Date / /
Page

- Mix questioner & guesser.
- Have a probability heatmap over graph & take a loss.
  Heat map must take error in each step of question.
- 2 eye attention. Attention over images.
- See how can you get the dialogs in hierarchy & use attention on image based on history & que
- use features from image Memory bank. using RCNN.
- See if you can use VAE

Date / /
Page

raw doing Preprocessing

1) If using img ~~preflores~~ ,
2) Change json dim file when using raw [img_loader];

☆ Change:

~~If config~~ config [model]
[img] == raw,

☆ Check image dimention every time img is raw.

Visualize graph later.

ToDo - reflect all changes you made recently using config file.

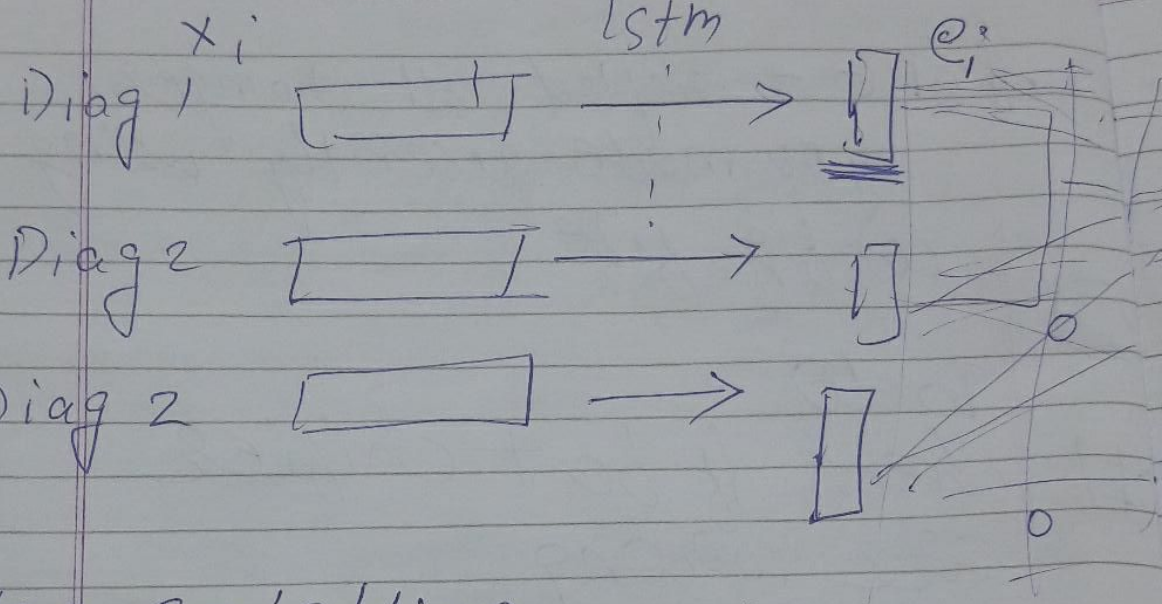
;)

To do

check if batchifier are wrong.

track val while training [change evaluator & instead of evaluating 2 times do it once)

