Introduction to Computer Vision

Pham Van Huy, Ph.D.

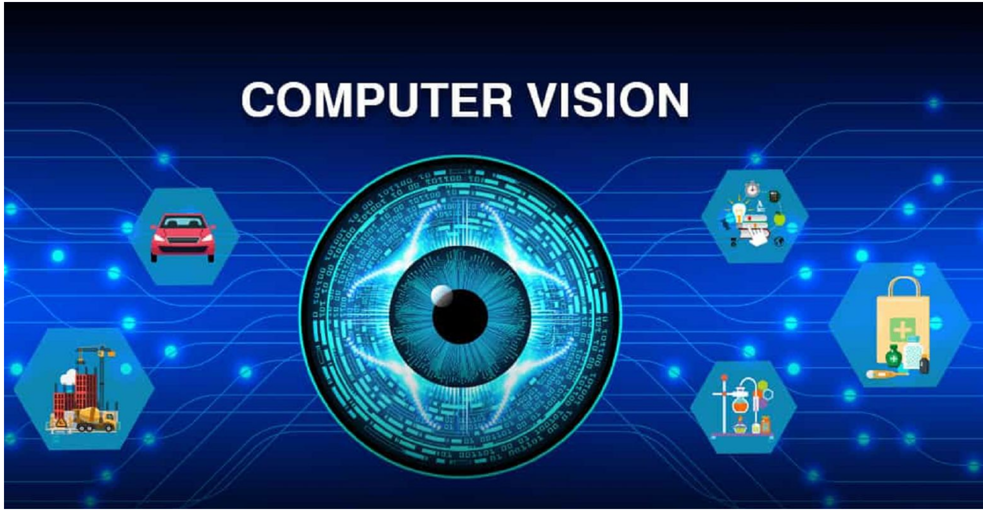
Faculty of Information Technology,
Ton Duc Thang University, Ho Chi Minh City

Semester 2024 - 2025

Introduction

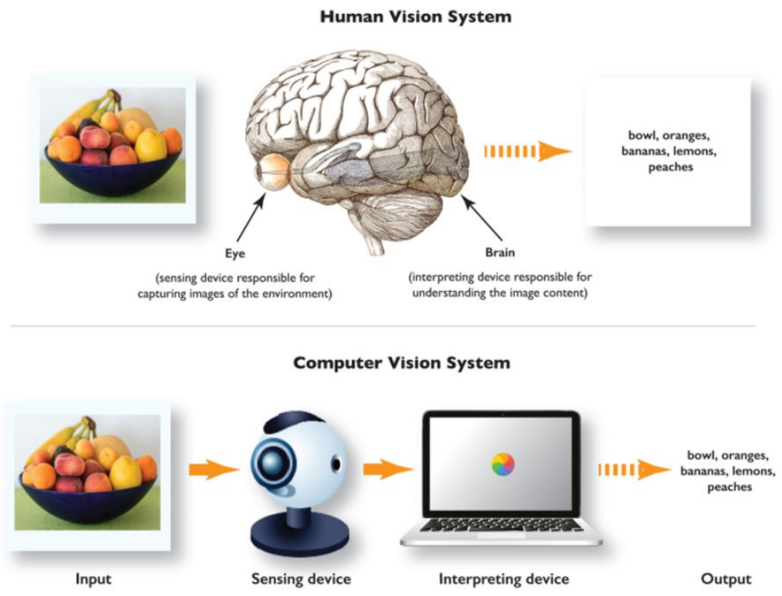
- Computer vision has significantly evolved, incorporating both **traditional techniques** and **deep learning.**

- This presentation explores **why traditional techniques remain valuable despite the advancements in deep learning.**



Computer Vision

Computer vision is a field of AI that enables machines to interpret and process visual data similarly to human vision.



Computer Vision

Computer vision is a field of AI that enables machines to interpret and process visual data similarly to human vision.

Applications

Facial Recognition:

- Security systems for identifying individuals.
- Mobile phone authentication and unlocking.
- Attendance and access control systems.
- Example: Apple's Face ID technology for secure authentication.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *IEEE Conference on Computer Vision and Pattern Recognition.*

Industrial Automation:

- Quality inspection and defect detection in manufacturing.
- Object sorting and assembly line automation.
- Inventory management and counting.
- Example: Automated inspection of circuit boards for defects.

Li, W., Liu, C., Xu, F., Zhang, Y., & Zhao, Y. (2018). Deep learning-based surface defect inspection system for industrial applications. *Sensors and Actuators A: Physical*, 295, 90-104.

Computer Vision

Computer vision is a field of AI that enables machines to interpret and process visual data similarly to human vision.

Applications

Medical Imaging:

- Detecting tumors in MRI and CT scans.
- Automated cell counting in microscopy images.
- Retinal image analysis for detecting diabetic retinopathy.
- Example: Using CNNs to classify skin lesions as benign or malignant.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

Autonomous Driving:

- Lane detection and keeping.
- Pedestrian and obstacle detection.
- Traffic sign recognition.
- Example: Tesla's Autopilot system using computer vision for navigation.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zhang, X. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316.*

Computer Vision

Computer vision is a field of AI that enables machines to interpret and process visual data similarly to human vision.

Applications

Agriculture:

- Crop monitoring and disease detection.
- Automated harvesting and sorting.
- Soil and plant health analysis.
- Example: Drones using computer vision to monitor crop health.

