GLOBAL ACADEMY OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Subject Name	Project Work Phase II	Subject Code	17CSP85
Student Name	C.P YASHWANTH	USN	1GA17CS035
	MADDALI SOWMYA		1GA17CS080
	RAKSHITHA MURTHY		1GA17CS118
	SWARAJ PARIDA		1GA17CS163
Domain	DEEP LEARNING	Group No:	11
Project Title	“LANDMARK RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS”		
Under taken at	GLOBAL ACADEMY OF TECHNOLOGY		
Guide Name	Prof. KAMLESHWAR KUMAR YADAV		

Agenda

- Introduction about the domain
- Literature Survey
- Problem Statement
- Objectives
- Architecture
- Implementation Modules
- Results – Presentation & Demonstration
- Conclusion and Future work
- References

Introduction about the domain

- **Deep Learning:**
  - Deep Learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning.
  - Learning can be supervised, semi-supervised or unsupervised.
  - Depp Learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, bioinformatics, drug design, medical image analysis, and others, where they have produced results comparable to and in some cases surpassing human expert performance.

Introduction about the domain

- **Computer Vision:**
  - Computer Vision (CV) is a field of Machine Learning that deals with how computers can gain high-level understanding from digital images or videos.
  - Some of its tasks include methods for acquiring, processing, analysing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information.

Literature Survey

Title of the paper and year	Methodology	Advantages	Disadvantages
Rich feature hierarchies for accurate object detection and semantic segmentation - 2014	The system • takes an input image • extracts around 2000 bottom-up region proposals • computes features for each proposal using a large convolutional neural network (CNN), and then • classifies each region using class-specific linear SVMs	• provides higher accuracy than CNNs (R-CNN achieves a mean average precision (MAP) of 53.7% on PASCAL VOC 201 for comparison, reports 35.1% MAP)	• training is multi-stage pipeline • training is expensive in time and space • object detection is slow
Fast R-CNN - 2015	• the image is processed with several convolutional and max pooling • then region of interest has been extracted from feature map • each feature vector is fed into a fully connected layers that finally branches into two outputs	• training is single-stage, using a multi-task loss • training can update all network layers • no disk storage is required for feature caching	• most of the time taken by Fast R-CNN during detection is a selective search region proposal generation algorithm. Hence, it is the bottleneck of this architecture.

Literature Survey

Title of the paper and year	Methodology	Advantages	Disadvantages
A Large-Scale Image Retrieval with Attentive Deep Local Features – 2017	• extract dense features from an image by applying a fully convolutional network • using RANSAC and employ the number of inliers as the score for retrieved images	• DELF clearly outperforms all other techniques significantly • DELF has higher recall • Attention helps more than fine-tuning	• pipeline requires less than 8GB memory to index 1 billion descriptors • challenges in query image with no correct match
1st Place Solution to Google Landmark Retrieval - 2020	• To use metric learning to classify numerous landmark classes and uses transfer learning with two train datasets. • Efficient pooling extract features from images and a deep neural network is followed to squeeze features into small dimensions.	• Google landmark dataset version 2 is the biggest landmark dataset containing human made and natural landmarks. • The model presents a cleaner version of Google Landmark dataset version 2 using automatic data cleaning system.	• The Google Landmark dataset version 2 is constructed by mining web landmark images, so it becomes very noisy. • In the cosine softmax method, the margin value was set to 0 so trying to cluster more between same classes could make training more difficult.

Literature Survey

Title of the paper and year	Methodology	Advantages	Disadvantages
Google Landmark Dataset v2 A Large-Scale Benchmark for Instance-Level Recognition and Retrieval - 2020	• Uses Google Landmark dataset for large-scale, fine-grained, instance recognition and image retrieval in the domain of human made and natural landmarks. • The dataset used consists of over 5M images and 200k distinct instance labels.	• It addresses the following challenges of industrial landmark recognition systems, like large-scale	• Visual inspection of retrieval and recognition results. • It showed that many errors were dude to missing ground truth annotations. • Training consumes a lot of time.

Problem Statement

- Landmark recognition on Google landmark dataset using various algorithms. The goal is to efficiently recognize objects in an image at an instance level, just not at the base level.

Objective

- To make a model which recognizes landmarks from an image using different algorithms such as Visual Geometry Group (VGG) and Deep Local Feature (DeLF)

Implementation Modules

1. Image Acquisition
2. Image Pre-processing
3. Feature Extraction
4. Geometric Verification
5. Final Predictions

Image Acquisition

- Functionality:- Fetching images from the given Universal Resource Locators (URLs)
- Input:- Universal Resource Locators (URLs)
- Output:- Customized images

Image Pre-processing

- Functionality:- Re-sizing images and converting images into NumPy arrays
- Input:- Images of different sizes
- Output:- Uniform sized images

Feature Extraction

- Functionality:- Obtaining the location and feature vectors
- Input:- NumPy array
- Output :- Array of location and feature vectors

Geometric Verification

- Functionality:- Verifying query image with database image
- Input:- Query image and database images
- Output:- Number of inliers among the matched images

Final Prediction

- Functionality:- To retrieve the most similar image
- Input:- Number of inliers
- Output:- Image with the highest number of inliers

