicle types or analyze traffic patterns. In my analysis, I trained a CNN model to classify images into three classes. The model reached a maximum validation accuracy of approximately 54% after 10 training epochs. However, fluctuations in validation accuracy suggest challenges like potential overfitting, class imbalance, or limited data augmentation (Figure 2).

This dataset highlights the possibilities for advancing autonomous vehicle systems and traffic monitoring solutions. To further improve model performance, techniques such as leveraging pre-trained models, applying advanced augmentation strategies, or addressing dataset-specific challenges (e.g. class distribution) could be considered.

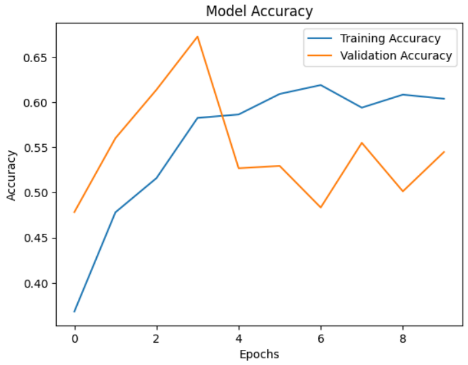

Fig. 2. Training and validation accuracy across 10 epochs

The model's loss behavior over training process is shown in (Figure 3). While the training loss decreases steadily, the validation loss fluctuates, further indicating potential overfitting or noise in the data.

The second dataset, **"The Car Detection dataset"** made by Parisa Karimi Darabi [3], is a comprehensive collection of traffic sign images designed for tasks like autonomous vehicle navigation and traffic rule compliance. The dataset contains 3530 training images, 801 validation images, and 638 test images, organized into 15 categories: Green Light, Red Light,

[2] Dataset available at https://www.kaggle.com/datasets/imtkaggleteam/city-intersection-computer-vision

[3] Dataset available at https://www.kaggle.com/datasets/pkdarabi/cardetection/data

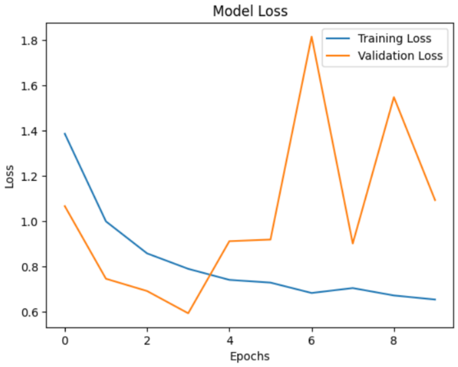

Fig. 3. Training and validation loss across 10 epochs

Stop, and various Speed Limit signs ranging from 10 to 120 km/h.

In my analysis, I trained a CNN to classify images into these 15 classes. The model achieved a maximum validation accuracy of approximately 40% after 10 training epochs. The training and validation accuracy trends are shown in Figure 4. Even though the model showed consistent improvement in training accuracy, the validation accuracy indicated possible challenges, such as class imbalance or insufficient data augmentation.

The loss behavior during training, depicted in Figure 5, reveals a gradual reduction in training loss, while the validation loss fluctuates, indicating potential overfitting or noisy data. To improve model performance, similar techniques to those mentioned earlier for the first dataset can be explored.

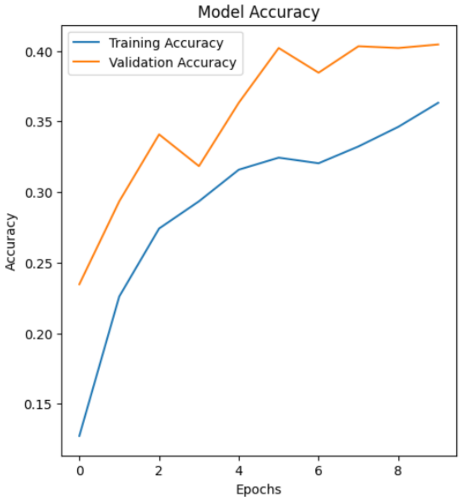

Fig. 4. Training and validation accuracy across 10 epochs

This dataset highlights the potential for advancing traffic sign recognition models, contributing to safer roads and more efficient autonomous vehicle systems.

