Deep Learning @ CVPRU, ISI Successes and Limitations

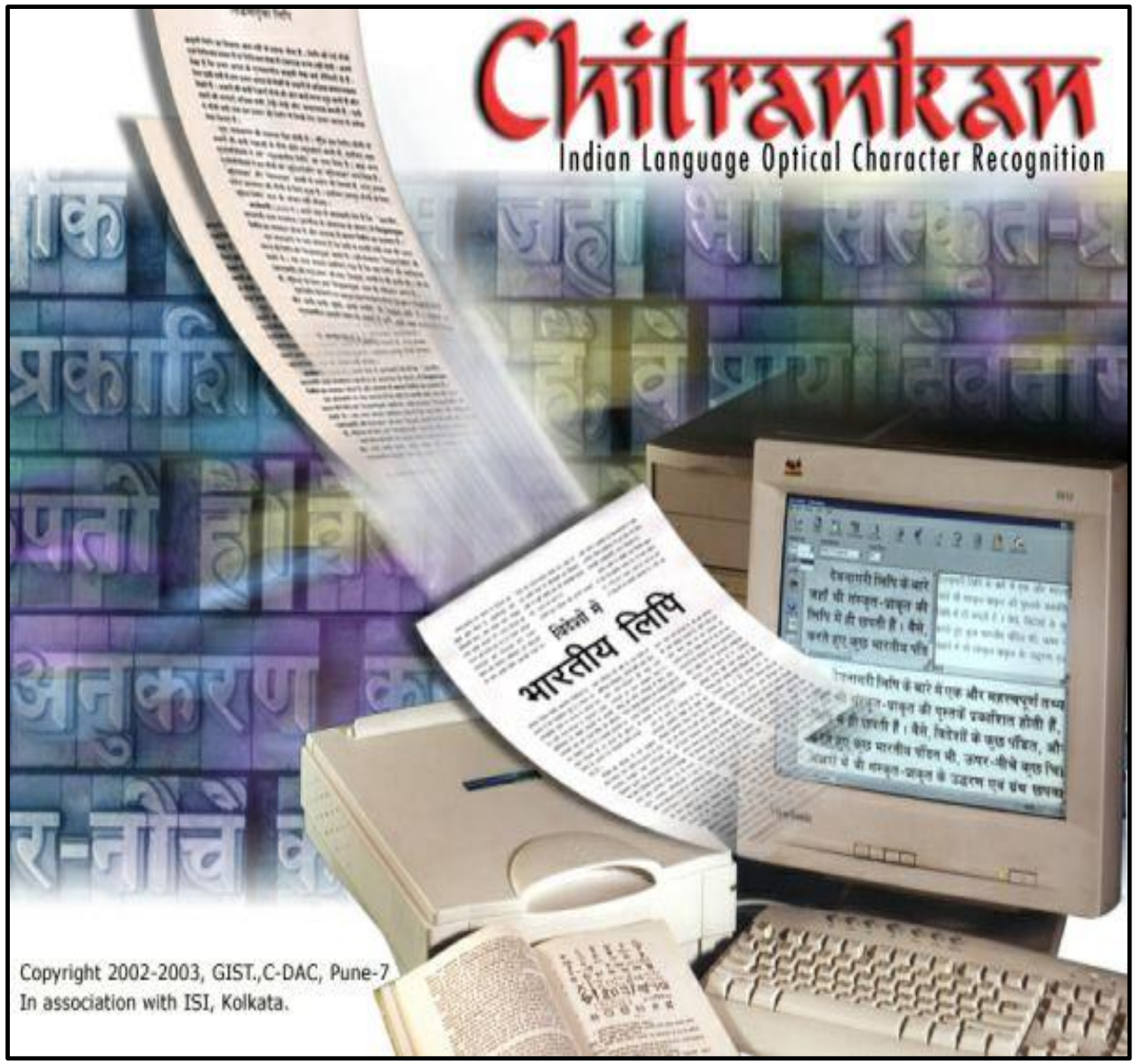
Utpal Garain

Professor, ISI-Kolkata

https://www.isical.ac.in/~utpal/

https://www.linkedin.com/in/utpal-garain-7b365943

The beginning: OCR technology



ISI's OCR Technology Recognizes 6 Indian Scripts

by Amit Tripathi Jan 22, 2004

The Indian Statistical Institute (ISI) has developed a multi-lingual character recognition technology, that recognizes at least six Indian languages based on the Devnagri script.

The Kolkata based institution devoted to the research, teaching and application of statistics, has developed the technology, called Multiple Script Segmentation Technology (MSST), to meet its own internal requirements, and is now offering the same commercially.

Says Dr. Bidyut Baran Chaudhuri, Head, CVPR unit, ISI, "The technology allows recognition of five Devnagri based languages including Bengali, Tamil, Telugu, Malyalam and Gujrathi. These technologies are definitely going to find commercial use in libraries, government archives etc."

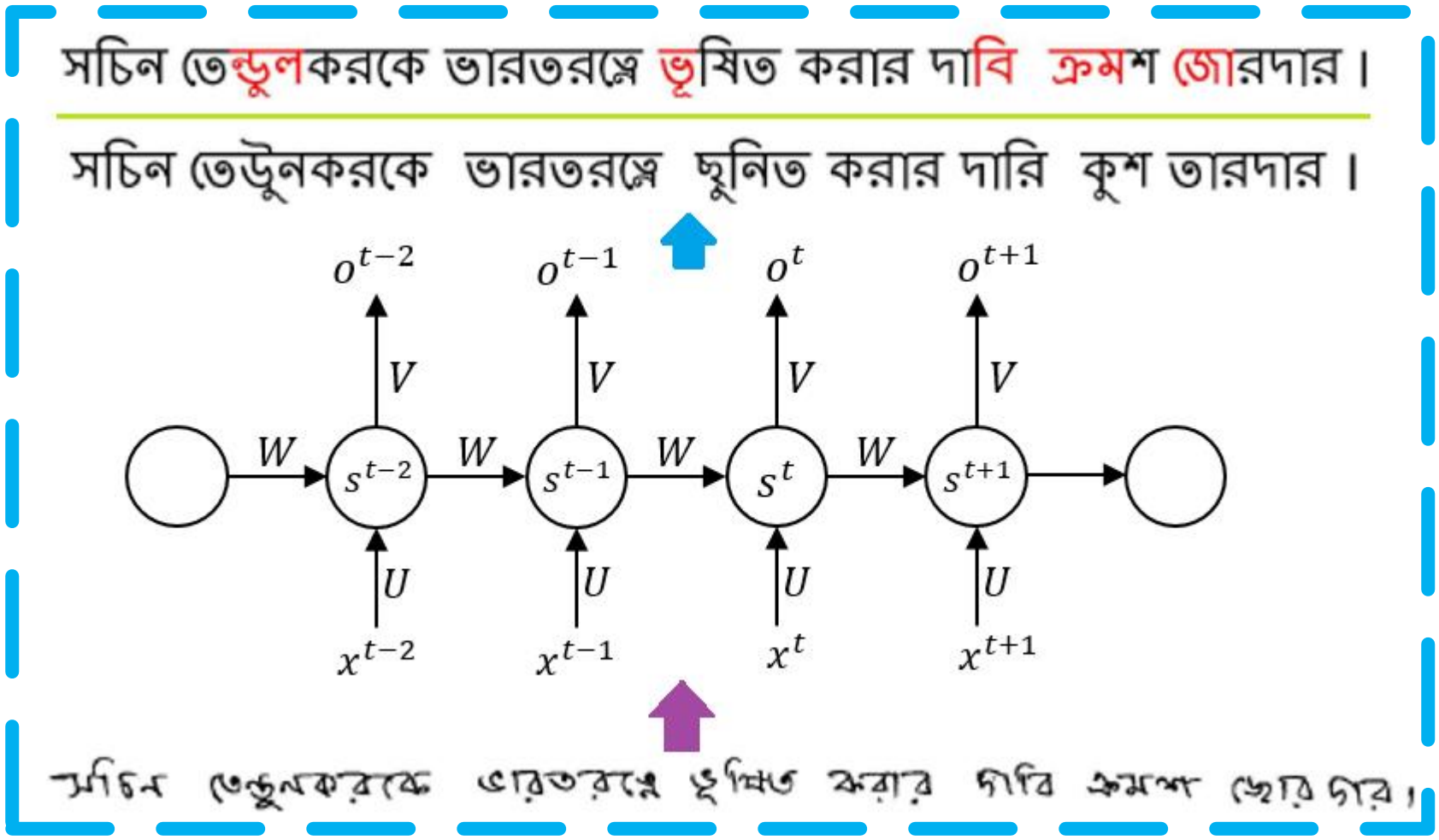
And though the accuracy factor of MSST is still below 95 percent, the institute is aiming to hit the 100 percent mark within a year.

