

OCR with MXNet Gluon

Work of Jonathan Chung (& Thomas Delteil)

Presented by Thom Lane

AWS MXNet Applications

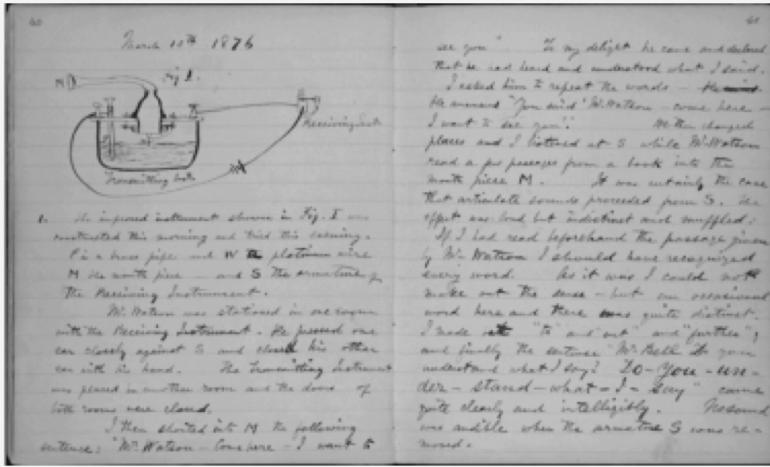
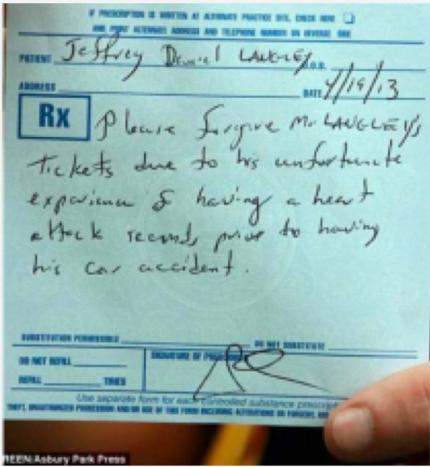
16th August 2018



Problem

Optical Character Recognition

Mom, I just wanted
 to tell you that Mother's
 Day wouldn't be
 possible without
 me. I'll be waiting
 for my present in the
 living room. Love,
 Joshua



Dataset

IAM Handwriting Database

Dataset

IAM Handwriting Database:

657 writers

1,539 pages of scanned text

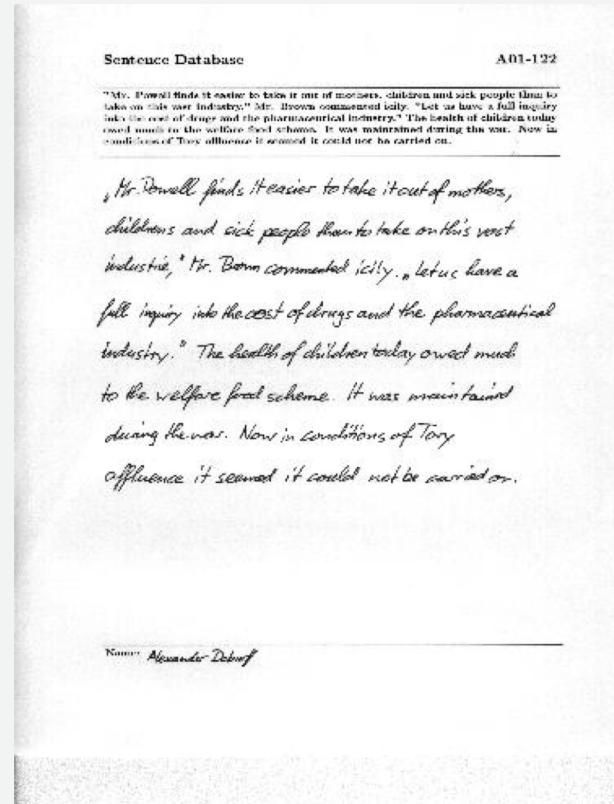
5,685 isolated and labeled sentences

13,353 isolated and labeled text lines

115,320 isolated and labeled words

<http://www.fki.inf.unibe.ch/databases/iam-handwriting-database>

(Register for username and password)



Solution

Handwriting Recognition Pipeline

Pipeline

Input image

Sentences Database **B03-023**

The discussion on current goals of multi-level councils could not go much further than, for example, the debate that policies of science and cultural development cannot be harmonized. What the council has learnt is that an achievement is to make religious content across the greatest political parties in what is not yet a unified world.

