DeepVision Tool

An Interative GUI & Server Solution for Information Extraction from Image

MENTOR
DR. AVLEEN KAUR MALHI
CPG 14

101512043 SAGAR SHIVANI
101690011 ASHISH RANA
10152062 YUVRAJ VERMA
101562009 SHAUNAK DIXIT

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

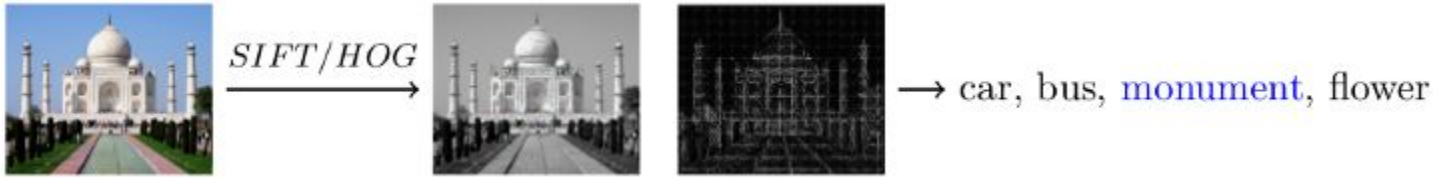
Table Of Contents

- Background Of Project
- Scope And Utility
- Problem Statement & Goals
- Solution & Progress Workflow
- Architecture Used
- Implementation Details
- Tools & Technology Used
- Working Screenshots of Project
- Key Highlights
- Assumptions & Constraints
- Model Architectures
- Video Attachment
- Literature and Standard Used
- Individual Roles & Outcomes

Background For Project

DEEP VISION TOOL

Machine Learning requires domain knowledge and feature engineering for computer vision tasks. For large data they tend to become very complex..



Deep Learning models learns feature engineering concepts in hierarchical manner. But, still software development in this domain quite unexplored.

Scope & Utilities

1

- Visual Question Answering module answers objective questions. For example,
  - What is the color of bottle in image ?
  - How many number of sheeps are there ?
- Image Captioning module answers relational and context questions. For example,
  - What is that man doing in image ?
  - Describe the image ?
- Chatbot module answers normal, information and personal assistance questions with some memory. For example,
  - What is your name or my name ?
  - Can you book my flight tickets ?
- Integration and Deployment as Open source Software.
- Chat assistant incorporated to GUI application.
- Client Server model implementation methodology documented.

Problem Statements

2

- Preprocessing Steps Required With ML as feature engineering.
- Are GUI deployment solution for Deep Learning models effective.
- Visual QA & Activity Classification information extraction capabilities needed.
- Applications often requires chatbot as a helper tool.

Goals

- Fully deployable solution for information extraction with scalability.
- Incorporating helper chabots and its effective practices.
- Optimal model processes for eliminating feature engineering.

Solution

- A GUI application with chatbot capabilities for information extraction from images with Deep Learning Models.
- A server based online solution for less powerful machines.

Progress Workflow

1

NLP Chatbot Module

General Purpose with personality, pre-coded rules & functionalities.

2

Visual QA Module

A CNN network trained on MS-COCO dataset for VQA.

3

GUI & Live Website

Interactive GUI app | live website for results & informative purposes.

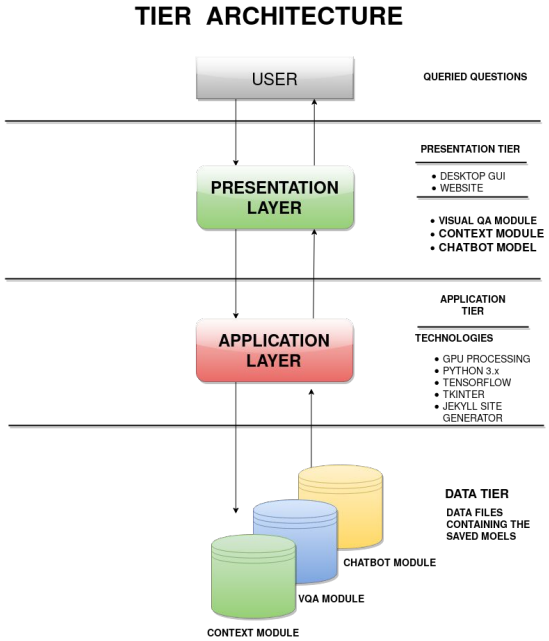
4

Activity Classification Module

An encoder VGG16 & decoder LSTM trained on Flickr dataset.