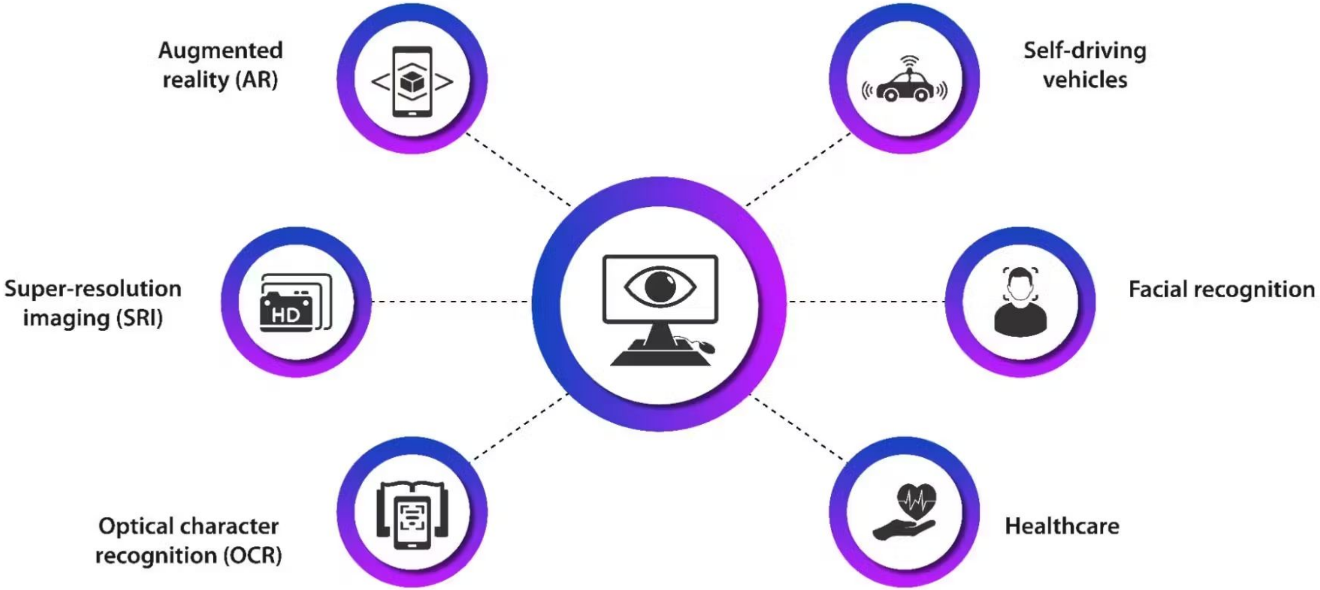
Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70-90.

Security and Surveillance:

- Automated monitoring of CCTV footage.
- Intrusion detection systems.
- License plate recognition.
- Example: Smart security cameras that can detect and alert on suspicious activity.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*.

Computer Vision Applications

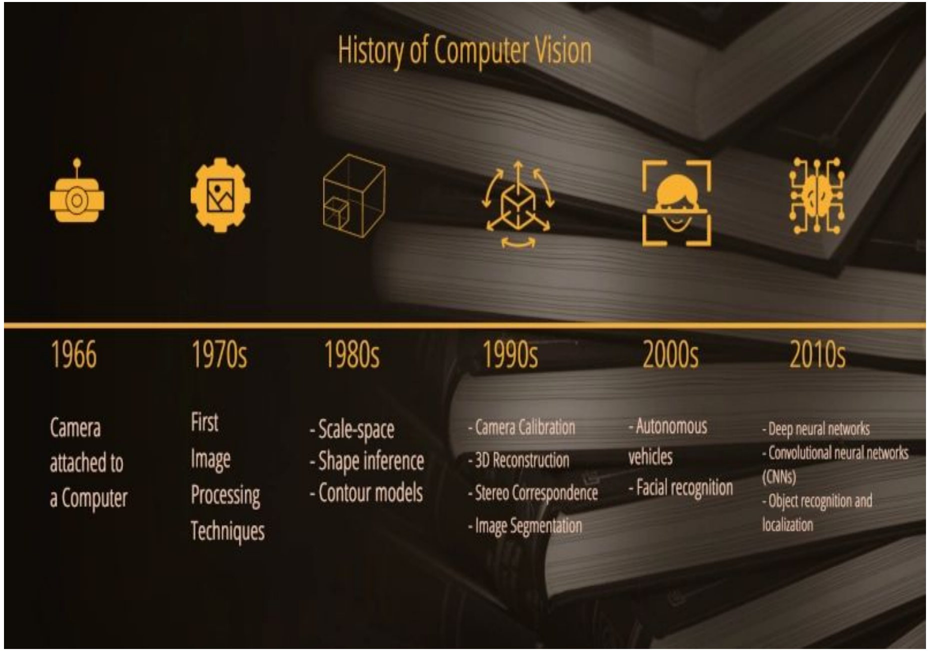


Turing

Evolution of Computer Vision

Timeline:

- **1960s-1980s:** Early experiments with image processing and simple feature extraction.
- **1990s-2000s:** Development of sophisticated algorithms like SIFT, HOG, and advancements in edge detection.
- **2010s-Present:** Emergence of deep learning models like CNNs, GANs, and transformers.



Computer Vision Techniques

Traditional Techniques

Definition: Methods that rely on mathematical models and handcrafted features.

Algorithms:

- **Edge Detection:** Canny Edge Detector, Sobel Operator.
- **Feature Detectors:** Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF).
- **Descriptors:** Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP).

Deep Learning in Computer Vision

Definition: Use of neural networks to automatically learn features from large datasets.

Key Models:

- **Convolutional Neural Networks (CNNs):** Used for image classification, object detection, and segmentation.
- **Generative Adversarial Networks (GANs):** Used for image generation and style transfer.
- **Recurrent Neural Networks (RNNs) and Transformers:** Used for video analysis and sequential image processing.

Computer Vision Techniques

Advantages of Traditional Techniques

- **Explainability:** Algorithms are transparent and easier to understand.

  Example: The steps in edge detection are well-defined and easily interpretable.

- **Resource Efficiency:** Require less computational power and memory.

  Example: Embedded systems and mobile devices benefit from low-resource algorithms.

- **Robustness:** Often more resilient to variations and noise in data.

  Example: Traditional methods can effectively handle varying lighting conditions in images.

- **Speed:** Typically faster in real-time applications due to simpler computations.

  Example: Real-time video processing in surveillance systems.

Advantages of Deep Learning

- **Performance:** Achieves high accuracy on complex tasks with large datasets.

  Example: ImageNet competition results showcasing deep learning models outperforming traditional methods.

- **Automation:** Eliminates the need for manual feature extraction.

  Example: Automated feature learning in CNNs for object detection.

- **Scalability:** Can handle large-scale datasets and complex models.

  Example: Google Photos using deep learning for image organization and search.

- **Versatility:** Applicable to a wide range of tasks beyond traditional vision.

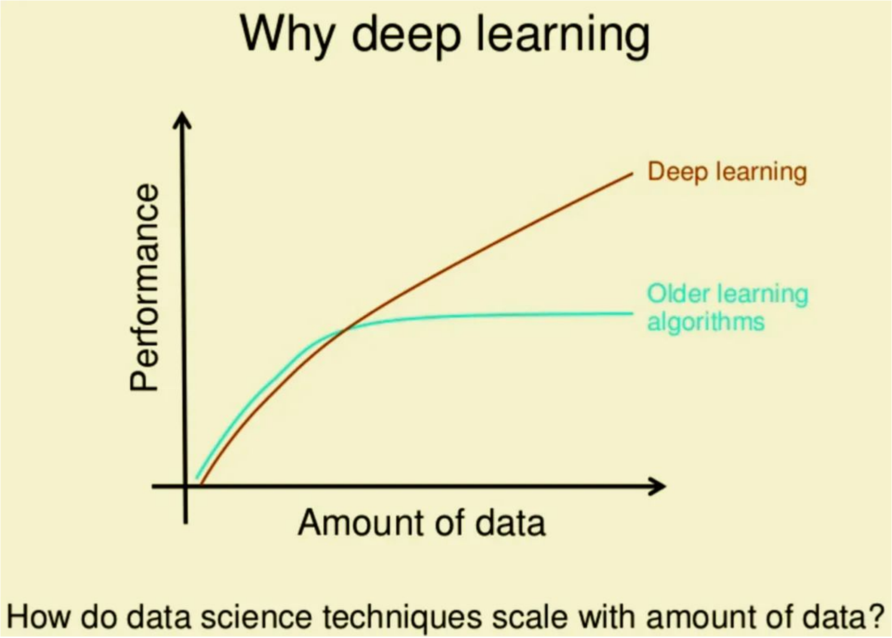
  Example: GANs used for creative applications like artwork generation.