Results

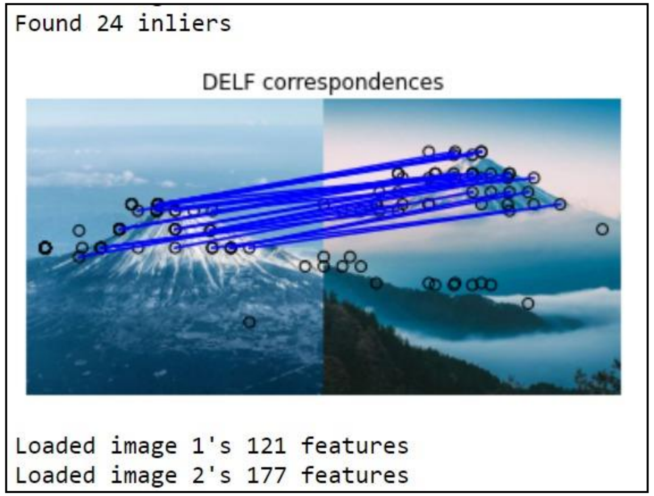

Figure 2: Snapshot of Mount Fuji being compared to itself

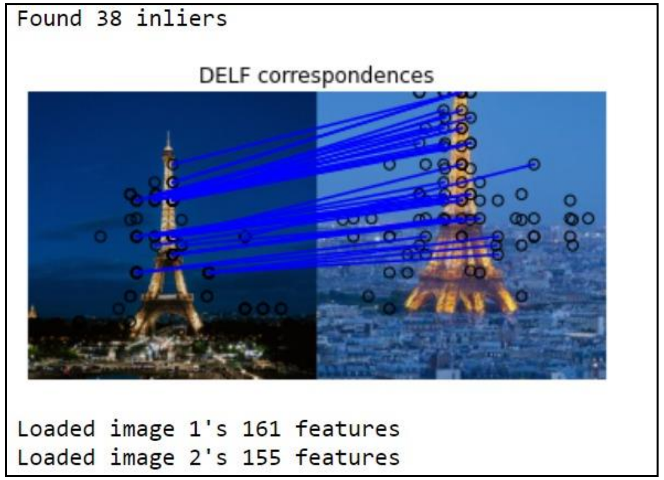

Figure 3: Snapshot of Eiffel Tower being compared to itself

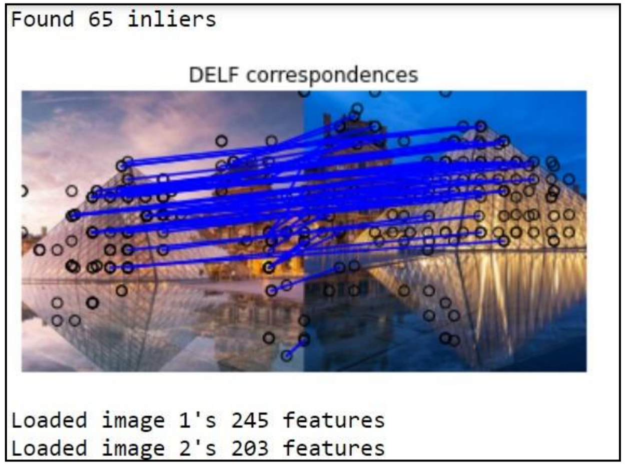

Figure 4: Snapshot of Louvre Museum being compared to itself

Conclusion and Future work

- In image processing there is always a trade-off between scalability and accuracy.
- Hence, this technology can be a solution to predict landmark labels directly from image pixels, to help people better understand and organize their photo collections.
- For further research, we can invariably increase the dataset by adding monuments and structures from all over the globe, as of right now there are 200 different landmarks with 2000 different images of the same.
- A full-featured smartphone application can be prepared using the algorithm and pre-trained dataset.
- This application can act as a guide to tourists visiting these monuments or famous architectures.

References

[1] André Araujo, Bingyi Cao, Jack Sim, Tobias Weyand. (2020). *Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval*. IEEE. pp. 2575-2584. DOI: 10.1109/CVPR42600.2020.00265

https://ieeexplore.ieee.org/document/9157053

[2] André Araujo, Bohyung Han, Hyeonwoo Noh, Jack Sim, Tobias Weyand. (2017). *A Large-Scale Image Retrieval With Attentative Deep Local Features.* IEEE. pp. 3456-3465. DOI: 10.1109/ICCV.2017.374

https://ieeexplore.ieee.org/document/8237636

[3] Christof Henkel, Philipp Singer. (2020). *Supporting Large-scale Image Recognition with out-of-domain Samples*. eprint arXiv:2010.01650

https://arxiv.org/abs/2010.01650

[4] Jeff Donahue, Jitendra Malik, Ross Girshick, Trevor Darrell. (2014). *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.* IEEE. DOI: 10.1109/CVPR.2014.81

https://ieeexplore.ieee.org/document/6909475

[5] Ross Girshick. (2015). *Fast R-CNN*. IEEE. pp. 1440-1448. DOI: 10.1109/ICCV.2015.169

https://ieeexplore.ieee.org/document/7410526

[6] SeungKee Jeon. (2020). *1st Place Solution to Google Landmark Retrieval.* eprintarXiv:2009.05132

https://arxiv.org/abs/2009.05132#:~:text=The%20solution%20is%20based%20on,enhance%20the%20model's%20performance%20further.

Thank You

Q & A