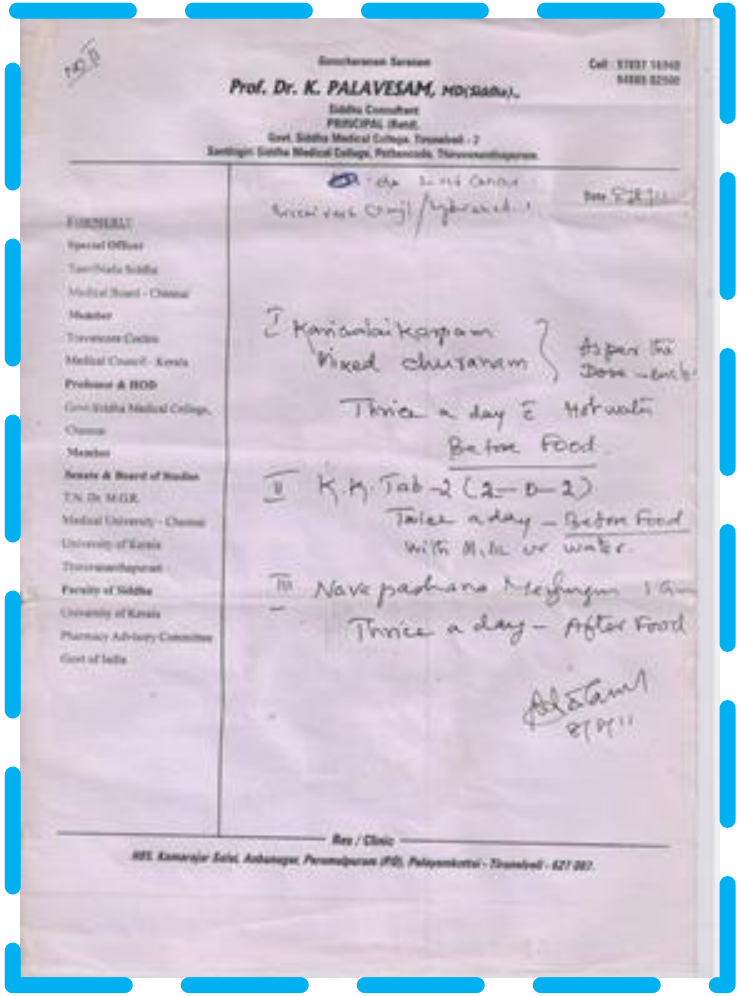
A typical MSST based solution would, besides the obvious scanner, include a software that recognizes Indian scripts through ISCII, like iLEAP to capture the recognized output from MSST.

According to Suranjit Sinha, Project Personnel, CVPR (Computer Vision Pattern Recognition) Section, ISI, the technology will be particularly useful in situations where a document with multiple languages needs to be recognized.

DL @ Indian Statistical Institute

- Machine Recognition of handwritten text





Doctor's Prescription: Vocabulary based HOCR

Unconstrained Handwriting: use of RNN

2015 13th International Conference on Document Analysis and Recognition (ICDAR)

Unconstrained Bengali Handwriting Recognition with Recurrent Models

Utpal Garain[1] Luc Mioulet[2] B. B. Chaudhuri[1] Clement Chatelain[2,3] Thierry Paquet[2]

1. CVPR Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, INDIA
2. Laboratoire LITIS - EA 4108, Universite de Rouen, 76800, FRANCE
3. Laboratoire LITIS - EA 4108, INSA Rouen, 76800, FRANCE

Abstract

This paper presents a pioneering attempt for developing a recurrent neural net based connectionist system for unconstrained Bengali offline handwriting recognition. The major challenge in configuring such a classification system for a complex script like Bengali is to effectively define the character classes. A novel way of defining character classes is introduced making the recognition problem suitable for using a recurrent model. Indeed, it has to deal with more than nine hundred character

word recognition has also been attempted [4]. The state of the art for the recognition of isolated characters is somehow clear as public domain datasets are available [5] but in case of word recognition it is still difficult to understand the progress as no common dataset is available for evaluation and the researchers have reported their research results using private datasets. There is a very good collection of handwritten samples as reported in [6] but as this dataset does not include annotation for handwritten words, it is not suitable for evaluation of handwriting recognition system.

- BLSTM
  - 2 hidden layers
  - 200 neurons in each layer
  - CTC layer consisting of 917 nodes
- 2300 lines for training
- Character recognition accuracy: 75.4%

In which script are you writing?