Computer Vision Techniques

Comparison Framework

Criteria:

- **Performance:** Accuracy and precision in various tasks.
- **Interpretability:** Ease of understanding how the algorithm works.
- **Resource Requirements:** Computational power, memory, and data needs.
- **Task Suitability:** Appropriate use cases for each approach.



Computer Vision Techniques

Traditional Techniques

- **Task-Specific Strengths**
  - **Simple Tasks:** Edge detection, basic image processing.
  - **Real-Time Processing:** Robotics, industrial applications with low latency.
  - **Low-Resource Environments:** Embedded systems, mobile devices.

- **Real-Time Processing**
  - Often faster due to simpler calculations.
  - Suitable for applications where quick decision-making is crucial, e.g., robotics, real-time video surveillance.

Deep Learning

- **Task-Specific Strengths**
  - **Complex Tasks:**Image segmentation, object detection, facial recognition.
  - **Large-Scale Applications:** Automated tagging in social media, medical diagnostics requiring high precision.

- **Real-Time Processing**
  - Requires significant computational resources but advancements like TensorRT and Edge TPU are improving real-time capabilities.
  - Still generally slower compared to traditional techniques due to complexity.

Computer Vision Techniques

Robustness and Noise Handling

- **Traditional Techniques**
  - Generally robust to noise and variations due to simpler, well-understood models.
  - Examples: Edge detection methods can effectively work even with noisy images.

- **Deep Learning**
  - Can be sensitive to noise and adversarial attacks.
  - Requires large, well-annotated datasets to perform robustly.
  - Ongoing research is improving robustness with techniques like data augmentation and adversarial training.

Explainability and Interpreta

- **Traditional Techniques**
  - Easier to understand and debug due to straightforward algorithms.
  - Engineers can directly see and adjust parameters affecting performance.

- **Deep Learning**
  - Often seen as a "black box" with many parameters and complex architectures.
  - Explainability is a growing research area with methods like attention mechanisms and interpretable AI models.

Computer Vision Techniques

Use Cases Where Traditional Methods Excel

- **Medical Image Analysis:**
  - Simple algorithms for detecting edges, shapes, and textures.
  - Example: Basic lesion detection in skin images.
- **Real-Time Vision Systems:**
  - Applications requiring immediate processing with low latency.
  - Example: Automated assembly line inspection.
- **Applications with Limited Data:**
  - When only small datasets are available, traditional methods can perform better.
  - Example: Niche scientific imaging tasks.

Hybrid Approaches

- **Integration:**
  - Combining traditional methods with deep learning can enhance performance.
  - Example: Using SIFT features as input to CNNs for more robust feature extraction.
- **Benefits:**
  - Leveraging the strengths of both approaches can improve accuracy, robustness, and computational efficiency.
  - Examples: Hybrid models for improved image classification and object detection.

Computer Vision Techniques

- **Case Study 1: Traditional Methods in Automated Driving Systems:**
  - **Details:** Lane detection using edge detection algorithms.
  - **Impact:** Real-time processing, robustness to varying lighting conditions.
  - **Citations:**
    - Chen, W., Xi, J., Zhang, K., Chen, L., & Wang, H. (2017). Lane detection for structured and unstructured road images based on line segment detector. *IEEE Intelligent Vehicles Symposium*

- **Case Study 2: Deep Learning in Facial Recognition and Object Detection:**
  - **Details:** Using CNNs for high-accuracy facial recognition in security systems.
  - **Impact:** Improved accuracy and scalability for large databases.
  - **Citations:**
    - Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *IEEE Conference on Computer Vision and Pattern Recognition*
    - Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*

Computer Vision Techniques

Current Trends and Research

- **Research Focus**
  - Hybrid models combining traditional techniques with deep learning.
  - Improving the efficiency and explainability of deep learning models.
  - Developing new algorithms in traditional vision that can complement deep learning.
- **Industry Examples**
  - Robotics: Using a mix of traditional and deep learning for navigation and object manipulation.
  - Medical Diagnostics: Hybrid approaches for more accurate and explainable results.
- **Reference:**
  - Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. *Facebook AI Research*
  - Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *IEEE International Conference on Computer Vision*

Computer Vision Techniques

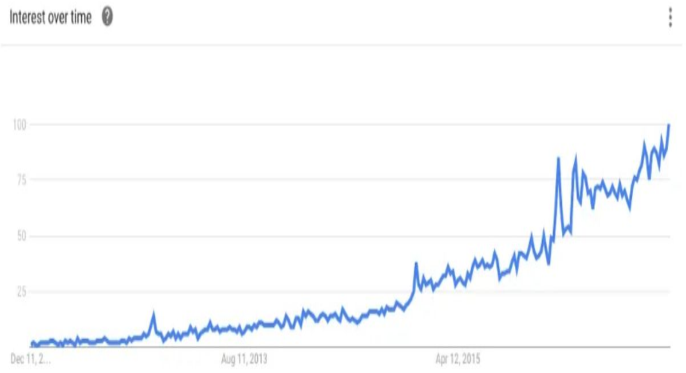
Challenges and Future Directions

- **Challenges**
  - Data requirements: Deep learning models need large, annotated datasets.
  - Computational costs: High resource demands for training and inference.
  - Interpretability: Understanding and trusting AI decisions.

- **Future Trends**
  - Advances in explainable AI for both traditional and deep learning methods.
  - Edge computing to bring powerful AI to resource-constrained environments.
  - Continued integration of traditional and deep learning methods for enhanced performance.

