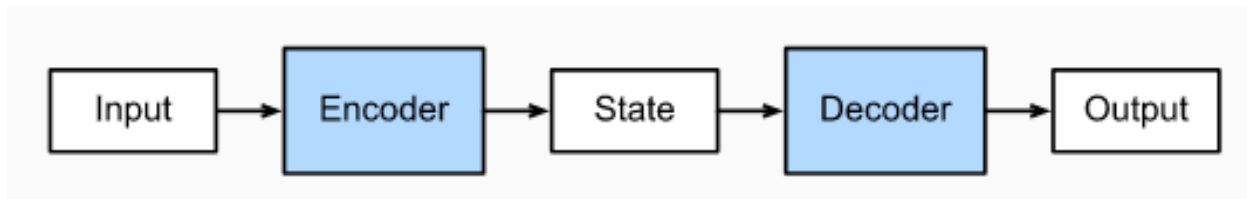# Lab13-2:Image Captioning
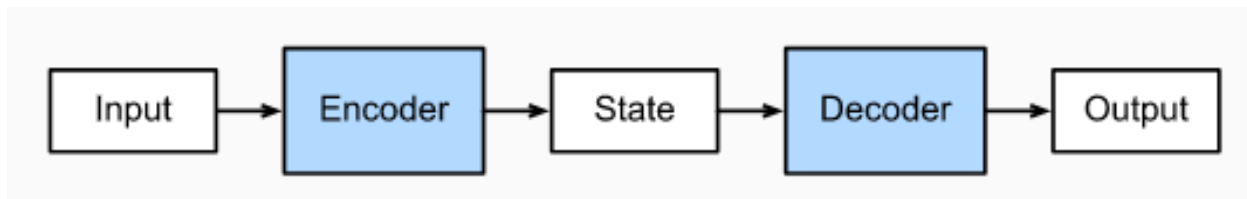
Encoder-Decoder

- ## Neural Machine Translation
  - ### Encoder RNN
    - reads the source sentence and transforms it into a rich fixed-length vector representation
  - ### Decoder RNN
    - uses the representation as the initial hidden state and generates the target sentence

Encoder-Decoder

- ## Image Captioning
  - ### Encoder CNN
    - reads the images and transforms it into a rich fixed-length vector representation
  - ### Decoder RNN
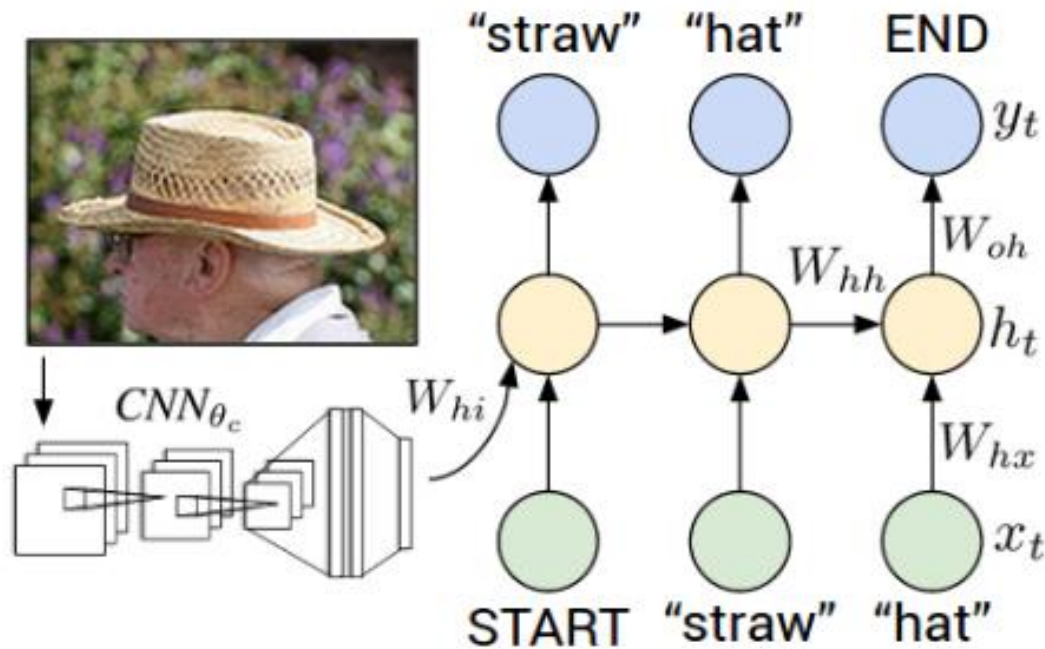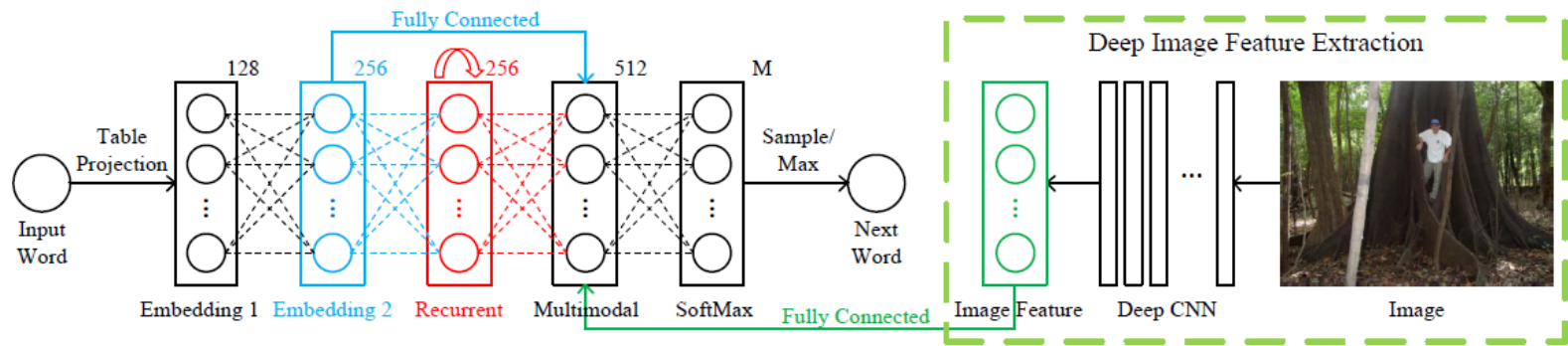   - uses the representation as the initial hidden state and generates the target sentence