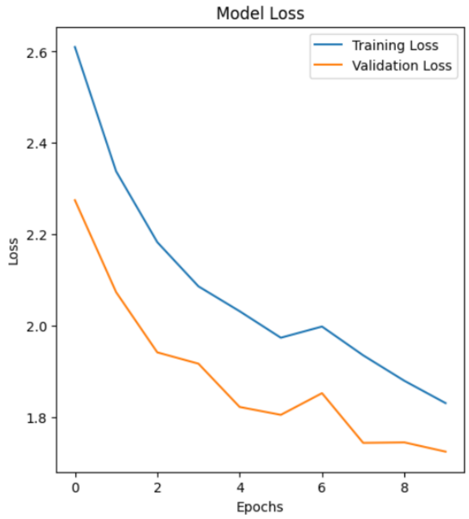

Fig. 5. Training and validation loss across 10 epochs

The third dataset, ”The Acne Dataset” [4], consists of 2436 annotated images, categorized into classes such as acne, pimples, and spots. The dataset is organized into three subsets: training, validation, and test sets, with a notable imbalance in the number of images across these subsets (Figure 6). It is noticeable that the training set contains the majority of the data, while the validation and test sets are relatively smaller.

This dataset provides a foundation for developing models to classify and detect skin conditions. The annotations were carefully refined over several iterations to capture six key classes, including acne, pimples, spots, moles, and scars. Specific annotation rules were followed to ensure consistency, such as excluding images with excessive freckles or ambiguous features.

This dataset can be analyzed using supervised learning techniques to classify skin conditions. The class imbalance visible in the distribution plot suggests the need for techniques like oversampling, undersampling, or cost-sensitive learning to address potential bias in the training process.

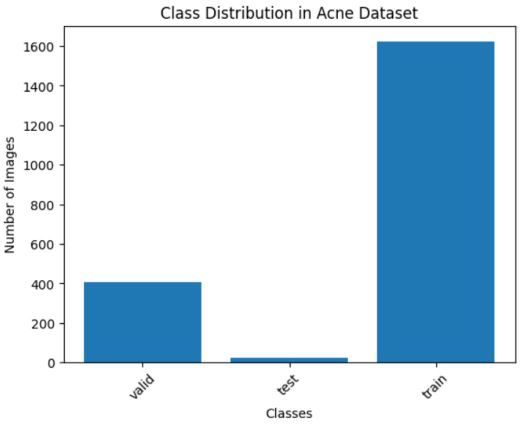

Fig. 6. Class distribution in the Acne Dataset, highlighting imbalances between training, validation, and test sets

[4] Dataset available at: https://www.kaggle.com/datasets/imtkaggleteam/acne-computer-vision

The Acne Dataset serves as an essential resource for advancing machine learning models in dermatology, with potential applications in automatic diagnosis and treatment planning.

The last dataset, ”Bone Fracture Detection Dataset” [5] made by Parisa Karimi Darabi [15], includes 5,006 X-ray images across three splits: 3,530 for training, 801 for validation, and 675 for testing (Figure 7). The dataset features six annotated classes of fractures, including Elbow Positive, Fingers Positive, and Humerus Fracture, designed for applications like automated fractured detection.

Even though the dataset provides a good basis for developing computer vision models, the imbalance in data distribution could pose challenges. Addressing these with techniques such as augmentation or transfer learning could significantly enhance model performance and diagnostic accuracy.

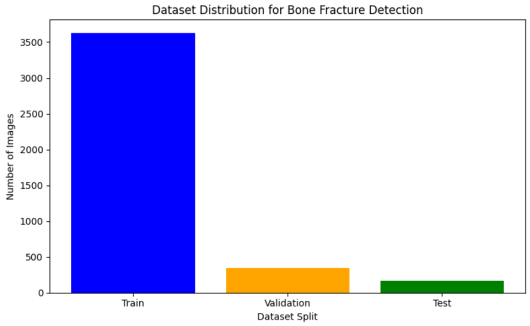

Fig. 7. Dataset distribution for Bone Fracture Detection Dataset

III. Background

The evolution of computer vision is remarkable. Each ”pawn” in this field - whether algorithms, datasets, mathematics, statistics, physics, or related disciplines - has played a critical role in advancing machine vision step by step to the sophisticated system we have today.

Figure 8 highlights the **evolution** of key research topics in computer vision, showing the transition from early geometric approaches to modern techniques like deep learning and SLAM.