More information:

- Li, Z., & Liu, F. (2019). Edge Computing: Enabling Technologies and Use Cases. *IEEE Internet of Things Journal*



Trend of "Deep Learning" in google

Assessment

In-class presentation (20-30%)

- Using books
- Slides, coding for illustrations

Midterm essay (20%)

- Simple applications
- Essay report

Final report (50%)

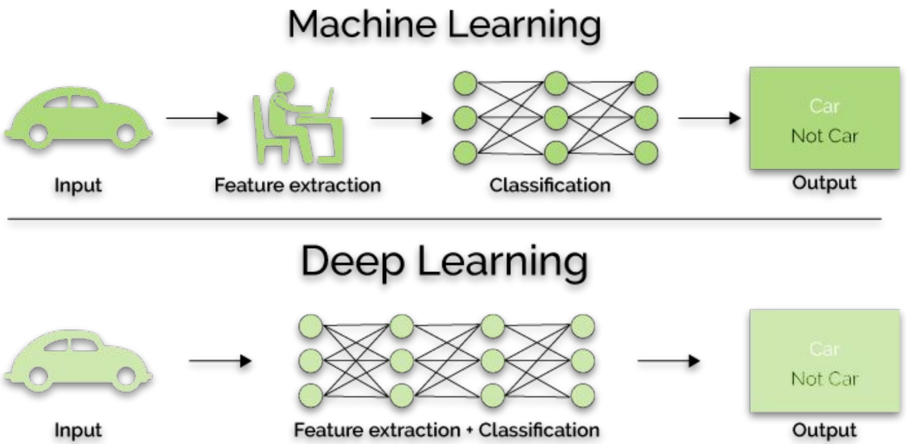
- Application app OR
- Research papers (paperswithcode.com)

Computer Vision Techniques

Conclusion

- **Summary:**
  - Traditional techniques offer explainability, resource efficiency, and robustness.
  - Deep learning excels in performance, automation, and scalability.
  - Both approaches have unique strengths; their synergy can drive future innovations.

- **Final Thought:** Balancing and integrating both traditional and deep learning methods can lead to more effective and versatile computer vision systems.



Continued integration of traditional and deep learning methods for enhanced performance: An example

Video stabilization

How does it work?

A traditional computer vision method called **Feature Matching**

Feature matching

How an Image Feature Looks

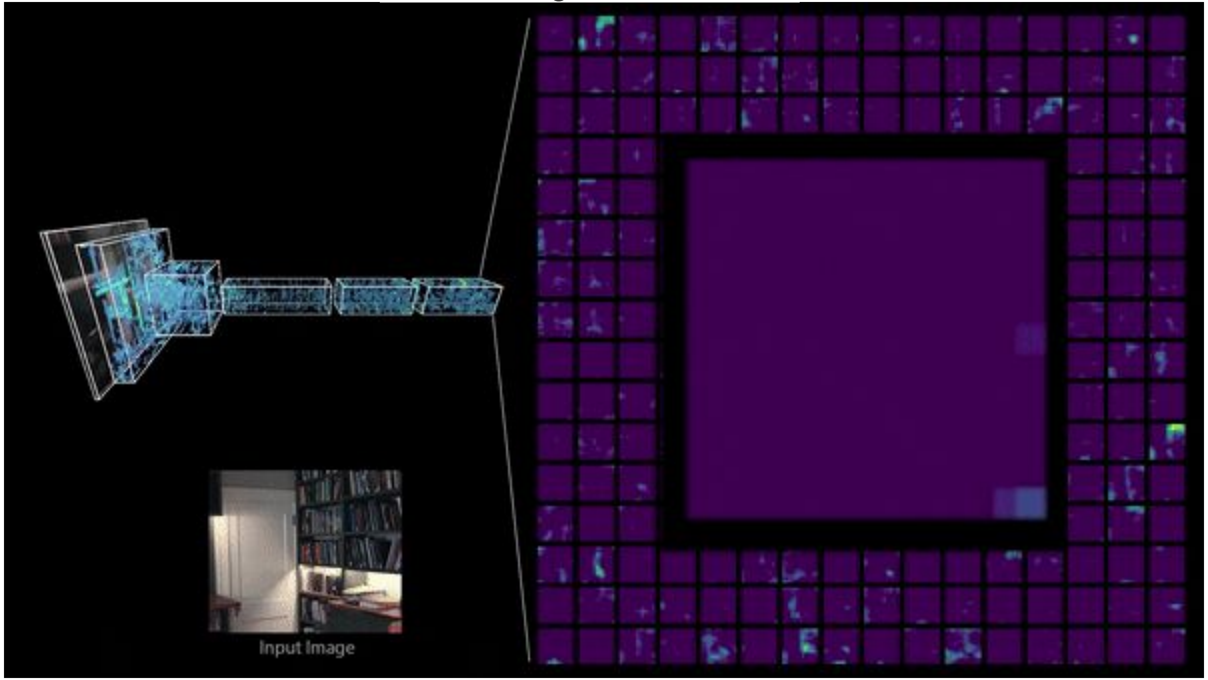


Image Feature

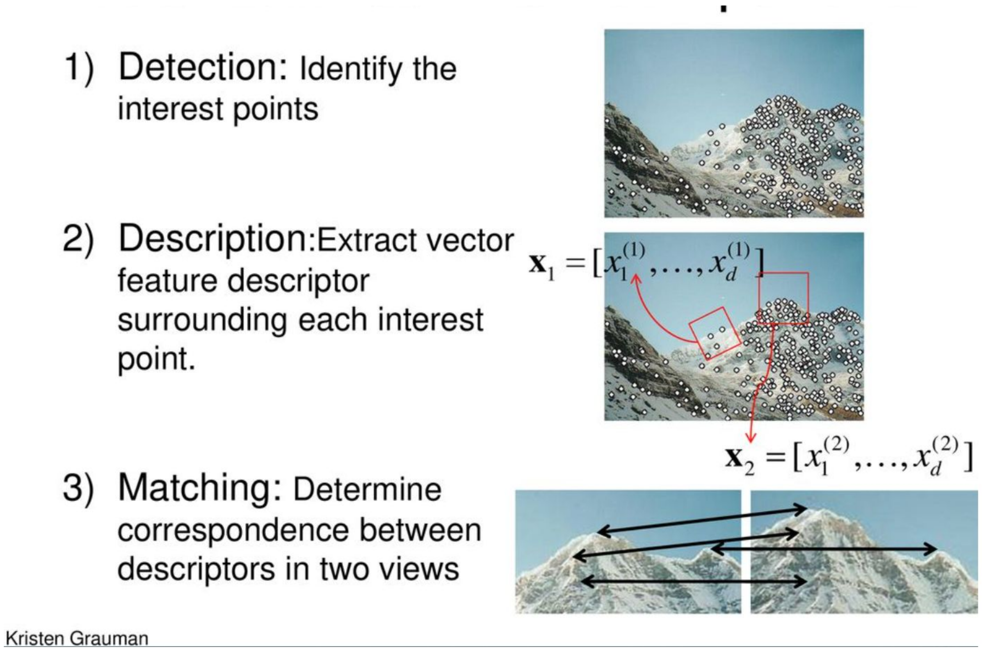
An image feature is a piece of information that describes the objects with a unique quality.

