6252-ITAI-1378-Comp Vision-Artificial Intel-RT-15698 - Spring 2025

Professor Anna Devarakonda

Team EyeConic

Brooke Broderick - W217065417
Matthew Choo - W206744145
Cameroun White-W216530571
Melvis Maduagwu-W216890463
Erick Banegas - W216200956
Nancy Chieu-W000098778
Yoana Cook - W215913890

Computer vision is a type of artificial intelligence that focuses on enabling computers to identify and understand objects, people, images and videos (Microsoft). This type of AI uses machine learning to adapt and mimic human capabilities by breaking down and analyzing the images and videos it is being received. By using data interpretation to create new data based on existing data, computer vision can aid the user to document, predict, and monitor future errors in jobs such as cybersecurity and retail. This paper will explain how computer vision was developed and how technology is able to advance using computer vision.

In the field of computer vision, one of the key steps to improve a model's performance is to apply data preprocessing and augmentation techniques. These include scaling and normalizing images so that pixel values fall within a specified range, reducing unwanted variance. Data augmentation through rotations, flips, random crops, and color adjustments increases image diversity, helping prevent models from becoming "lazy" and memorizing overly specific patterns that fails to generalize.

When facing imbalanced datasets—where certain classes are poorly represented—oversampling, undersampling, or data synthesis methods (e.g., SMOTE for images) can help balance the number of examples. Once the data is prepared, training the model efficiently becomes essential, harnessing the power of transfer learning and leveraging pre-trained models such as ResNet, MobileNet, or VGG, which already possess initial knowledge acquired from large databases like ImageNet. This drastically cuts down on training time and resources for specific tasks. However, refining results often requires hyperparameter tuning, experimenting with various learning rates, batch sizes, and optimizers. Tools like Grid Search or Bayesian Optimization can automate this process, speeding up the search for optimal parameter combinations.

After achieving acceptable accuracy, the next challenge is optimizing the model for deployment across various environments, such as mobile devices or embedded systems. Techniques like quantization reduce the precision of floating-point weights to lighter formats (e.g., int8), minimizing model size and memory consumption; pruning removes connections or neurons with little contribution to the final outcome, while knowledge distillation transfers knowledge from a large model to a smaller one without significant loss in predictive power. The main goal is to balance accuracy with efficiency, which is particularly critical on edge devices with limited resources. Quality data and labels directly impact a model's generalization capabilities. Tools like Labeling enable quick and intuitive object annotation, while semi-supervised methods help when only a small subset of labeled data is available alongside a larger unlabeled pool. In situations requiring vast amounts of information, crowdsourcing—via platforms like Amazon Mechanical Turk or internal organizational tools—facilitates the collection of large annotation volumes, provided adequate quality checks are in place. No model is perfect from the start, and errors present opportunities to better understand the problem and fine-tune the model. An error analysis approach involves examining the cases in which the model fails, categorizing them, and identifying patterns for improvement—such as refining segmentation techniques, adjusting data augmentation strategies, or reviewing label quality. In this context, explainable AI (XAI) becomes an indispensable tool, allowing developers to visualize which parts of the image most influenced the model's decision. This transparency fosters user trust and aids in identifying system weaknesses, fueling a continuous cycle of improvement in computer vision solution development.

Computer vision has made incredible strides, thanks to advancements in deep learning and AI-driven architectures. Neural networks are incredibly powerful when it comes to recognizing patterns, making predictions, and generating new data, and particularly: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs), have played a key role in revolutionizing pattern recognition, prediction, and data generation across multiple industries. (IBM, 2024). CNNs are particularly effective at image analysis, making them essential for facial recognition, medical imaging, and self-driving cars.   They break down images layer by layer, detecting everything from edges to complex objects.   Meanwhile, RNNs excel at handling sequential data, which makes them ideal for applications like speech recognition, language translation, and stock market predictions. LSTMs, a type of RNN, take this a step further by retaining long-term dependencies in data, which is why they power chatbots, voice assistants, and predictive text models. On the other hand, GANs generate hyper-realistic data by having two networks, one creating synthetic images and the other trying to distinguish them from real ones, leading to groundbreaking applications in deepfake technology, AI-generated art, and medical image synthesis. (LeCun, 2020).