Page

But discussion on current goals of east-west integration could not go much further than, for example, the decision that policies of rescue and mutual disarmament cannot be followed together. What the council says down and it's an achievement - to make religious conflict across the greatest political barrier in what is not yet a ~~united~~ world.

Line segmentation

is not yet a unitary world.

Handwriting recognition

Bad discussion on current goals of east-west co-

It could not go much further than, for example, the
time when policies of resource and cultural disengagement
cannot be followed together. What the council has done
and it is an achievement - is to make religious con-
tact across the greatest political barrier in what

But discussion on current points of east-west conflict could not go much further than, for example, the truism that policies of menace and mutual disarmament cannot be followed together. What the council has done - and it is an achievement - is to make religious contact across the greatest political barrier in what is not yet a unitary world.

1) Page Segmentation

Page Segmentation

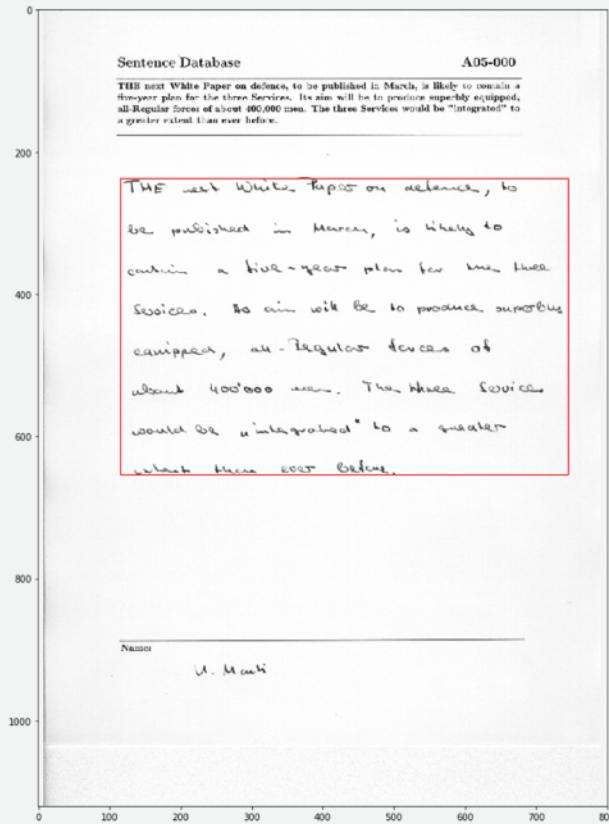
Input: Image with handwritten text region

Output: Bounding box of handwritten text

Problem: Single object detection

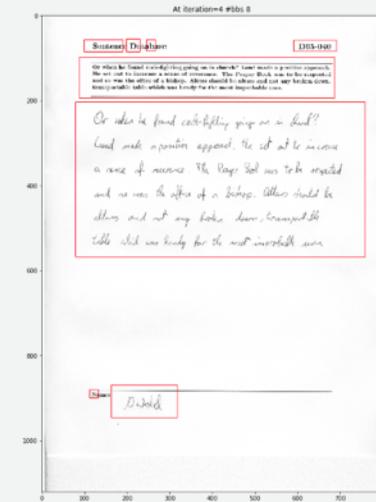
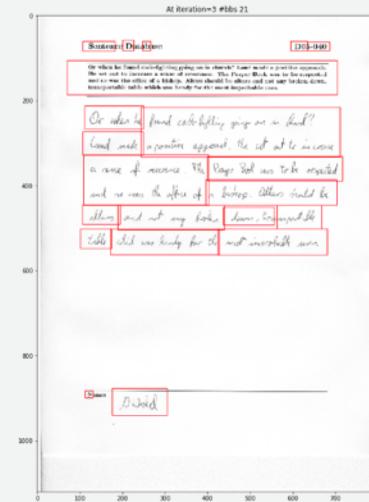
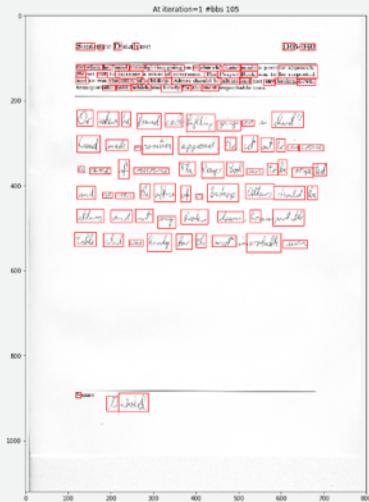
Solutions:

- 1) Using MSERs
- 2) Using CNNs

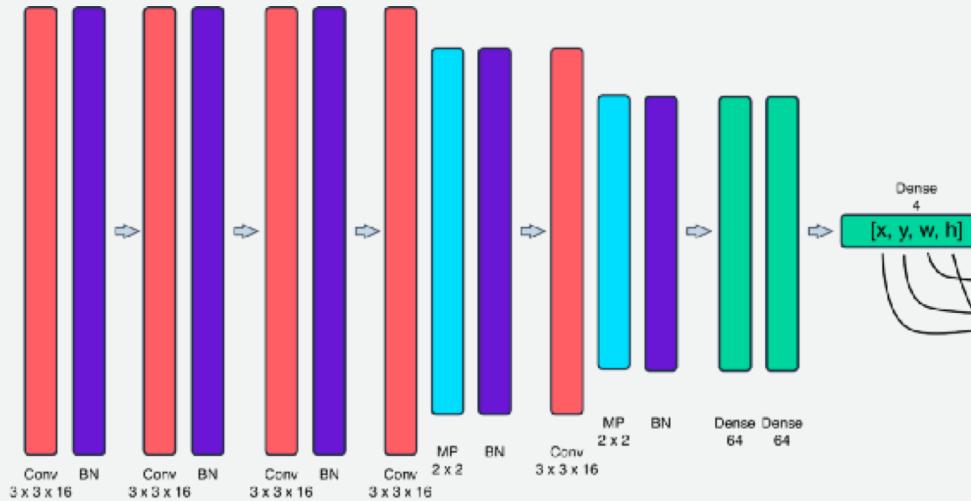


Using MSERs

Maximally Stable Extremal Regions (MSERs) algorithm



Using CNN



CNN: Data Augmentation

Kontext: Diskussion	G96-012
<p>Der Hörer ist während eines sozialen Treffens mit seinen Freunden. Er spricht über die politische Partei, die er für die nächsten Wahlen unterstützen möchte. Ein anderer Teilnehmer ist ebenfalls dabei und reagiert auf das Gesagte des Hörers.</p>	
<p>Meine Freunde haben sich darüber unterhalten, ob wir diese Partei unterstützen sollten. Sie diskutieren, ob sie sie unterstützen sollten, Aber ich habe mich selbst für sie entschieden, die sie besser versteht, was wir wollen, als wir es selbst tun.</p>	
<hr/> <p>Franziska Schmidlin</p>	

Section One: Database	F97-019
<p>Reference documents, used in the development of database, were used and are listed below. The following is a list of the sources used to develop the database and the source, and list of people. Other publications have been used in several sections, references and sources are available through literature that shall be made available upon request.</p>	
<p>Figures and agreements, which by this sounding:</p> <p>the situation of African music from ancient times until the eight century - such recording exists all of Africa with different - enough information at least three thousand of the recordings instead of trying the task with complete range to undertake. The idea of analytical, absolute and comparative simplicity, though obviously there conflicts could exist among the recordings.</p>	