2015 13th International Conference on Document Analysis and Recognition (ICDAR)

Language Identification from Handwritten Documents

Luc Mioulet[†], Utpal Garain[*], Clément Chatelain[‡], Philippine Barlas[†] and Thierry Paquet[†]
[†]Laboratoire LITIS - EA 4108, Universite de Rouen, FRANCE 76800
[*] CVPR Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India.
[‡]Laboratoire LITIS - EA 4108, INSA Rouen, FRANCE 76800

***Abstract*—This paper presents a novel approach for language identification in handwritten documents. The approach is based on script identification followed by character recognition. BLSTM-CTC based handwriting recognizers are used and the OCR output is fed to a statistical language identifier for detecting the language of the input handwritten document. Documents in two scripts (Latin and Bengali) and four languages (English, French, Bengali and Assamese) are considered for evaluation. Several alternative frameworks have been explored, effects of handwriting recognition and text length on language detection have been studied. It is observed that with some empirical restrictions it is very much possible to achieve more that 80% language detection accuracy and based on the current research**

purpose, script identification followed by character recognition was sufficient. However, with the newer needs, language identification of a document is required for many applications like machine translation (MT), information extraction (IE), etc.

Among the previous works on language identification the research by Spitz [4] is worth mentioning. Printed documents are considered, two scripts (Latin-based and Han-based) and for each script several languages are considered. At first scripts are recognized and then within that script a specific language is identified. Character shape codes were used as features for such identification. Another notable work is due to Hochberg

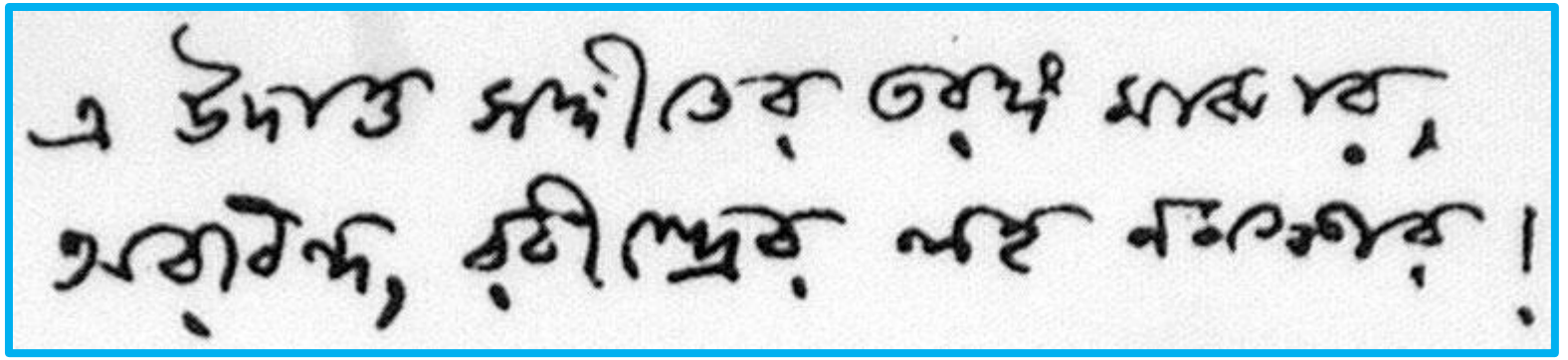
Present issues

- Indic script rendering

  हिंदी => ह + ि + द + ी

  - Post-processing rules
  - Definition of class
    - Idea of **akshara**
  - Use of attention
- Data
  - Transfer learning
    - Text in different fonts
      - Image + text

এ উদাত্ত সঙ্গীতের তরঙ্গ মাঝার,
অরবিন্দ, রবীন্দ্রের লহ নমস্কার !

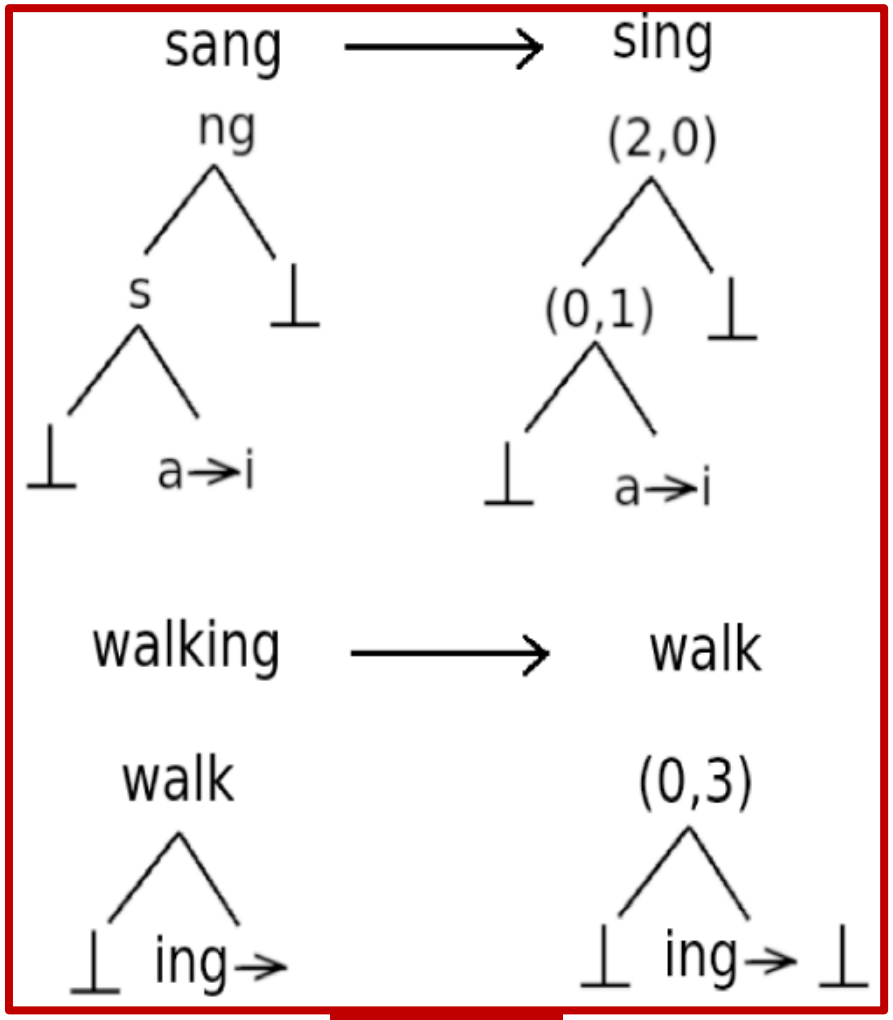


  - Handwriting generation (Ref. Alex Graves)
    - Challenges: modelling handwriting in Indic scripts (Not yet done)

Lemmatization (ACL, 2017)