Beyond these networks, the rise of edge computing has brought real-time computer vision to our devices. Smartphones, drones, and smart cameras can now process images and videos locally without relying on cloud computing, allowing for faster and more secure AI-powered applications. This is particularly useful in real-time road monitoring, where vehicles equipped with high-resolution cameras and GPS systems continuously scan road conditions. (Ruseruka, 2023). These systems use CNNs for image classification, deep learning models for damage detection, and GPS for precise location tracking. Cities worldwide are using such technology to analyze cracks, potholes, and surface conditions, enabling proactive maintenance decisions. Laser scanners and LiDAR sensors further enhance the accuracy by generating detailed 3D models of road surfaces, while *infrared sensors* detect temperature variations that could indicate underlying issues. (Ruseruka, 2023).

Generative AI is also transforming the field by creating synthetic datasets for training models, which is especially useful in areas like virtual try-ons and gaming. By simulating real-world conditions, AI can improve model accuracy and performance. Additionally, 3D computer vision, powered by LiDAR and advanced modeling techniques, is crucial for autonomous vehicles and augmented reality. Self-driving cars rely on LiDAR to create real-time depth maps of their surroundings, while AR applications use these advancements to overlay digital elements onto the physical world seamlessly. (Wasser, 2014)

From road maintenance to autonomous driving and medical imaging, deep learning and AI-driven computer vision is shaping industries worldwide. Whether it's identifying road distresses, improving urban infrastructure, or advancing hyper-realistic AI-art, these technologies are pushing boundaries and making real-world applications smarter, faster, and more efficient.

Computer vision is revolutionizing industries by enhancing efficiency, safety, and accessibility through AI-driven automation. In healthcare, advanced imaging techniques like CT scans and X-rays, combined with AI-powered prognosis, enable early disease recognition and more exact medical conclusions, leading to higher quality patient care. Transportation has also seen major improvements, with autonomous vehicles and advanced driver assistance systems (ADAS) minimizing human error, while smart traffic management helps prevent congestion and accidents. Object detection also reinforces safety by mapping out hazards in real time to avoid fatalities while traveling.

Retail and e-commerce leverages computer vision for more individualized shopping experiences, such as virtual try-on tools and AI-backed product suggestions. Self-operating checkout systems and inventory management streamline operations, making shopping hastened and more favorable. In fabrication, computer vision ensures exceptional production by detecting flaws in real-time, reducing waste, and increasing efficiency on production lines.

Security and surveillance have become more advanced with smart monitoring systems that detect impending threats instantly, alongside facial recognition technology that improves authentication and access control. Additionally, computer vision is improving accessibility, providing visually impaired individuals with object detection apps to maneuver their environment and enable real-time sign language translation for more effective communication.

As AI-powered vision technology continues to evolve, its value on daily life grows stronger, bettering processes across healthcare, transportation, retail, manufacturing, security, and accessibility. These developments not only boost efficiency but also create a safer, more inclusive world, shaping the future of many different industries in groundbreaking ways.

The challenges and ethical implications in computer vision technology sparks concern about the intrusion into private spaces and the collection of personal data without explicit permission and discriminatory outcomes, such as misidentification or differential treatment based on race, gender, or other characteristics. The bias in AI models creates an imbalance in the algorithm resulting in unfair and discriminatory outputs. Ongoing regulation of the language prediction tools would result in mitigating online hatred, offensive speech, hatred towards groups of specific ethnic origins. In data privacy and security, privacy-preserving techniques are used to protect individual privacy and prevent misuse of data. For example, the use of differential privacy is applied to visual data to prevent personal information from being inferred from AI models. Another privacy-preserving technique, a PPE algorithm, used for facial recognition encodes facial data in such a way that it can be used for recognition purposes but does not expose sensitive personal characteristics. The downside to privacy-preserving algorithms are the reduction of model accuracy and slower processing times. The aspect of informed consent when it comes to the collection of personal data has raised concerns about privacy violations, lack of transparency, and the potential misuse of sensitive information, especially in AI-driven computer vision applications.

The advancements in computer vision come with considerable environmental costs, due to significant energy consumption as a result of large computational power required to train and run complex models, resource depletion, and electronic waste, leading to high carbon emissions from data centers.

In conclusion, building efficient computer vision systems requires mastering key "tricks of the trade," such as optimizing data efficiency through augmentation, active learning and transfer learning. It's also important to select the right model architectures, leverage hardware acceleration and parallel processing to maximize speed, efficiency, and performance while minimizing power consumption and latency. By combining these strategies, computer vision systems can run faster, consume less power, and handle larger workloads efficiently, whether on mobile devices, embedded systems, or high-performance computing clusters and cloud computing systems. Technological advancements have played a huge role in simplifying complex tasks, making automation more accessible, and significantly improving the quality of life across various areas. One of the most noticeable impacts is in communication and connectivity. The internet, smartphones, and social media platforms have made it easier to stay in touch with people globally in real-time. Video calls, instant messaging, and email have replaced slower, traditional methods of communication, making interactions more immediate and seamless. It's now nearly effortless to contact family and friends. Moving forward, the potential of computer vision extends far beyond its current applications in healthcare, surveillance or robotics. with promising contributions to solving global challenges, including environmental monitoring e.g. air pollution monitoring and climate change impact monitoring, healthcare diagnostics e.g. CT scans and MRIs, and intelligent automation e.g. self driving cars and online maps, paving the way for a smarter and more connected world. Computer Vision is evolving towards more autonomous, intelligent and human-like vision systems that would be capable of making complex decisions with minimal human intervention.