EXHIBIT D

AM-617

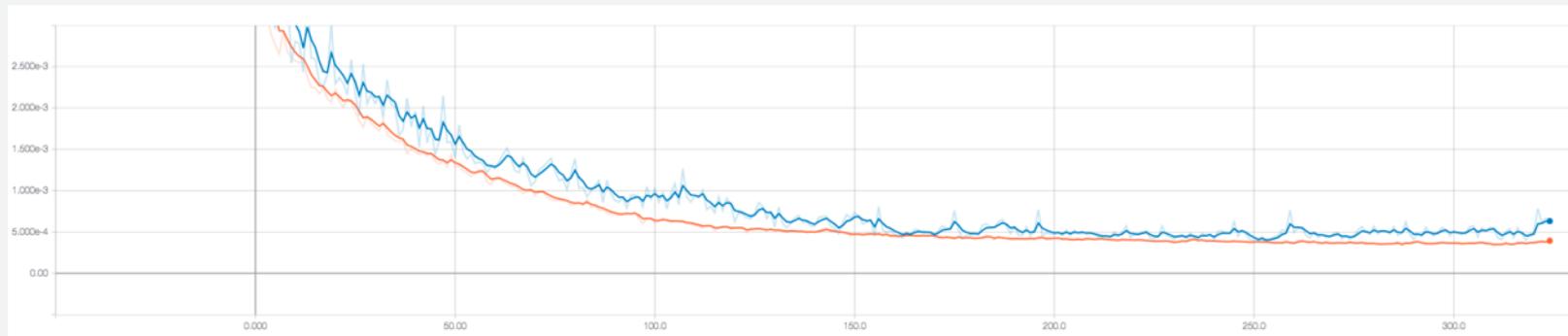
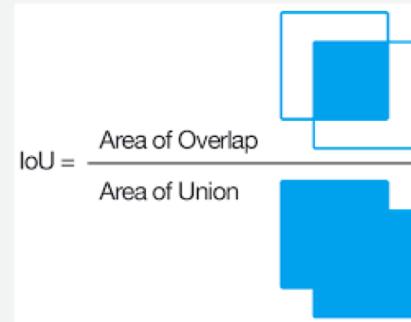
From: Chairman Pinochet to the Director General of Education
Date: 1983-03-23 10:00 AM
Subject: Re: The proposal of the Ministry of Education to close the
higher schools, change in the cost of studied subjects of the society to its
protectionism and the present situation.

Re: General Pinochet has not yet replied to
the memo, but recent documents made by his
team show that he still regards
himself as the only legal Prince Charles of
UK. His policy of strict neutrality from 1982
to 1983 kept the kingdom in peace, though
at the cost of virtual collapse of the country
for the pro-communist north and the
pro-western south and east.

Epoch 40

CNN: Training

Start by optimizing Mean Square Error.
When consistent overlap, then fine-tune IOU.



CNN: Results

2) Line Segmentation

Line Segmentation

Input: Image containing only handwritten text

Output: Bounding boxes for each handwritten line

Problem: Multiple object detection

Solutions:

- 1) Using SSD for lines
- 2) Using SSD for words

Using SSD

SSD: Single Shot Multi-box Detector

Customized:

1. Data augmentation
2. Anchor boxes
3. Network architecture
4. Non-maximum suppression

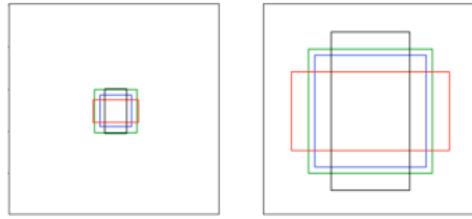
Line SSD: Data Augmentation

Many men with only limited accommodation
have to do their woodwork on the kitchen table.
Providing this is sound, some perfectly good
work can be done on it, but the usual problems
are those of the vice, the bench stop, and storage
place for tools. The combined bench top and
tool cupboard shown here has been specially
designed and made for woodworkers.