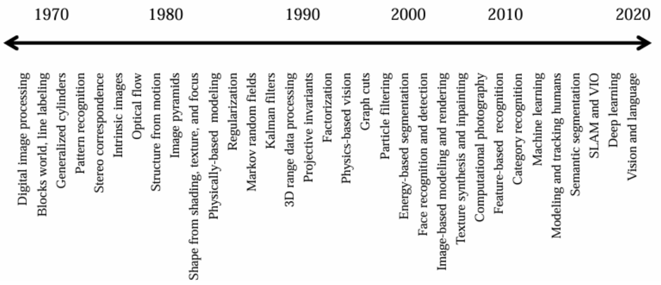

Fig. 8. ”A rough timeline of some of the most active topics of research in computer vision.”. Adapted from [2].

A. *Early beginning*

Computer vision as a concept began in the 1960s at universities pioneering artificial intelligence. Early experiments aimed to interpret simple images as geometric shades, such as lines, edges, and regions. [2]

This was inspired by an experiment by a neurophysiologist, which demonstrated that visual processing in cats begins with recognizing simple shapes. [1] These findings laid the foundation for using geometric and low-level features in computer vision.

One of the most ambitious projects of this era was Seymour Papert’s *Summer Vision Project* at MIT in 1966. The goal was to get a computer to ”describe what it saw” by attaching a camera to it. This highlighted the challenge of interpreting raw visual data and shaped the direction of future research. [7], [8]

B. *1970s: Advancing Geometric and quantitative approaches*

The 1970s marked a significant shift in computer vision research, as detailed in [2], emphasizing geometric and quantitative methods to analyze visual data. Researchers focused on understanding images by extracting low-level features such as edges, shapes, and regions, laying the foundation for scene understanding. Key developments included early attempts to interpret the three-dimensional structure of the world from two-dimensional images.

During this time, techniques such as line labeling, edge detection, and stereo correspondence gained popularity. For instance, line labeling algorithms were used to identify and label structural elements in images, as shown in Figure 9. In addition, pictorial structures and articulated bodily models provided frameworks for understanding the relationship with objects and movement.

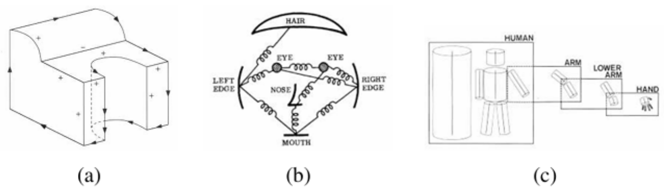

Fig. 9. Examples of early computer vision algorithms from the 1970s, recreated and published in later works: (a) line labeling (Nalwa 1993) ©1993 Addison-Wesley, (b) pictorial structures (Fischler and Elschlager 1973) © 1973 IEEE, (c) articulated body model (Marr 1982) © 1982 David Marr. Adapted from [2].

This decade also introduced foundational ideas like shapeform shading and intrinsic images. They explored how lighting and surface orientation affect intensity variations. These mathematical and quantitative advancements laid the foundation for future developments in motion analysis and 3D object recognition.

C. *1980s: Mathematical Foundations*

In the 1980s, mathematical techniques became central to computer vision as presented in [2]. The focus shifted toward

[5] Dataset available at https://www.kaggle.com/datasets/pkdarabi/bone-fracture-detection-computer-vision-project

analyzing motion, reconstructing 3D structures from motion, and using multi-resolution representations such as image pyramids.

Key developments during this time included:

- Algorithms for optical flow, which calculate motion in images
- Techniques like shape-form-shading that I mentioned earlier, which infer 3D shapes from lighting
- Probabilistic methods such as Markov random fields and Kalman filters, which improved tracking and reconstructing tasks

D. 1990s: Practical applications

By the 1990s, computer vision started transitioning from foundational research to practical applications. Face detection and recognition became viable, driven by advancements in feature extraction and statistical modeling, as explored in [2].

Other notable developments included:

- 3D range data processing, enabling precise object measurements
- Projective invariants for object recognition, which provided robustness against viewpoint changes
- Early applications in video analysis and texture mapping

E. 2000s: Learning-based methods

The 2000s marked a transformative period for computer vision as shown in [2], with the integration of learning-based approaches and computational photography. Techniques like support vector machines (SVMs) and other statistical methods became standard tools for tasks such as image classification and segmentation. This decade also saw an increasing overlap between computer vision and computer graphics.