Bidirectional Lstm

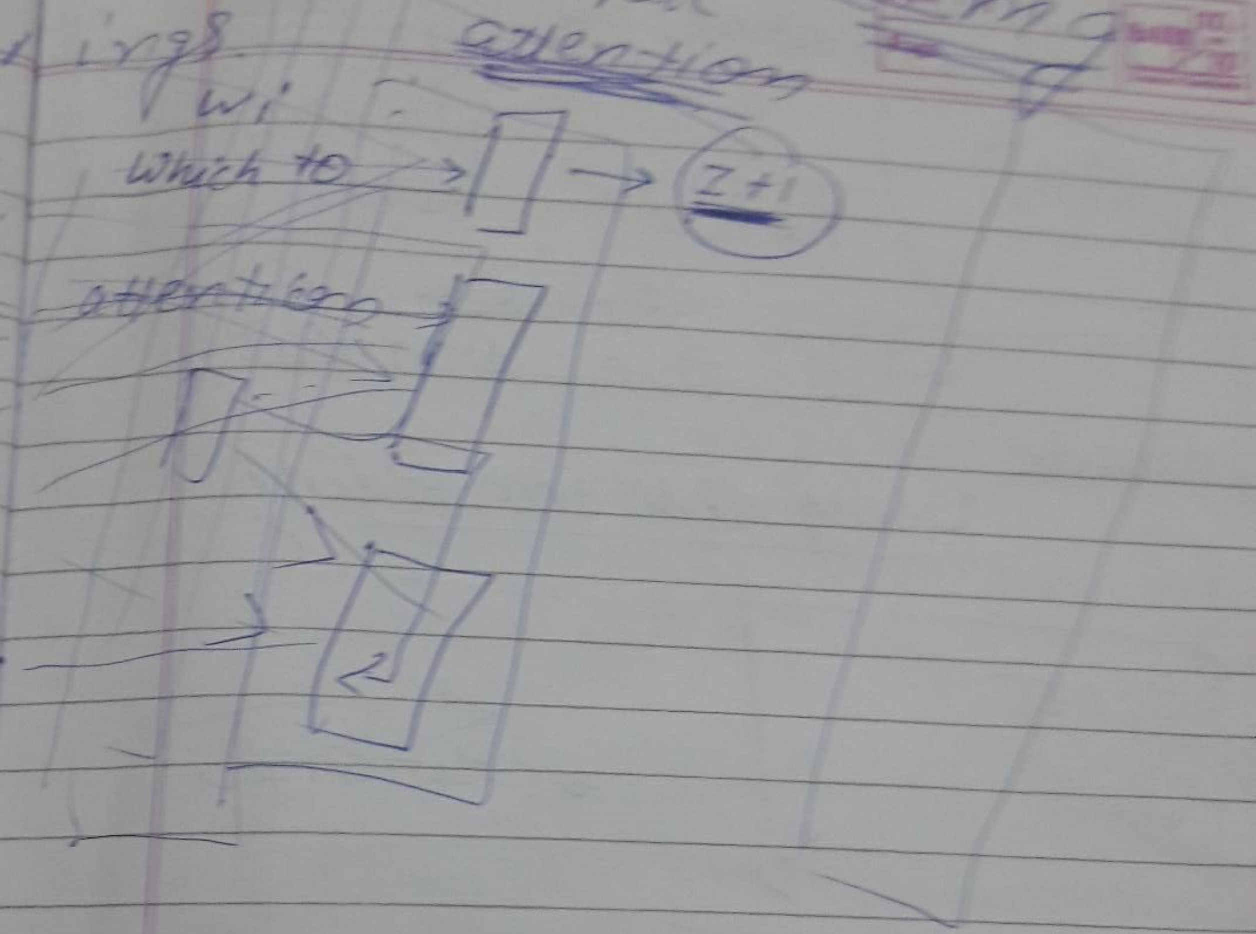


New embeddings

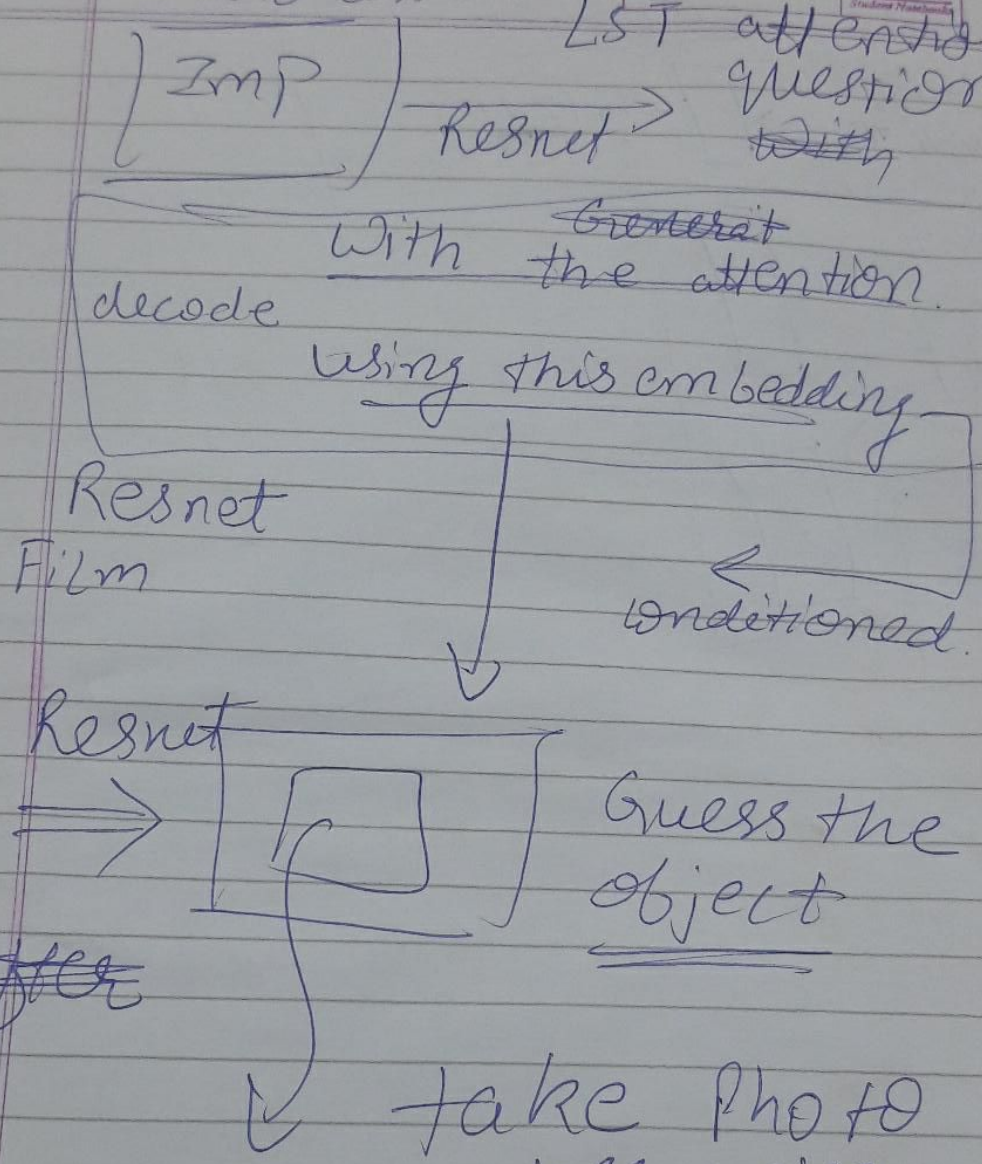
Use GRU & LSTM

Compute attention

New Img

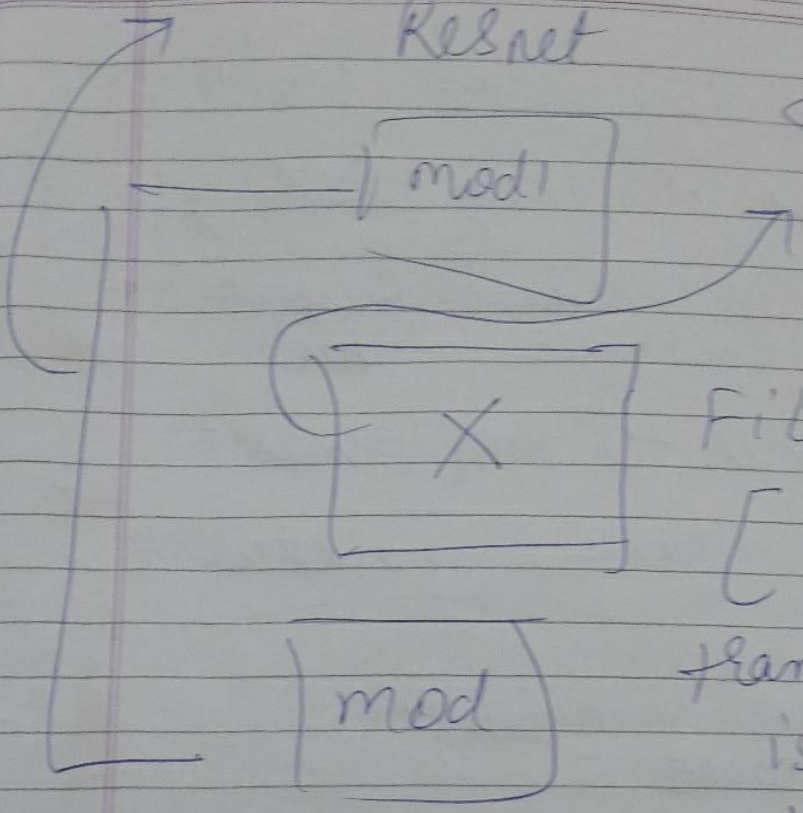


2 Main Idea:
Image & word correlated
Photo loss novel



Share these

Resnet



Film layer
[Make 2 Non linear transformation instead of 1 linear]

For ygen [X for first resnet should be Identity.