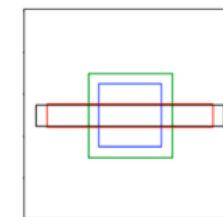
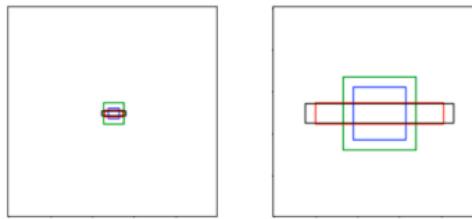
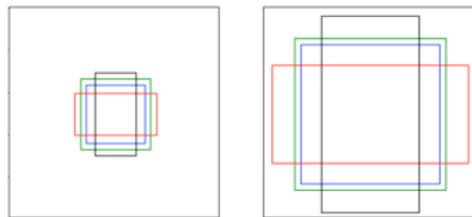
Mrs Brown, passionate and warm-hearted,
and Labour's attack on the higher health
outwardly meekish, replied with a
statistical statement - and ended by inviting
Labour MPs to angry up roar. One dealt
Service; the others tried to show that the
balance-sheets must always come first.

Line SSD: Anchor boxes

Standard

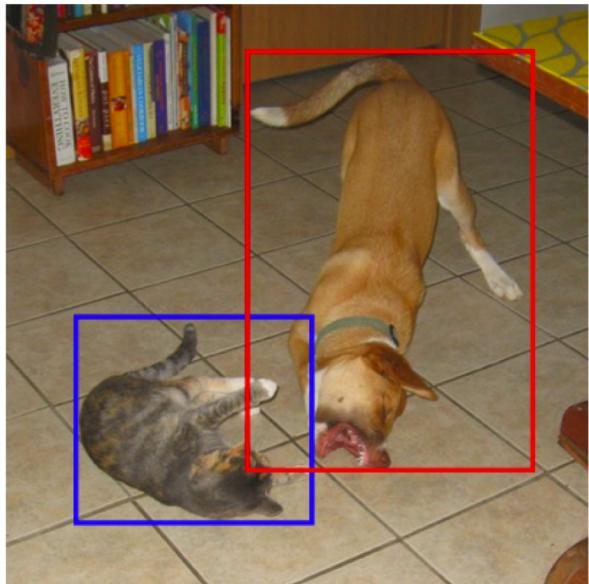


Custom

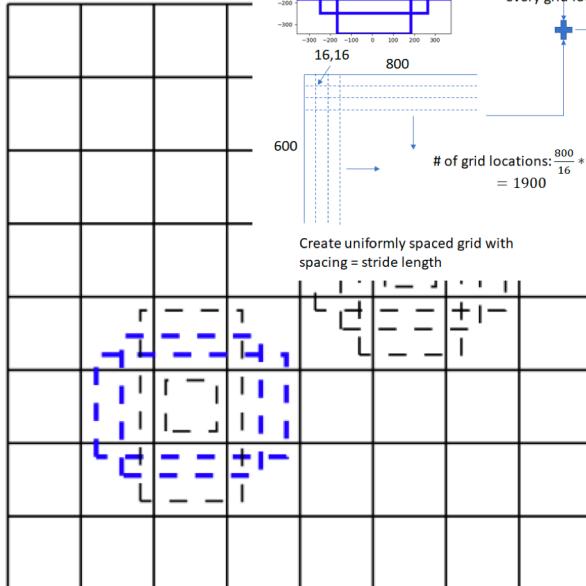


``mxnet.ndarray.contrib.MultiBoxPrior``

Line SSD: Anchor boxes



(a) Image with GT boxes



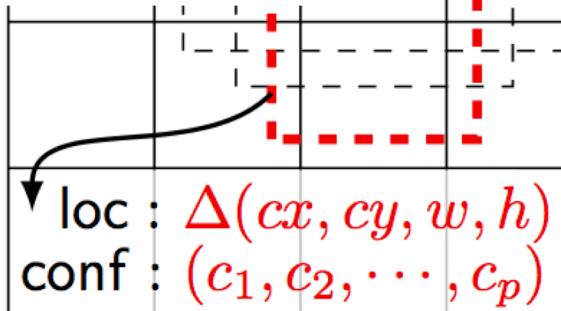
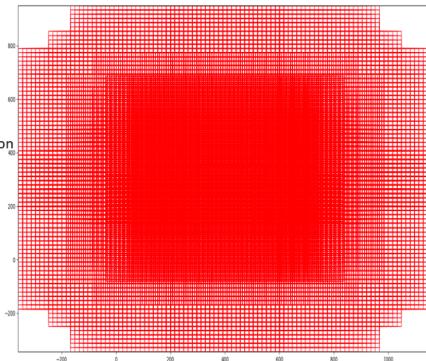
(b) 8×8 feature map