Key advancements included:

- Techniques like graph cuts and normalized cuts significantly improved object boundary detection by optimizing energy functions
- In computational photography, important innovations included:
  - High dynamic range (HDR) imaging, merging multiple exposures to capture wider light intensities
  - Texture synthesis, using methods like path-based algorithms for realistic image manipulation
  - Panoramic image stitching, seamlessly merging multiple images for wide-field views
- Learning-based techniques, such as the constellation model, advanced object recognition and scene understanding, with interest-point-based methods becoming prominent for panorama creation and object detection

F. 2010s: Deep learning revolution

The 2010s were rvolutionary for computer vision as highlighted in [2], driven by emergence of deep learning. Convolutional neural networks became the cornerstone of modern computer vision, enabling significant advancements in tasks like image classification, object detection, and semantic segmentation.

The availability of large datasets, such as ImageNet, allowed for the training of deep networks with millions of parameters, leading to substantial improvements in model accuracy. Applications including:

- Self-driving cars, leveraging CV for lane detection, object avoidance, and traffic sign recognition. An example is depicted in Figure 10
- Facial recognition, used in security, social media, and consumer devices
- Augmented reality, blending virtual and real-world environments

Fig. 10. Self-driving car technology. Adapted from Montemerlo et al. (2008) and [2].

G. 2020s and beyond

In the 2020s, research continues to push the boundaries of computer vision. Notable advancements remarked in [2] included:

- 3D vision and applications in robotics and autonomous navigation.
- SLAM (Simultaneous Localization and Mapping), enabling robots to map environments in real-time.
- Integration of vision and language, powering systems like image captioning and visual question answering.

When the idea of computer vision first emerged, it was underestimated and seen as ”an easy step along the path to solving more difficult problems such as higher-level reasoning and planning” [2]. However, history has shown that the field is much more difficult. Imagine how complex visual data is, with its variations in lighting, shapes, colors, textures, and perspectives. Even with all these challenges, innovations continued to emerge and lead to the technological advancements we have today.

IV. Approach

A. Methodology

To achieve a comprehensive understanding of the field of computer vision, I decided to use a dual approach combining **qualitative** and **quantitative** methods. The goal of this approach, was to identify gaps in dataset diversity and provide actionable insights into improving model performance in challenging edge cases. By exploring both perspectives, I aimed to gain a balanced view of the datasets, methods, and challenges in this domain.

Supervised learning was chosen as the focus of this study, aligning with my coursework this semester, where I had the opportunity to dive deeper into its applications. Using methods and tools I was already familiar with, I explored the datasets, analyzed their characteristics, and identified potential improvements.

B. Dataset Selection

Datasets were chosen based on two criteria:

- Larger datasets with clear class distributions were prioritized
- Preference was given to recently published datasets

C. Literature and Resources

To support this analysis, I selected:

- One book on computer vision for its structured and clear presentation of concepts
- Research papers accessed from various digital libraries, providing detailed insights into current advancements and challenges
- Articles from IBM, offering practical perspectives on applying computer vision in real-world scenarios

D. Tools and Techniques

The following tools were used in this study:

- **Programming languages**: Python
- **Libraries and frameworks**: TensorFlow, Pandas, NumPy, Matplotlib, Seaborn
- **Platforms**: Kaggle for accessing datasets and Jupyter Notebooks for coding and analysis.
- **Others**: Scikit-learn for data preprocessing and machine learning tasks.

E. Quantitative and Qualitative Analysis

To better structure my findings, I assessed the presence of specific themes or challenges across the two approaches. The following table highlights whether each challenge or theme appeared in quantitative and/or qualitative analyses:

TABLE III
QUANTITATIVE VS QUALITATIVE

Aspect	Quantitative	Qualitative	Remarks
Dataset challenges	✓	✓	Imbalance, annotation inconsistencies, and dataset size were observed
Ethical concerns		✓	Explored through literature review, focusing on interpretability issues
Bias and fairness	✓	✓	Present in discussions of imbalanced data and fairness in dataset usage
Edge cases	✓		Quantitative analysis highlighted model struggles with rare scenarios
Dataset diversity	✓	✓	Identified as a key issue in both analyses, affecting generalization
Model performance	✓		Metrics like accuracy, loss, and overfitting provided measurable insights

V. CONCLUSION

In this paper, I aim to answer several research questions that are crucial for advancing the field of computer vision:

1) **How does dataset diversity affect the performance of computer vision models?**
By analyzing multiple datasets, I explored how variations in data, such as class imbalance and dataset size, impact model performance. This was a central focus, especially in terms of how diverse datasets help improve generalization in real-world applications.