For guesser part: Make it non lin

Date / /
Page

To do: see clever [& do they classify]

How to seperate guesser & questioner in this.

how to implement ~~the~~ Memory network.

decide model for Oracle

Reduce the number of question.

☆ Arrange ~~quest~~ games by length of questions in batchifier.

Date / /
Page

③ — change the config file.

① - Redefine the model.
② - include the resnet model
- change the train file

last. [Add dropout & regularization & other good basic ML tricks]

change softmax to logistic

train last layer of resnet with LSTM

[IMP]

I must be L2 normalized ⑥

Image → word.

(VQA) ← use as the first state

try deeper LSTM.

Date / /
Page

take 2D input

LSTM on A [Q | A] B [Q | A]

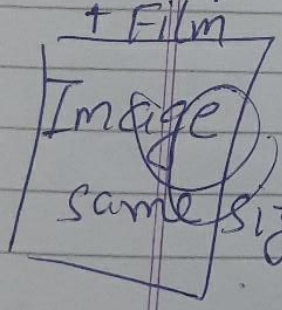
A significantly determines what type of new Ques

Single LSTM
A. B. C D . . X

For each Question you have a Representation
W_1 W_2 . . . last state

Let there be a learnable representation of A_1 A_2

[Append to all LSTM inputs]



Feed the last state of previous LSTM & Resized Image

W_1 W_2 to a LSTM
A_1 A_2

Pass this in LSTM to get the New state PTA using attention

Attention wrapper

Resized

have to tweak LSTM cell A bit

NOW in each step use what was done in ~~VQA~~ [ie attent extra Visual dialog size] [included in $W_1 A_1$

If You have to generate Q_2 use Image + [output of $W_1 A_1$]

Dual Attention Mechanism

image grounding +

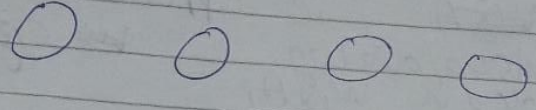
bringing things to same space seems to be a better option

To generate An Answer use [output of $W_{last} A_{last}$] using Resnet film

merging of different features.

[Merging has to be decided]

Different LSTM cell proposals



s_i = Att over different states s_j $j<i$

each time with

RNN

Predict a state s_j°

s_i → (s_j) $|s_j^\circ| < |s_i|$

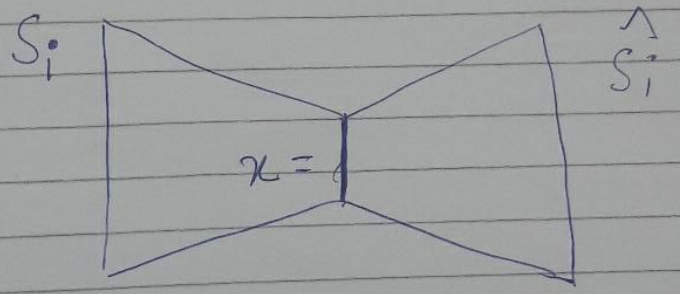
size [no of units

append ⊗ × so s_j such that

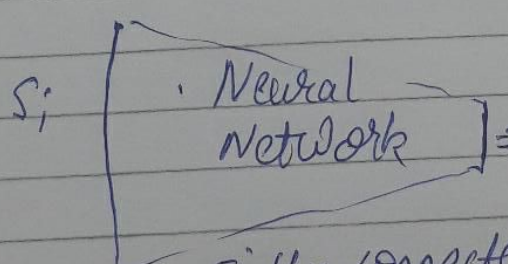
$|x : s_j| = |s_i|$

Appending can be done in several ways.

a) Autoencoder (Rigidual networks)



b)



Fully connected or convolutional

c) Use attention over $s_i, s_{i-1} \cdots$ to get a context state s_i then apply b or a to get x

d)

Film also has heatmaps

Memory Bank:
a) object embeddings & bounding boxes
b) Image features

Attention over

For generating Q: use Film with each question use film on image (with adding residue) & (conditioned on this generate new question

using:- 1) the RL NIPS Paper
2) Simple attention over Entire image + dig + bounding box
3) already implemented thing.

The one used for segmentation [U Net style] may be conditioned.

use the features of Pro[illegible]net for guesser.

Prob. 1 at Map: 1) heat map over bounding box & embeddings
→ embedding softmax
2) [illegible] (see paper)
3) over Entire pixels.