Lemmatization

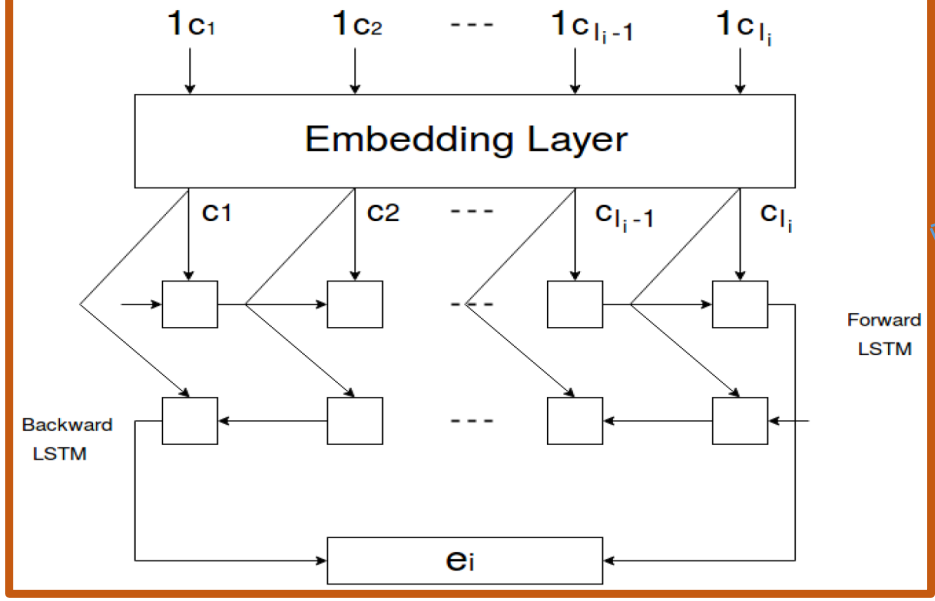
- It is a process that returns the base or dictionary form of a word in context
- Examples
  - sang => sing (rang => ring)
  - walking => walk (talking => talk)
- Challenges
  - Presence of large number of roots
  - Irregular word forms – e.g. went ; go
  - Morphologically complex patterns
- Previous model
  - Lemming: second order CRF
  - Viewed as edit tree classification task



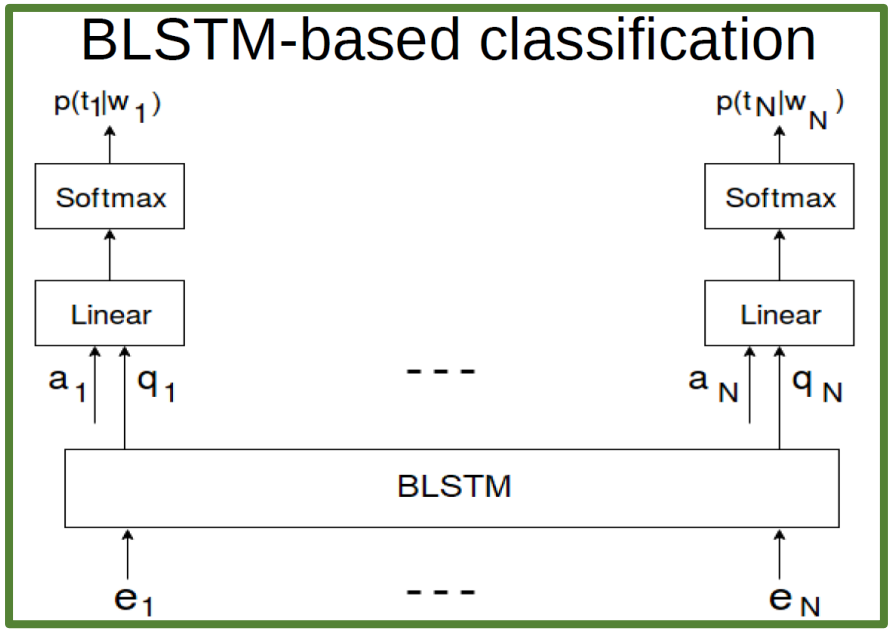
Edit Trees

Two Models (ACL, 2017 and CODS-COMAD, 2019)

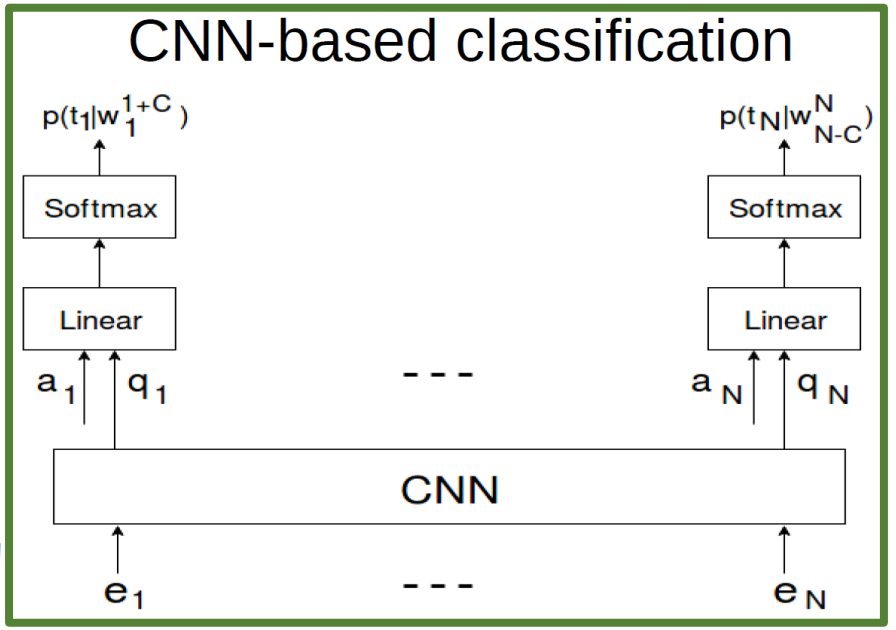
Building character-level embeddings



BLSTM-based classification



CNN-based classification



Take Home Message

	BLSTM-CNN Best results are at context ±1	BLSTM-BLSTM	CRF-based Model
French	**95.47%**	94.67%	93.09%
Hindi	**97.07%**	96.57%	95.38%
Marathi	**84.95%**	75.48%	79.13%
Spanish	94.94%	94.64%	**98.42%**

- Inherently complex morphology of Indic languages
- Annotated data is an issue
- Hand-crafted features still may give good accuracy (e.g. Spanish)

Textbook Question Answering (TQA)

Evaluating Machine Comprehension

- Is machine smarter than grade-VI students?
  - A textbook article
  - Multiple choice questions
    - 2 (true/false) to 7 options

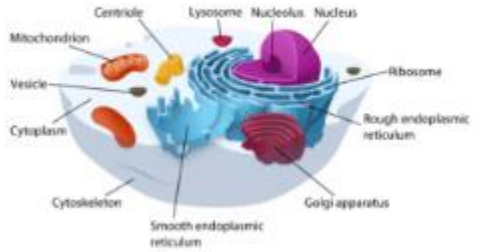
Cell Structures

Introduction

In some ways, a cell resembles a plastic bag full of Jell-O. Its basic structure is a cell membrane filled with cytoplasm. The cytoplasm of a eukaryotic cell is like Jell-O containing mixed fruit. It also contains a nucleus and other organelles.