2) **How do object detection models handle rare or unseen scenarios?**
In this research, I acknowledged that handling rare or unseen scenarios is critical, especially in applications like autonomous driving. However, the datasets I used in this study did not specifically address these types of edge cases.

3) **What are the current limitations of LIDAR-based systems in autonomous vehicles?**
Through the literature review, I examined technologies behind LIDAR, focusing on its strengths in mapping and object detection, as well as its limitations, such as challenges with scalability and performance in harsh conditions.

4) **Are there ways to reduce bias in medical imaging datasets?**
This question was explored through both qualitative and quantitative analyses. In the qualitative analyses, I reviewed literature that discussed the challenges of bias in medical imaging datasets. It highlighted the need for more diverse data collection to ensure fairness in model predictions. Many studies emphasize that diverse datasets can improve model accuracy and prevent biases toward underrepresented groups.
In quantitative analysis, we observed class imbalance in the acne dataset, which can introduce bias during model training. This issue was addressed by exploring techniques such as oversampling, undersampling, and cost-sensitive learning. In addition, the use of data augmentation and pre-trained models can improve model performance, reducing bias and improving generalization across the dataset.

Computer vision is by far an important domain nowadays, although it faces many challenges. Whether it is self-driving cars or whether we ask a chat bot to generate an image or video, AI is not entirely capable of handling visual input perfectly.

The complexity of what goes on behind the scenes, considering all the factors that enable this field to function, is incredible! Just image the number of sensors and cameras an autonomous car uses to gather vast amount of real-time information and then respond appropriately depending on the situation.

What we have managed to achieve so far is remarkable and can certainly be improved.

REFERENCES

[1] IBM, “What is computer vision?” IBM, 2021. Available: https://www.ibm.com/think/topics/computer-vision

[2] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed. Springer, 2021. Available: https://szeliski.org/Book

[3] IBM, “What is deep learning?” IBM, 2024. Available: https://www.ibm.com/think/topics/deep-learning

[4] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, “A review of convolutional neural networks in computer vision”, Available: https://doi.org/10.1007/s10462-024-10721-6

[5] Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2021). ”Dive into deep learning.” release 0.16.1, 2021. [Online]. Available: https://d2l.ai.

[6] P. Flori, T. Leduc, Y. Sutter, R. Brémond, ”Visual complexity of urban streetscapes: human vs computer vision”, Available: https://doi.org/10.1007/s00138-023-01484-1

[7] Papert, Seymour (1966-07-01). ”The Summer Vision Project”. MIT AI Memos (1959-2004). Available: https://hdl.handle.net/1721.1%2F6125

[8] Boden, Margaret Ann (2006). ”Mind as a Machine: A History of Cognitive Science”. Clarendon Press, p. 781. ISBN 978-0-19-954316-8.

[9] Y. Ren, ”Intelligent Vehicle Violation Detection System Under Human–Computer Interaction and Computer Vision”, Available: https://doi.org/10.1007/s44196-024-00427-6

[10] P. Pochelu, O. Castro-Lopez, ” Mastering Computer Vision Inference Frameworks”,. Available: https://doi.org/10.1145/3629527.3651430

[11] A. Maihemuti, B. Zhang, J. Zhang, ”Global Terrain Registration of LiDAR and Camera Fusion Using Multiple Calibrators”, Available: https://doi.org/10.1145/3697355.3697379

[12] S. Susan, ”Neuroscientific insights about computer vision models: a concise review”, Available: https://doi.org/10.1007/s00422-024-00998-9

[13] P. Gupta, J. Nirmal, N. Mehendale, ”A survey on computer vision approaches for automated classification of skin diseases”, Available: https://doi.org/10.1007/s11042-024-19301-w

[14] Y. Xu, T. M. Kjan, Y. Song, E. Meikering, ” Edge deep learning in computer vision and medical diagnostics: a comprehensive survey”, Available: https://doi.org/10.1007/s10462-024-11033-5

[15] P. Karimi Darabi, ”Bone Fracture Detection Dataset,” DOI: 10.13140/RG.2.2.14400.34569