Input → Encoder → State → Decoder → Output

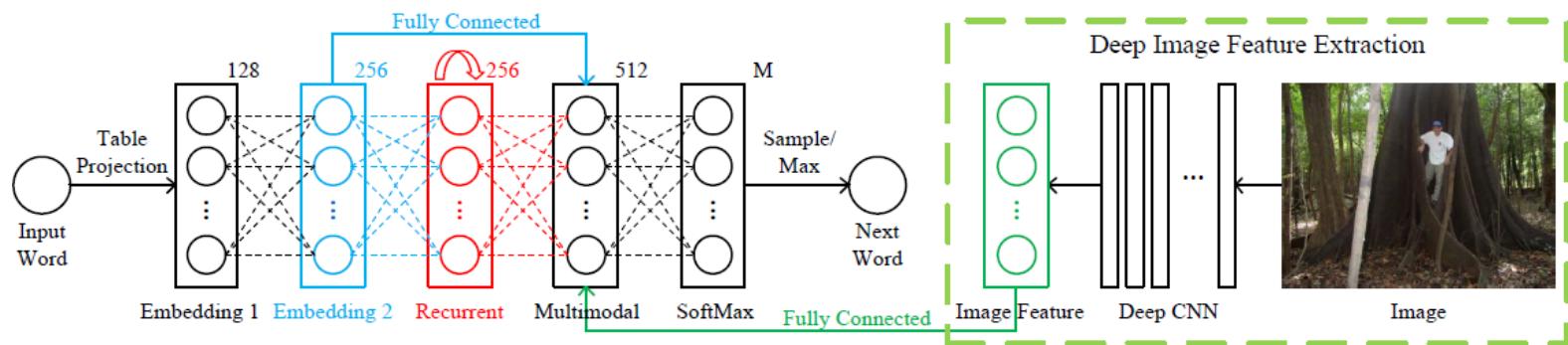# Encoder-Decoder

- Image Captioning

Image Captioning

- ## m-RNN

  - ### a multimodal Recurrent Neural Network model

    - language model part

   - image part

    - multimodal part

Image Captioning

- ## m-RNN
  - ### It must be emphasized that
    - The image part is AlexNet, which connects the seventh layer of AlexNet to the multimodal layer.
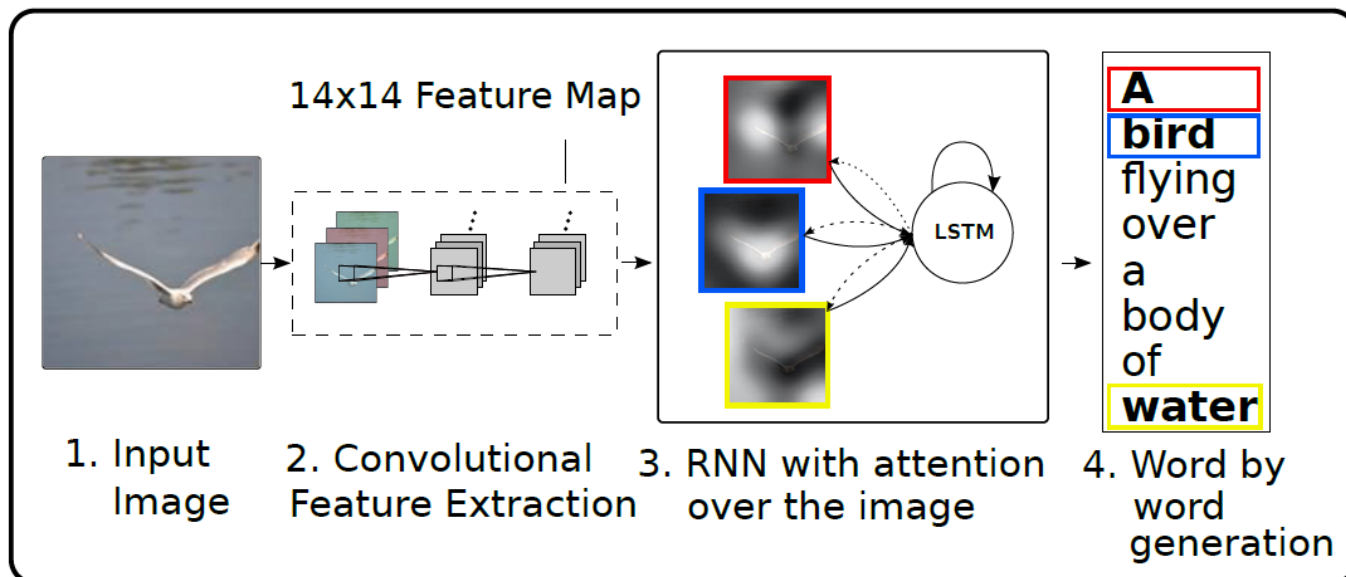   - This model feeds the image at each time step.

Image Captioning

- NIC

  – a generative model combining recent advances in computer vision and machine translation

    - The image is only input once, at t = -1, to inform the LSTM about the image contents.
