UE18CS390A – Capstone Project Phase – 1

SEMESTER - VI

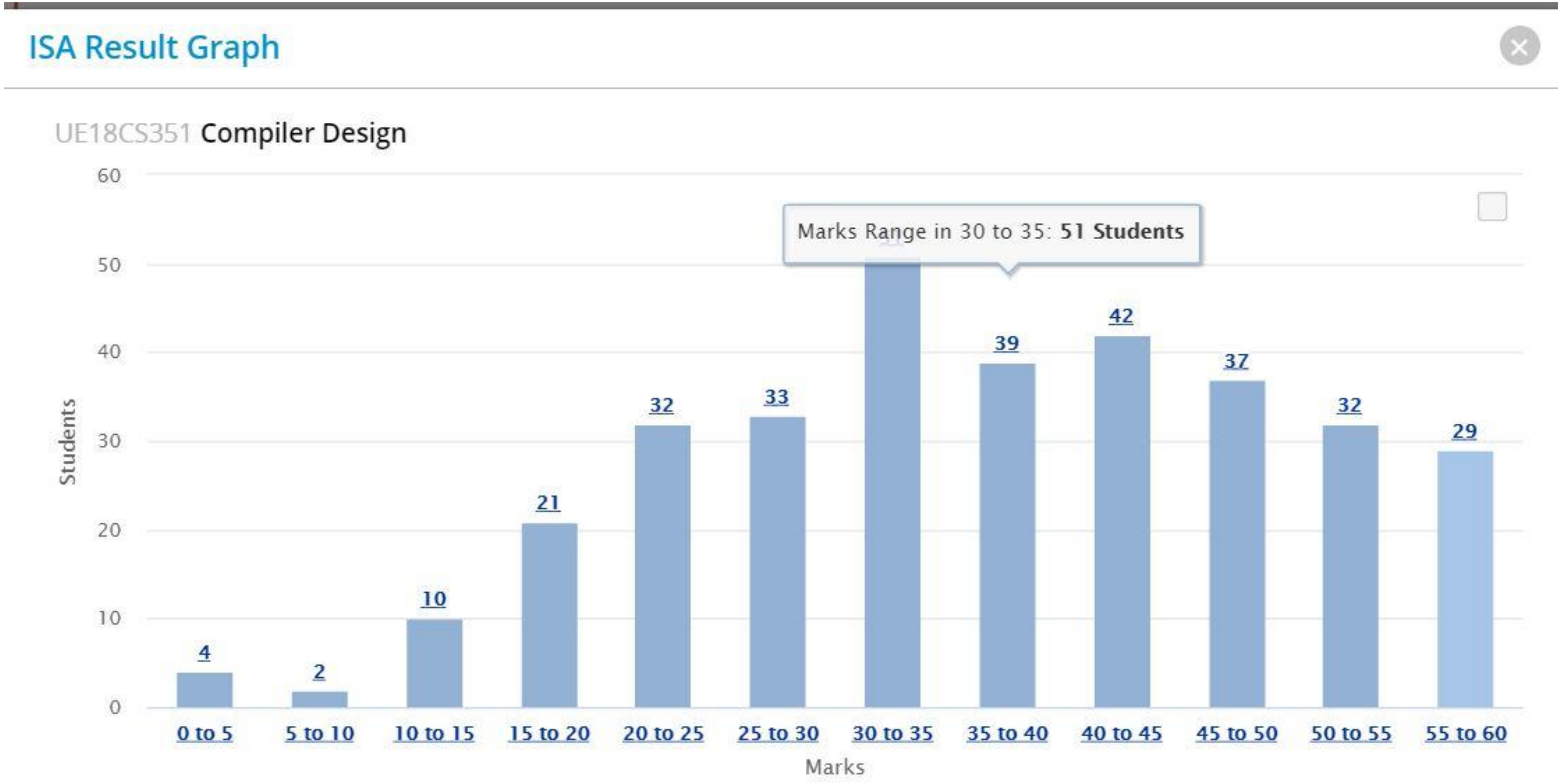
END SEMESTER ASSESSMENT

Project Title : Visual Question Answering On Statistical Plots
Project ID : PW22MHR02
Project Guide : Dr. Mamatha H.R.
Project Team : Sneha Jayaraman PES1201802825
Sooryanath I T PES1201802827
Himanshu Jain PES1201802828

OUTLINE

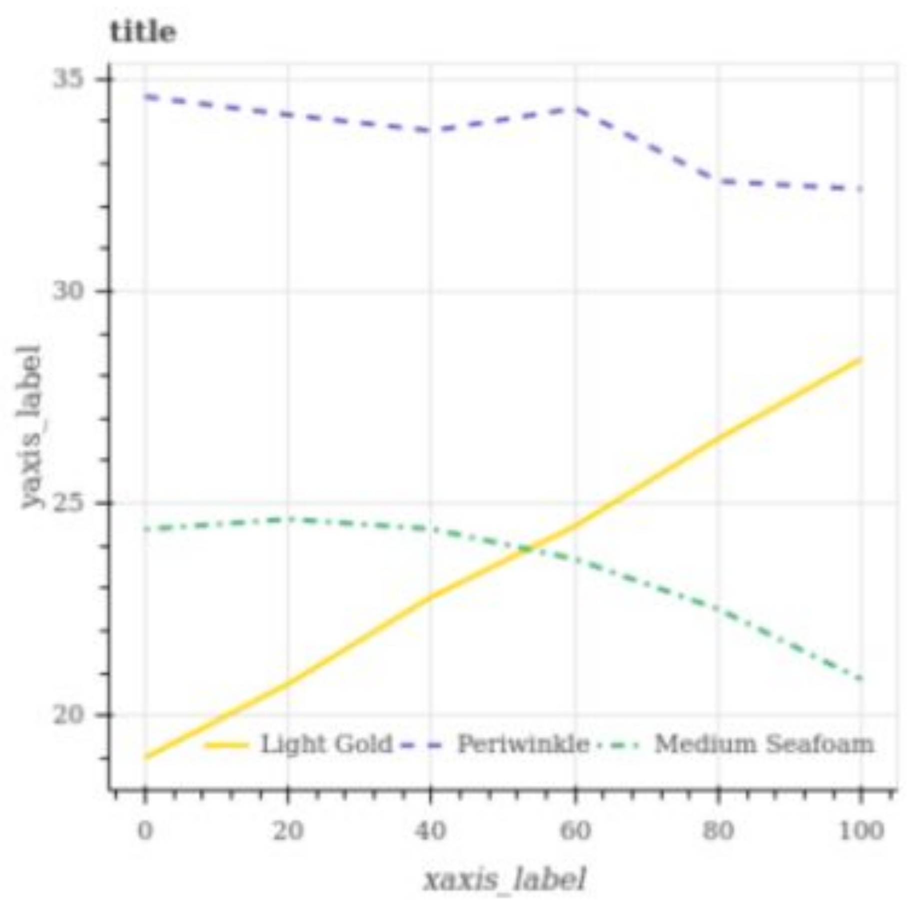
- Problem Statement
- Abstract and Scope
- Literature Survey
- Suggestions from Review – 3
- Design Approach
- Design Constraints, Assumptions & Dependencies
- Proposed Methodology / Approach
- Architecture
- Design Description
- Technologies Used
- Project Progress
- References

MOTIVATION



ISA Result Graph
UE18CS351 Compiler Design
Marks Range in 30 to 35: 51 Students
60
50
40
30
20
10
0
Students
4
2
10
21
32
33
39
42
37
32
29
0 to 5
5 to 10
10 to 15
15 to 20
20 to 25
25 to 30
30 to 35
35 to 40
40 to 45
45 to 50
50 to 55
55 to 60
Marks

PROBLEM STATEMENT



Q: Does Medium Seafoam intersect Light Gold?
A: Yes

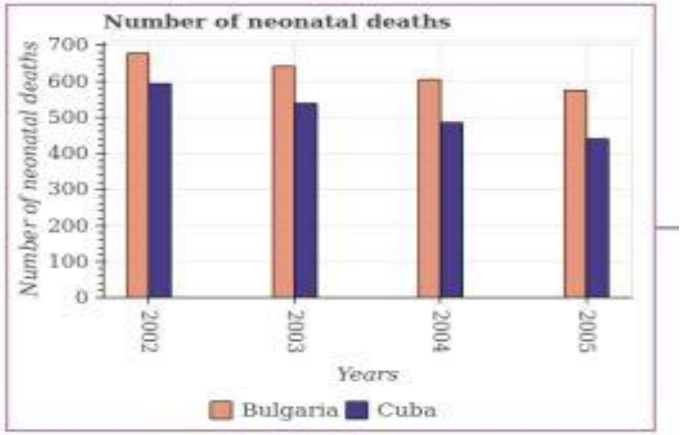
Q: Is Medium Seafoam the roughest?
A: No

Q: Is Light Gold less than Periwinkle?
A: Yes

Q: Does Periwinkle have the maximum area under the curve?
A: Yes

Q: Does Medium Seafoam have the lowest value?
A: No

PROBLEM STATEMENT



Q: How many bars are there in the image?
Q: What is the average number of neonatal deaths in Cuba per year?

- **This is a Grouped bar graph**
- **X axis : Years , Y axis : Count of Neonatal deaths**
- **We have 2 keys (Purple : Cuba , Peach : Bulgaria)**
- **We now have queries based on the elements within the graph**
- **They may be directly extracted from the statistical plot image or may need an out of vocabulary treatment**

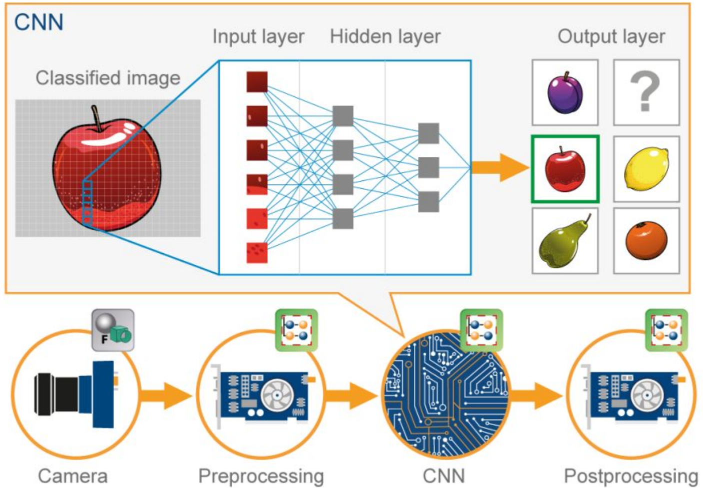
PROBLEM STATEMENT

The aim of the project is to build a **Visual Question Answering system** which accepts **statistical plots** along with **questions** on the plot with respect to the elements of the plot (such as intersection of the curves, area under the curve, median value and few other varieties of such relational queries) and provide answers to the questions posed.