These features include anything from simple edges and corners to more complex textures like intensity gradients or unique shapes like blobs.

Why Feature Matching in 2024?

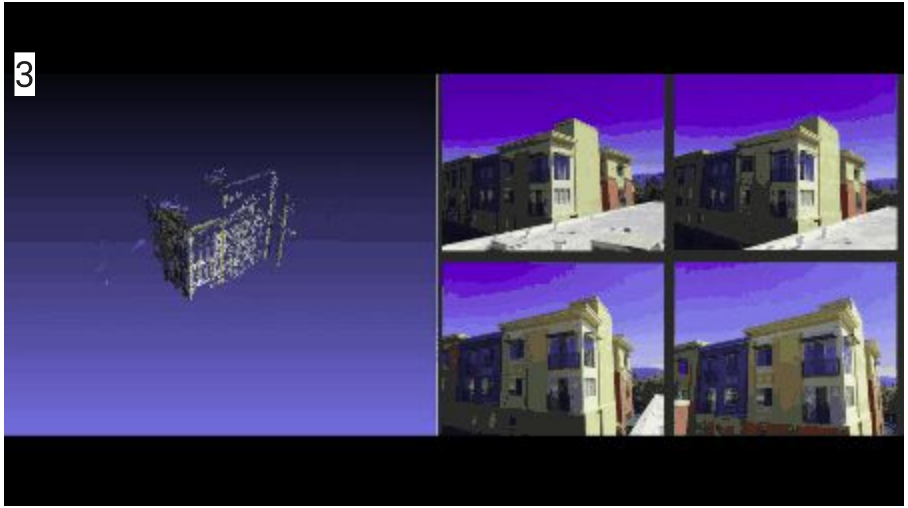
Feature matching is an old computer vision technique that started in the late 1990s with edge detection algorithms like *Sobel, Canny, and the corner detection algorithm Harris Corner Detection.*

Feature matching is heavily used in some of the mainstream computer vision tasks.

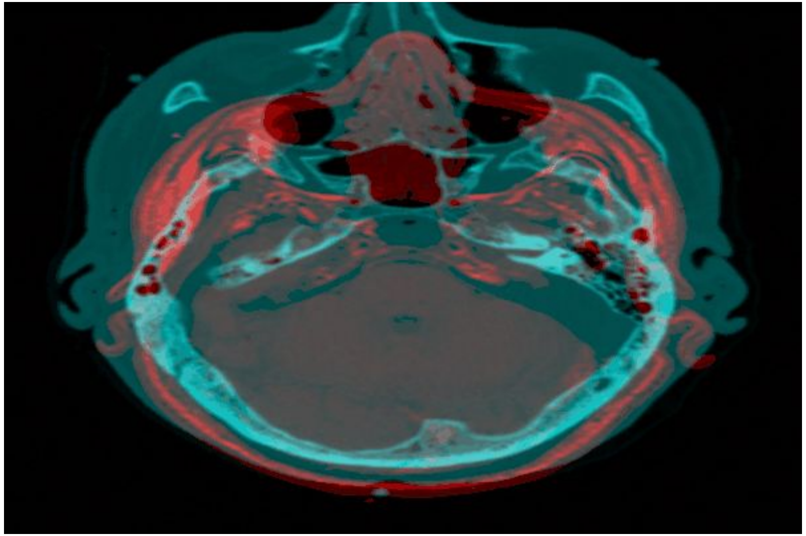


Feature matching Applications

Structure from Motion(3D Reconstruction)



Medical Image Registration



Feature matching Applications

Face Recognition

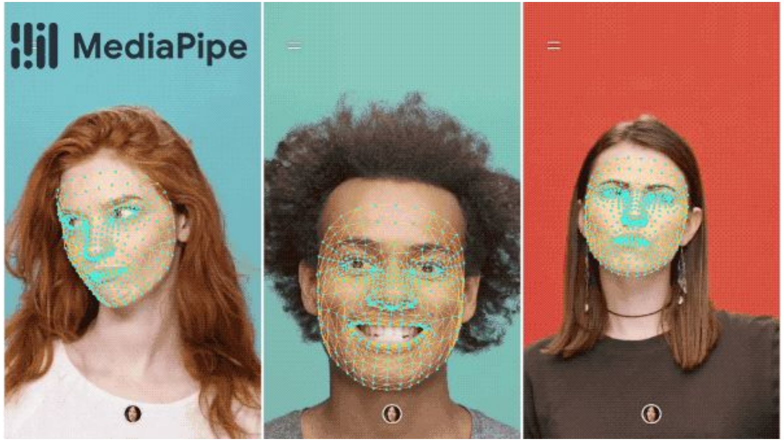
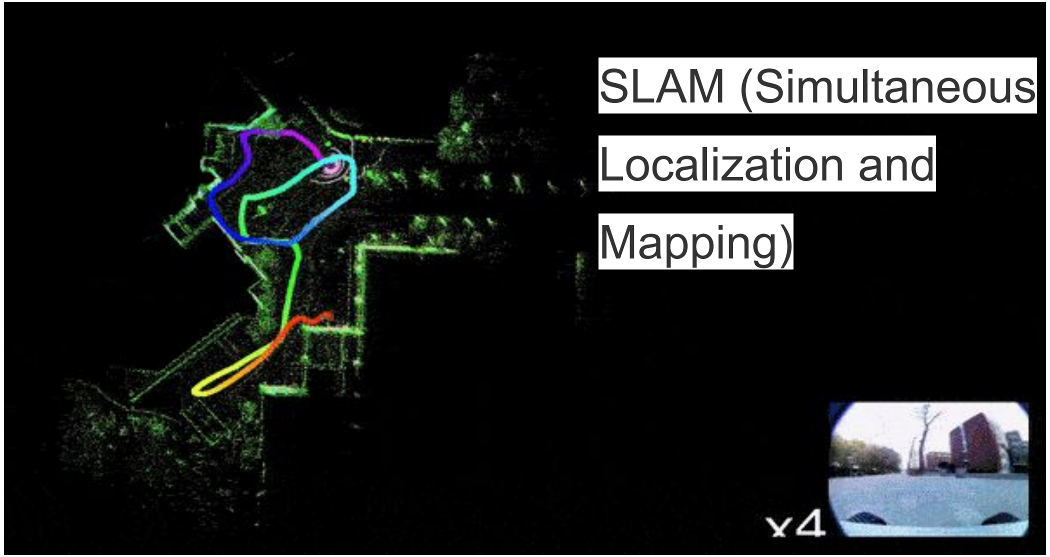