Cell Membrane

The cell membrane is like the bag holding the Jell-O. It encloses the cytoplasm of the cell. It forms a barrier between the cytoplasm and the environment outside the cell. The function of the cell membrane is to protect and support the cell. It also controls what enters or leaves the cell. It allows only certain substances to pass through. It keeps other substances inside or outside the cell.



Cell Membrane Structure

Cytoplasm

Organelles

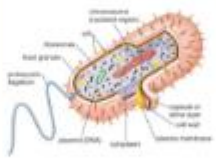
Lesson Summary

- The cell membrane consists of two layers of phospholipids.
- The cytoplasm consists of watery cytosol and cell structures.
- Eukaryotic cells contain a nucleus and other organelles

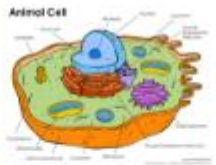
Vocabulary

Cell Wall	rigid layer that surrounds the cell membrane of a plant cell or fungal cell and that supports and protects the cell
Cyto-skeleton	structure in a cell consisting of filaments and tubules that crisscross the cytoplasm and help maintain the cells shape
Central Vacuole	large storage sac found in the cells of plants

Instructional Diagrams



The image below shows the Prokaryotic cell. A prokaryote is a single-celled organism that lacks a membrane-bound nucleus (karyon), mitochondria, or any other membrane-bound organelle. In the prokaryotes, all the intracellular water-soluble components (proteins, DNA and metabolites) are located together in the cytoplasm enclosed by the cell membrane, rather than in separate cellular compartments.



This diagram shows the anatomy of an Animal cell. Animal Cells have an outer boundary known as the plasma membrane. The nucleus and the organelles of the cell are bound by this membrane. The cell organelles have a vast range of functions to perform like hormone and enzyme production to providing energy for the cells. They are of various sizes and have irregular shapes. Most of the cells size range between 1 and 100 micrometers and are visible only with help of microscope.

Questions

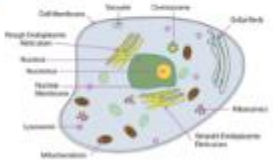
What is the outer surrounding part of the Nucleus?

a. **Nuclear Membrane**

b. Golgi Body

c. Cell Membrane

d. Nucleolus



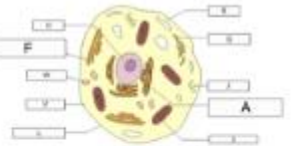
Which component forms a barrier between the cytoplasm and the environment outside the cell?

a. J

b. **L**

c. X

d. U



Which statement about the cell membrane is false?

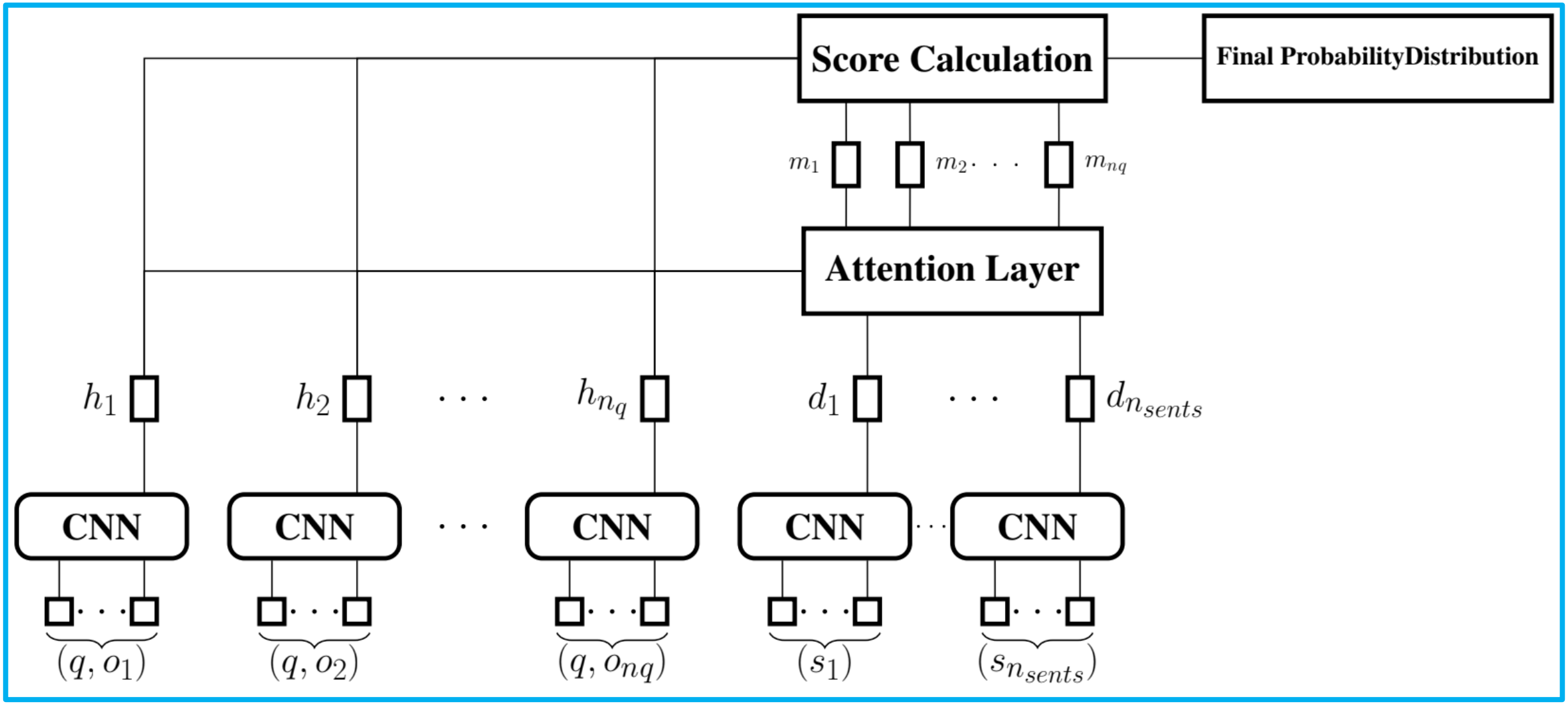
a. It encloses the cytoplasm

b. It protects and supports the cell

c. **It keeps all external substances out of the cell**

d. none of the above

Architecture



Where do we stand

- Competitors
  - Text Only (Jason Weston, Sumit Chopra, and Antoine Bordes, CoRR abs, 2014)
    - A variant of Memory network
    - Paragraph, question and options are embedded separately using LSTM
  - BIDAF (Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi, 2016, Bidirectional attention flow for machine comprehension. CoRR abs)
    - Character and word level embeddings are used to encode the question and the text followed by bidirectional attention mechanism.