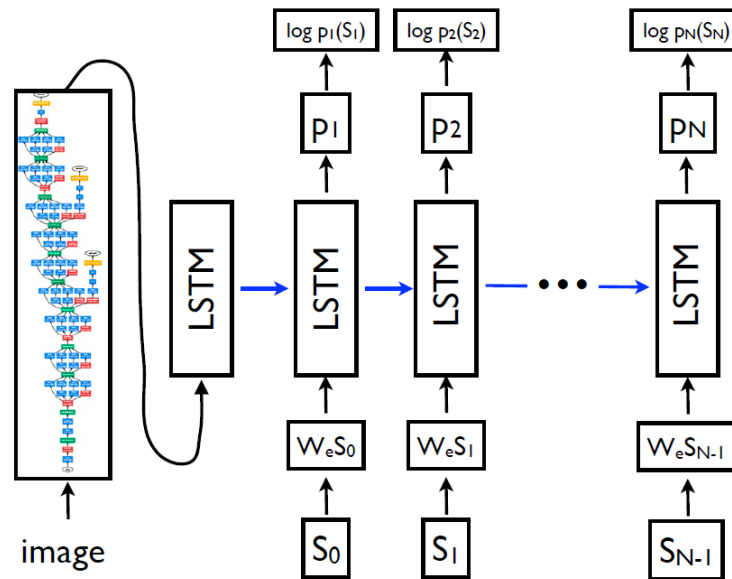
Image Captioning

- Attention-Based
  - 《Show, Attend and Tell: Neural Image Caption Generation with Visual Attention》
    - a generative model attempting to incorporate a form of attention with two variants

14x14 Feature Map

LSTM

A
bird
flying
over
a
body
of
water

1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

Image Captioning

- To get more related papers,
 - https://paperswithcode.com/task/image-captioning

Image Captioning

- Implementation
  - The model architecture is similar to 《Show, Attend and Tell: Neural Image Caption Generation with Visual Attention》
    - https://www.tensorflow.org/tutorials/text/image_captioning

Assignment

- CAPTCHA
  - an acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart"