Image Stitching (Panorama)

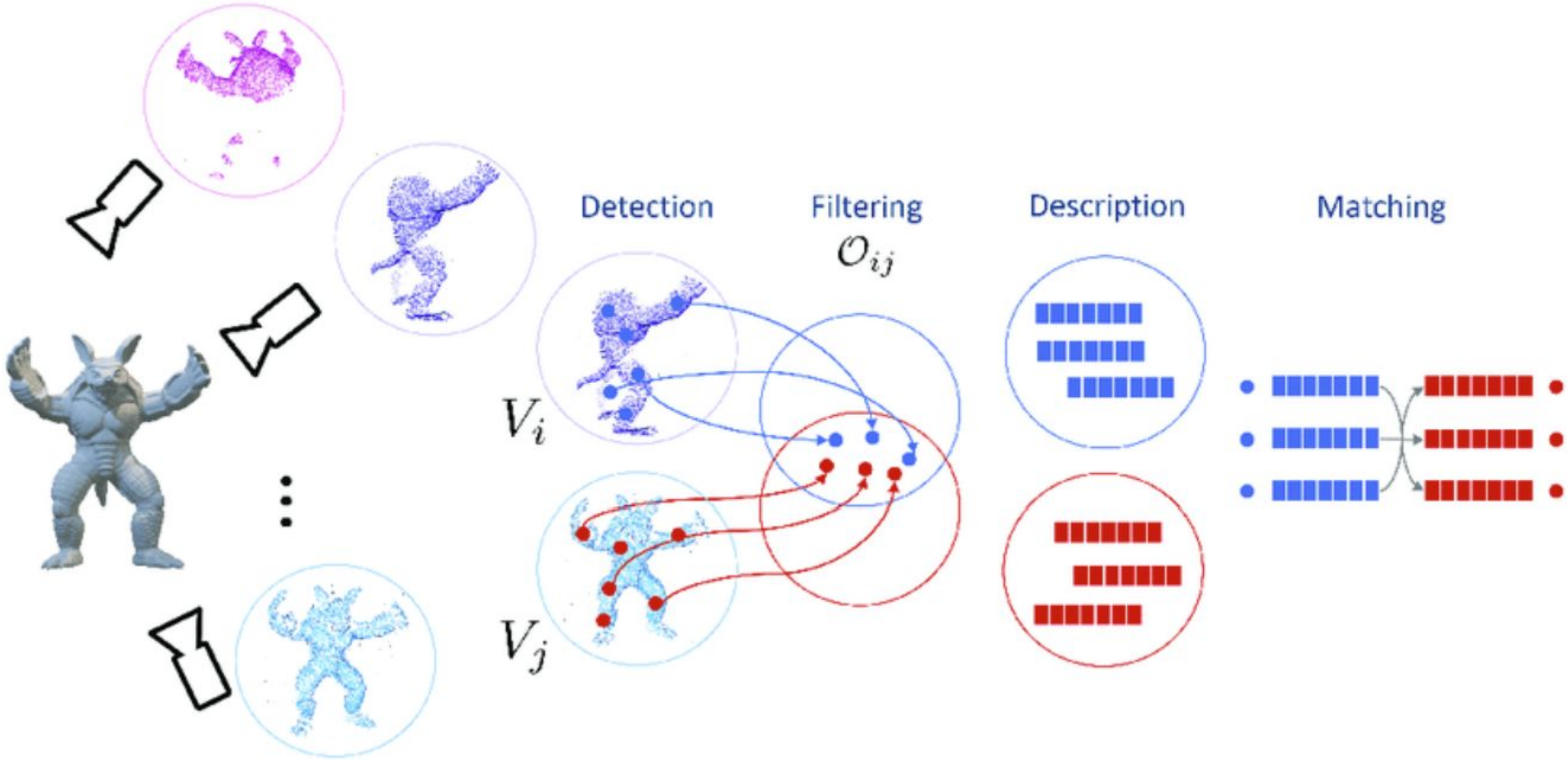
Feature matching Applications



SLAM (Simultaneous Localization and Mapping)

Video Stabilization

What are the Key Components of Classical Feature Matching?

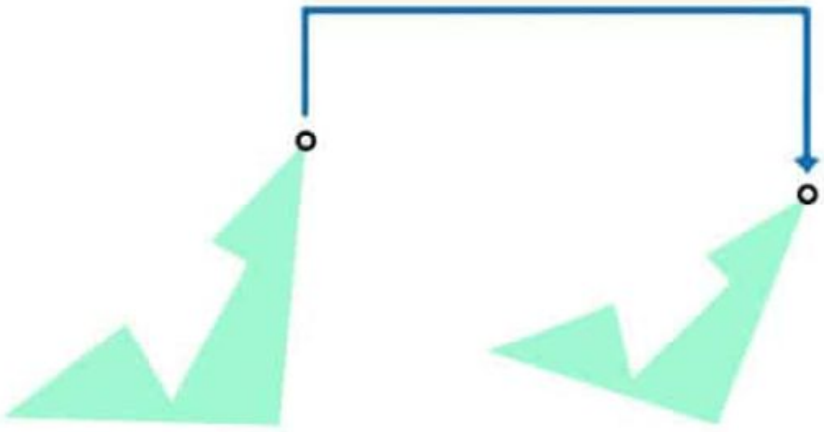


Feature Matching Method Overview

DETECTOR

Detects corners

DESCRIPTOR



Defines area around the corner
Used for feature matching

Traditional algorithms

David G. Lowe invented **Scale Invariant Feature Transform** or **SIFT** in 2004.

It had a huge impact on computer vision, and until the advent of Deep Learning, SIFT was the dominant handcrafted feature. It is based on the DoG (Difference of Gaussian Detector).

SIFT was a good descriptor invariant to the above transformations, but it was slow and was patented by the University of British Columbia. So you had to pay to use SIFT.

Fun fact: The SIFT patent expired in March 2020, and the good folks at OpenCV moved the algorithm from opencv-contrib to the main OpenCV repository.