Architecture Used

Python based Desktop GUI is simply event-driven architecture with independent module implementation.
Other architecture for deployment is client-server architecture. With live website and backend handled with Formspree.io
Individual model architectures are explained in the end.



Client Server Architecture

Implementation Details

Tkinter GUI

Event Driven Programming based GUI which triggers modules, help page as a subprocess.

Website

Facilitate test image uploading, Informatory website containing all theoretical & implementation details.

Visual QA Module

CNN network trained on MS-COCO dataset.

Gives probability of all possible solutions.

NLP Chatbot Module

Seq2Seq model trained on Reddit cleaned dataset. With personality, pre-coded information & general purpose usage.

Context Classification Module

CNN + LSTM encoder

-decoder trained on Flickr 8K dataset. Generates a summarizing statement.

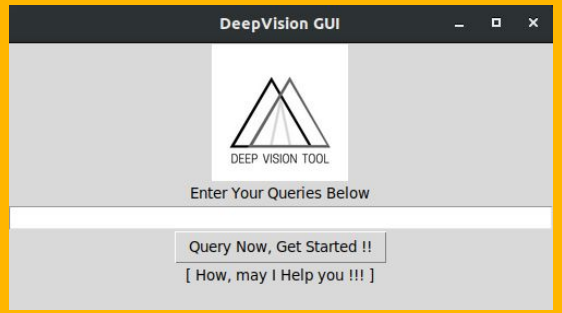
Tools & Technologies

Python, Both 2.x and 3.x is used for development in project
Tensorflow
Tkinter
Keras
Regex

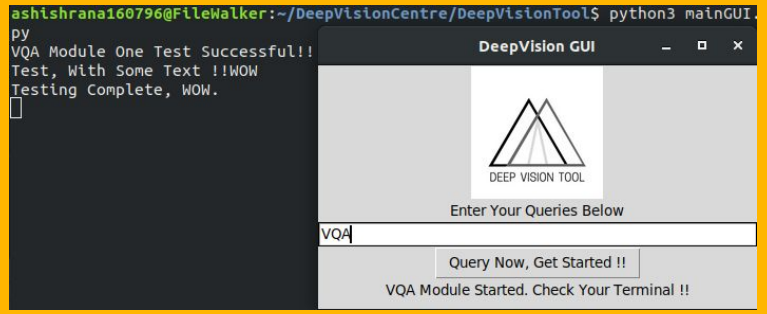
Jekyll
HTML5, CSS, JS
Formspree.io

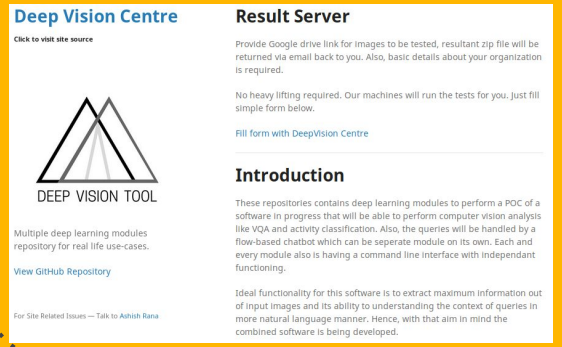
Convolutional Neural Network
Recurrent Neural Network
Seq2Seq Model
Shell Commands
Pyttsx Engine
Space Word2Vec library

Application ScreenShots

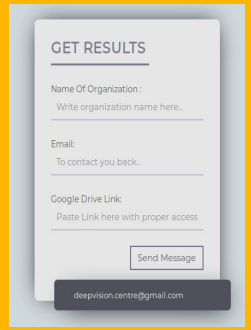


GUI Application





Website with Form for uploading Test Images.



Receive Results Via Email.

Key Highlights

Two deployment techniques explored and implemented, GUI & client-server based.
Three different deep learning models trained for three different modules.
Independent functionality of each module is achieved.
Awareness created for Open source projects with Hacktoberfest 2018.