Model	True-False	Multiple Choice
Random*	50.0	22.7
Text-Only*	50.2	32.9
BIDAF*	50.4	32.2
$CNN_{2,3,4}$	**53.7**	**35.8**

Shortcoming: Inability to Reason

http://data.allenai.org/tqa/absolute_ages_of_rocks_L_0010/

absolute ages of rocks

isotopes

An element is defined by the number of protons it contains. All atoms of a given element contain the same number of protons. The number of neutrons in an element may vary. Atoms of an element with different numbers of neutrons are called isotopes. Consider carbon as an example. Two isotopes of carbon are shown in Figure 11.15. Compare their protons and neutrons. Both contain 6 protons. But carbon-12 has 6 neutrons and carbon-14 has 8 neutrons. Almost all carbon atoms are carbon-12. This is a stable isotope of carbon. Only a tiny percentage of carbon atoms are carbon-14. Carbon-14 is unstable. Figure 11.16 shows carbon dioxide, which forms in the atmosphere from carbon-14 and oxygen. Neutrons in cosmic rays strike nitrogen atoms in the atmosphere. The nitrogen forms carbon- 14. Carbon in the atmosphere combines with oxygen to form carbon dioxide. Plants take in carbon dioxide during photosynthesis. In this way, carbon-14 enters food chains.

radioactive decay

Radioactive decay is the breakdown of unstable elements into stable elements. To understand this process, recall that the atoms of all elements contain the particles protons, neutrons, and electrons.

carbon14 dating

The best-known method of radiometric dating is carbon-14 dating. A living thing takes in carbon-14 (along with stable carbon-12). As the carbon-14 decays, it is replaced with more carbon-14. After the organism dies, it stops taking in carbon. That includes carbon-14. The carbon-14 that is in its body continues to decay. So the organism contains less and less carbon-14 as time goes on. We can estimate the amount of carbon-14 that has decayed by measuring the amount of carbon-14 to carbon-12. We know how fast carbon-14 decays. With this information, we can tell how long ago the organism died. Carbon-14 has a relatively short half-life. It decays quickly compared to some other unstable isotopes. So carbon- 14 dating is useful for specimens younger than 50,000 years old. Thats a blink of an eye in geologic time. But radiocarbon dating is very useful for more recent events. One important use of radiocarbon is early human sites. Carbon-14 dating is also limited to the remains of once-living things. To date rocks, scientists use other radioactive isotopes.

questions

How much percent of the parent isotope remains after 2 half-lives?

a. 100%

b. 50%

--> c. 25%

d. 75%

The half-life of a radioactive element is

a. half the estimated age of Earths crust

--> b. the time it takes for half a parent isotope to decay into the daughter isotope

c. half the weight of the original radioactive element

d. the time it takes for half of a daughter isotope to decay into a parent isotope

Carbon dating is useful for

a. igneous rocks

b. sedimentary rocks

--> c. organic materials

d. none of the above

Medical Image Analysis (Diagnosis of Psoriasis)

Camera Image

Visualization of psoriasis plaques with different severity scores

Severity Factor	Absent (0)	Mild (1)	Moderate (2)	Severe (3)	Very Severe (4)
Erythema (Redness)					
Scaling (Silveryness)					
Induration (Elevation)					

Ordinal Classification, MICCAI 2018

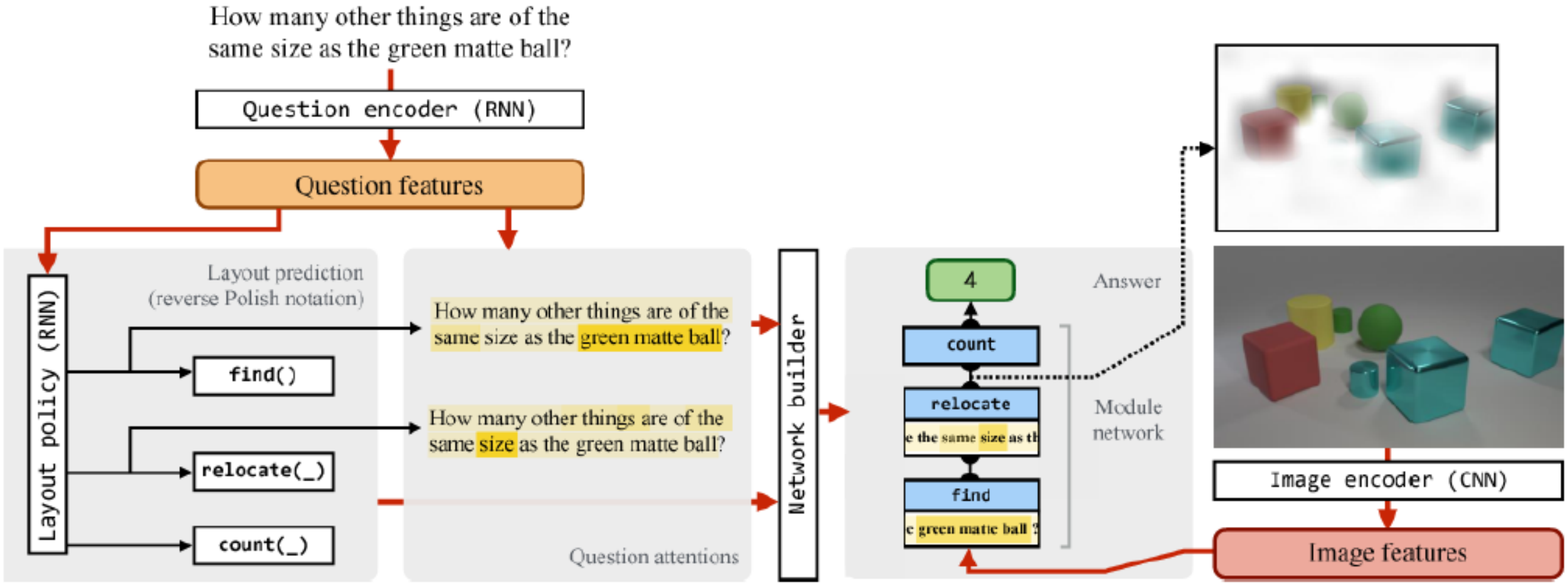
- Image classification by using pre-trained CNN models (ResNet-50 and Mobile Net)

- Commonly used loss functions :
  - $\mathcal{L}_{CCE} = -\sum_{i=1}^{C} G_i \, ln(P_i)$,
  - $\mathcal{L}_{MSE} = \sum_{i=1}^{C} (P_i - G_i)^2$,
    where G and P represent actual and predicted class probability distribution respectively
- Ordinal classifier
  - Conversion to regression Problem
  - Decomposition of the C class classification problem into $C-1$ binary classification problems