Bibliography

Augmented Startups. (2023, June 13). Exploring the ethical implications of computer vision technology in 2023. Augmented Startups. https://www.augmentedstartups.com/blog/exploring-the-ethical-implications-of-computer-vision-technology-in-2023

Boesch, G. (2024, December 19). Applied Computer Vision: Which Industries Benefit the most?. viso.ai. https://viso.ai/computer-vision/applied-computer-vision/

Brown. (2020, June). *Introduction to A.I. - Course Setup*. *Module 10*. Houston.

Conti, M. (2017, February 28). *The incredible inventions of intuitive AI | Maurice Conti*. YouTube. https://www.youtube.com/watch?v=aR5N2Jl8k14.

Cook, B. (n.d.). The process behind pavement condition assessments. Galveston; City of Galveston.

Day 55: Transfer Learning — Using Pre-trained Models. (2025). recuperado February 1, 2025, https://medium.com/@bhatadithya54764118/day-55-transfer-learning-using-pre-trained-models-7ce1d81712e5

Dilek E, Dener M. Computer Vision Applications in Intelligent Transportation Systems: A Survey. Sensors (Basel). 2023 Mar 8;23(6):2938. doi: 10.3390/s23062938. PMID: 36991649; PMCID: PMC10051529.

European Union Agency for Fundamental Rights. (2022). Bias in algorithms – Artificial intelligence and discrimination. FRA. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

Fridman, L. (2018a, January 27). *MIT 6.S094: Computer vision*. YouTube.

https://www.youtube.com/watch?v=CLOAswsxudo.

Goodfellow, I. (2019, April 18). *Ian Goodfellow: Generative Adversarial Networks (Gans): MIT: Artificial Intelligence Podcast*. Lex Fridman. https://lexfridman.com/ian-goodfellow

Goodfellow, I. (2017, May 8). *Deep learning Chapter 1 introduction presented by Ian Goodfellow*. YouTube. https://www.youtube.com/watch?v=vi7lACKOUao&t=1663s

Holdsworth, J., & Scapicchio, M. (2024, June 17). *What is deep learning?*. IBM. https://www.ibm.com/topics/deep-learning.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … Adam, H. (2017).

MobileNets: Efficient convolutional neural networks for mobile vision applications.

arXiv preprint arXiv:1704.04861.

https://arxiv.org/abs/1704.04861

LeCun, Y. (2020, June). *Deep learning, neural networks and the future of ai*. Yann LeCun: Deep learning, neural networks and the future of AI | TED Talk. https://www.ted.com/talks/yann_lecun_deep_learning_neural_networks_and_the_future_of_ai?subtitle=en.

Patel, M. (2025, January 28). Applications of computer vision in Healthcare. Kritikal Solutions Pvt. Ltd. https://kritikalsolutions.com/applications-of-computer-vision-in-healthcare/?

Ruseruka, C., Mwakalonge, J., Comert, G., Siuhi, S., & Perkins, J. (2023, August 7). *Road condition monitoring using vehicle built-in cameras and GPS sensors: A deep learning approach*. MDPI. https://www.mdpi.com/2624-8921/5/3/51.

Snoek, J., Larochelle, H., & Adams, R. P. (2012).

Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2951–2959.
https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Wasser, L. (2014, November 24). *How does lidar remote sensing work? light detection and ranging*. YouTube. https://www.youtube.com/watch?v=EYbhNSUnIdU.

"What Is Computer Vision?: Microsoft Azure." What Is Computer Vision? | Microsoft Azure,
azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-computer-vision#object-classification. Accessed 27 Jan. 2025.

UNESCO. 2023, February 6. Ethics of AI: Challenges and Governance. YouTube
https://youtu.be/VqFqWIqOB1g?si=5BJluS0CDJ_0VybC

XenonStack. (n.d.). Ethical considerations in computer vision. XenonStack.
https://www.xenonstack.com/blog/ethical-considerations-in-computer-vision