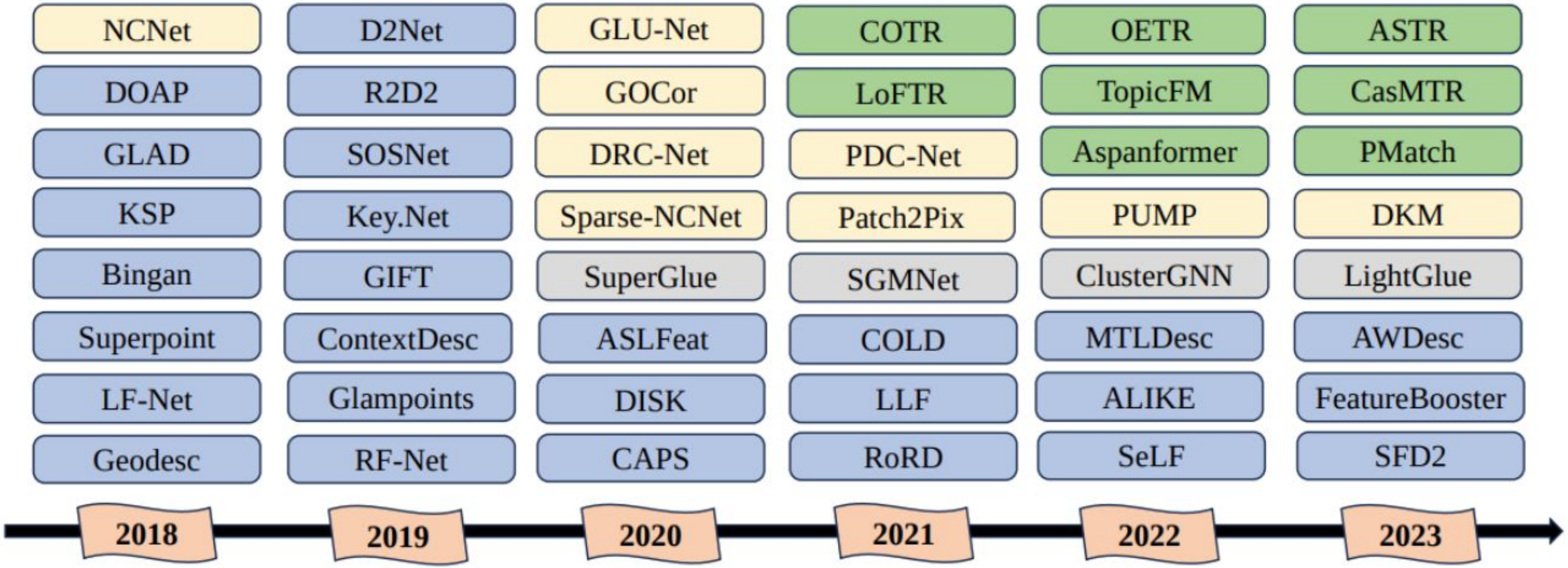
SURF stands for **Speed Up Robust Features.** Released in 2006. It is faster than **SIFT** but again it has the disadvantage of being patented. You need to compile OpenCV from scratch, enabling a certain flag to use these algorithms, i.e., **OPENCV_ENABLE_NONFREE = ON.**

ORB, BRISK, and FREAK are faster and most importantly, not patent-encumbered. **BEBLID** (Boosted Efficient Binary Local Image Descriptor) is another improved algorithm, released recently with OpenCV 4.5.1.

Now, let's understand how **ORB** (Oriented FAST and Rotated BRIEF) detects and computes features. Invented by Ethan Rublee, **ORB** was released in 2011. It uses **FAST** (Features from Accelerated Segment Test) to detect features and to compute descriptors it uses **BRIEF** (Binary Robust Independent Elementary Features). It's been 11 years and ORB still stands strong. Recently, the paper received ICCV 2021 Helmholtz Prize for creating such an impact on computer vision research.

https://learnopencv.com/how-to-build-chrome-dino-game-bot-using-opencv-feature-matching/

Why do we need Deep Learning in Feature Matching?



Evolution of Feature Matching in Deep Learning

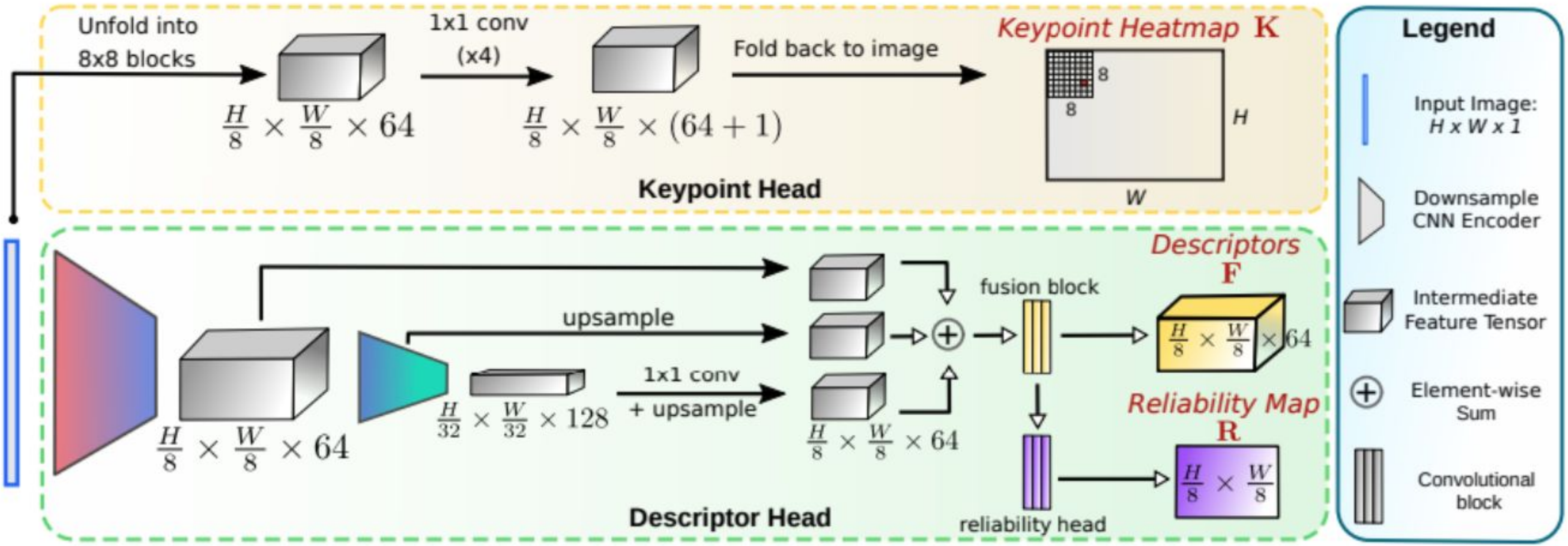
Latest trends in Feature Matching with Deep Learning

- **XFeat:** an optimized deep neural network for feature matching that only uses CPU

- **OmniGlue**: a perfect model that uses Transformers and CNN in the feature matching pipeline.