Generate Anchors

Given:

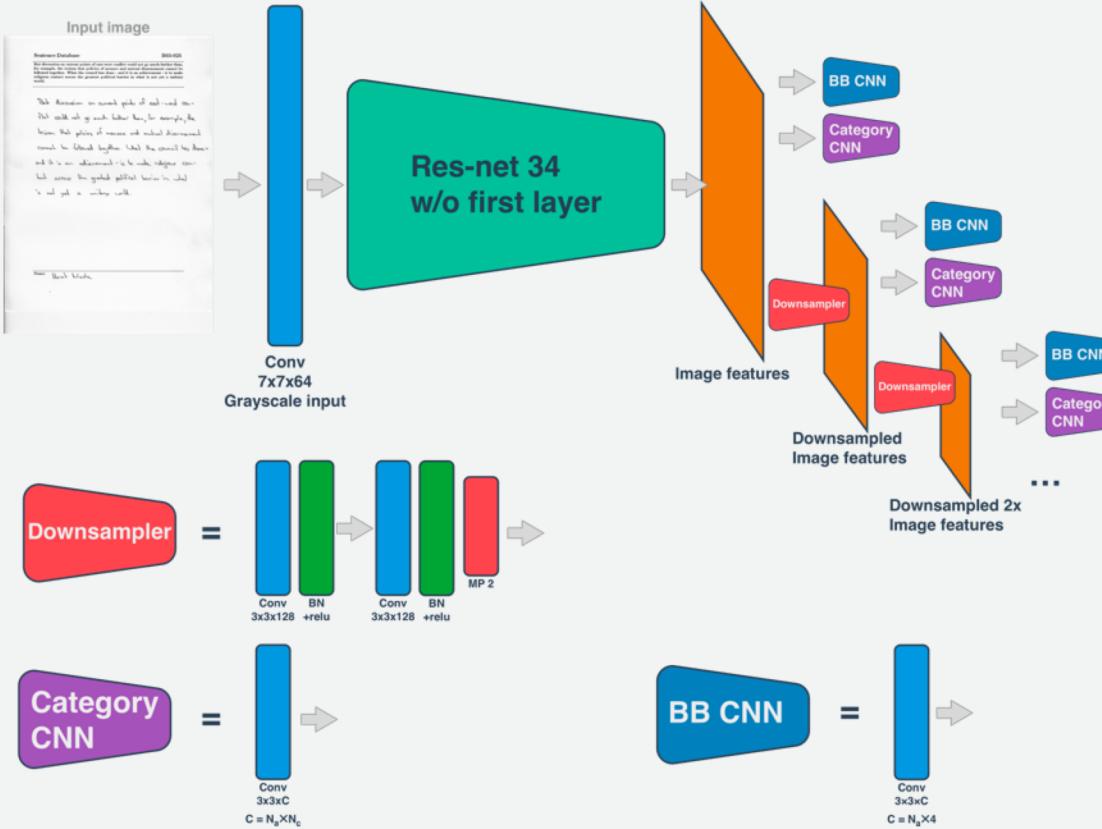
- Set of aspect ratios (0.5, 1, 2)
- Stride length (downscaling performed by resnet head: 16)
- Anchor Scales (8, 16, 32)

Total number of anchors: $1900 * 9 = 17100$
Some boxes lie outside the image boundary