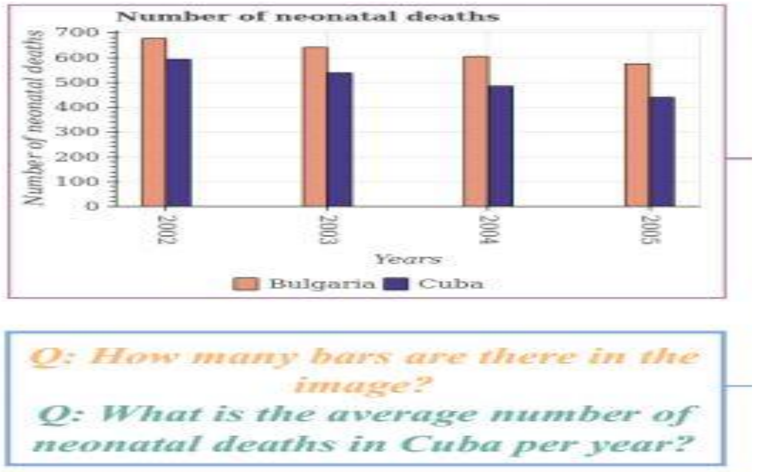
In short, the aim is to build a system that can discover relationships between elements of a plot and provide results to answer questions on the plot.

ABSTRACT

We have Q/A systems becoming prominent

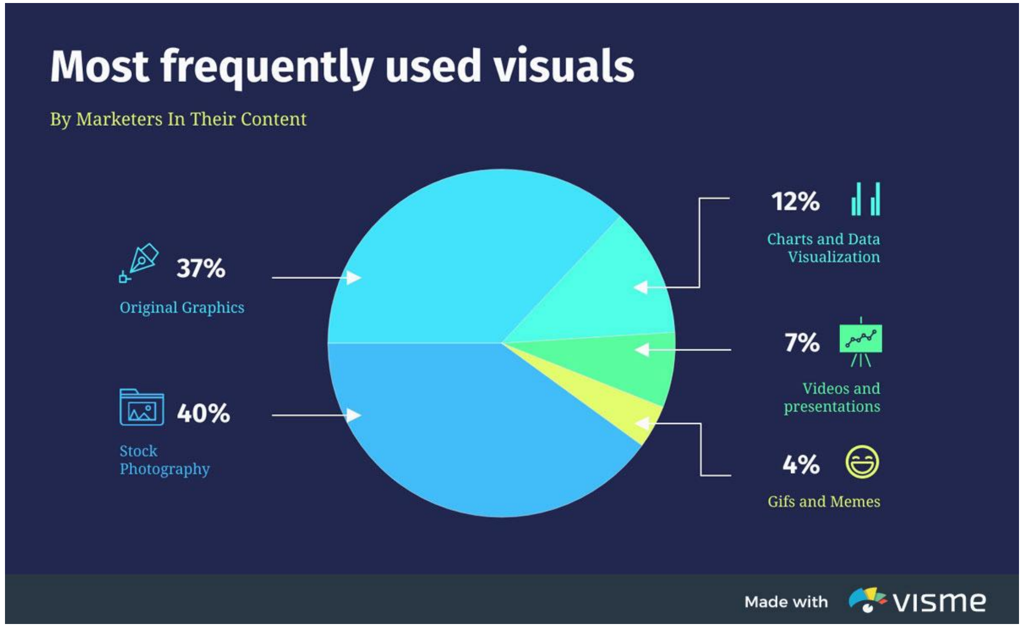


We have image classification ,
information extraction from images

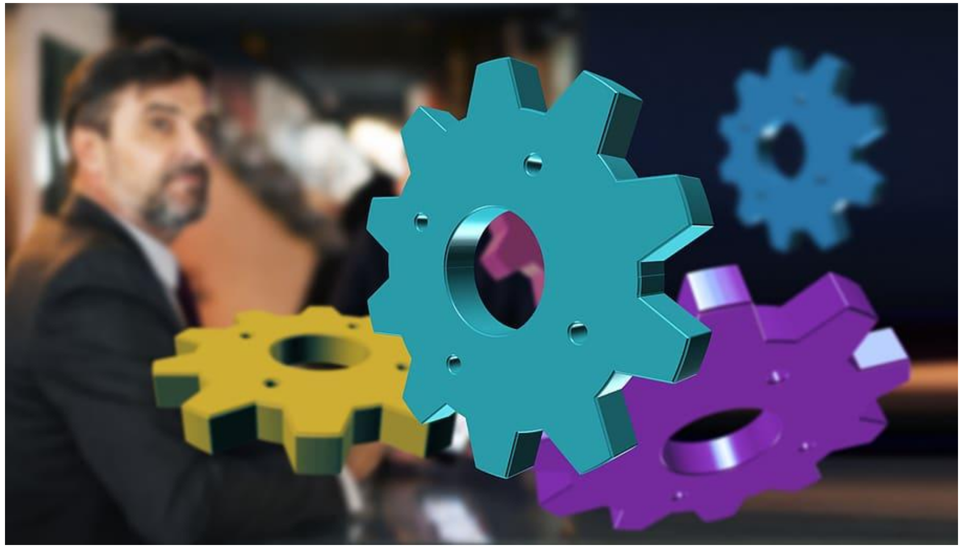


ABSTRACT

Statistical charts are used on a regular basis for data visualization to interpret the data and derive meaningful inference from them.



This process can be automated using Question answering systems. Users can impose a question to the system, for a particular statistical chart, and the system must provide the answer to it in the most accurate manner.

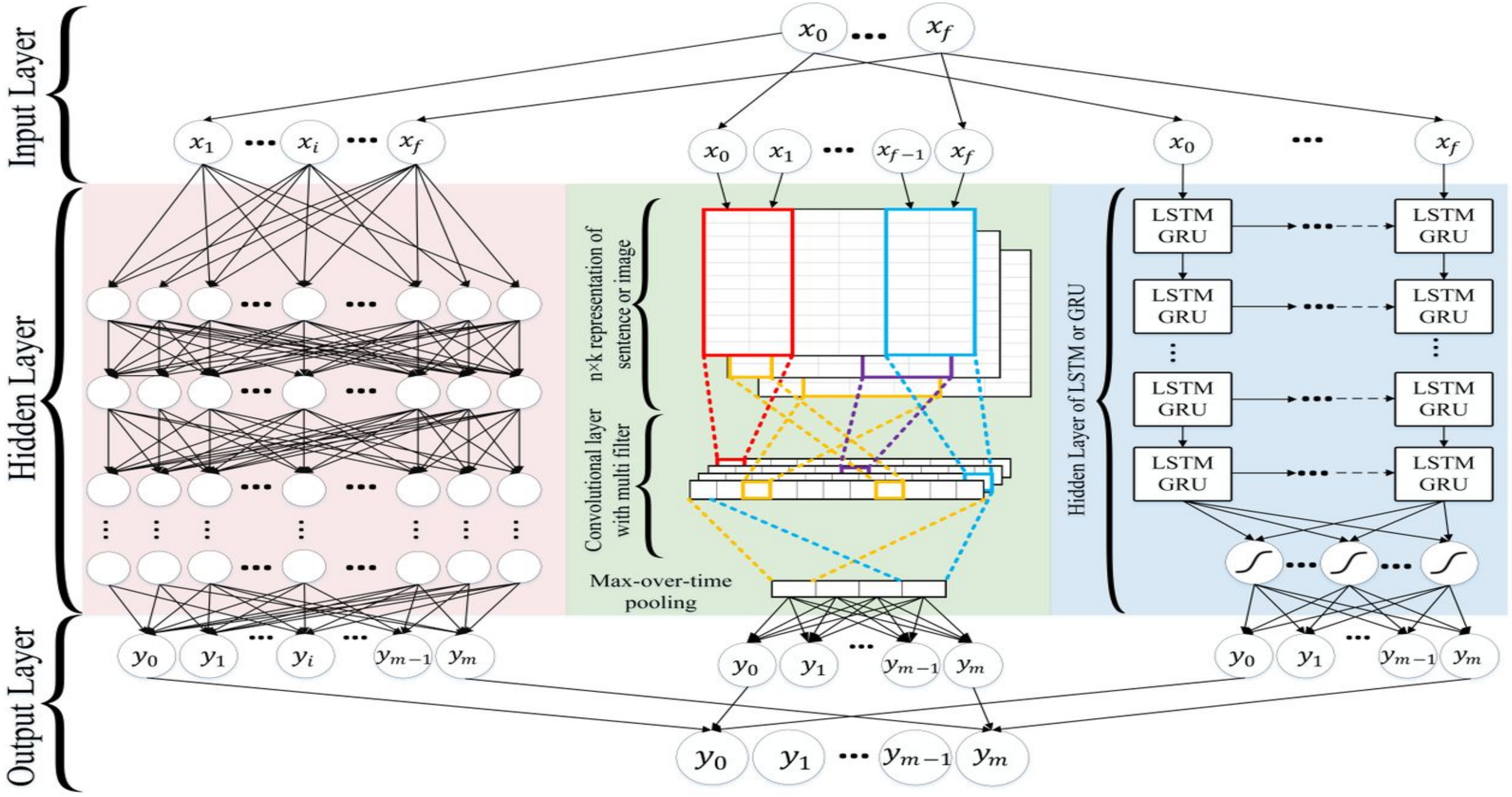


ABSTRACT

- Statistical plots used on a regular basis to interpret data and derive meaningful inferences.
- This process can be automated using Question answering systems.
- Users can impose a question to the system, for a particular statistical plot, and the system must provide the answer to it in the most accurate manner.
- Help data analysts question and reason plots on a large scale, and automate the decision-making capabilities.
- Building such a system requires the usage of right architecture, right frameworks and huge amounts of data. Once the model is built that satisfies the requirements, it can be deployed to a web application where a user can upload an image and, input a question, to generate the expected answer.

SCOPE

- Visual Question Answering involves Computer Vision, Natural Language Processing and Deep Learning.



SCOPE

- The chosen problem statement is based on Visual QA specific to statistical plots , which has less prevalent work done but has opened the door towards visual reasoning with respect to patterns in statistical plots.