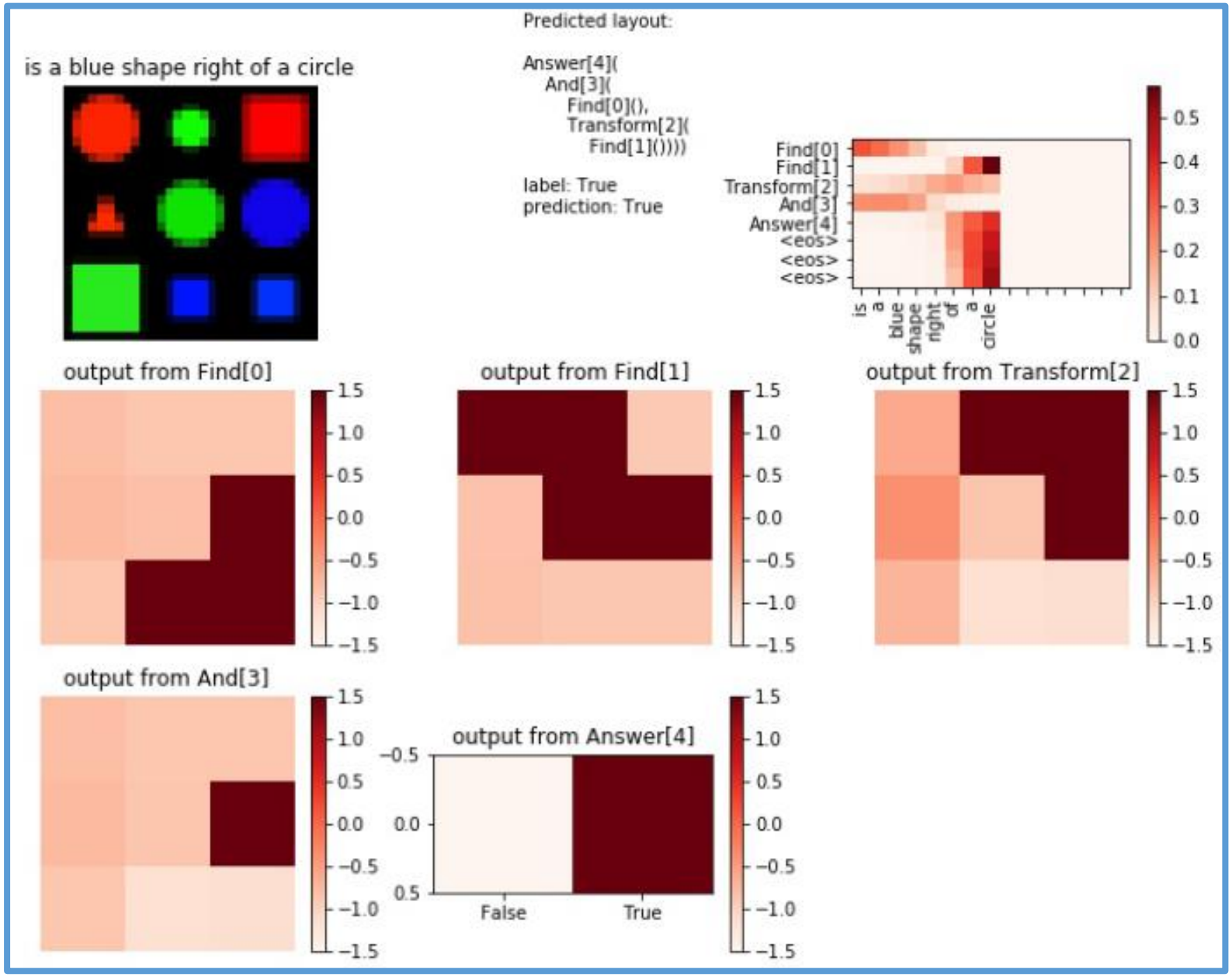
$$
\begin{aligned}
\mathcal{L}_{EMD} &= \sum_{i=1}^{C} (P_i^{CDF} - G_i^{CDF})^2 \\
&= \sum_{i=1}^{C} (\sum_{j=1}^{i} P_j - \sum_{j=1}^{i} G_j)^2 \\
&= \underbrace{\sum_{i=1}^{k-1} \left(\sum_{j=1}^{i} P_j \right)^2}_{\mathcal{A}} + \underbrace{\sum_{i=k}^{C} \left(\sum_{j=1}^{i} P_j - 1 \right)^2}_{\mathcal{B}}
\end{aligned}
$$

Visual Question Answering

End to End Neural Module Networks (N2NNMN)