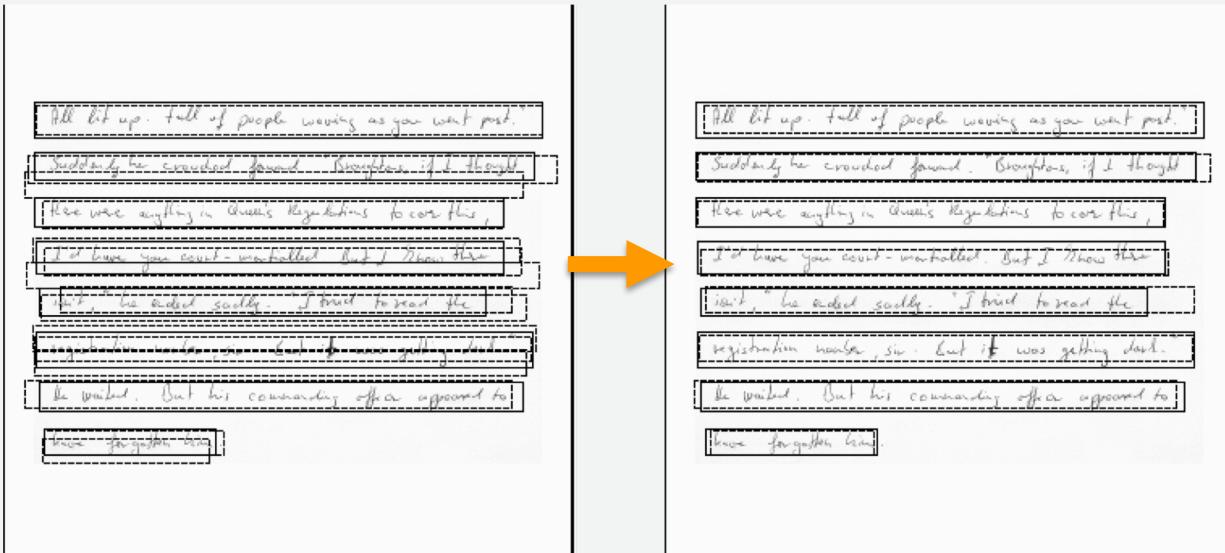


(c) 4×4 feature map

Line SSD: Network architecture



Line SSD: Non-Maximum Suppression



a) without non-maximum suppression

b) with non-maximum suppression

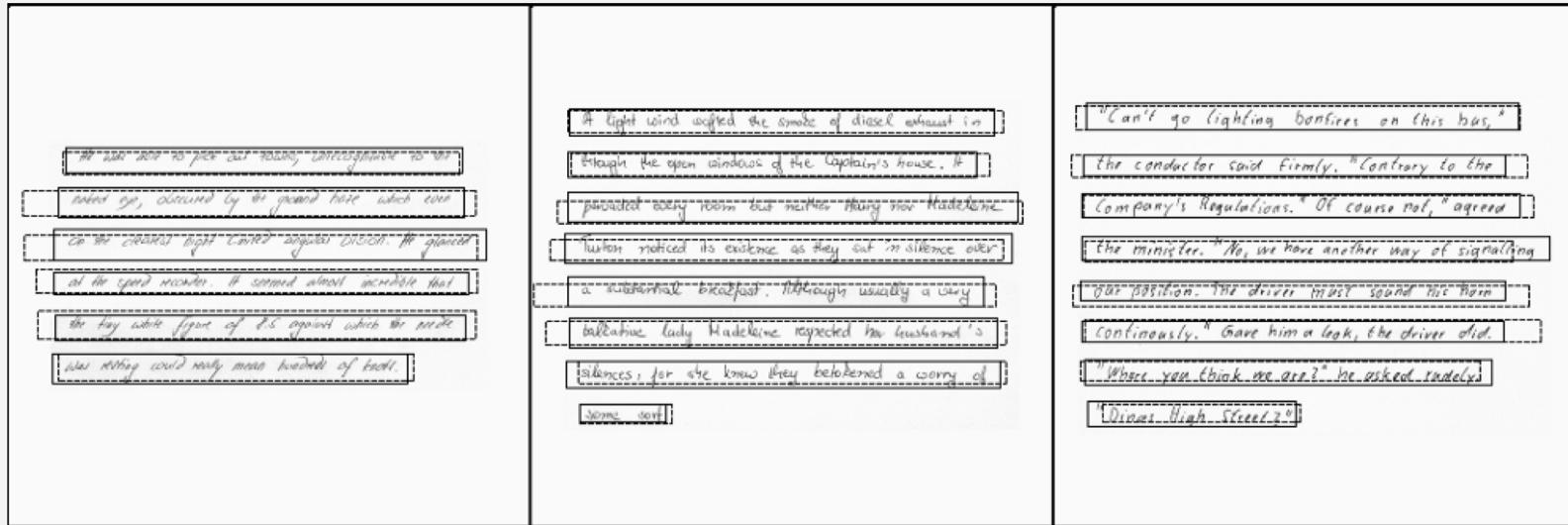
`mxnet.ndarray.contrib.box_nms`

Line SSD: Training

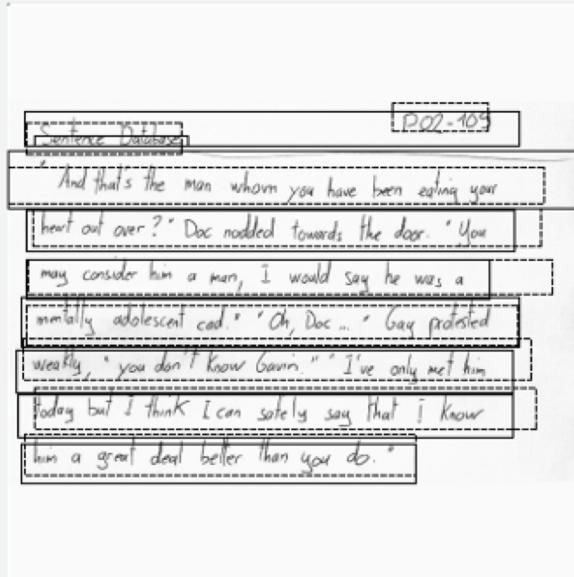
	<p>All quiet, and all expect, we a message did, and He advanced his lips, which the more my nose inclined to Touch his hair. I want now if you didn't be too no party, and begged to party with an arm. Now it right, yet I do like this book. I'm old enough and every are going out with each others. I don't believe you.</p> <p>But the first aircraft was already miles away. Coming down to the destination with the passed me at 4000 feet forward flight, over 37 was steering sharply away before descending, just after Community R.R. Station, Davis, started slowly at this Flight off with "Wait" just so inconveniently, and this off. And he had said that he had been flying back to you.</p> <p>Sort of incident.</p>	<p>At 1000 feet to pick out today's correspondence to 900 Picked up delivered by the express boat about 1000 On the way to the station, I saw a large crowd of and a group of spectators - the crowd - who had been waiting the big, tall, friend of his opinion which he needed was sitting could easily make hundreds of books.</p>