- Given an visual image of a statistical plot and a corresponding question, the model must be able to generate a representation of the image, parse and understand the query, and generate a suitable reply.

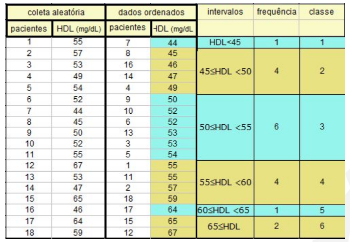
- It involves an understanding of image and the query language to be able to provide for visual reasoning.

- The system will not be able to answer questions on the smoothness or the roughness of the plots. It can only answer relational queries - queries with respect to the other elements of the plot.

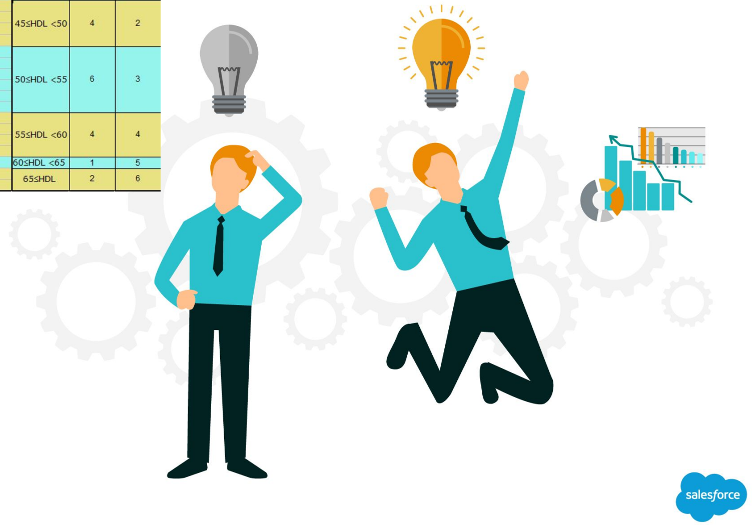
USE CASES/ APPLICATIONS

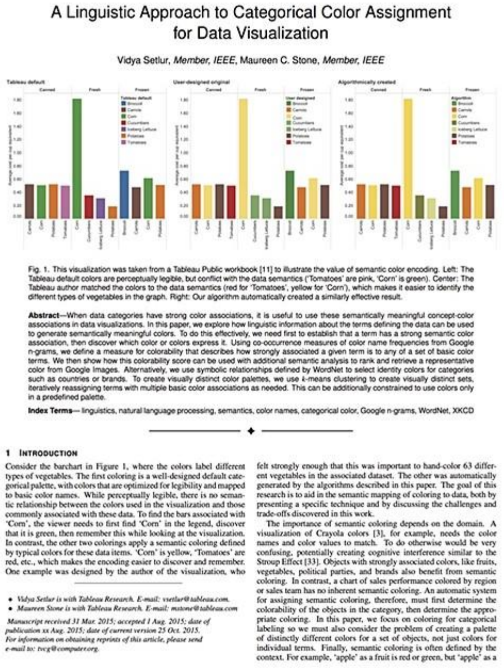
Statistical charts are an intuitive and simple way to represent data. Deep Learning focuses on emulating human intelligence to develop new models that can reason figures and understand relationships .

Visual plots are commonly found in research papers, scientific journals, business records e.t.c. Therefore, automation of plot analysis through the means of question-answering aids an individual to draw statistical inferences quickly from them.



coleta aleatória		dados ordenados		intervalos	frequência	classe
pacientes	HDL (mg/dL)	pacientes	HDL (mg/dL			
1	55	7	44	HDL<45	1	1
2	57	8	45	45≤HDL <50	4	2
3	53	16	46			
4	49	14	47			
5	54	4	49			
6	52	9	50	50≤HDL <55	6	3
7	44	10	52			
8	45	6	52			
9	50	13	53			
10	52	3	53			
11	55	5	54			
12	67	1	55	55≤HDL <60	4	4
13	53	11	55			
14	47	2	57			
15	65	18	59			
16	46	17	64	60≤HDL <65	1	5
17	64	15	65	65≤HDL	2	6
18	59	12	67			





A Linguistic Approach to Categorical Color Assignment for Data Visualization

Vidya Setlur, *Member, IEEE*, Maureen C. Stone, *Member, IEEE*

Fig. 1. This visualization was taken from a Tableau Public workbook [11] to illustrate the value of semantic color encoding. Left: The Tableau default colors are perceptually legible, but conflict with the data semantics ('Tomatoes' are pink, 'Corn' is green). Center: The Tableau author matched the colors to the data semantics (red for 'Tomatoes', yellow for 'Corn'), which makes it easier to identify the different types of vegetables in the graph. Right: Our algorithm automatically created a similarly effective result.

Abstract—When data categories have strong color associations, it is useful to use these semantically meaningful concept-color associations in data visualizations. In this paper, we explore how linguistic information about the terms defining the data can be used to generate semantically meaningful colors. To do this effectively, we need first to establish that a term has a strong semantic color association, then discover which color or colors express it. Using co-occurrence measures of color name frequencies from Google n-grams, we define a measure for colorability that describes how strongly associated a given term is to any of a set of basic color terms. We then show how this colorability score can be used with additional semantic analysis to rank and retrieve a representative color from Google Images. Alternatively, we use symbolic relationships defined by WordNet to select identity colors for categories such as countries or brands. To create visually distinct color palettes, we use *k*-means clustering to create visually distinct sets, iteratively reassigning terms with multiple basic color associations as needed. This can be additionally constrained to use colors only in a predefined palette.

Index Terms— linguistics, natural language processing, semantics, color names, categorical color, Google n-grams, WordNet, XKCD

1 INTRODUCTION

Consider the barchart in Figure 1, where the colors label different types of vegetables. The first coloring is a well-designed default categorical palette, with colors that are optimized for legibility and mapped to basic color names. While perceptually legible, there is no semantic relationship between the colors used in the visualization and those commonly associated with these data. To find the bars associated with 'Corn', the viewer needs to first find 'Corn' in the legend, discover that it is green, then remember this while looking at the visualization. In contrast, the other two colorings apply a semantic coloring defined by typical colors for these data items. 'Corn' is yellow, 'Tomatoes' are red, etc., which makes the encoding easier to discover and remember. One example was designed by the author of the visualization, who felt strongly enough that this was important to hand-color 63 different vegetables in the associated dataset. The other was automatically generated by the algorithms described in this paper. The goal of this research is to aid in the semantic mapping of coloring to data, both by presenting a specific technique and by discussing the challenges and trade-offs discovered in this work.

The importance of semantic coloring depends on the domain. A visualization of Crayola colors [3], for example, needs the color names and color values to match. To do otherwise would be very confusing, potentially creating cognitive interference similar to the Stroop Effect [33]. Objects with strongly associated colors, like fruits, vegetables, political parties, and brands also benefit from semantic coloring. In contrast, a chart of sales performance colored by region or sales team has no inherent semantic coloring. An automatic system for assigning semantic coloring, therefore, must first determine the colorability of the objects in the category, then determine the appropriate coloring. In this paper, we focus on coloring for categorical labeling so we must also consider the problem of creating a palette of distinctly different colors for a set of objects, not just colors for individual terms. Finally, semantic coloring is often defined by the context. For example, 'apple' as a fruit is red or green, but 'apple' as a

- *Vidya Setlur is with Tableau Research. E-mail: vsetlur@tableau.com.*
- *Maureen Stone is with Tableau Research. E-mail: mstone@tableau.com*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

LITERATURE SURVEY

Research Papers Reviewed

1. Answering Questions about Charts and Generating Visual Explanations
2. FigureNet: A Deep Learning model for Question-Answering on Scientific Plots
3. ChartNet: Visual Reasoning over Statistical Charts using MAC-Networks
4. PlotQA: Reasoning over Scientific Plots

LITERATURE SURVEY - Paper 1

Paper Title: Answering Questions about Charts and Generating Visual Explanations

Year of Publication: April 2020

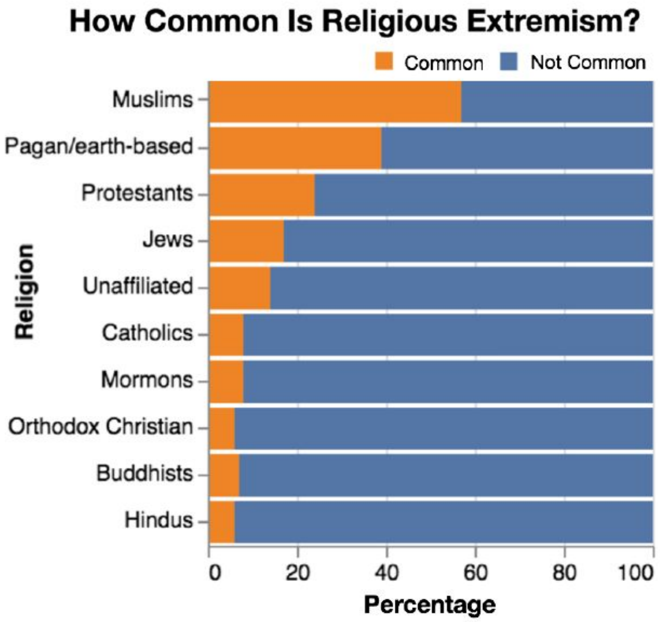
Authors: Dae Hyun Kim, Enamul Hoque, Maneesh Agrawala

Data: https://dl.acm.org/doi/pdf/10.1145/3313831.3376467

Summary:

Paper proposed the chart question answering system that generates chart specific answers along with the explanation on how the answer was obtained. The visual attributes of the charts are transformed into references to the data. State-of-the-art ML algorithms are used to generated answers and its corresponding explanation.