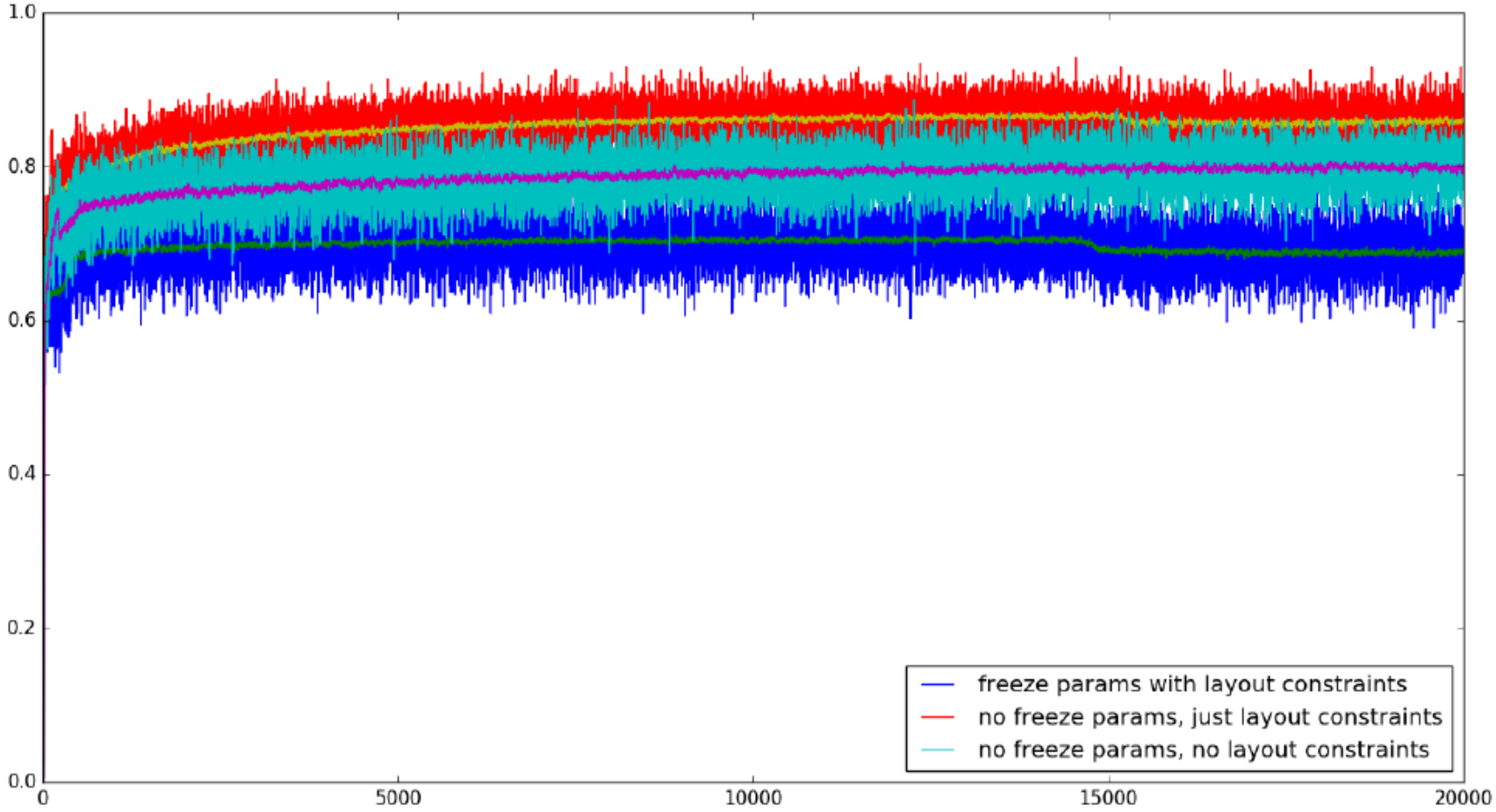


N2NNMN on SHAPES



Current drawbacks and new results

- The set of modules need to be defined beforehand
- Ground truth layout is needed during training
- **New experiments**
  - Giving set of layout rules instead of layout
  - Freezing the parameters of modules during training
- **Observations**
  - Despite the gain in accuracy, the layout accuracy remains very low
  - Incorrect layout often leads to correct prediction
  - This is unlike typical reinforcement learning problem



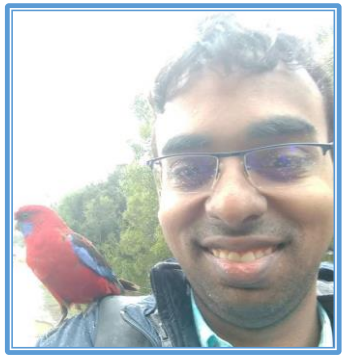
Looking for collaboration on

- ISI Kolkata has setup an AI Centre
  - Dedicated to do theoretical research and provide engineering solutions for the problems of National Importance
    - Through funded projects and consultancy services (*public* and *private*)
  - Generating manpower through short term courses, training for Industry, Govt. offices and scholars/students
- **Personal interest**
  - Harnessing Deep models with logic
  - Ordinal classification problems
  - Adversarial examples for NLP
  - Deep models for sensor data
    - **Eye gaze** data for human behavior analysis

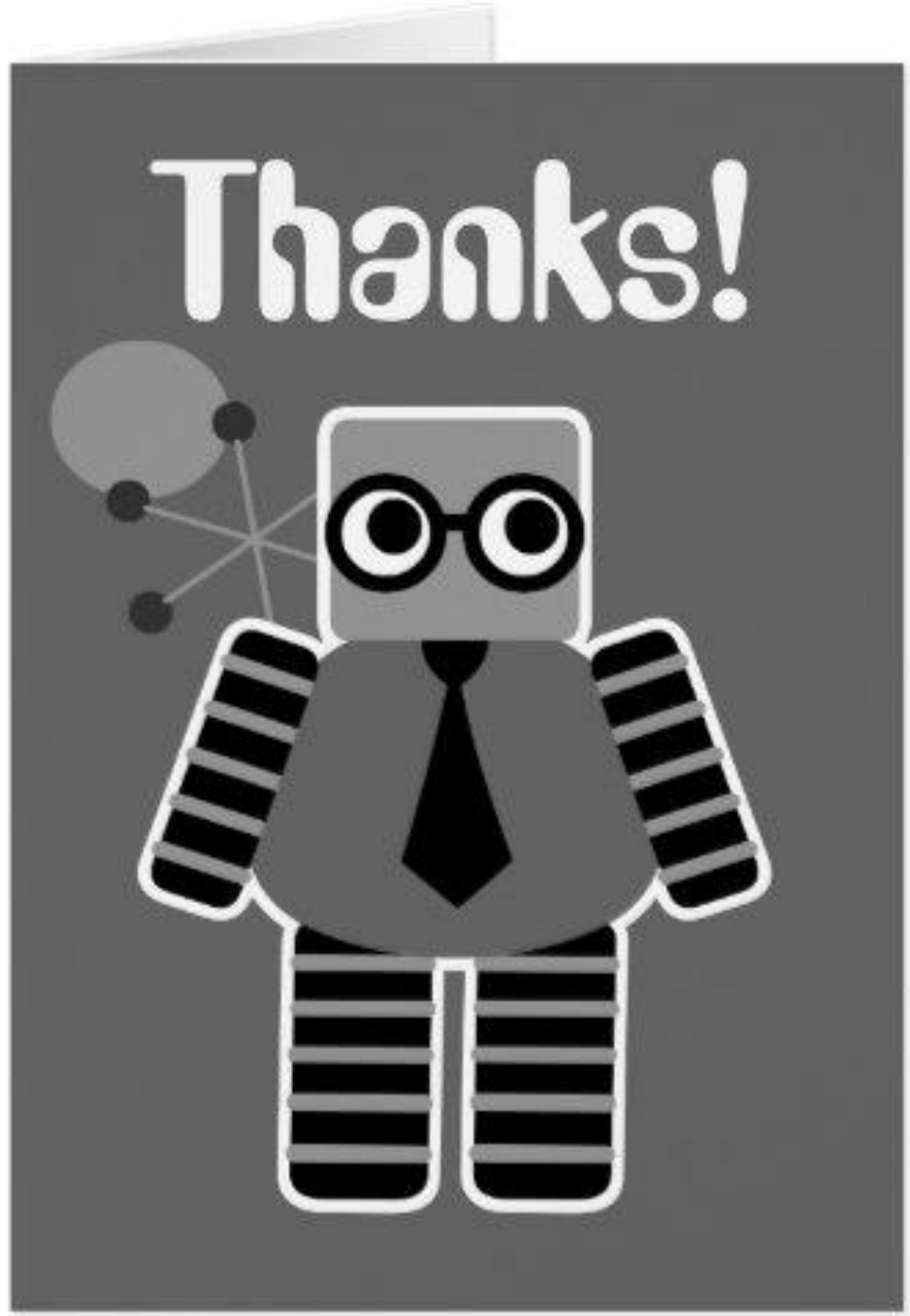
Acknowledgement

- **Abhisek Chakarabarty**, SRF
  - Morphological analysis (Neural NLP)

- **Anabik Pal**, SRF
  - Medical Image Analysis



- **Akshay Chaturvedi**, SRF
  - TQA and VQA
  - Adversarial attacks


Thanks!