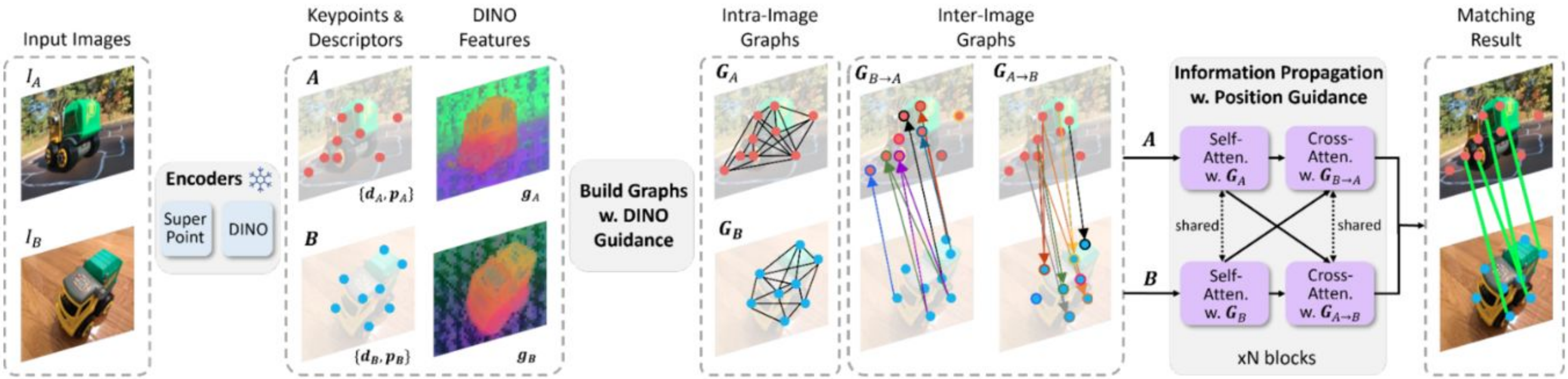
- **LoFTR**: the model that introduced the Transformer architecture into the feature matching pipeline.

XFeat



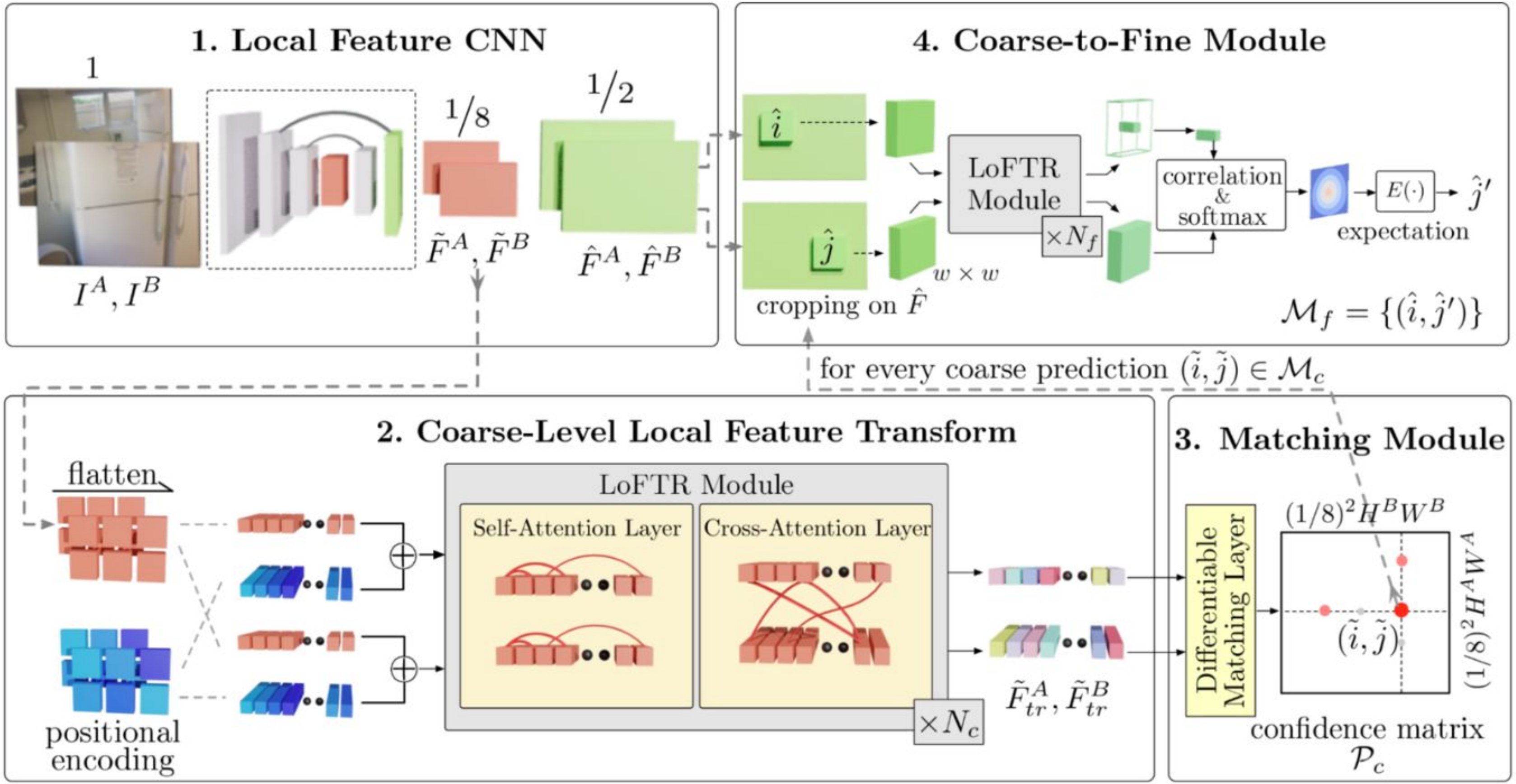
XFeat Model Architecture

OmniGlue



OmniGlue Model Architecture

LoFTR



LoFTR Model Architecture

References

1. Deep Learning Vs. Traditional Computer Vision – A Comparison
2. HCL whitepaper
3. Deep Learning vs. Traditional Computer Vision PDF
4. Traditional machine learning methods
5. Conventional computer vision coupled with deep learning makes AI better
6. What are the pros and cons of using deep learning vs. traditional ML methods?
7. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning (Website)
8. Why Traditional Computer Vision Thrives Alongside Deep Learning: A Counterpoint to Deep Learning Dominance (Website)
9. Deep Learning Architectures for Computer Vision Applications: A Study
10. Introduction to Feature Matching Using Neural Networks (Website)

Questions

Deep learning vs Machine Learning?

Feature Extraction?