DATA AND OUTCOME



Q1: *What is the percentage of response 'Common' for Catholics?*

A(Sempre): ***92***

A(Ours): ***8***. *I looked up the length of the orange bar for 'Catholics'.*

Q2: *Which religion has the longest orange component?*

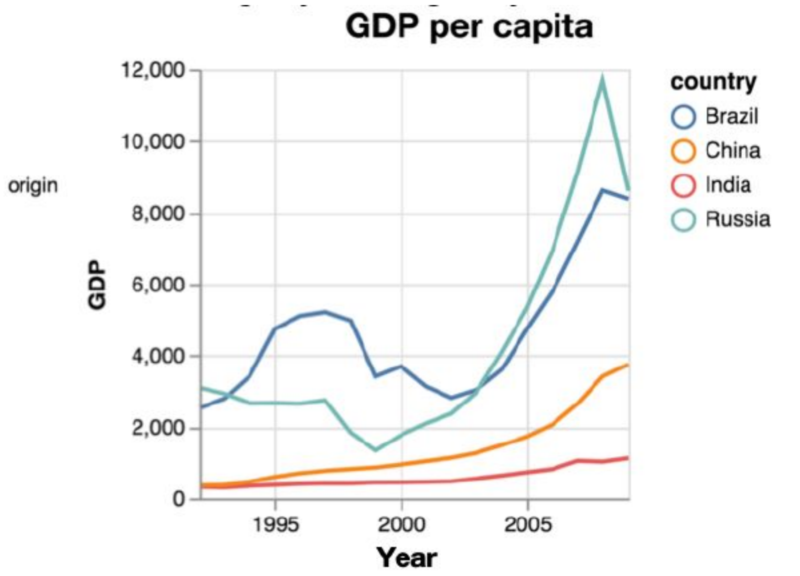
A(Sempre): ***Hindus***

A(Ours): ***Muslims***. *I looked up 'Religion' of the longest orange bar.*

Q3: *What does the blue field represent?*

A(Sempre): ***24***

A(Ours): ***Not Common***. *I looked up what blue represents by looking at the legend.*



Q15: *Which country has the highest GDP per capita in year 2005?*

A (Sempre): ***Russia***

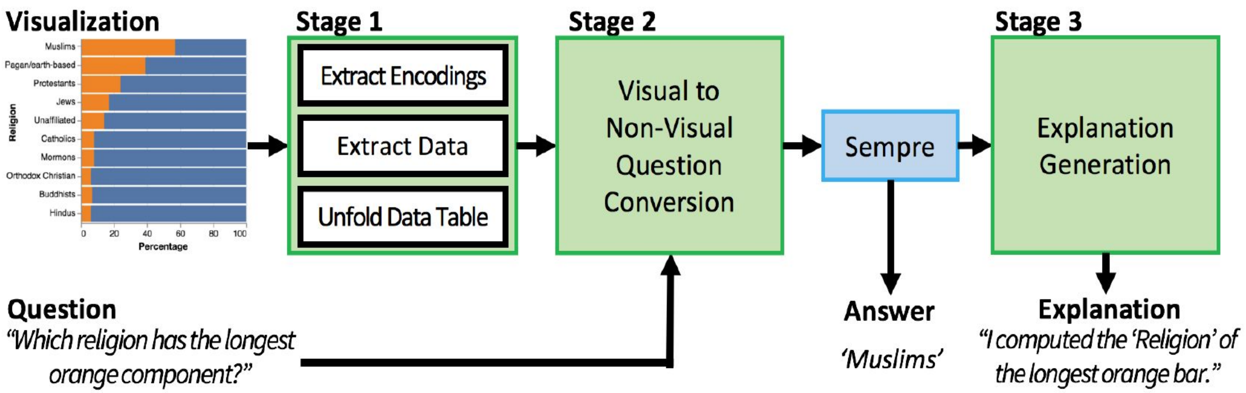
A (Ours): ***Russia***. *I looked up 'country' of the highest line for '2005'.*

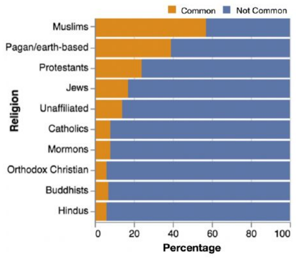
Q16: *How many times do Brazil and Russia flip in terms of GDP ranking?*

A (Sempre): ***18***

A (Ours): ***2***. *I counted the number of the blue line or the cyan line.*

MODEL






```
"data": {"url": "data/kong/data/3.csv"},
"transform": [
    {"filter": "datum.year == 2000"},
    {"filter": "datum.question == 'Extremism'"}],
"mark": "bar",
"encoding": {
    "x": {"field": "Percentage", "type": "quantitative"},
    "y": {"field": "Religion", "type": "nominal"},
    "color": {
        "field": "Response", "type": "nominal",
        "scale": {
            "domain": ["Common", "Not common"],
            "range": ["#EE8426", "#5376A7"]}}}
```

Religion	Response	Percentage
Muslims	Common	57
Muslims	Not common	43
Pagan/earth-based	Common	39
Pagan/earth-based	Not Common	61
⋮	⋮	⋮
Hindus	Common	6
Hindus	Not Common	94

(a) Flat data table

Religion	Common	Not common
Muslims	57	43
Pagan/earth-based	39	61
Protestants	24	76
Jews	17	83
⋮	⋮	⋮
Buddhists	7	93
Hindus	6	94

(b) Unfolded data table

METHODOLOGY

- Firstly, visual encodings like the height of the bar, color of the line, etc. are extracted from the charts.
- The input question is transformed, replacing any visual references to chart elements with non-visual references to data fields and data values.
- The Unfolded table, the transformed non-visual question in passed through Sempre (question answering algorithm that works with relational data tables instead of charts) to generate the answer.
- Sempre converts the input natural language question into a logical query called a lambda expression, and then executes the query on the data table to generate the answer.
- Finally, we convert the lambda expression from Sempre into a visual explanation for the answer, using template-based translation.

MERITS & DEMERITS

Merits

- The paper not only provides accurate answers to the questions but also provides an explanation on how the answer was obtained
- The system generates correct answers and explanations for many questions that **Sempre** cannot answer correctly

Demerits

- The system cannot handle certain types of questions that involve synonyms of the features present in the chart
- Improving the explanation provided for the answers

LITERATURE SURVEY - Paper 2

Paper Title: FigureNet: A Deep Learning model for Question-Answering on Scientific Plots

Year of Publication: July 2019 (2019 International Joint Conference on Neural Networks (IJCNN))

Authors: Revanth Reddy, Rahul Ramesh, Ameet Deshpande, Mitesh M. Kapra

Data: https://arxiv.org/pdf/1806.04655

Summary: Use of CNN with depth-wise convolutions, LSTM and feed-forward NN to handle the task of visual question answering on bar plots and pie charts on the FigureQA dataset and give out a binary response.

END-TO-END MODEL ARCHITECTURE

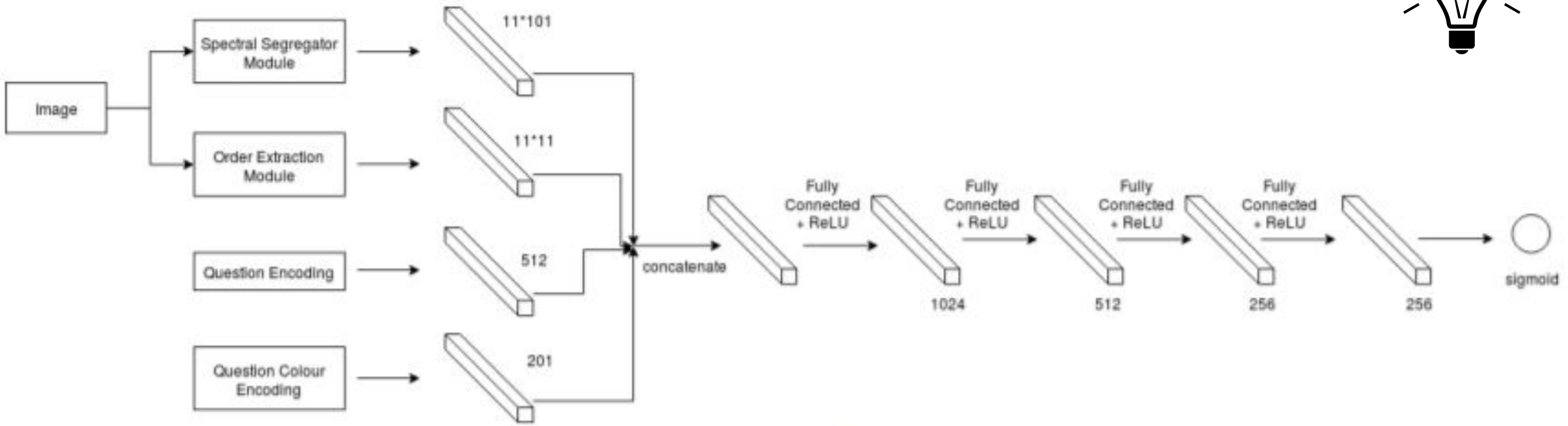


Fig. 7: Architecture of final feedforward network

ARCHITECTURE OF THE SPECTRAL SEGREGATOR MODULE

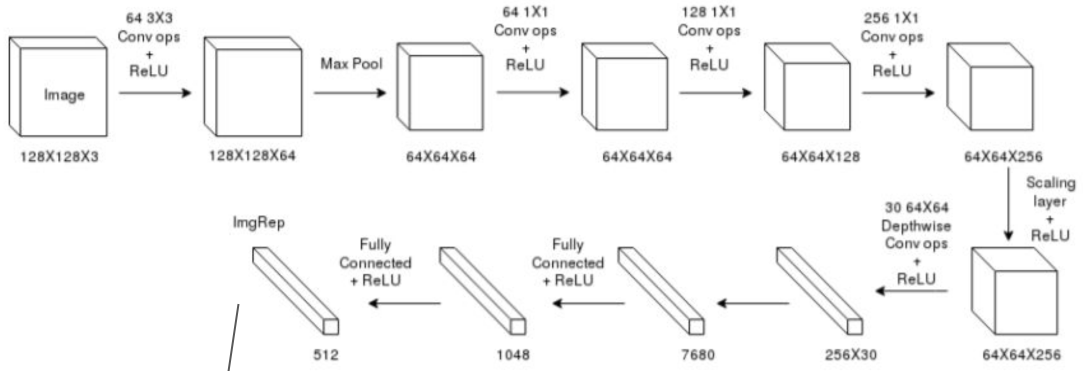


Fig. 4: Architecture of Spectral Segregator Module - Image visualizes the sequence of convolution operations

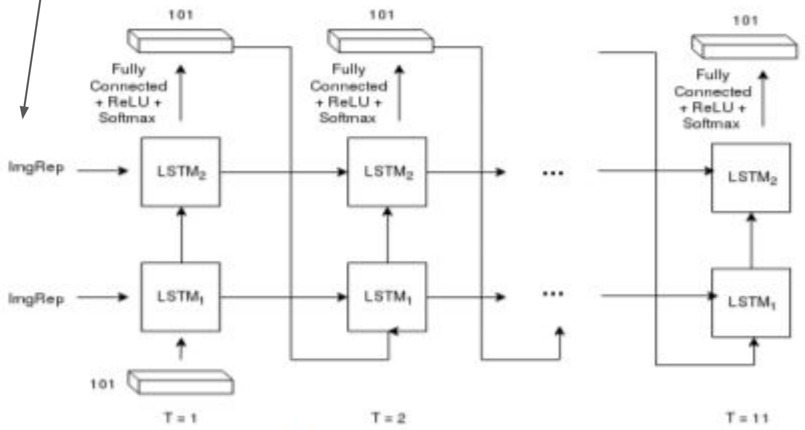


Fig. 6: Architecture of Spectral Segregator Module - Custom LSTM architecture

A Convolutional Neural network for obtaining a representation of the image

A Sequential LSTM Network that completes the Spectral Segregator Module

METHODOLOGY

Spectral Segregator Module:

- Identify plot elements and color of the plot elements
- 128 x 128 x 3 image is passed as input to a CNN that uses depth-wise convolutions to identify colors and separate channel information. This way we don't just get an aggregate map of the image. The output here is a 512 dimensional image representation.
- This image representation is passed to a 2-layer LSTM to get a probability distribution across the 100 images and an additional STOP label(used for no plot element)

Order Extraction Module:

- Identify and quantify the statistical values of each plot element and their relative order
- Similar to that of the previous module expect that now the output of the LSTM will be the ordering for each of the plot elements starting from 1.