Assignment

- CAPTCHA
  - a type of challenge–response test used in computing to determine whether or not the user is human
  - prevents spam attacks and protects websites from bots
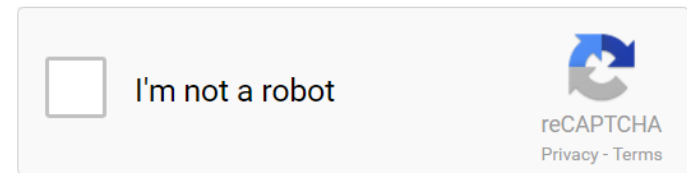
7 × **5** = [ ] ↻
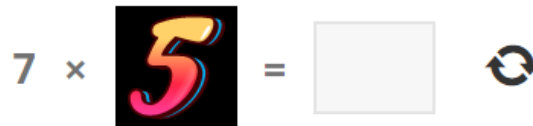
COMMENT
[   ]

Assignment

- ## reCAPTCHA
  - – a CAPTCHA-like system
    - establish that a computer user is human
    - assist in the digitization of books or improve machine learning
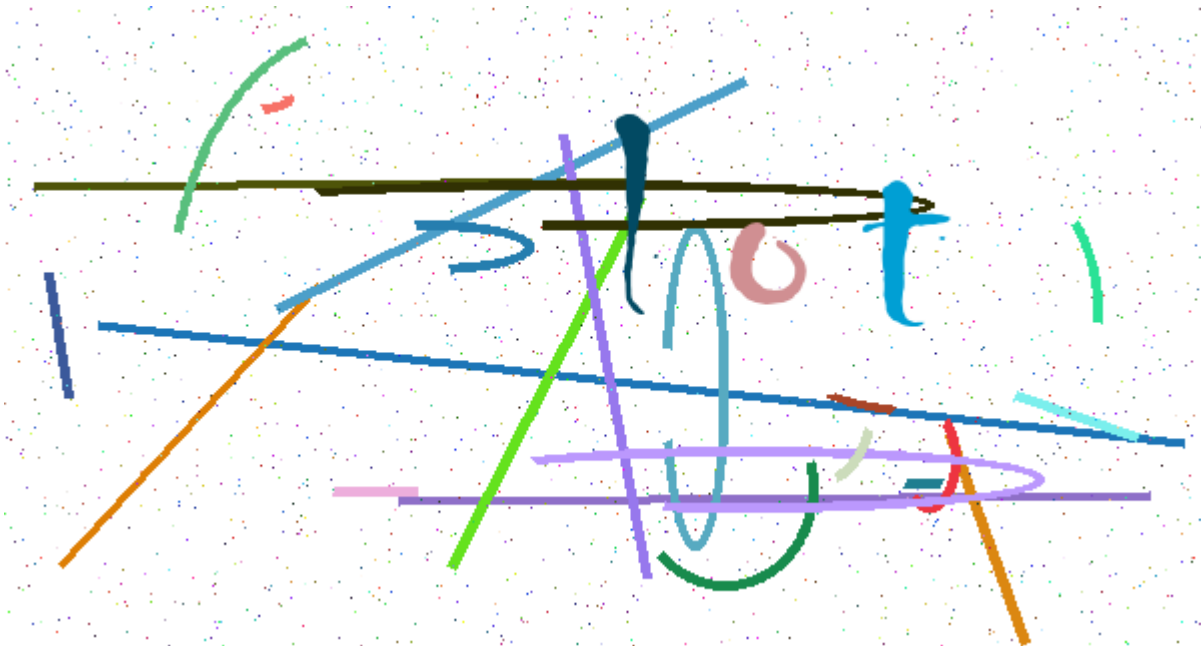
Assignment

- Homework
  - Train a captcha-recognizer to identify English words in images

Assignment

- Homework
  - Dataset
    - 140,000 CAPTCHAs
      - 2500 English words
      - 3-5 character lengths
      - four different fonts.

Assignment

- Homework
  - Requirements
    - Use any model architectures you want
    - Design your own model architecture
    - The first 100,000 as training data, the next 20,000 as validation data, and the rest as testing data
    - Only if the whole word matches exactly does it count as correct
    - Predict the answer to the testing data and write them in a file
    - Testing accuracy should be at least 90%

Assignment

- Homework
  - Notification
    - your code file (Lab13-2_{student id}.ipynb)
    - your answer file (Lab13-2_{student id}.txt)

```
a0 thus
a1 www
a2 tied
a3 ids
a4 jam
a5 zoo
a6 apple
a7 big
a8 lot
a9 above
a10 ooo
a11 dame
a12 dates
```

Assignment

- Homework
  - Loss/Acc