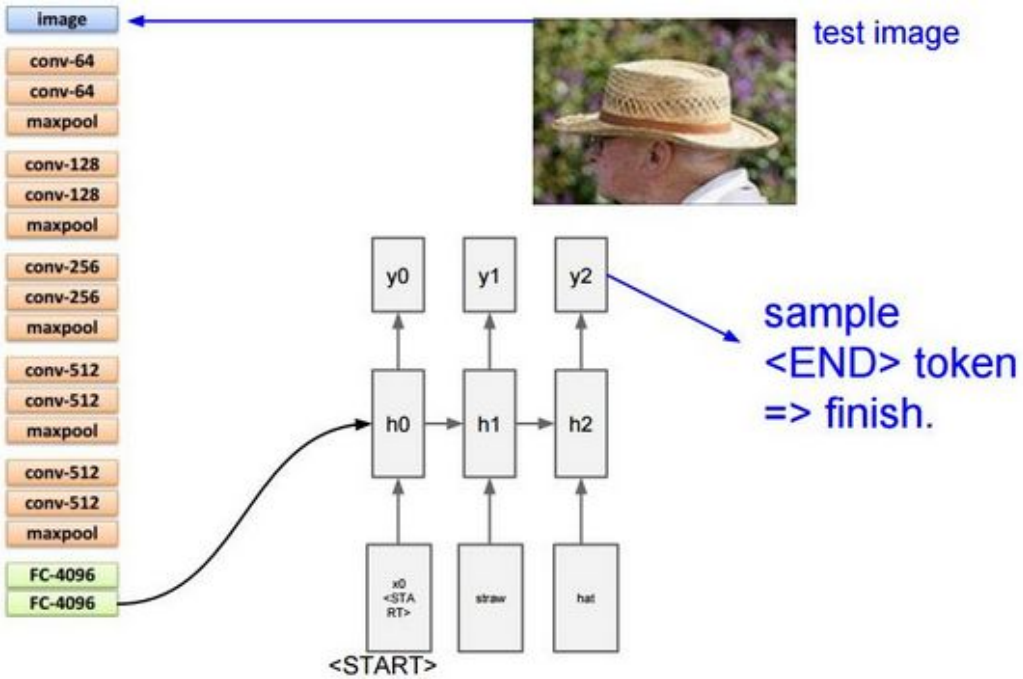
Assumptions & Constraints

- Models are scalable for large testing datasets.
- DOS & DDOS attacks are not considered while developing website.
- User will only extract one kind of feature at a time i.e. one module at a time from GUI.
- Single query at a time for given module.
- User is aware about opening help page.
- Standard dataset images with no random noises.
- Version consistency for APIs & libraries used amongst modules.
- Different regularization & initialization aren't tested with limited GPU access.

Model Architectures

VGG CNN Architecture with 138M parameters

A VGG Encoder with LSTM Decoders in a RNN.



Existing Literature

- ‘Visual7W’ dataset based on the MS COCO for Visual QA, largest gap in performance recorded.
- ‘Visual Madlibs’ dataset with fill-in-the-blanks approach focus on Natural Language Description, CNN+LSTM used but no compositional natural language understanding.
- DAQUAR dataset from NYU indoor dataset. For spatial relation finding of Visual QA.
- Neural Image-QA for caption generation, CNN + RNN trained end-to-end.
- 3-CNN architecture for encoding, composing question words & classification learning for candidate answer words.

Standards Used

Project is made in compliance to IEEE standards with documentation, SRS, SQA, V&V and design specifications being the main focus..

Project Video

https://www.youtube.com/watch?v=FJuqCjpXYAo&list=PLbgWYmmfedhBcD_7wv0CEyjR2fQEQyyEY

Individual Contributions

- **NLP Chatbot Module:** Ashish Rana & Shaunak Dixit
- **Visual Q/A Module:** Shaunak Dixit & Yuvraj Verma
- **Image Captioning Module:** Sagar Shivani & Yuvraj Verma
- **Desktop GUI | Website & Backend:** Ashish Rana
- **Open Source Deployment:** Ashish Rana & Sagar Shivani
- **Report Completion:** Everyone From Group
- **Specifications & Diagrams:** Everyone From Group
- **Video:** Everyone From Group

A little learning is dangerous thing. Drink Deep or taste not the Pierian Spring.

Thank you !!