Question Encoding and Question Color encoding:

- Uses LSTM

Final feed-forward NN:

- All the four modules are concatenated and passed onto a feed forward NN to produce a binary (Yes/No) answer (using Sigmoid Activation)

MERITS & DEMERITS

Merits:

- The model performs significantly better than the baseline models.
- This is because the architecture doesn’t use the traditional CNN, instead uses depthwise convolutions
- Lesser training time as articulated in the paper

Demerits

- The model works on only bar plots and pie charts
- It is capable of only binary reasoning, and not capable of answering open-ended questions
- It makes use of the Figure QA dataset, thereby making use of the property of the charts being color coded.

LITERATURE SURVEY - Paper 3

Paper Title: ChartNet: Visual Reasoning over Statistical Charts using MAC-Networks

Year of Publication: July 2019

Authors: Monika Sharmaa, Shikha Guptab, Arindam Chowdhurya, Lovekesh Vig

Data : https://arxiv.org/pdf/1911.09375.pdf

Summary:

The paper solves the problem of reasoning over statistical charts (only bar charts and pie charts) using MAC-Network (Memory, Attention, and Composition). The model is capable of answering open-ended questions and gives chart-specific answers. The classification layer of MAC is replaced by the regression layer and constructs a bounding box around the text of the answer. OCR is used to read the text and display the answer.

MODEL

ChartNet: Visual Reasoning over Statistical Charts using MAC-Networks

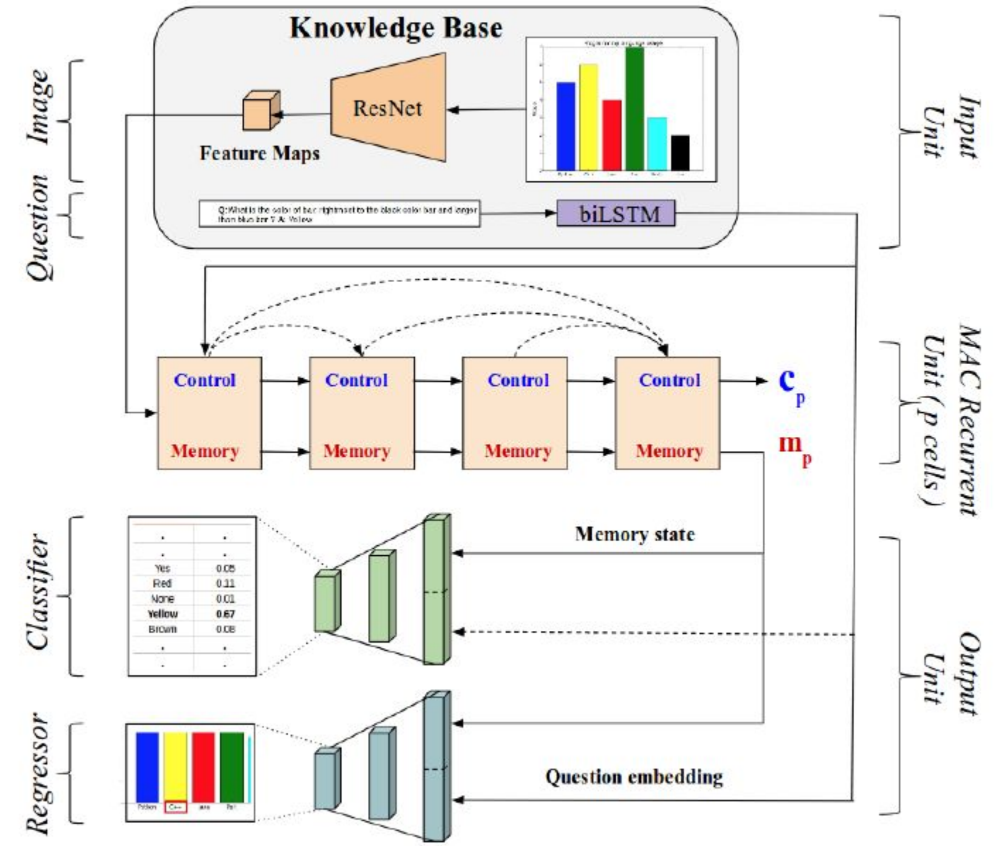


Fig. 2. Flowchart showing proposed architecture of *ChartNet* for visual reasoning over bar and pie charts. The Knowledge base consists of visual feature maps extracted using a ResNet-101 [5] pre-trained model. The question is encoded using a Bidirectional LSTM. A recurrent MAC layer is used to generate the reasoning output at each step, based on the question and two fully connected branches perform classification over a generic set of answers and regress the coordinates of image specific answers.

METHODOLOGY

ChartNet network consists of **three layers**:

- Input unit
- MAC Cell
- Output unit

Input Unit

- Bar or pie chart is given as input and corresponding question
- Features from the images are extracted using ResNet101 deep CNN architecture
- Knowledge base is defined to represent the height and width of the processed image
- Question is converted into a word embedding and further processed by biLSTM

MAC Cell

- Recurrent unit which consists of three components : Control, Read and Write
- Defined to learn fundamental reasoning operations and implement them

Output Unit

- This unit consists of two networks : Classifier and Regressor
- Classifier network predicts a probability distribution over the pre-defined set of generic answers through a softmax normalization
- Regressor network is used to provide chart-specific answers

MERITS & DEMERITS

Merits

- Automated method for question answering over open-ended questions
- MAC-Network appended with a regression layer helps the model make prediction over unseen answers

Demerits

- The model is not generic and works only for vertical bar charts and pie charts
- Model cannot answer questions that require numerical operations

LITERATURE SURVEY - Paper 4

Paper Title: “PlotQA: Reasoning over Scientific Plots”

Year of Publication: 2020

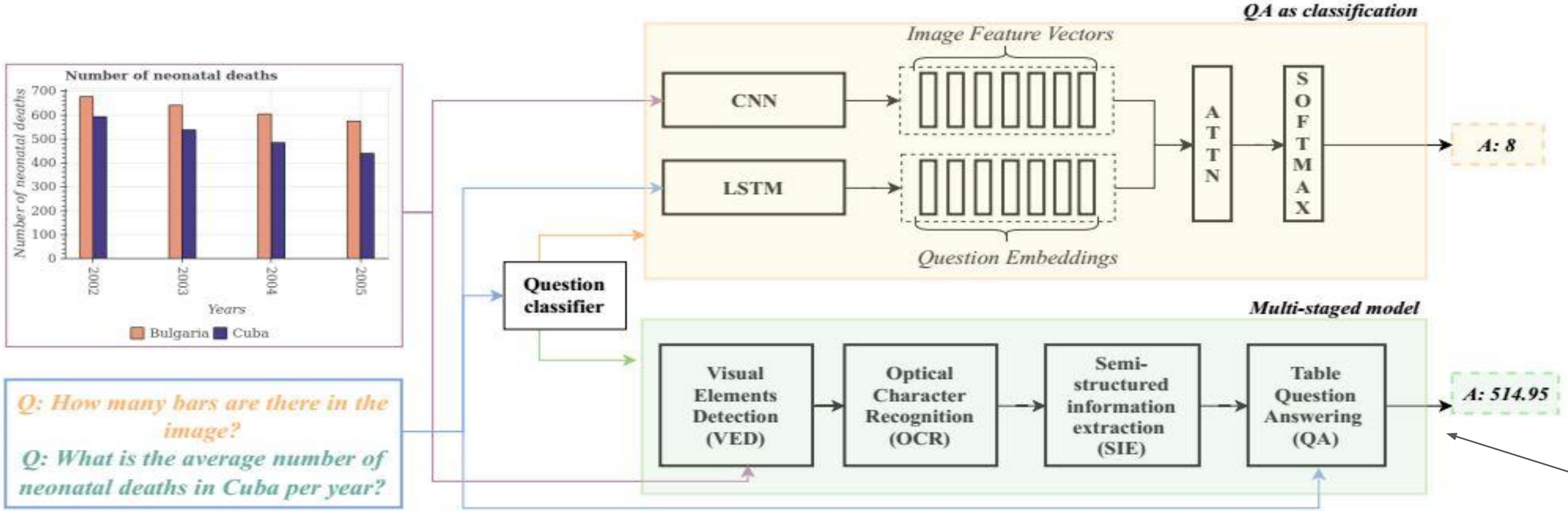
Authors: Nitesh Methani , Pritha Ganguly,
Mitesh M.Khapra and Pratyush Kumar.

Data: PlotQA : Link to Official Paper Work

Summary:

A step towards developing a holistic plot based visual question answering model , which can handle both in vocabulary and open ended queries using a hybrid approach.

MODEL



The final stage of the pipeline is to answer questions on the semi-structured table. As this is similar to answering questions from the WikiTableQuestions dataset , this work adopts the same methodology .

The first task is to extract all these visual elements by drawing bounding boxes around them and classifying them into the appropriate class.

Some of the visual elements such as title, legends, tick labels, etc. contain numeric and textual data. For extracting this data from within these bounding boxes, we use a stateof-the-art OCR model.

The next stage of extracting the data into a semistructured table.

LITERATURE SURVEY - Paper 4

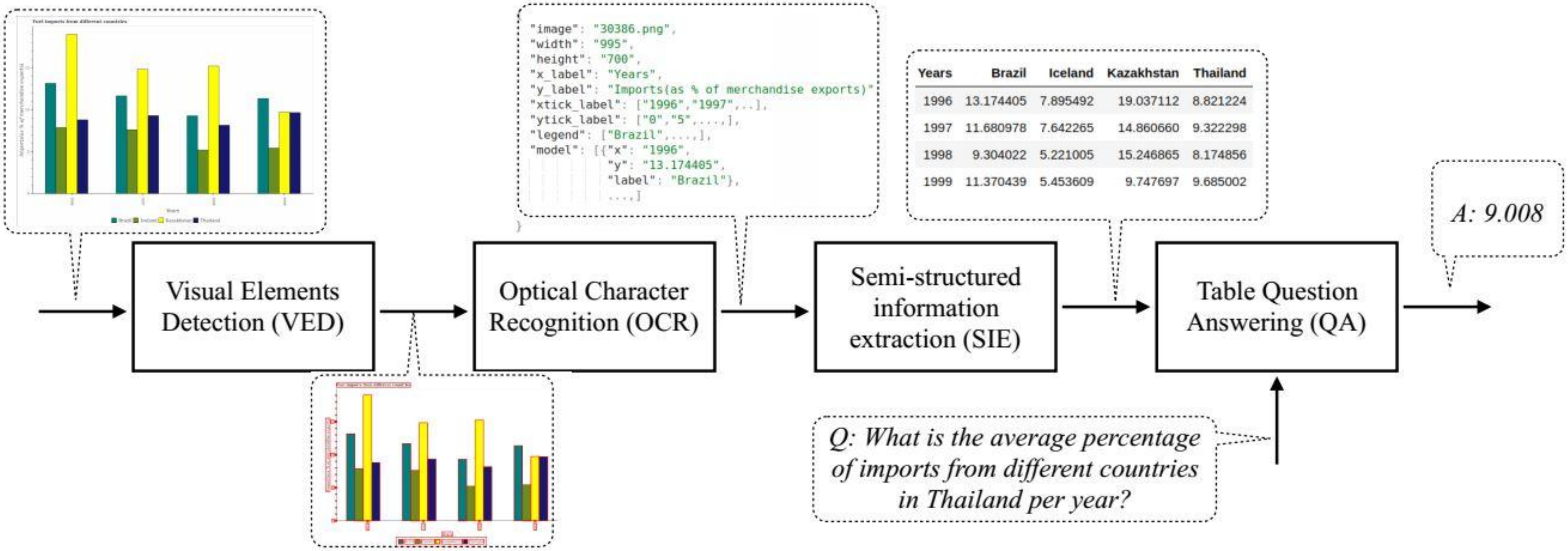


Figure 3: Our proposed multi-staged modular pipeline for QA on scientific plots.

LITERATURE SURVEY - Paper 4

This is a hybrid model containing the following elements: (i) a binary classifier for deciding whether the given question can be answered from a small fixed vocabulary or needs more complex reasoning, and (ii) a simpler QA-as-classification model to answer questions of the former type, and (iii) a multi-staged model containing four components as described below to deal with complex reasoning questions.

1. Visual Elements Detection (VED)

The first task is to extract all these visual elements by drawing bounding boxes around them and classifying them into the appropriate class.

Upon comparing all methods, it is found that Faster R-CNN model along with Feature Pyramid Network(FPN) performed the best and hence we used it as our VED module.

2. Object Character Recognition (OCR):

Some of the visual elements such as title, legends, tick labels, etc. contain numeric and textual data. For extracting this data from within these bounding boxes, the state of-the-art OCR model is used.

3. Semi-Structured Information Extraction (SIE)

4. Table Question Answering (QA)

The final stage of the pipeline is to answer questions on the semi-structured table. As this is similar to answering questions from the WikiTableQuestions dataset , we adopt the same methodology ..

MERITS & DEMERITS

Merits

- Can handle out of the vocabulary questions (OOV) along with in-vocabulary question types.
- The data collected to prepare graphs in the dataset are from various financial and business resources.
- Reduces the gap between existing synthetic plot datasets and real-world plots and question templates.

Demerits

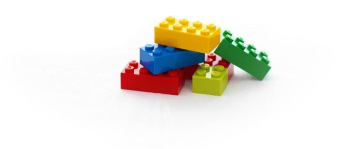
- The model is not generic and works only for bar charts , line charts and dot plots.
- There exists a need for more accurate visual element detection (VED) module to improve reasoning over plots.

PAPER COMPARISON TABLE

	Paper 1	Paper 2	Paper 3	Paper 4
Paper Title	Answering Questions about Charts and Generating Visual Explanations	FigureNet: A Deep Learning model for Question-Answering on Scientific Plots	ChartNet: Visual Reasoning over Statistical Charts using MAC-Networks	PlotQA:Reasoning over Scientific Plots
Similarities with other papers	Uses table based QA technique like paper 4	Forms a part of Paper 4, answering only YES/NO questions	Similar to paper 4 - since it can answer open-ended questions	Can handle OOV questions and in -vocabulary questions.
Differences with other papers	Generates answers along with the reasoning (as to how the answer is obtained).	Uses deep learning techniques such as CNN and LSTM, can only answer binary YES/NO questions	Uses compositional models.	Uses separate pipelines for answering OOV and in-vocabulary questions.The dataset contains 80% of reasoning based QA pairs.
Use case in our project	Dataset available for bar chart and line chart can be used to train our own model	Base paper, to test out the applicability of CNN for image representation, LSTM for question encoding	Paper helps us build and understand DL architecture for open-ended questions	Paper to help analyse how to deal with open-ended questions, given such a dataset

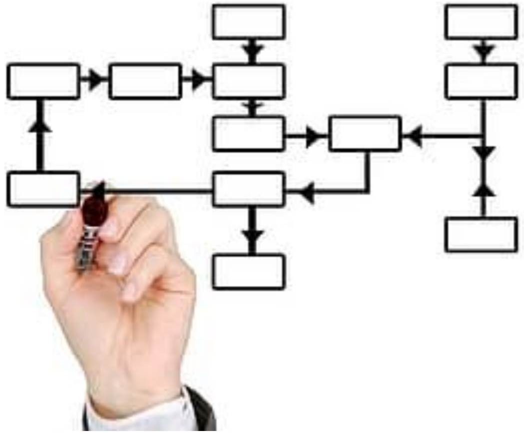
SUGGESTIONS FROM REVIEW-3

1. Usage of Illustrations

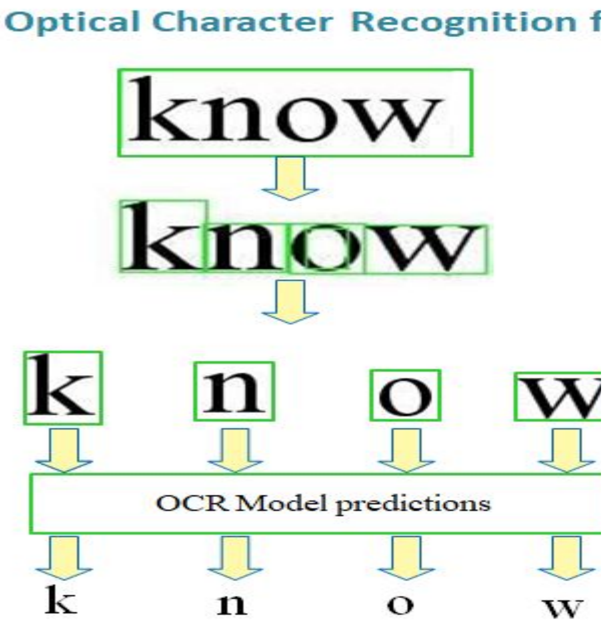


2. Intuitive representations have been included

3. Limited amount of verbose



DESIGN CONSTRAINTS, ASSUMPTIONS & DEPENDENCIES



- The design approach chosen is constrained by the existing OCR and VED (Visual Elements Detection) modules

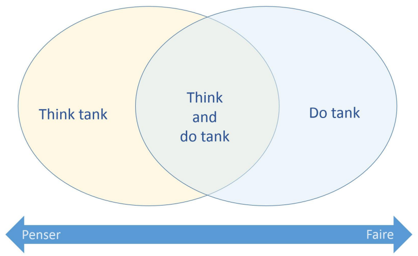
- They serve as the off the shelf components in the implementation

- It is constrained by the GPUs for model training.

DESIGN DETAILS

- NOVELTY

  - One step towards better machine reasoning capabilities.

  - Comprehensive automated model so as to improve accuracy of existing models, with focus on specific charts.

- INNOVATIVENESS

  - Statistical plots, though images, have different properties that make the use of simple object detection models insufficient.
  - Use of scientific deep learning models to explore further enhancements to the relatively new field/area.

- INTEROPERABILITY

  - Any platform irrespective of Operating system

DESIGN DETAILS

- PERFORMANCE

  - Accuracy as a metric to evaluate performance

  - Comparison parameter is with respect to human inference

- MAINTAINABILITY

  - Model is easy to maintain - uses a machine; browser interface

  - Project documentation and plan to support easy upgrade

- PORTABILITY

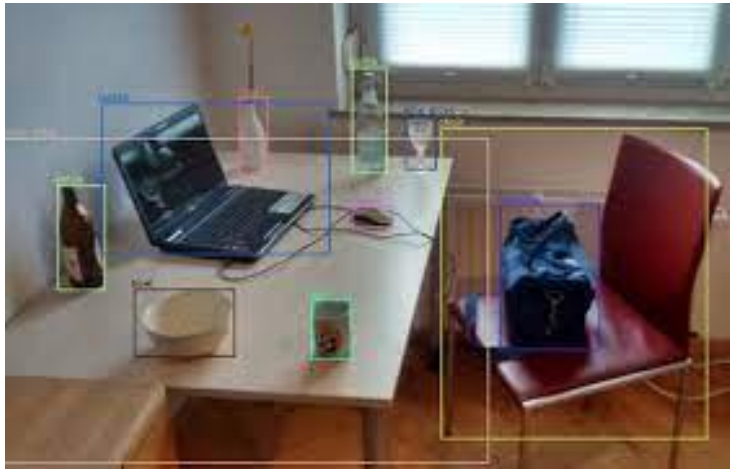
  - Code is packaged as a python unit with the only dependencies being python modules

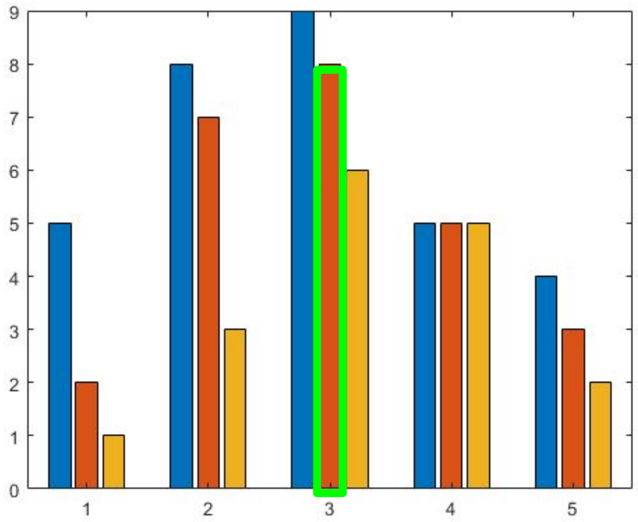
- RESOURCE UTILISATION

  - CPU/ GPU units only for model training

BASIC APPROACH

- OBJECT DETECTION IN IMAGES
  - Statistical plots are images
  - Each plot element can be thought of as an object
  - Detect all of the object boundaries
  - Make use of plain OCR modules for legend detection





WHY ARE STATISTICAL PLOTS DIFFERENT FROM NATURAL IMAGES?

NATURAL IMAGES	STATISTICAL PLOTS
Only visual elements	Visual (bars/ sectors) and textual elements (axis labels and ticks)
Consists mainly of unstructured data	Contain both structured and unstructured data
Size of objects constrained to small, medium and large variants. (constrained aspect ratio)	Size of objects (plot elements) can be varied (aspect ratio unconstrained)
Success criteria is over 50 percent for the accuracy measure IOU (Intersection over union). Adequacy criteria is low.	Higher adequacy criteria, for better accurate predictions

EXISTING DEEP LEARNING MODELS

- The baseline models as used by multiple papers is the RN model (Relation Networks).
- Papers have been published that propose CNN + LSTM pipeline models to tackle this problem.
- Additionally, one recent paper mentioned the use of OCR and semi-structured information extraction to tabulate data obtained from the plots.

OUR IDEA

- We intend to combine these approaches, to obtain a better overall accuracy across a variety of plot types.
- Additionally, we would want to explore on other variants of memory cells than the ones experimented with , that can better capture information.
- A potential case would be to investigate the usage of gated recurrent units over LSTM cells and usage of R-CNNs over the traditional CNNs.

PROPOSED METHODOLOGY/ APPROACH

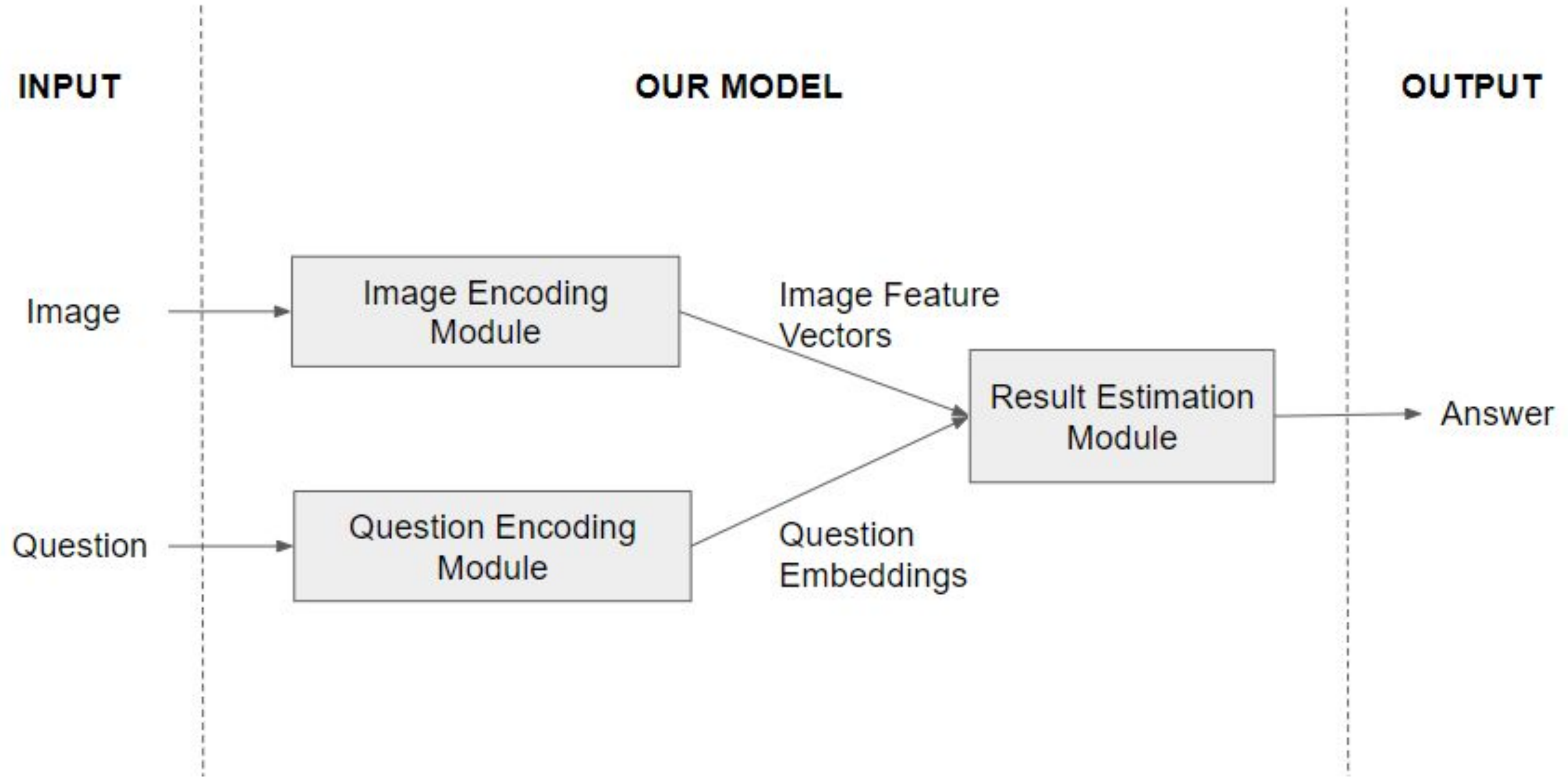
Encoder Decoder Architecture:

- CNN - To be used for obtaining an image representation
- LSTM/ variants such as GRU's - To be used for obtaining representation of the question
- Visual Elements Detection - OCR - Information Extraction Module - Additional module for extracting and tabulating data values.
- Table creation based on the VED and OCR results so that the input queries can be treated as relational queries.

Apart from this we would like to explore on R-CNN's and other variants for better overall accuracy for the various plots.

Ideally we would like to develop one model that works for chosen plots under our workframe.

HIGH LEVEL DESIGN OF THE SYSTEM



ARCHITECTURE

- **Input**
- **Our Model**

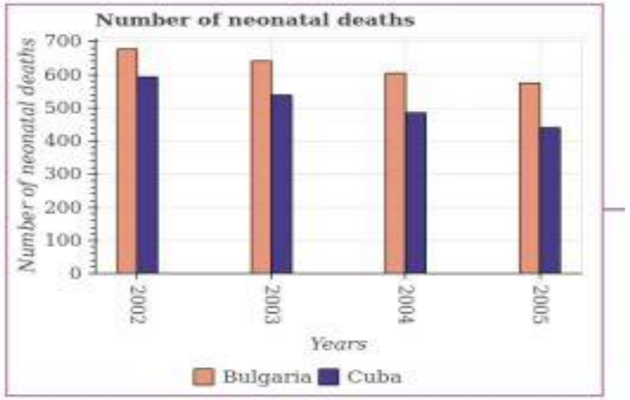
The design consists of 3 primary components.

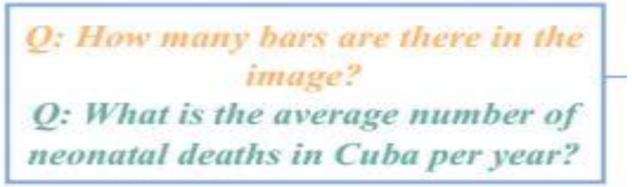
1. Image Encoding Module
2. Question Encoding Module
3. Result Estimation Module

- **Output**

ARCHITECTURE

Input: There are two inputs to our model that the user needs to provide. The first input is an image – that depicts a statistical plot and the second input is a relational question on the image.



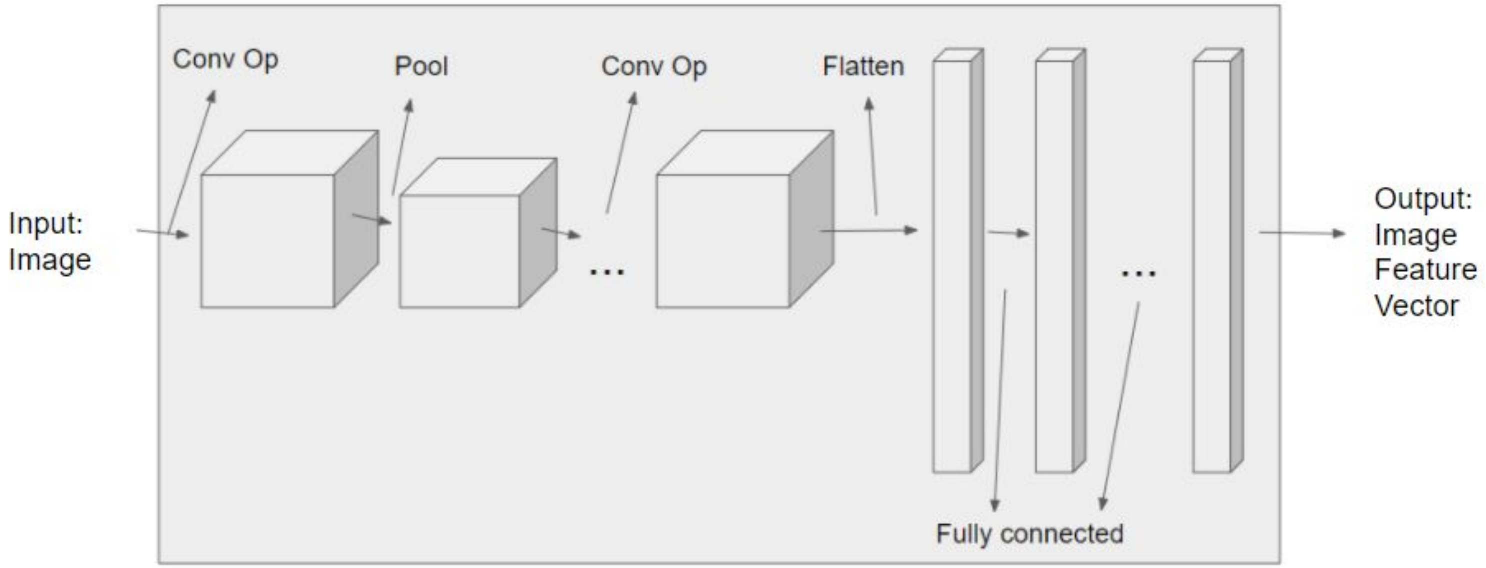


ARCHITECTURE

Our Model:

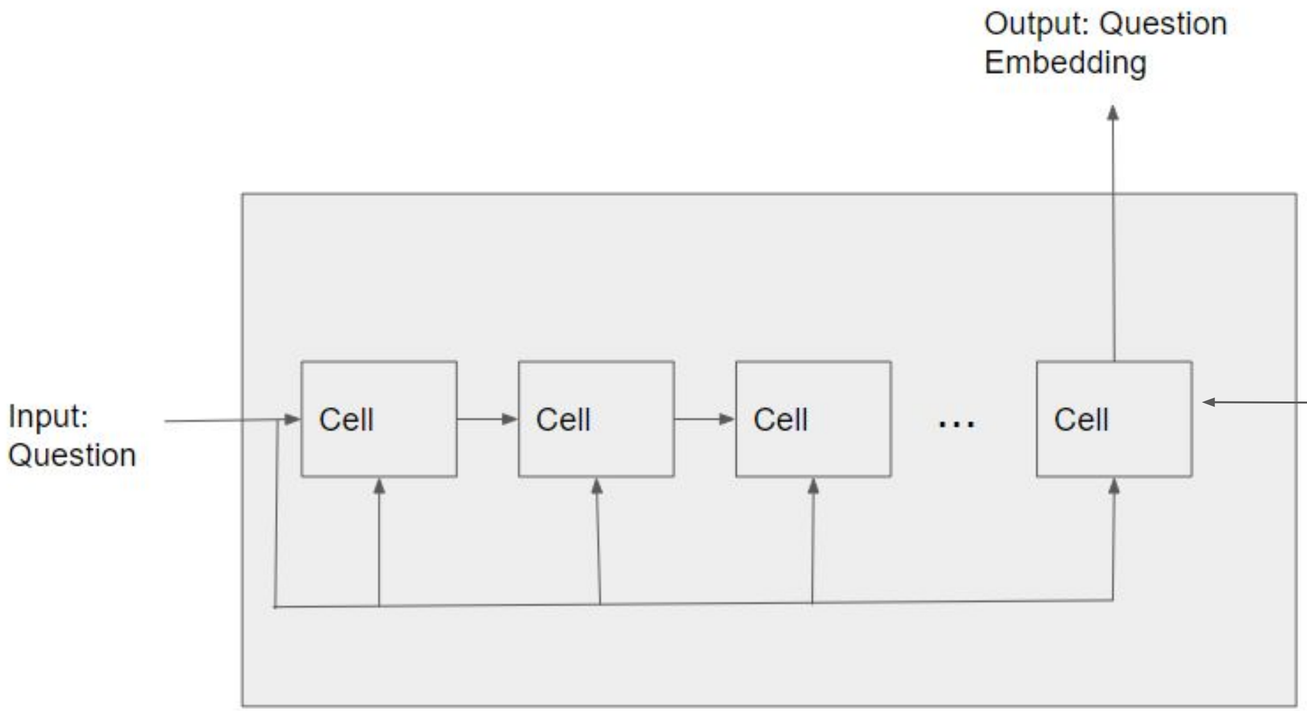
1. **Image Encoding Module**

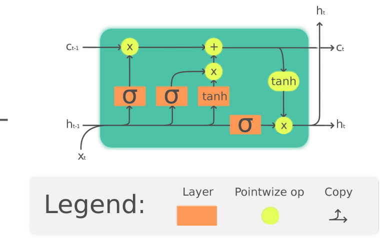
This consists of a sequence of convolution operations and pooling operations, followed by fully connected layers that would output a flattened image feature vector.



2. Question Encoding Module:

This consists of a series of recurrent cells, that would take each word of the input question as its input, to finally output a question embedding.





ARCHITECTURE

3. Result Estimation Module

This module takes in as input the output produced by both of the previous modules. It produces as an output the final answer to the question. Thereby, it finds the correlation between the image and the question based on previous training on a variety of image-question pairs.

Output: The output is an answer to the question that was posed on the image.

TECHNOLOGIES TO BE USED

- High level Deep Learning API such as Keras with Tensorflow backend
- Visual Element Detection modules
- OCR modules
- Google Colab Platform (Use of GPU)
- Flask web interface

PROJECT PROGRESS

- Problem statement is well-defined and thorough exploration is conducted in the associated domain.

- Thorough Literature survey pertaining to multiple variants of the problem is done to explore the innovation that can be brought to the project.

- Data to be used for the project is identified.

- The intermediary modules required are identified as well.

- Tools Identification .

PlotQA DATASET - DESCRIPTION

Dataset Split	Plot Types				Question Types			Answer Types		
	vbar	hbar	line	dot-line	Structural	Data-Retrieval	Reasoning	Yes/No	Fixed vocab.	Open vocab.
Train	52,463	52,700	25,897	26,010	871,782	2,784,041	16,593,656	784,115	3,095,774	16,369,590
Validation	11,249	11,292	5,547	5,571	186,994	599,573	3,574,081	167,871	600,424	3,592,353
Test	11,242	11,292	5,549	5,574	186,763	596,359	3,559,392	167,727	667,742	3,507,045

Detailed Statistics for different splits of the PlotQA dataset.

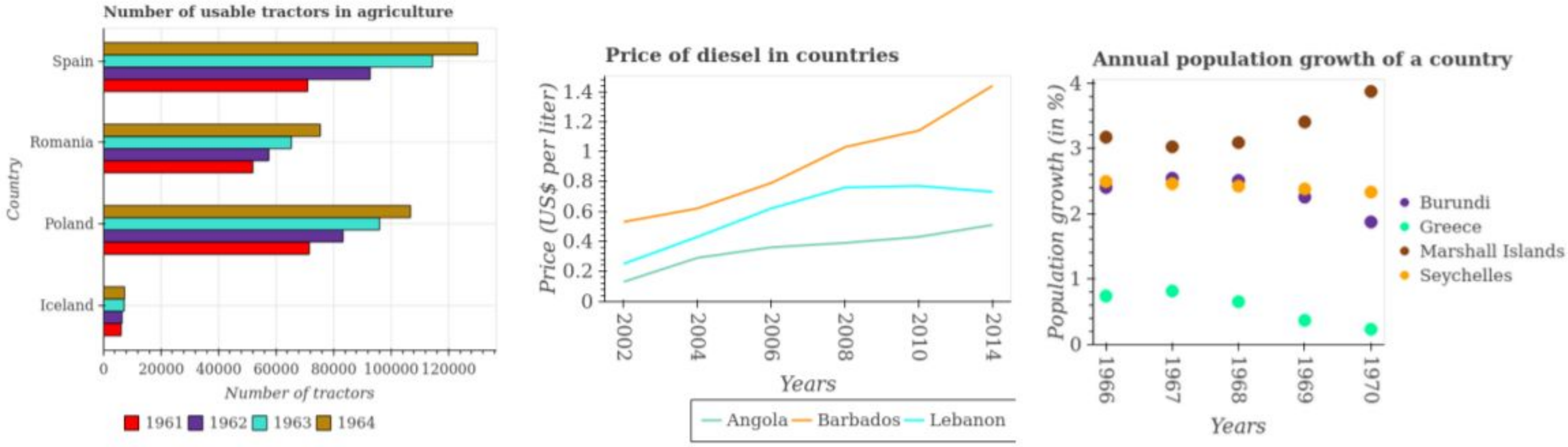
	Question (Q) Type		
Answer (A) Type	**Structure**	**Data Retrieval**	**Reasoning**
Yes/No	36.99%	5.19%	2.05%
Fixed vocabulary	63.01%	18.52%	15.92%
Open vocabulary	0.00%	76.29%	82.03%

Overall distribution of Q and A types in PlotQA.

Dataset Split	#Images	#QA pairs
Train	157,070	20,249,479
Validation	33,650	4,360,648
Test	33,657	4,342,514
Total	224,377	28,952,641

PlotQA Dataset Statistics

PlotQA DATASET - SAMPLE PLOTS



(a) Horizontal bar graph (b) Line plot (c) Dot-Line graph

Sample plots of different types in the PlotQA dataset.

WALKTHROUGH

- Dataset analysis - Demo

https://colab.research.google.com/drive/1DbmrMoa-q6JNt4z573fJ_rD2NHJpV5HZ?usp=sharing

SUMMARY OF WORK DONE IN CAPSTONE PHASE 1

- The abstract , scope , use cases of the current project have been identified. The Requirements are well understood and documented as a SRS.
- Literature survey is done on possible and existing solutions to the problem under consideration , hence opening up the doors for innovation .
- An extensive design document is prepared based on the design principles , architecture , components to be used .
- A detailed report encompassing the work done during the phase -1 is prepared in order to trace forward/backward when needed.

PROJECT PLAN FOR CAPSTONE PHASE 2

- Implement existing solutions
- Model fine-tuning to improve accuracy
- Synthesize new data for pie charts and line charts based on its feasibility

REFERENCES

Kim, Dae Hyun, Enamul Hoque, and Maneesh Agrawala. "Answering questions about charts and generating visual explanations." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

Reddy, Revanth, et al. "Figurenet: A deep learning model for question-answering on scientific plots." 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.

Sharma, Monika, et al. "ChartNet: Visual reasoning over statistical charts using MAC-Networks." 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.

Methani, Nitesh, et al. "Plotqa: Reasoning over scientific plots." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

Thank You