Epoch 40	<p>And as he spoke, the thought of Phillip (from a movie) room filled him with a new resolution. That was his son. Even in that his hands were tired. He dare not precipitate what might well be another Casanova. And so, first behind his back, being professional, Marble, and said with such deep politeness, such wisdom that</p> <p>SAT 111 P02:493</p> <p>"And that's the man whom you have been asking her out over?" Doc nodded towards the door. "You may consider him a man, I would say he was a possibly adolescent child." "Oh, Doc," - his mother whispered. "You don't know him." "I do - only met him today but I can safely say that I know him a great deal better than you do."</p>	<p>"Oh, she was done!" And then he giggled and the sound Circus like. She's lot returning to her family. "Doc's" going to the north (because of the break between), just Her side of Bellamy. This is the last time he will be concerned, but Lucy's said nothing. She was holding him and he was not even her best boy - that because her unfortunately would continue. Yes was surrounded with anger against his she had worked.</p>

Line SSD: Results

Train IOU = 0.593 & Test IOU = 0.573



Line SSD: Results



b) Ignored incorrectly labelled smudge
(second line from the top)



c) Failure cases: misaligned predicted box
(forth line from the bottom)

Word SSD



3) Handwriting Recognition

Handwriting Recognition

Input: Image of single line of handwritten text

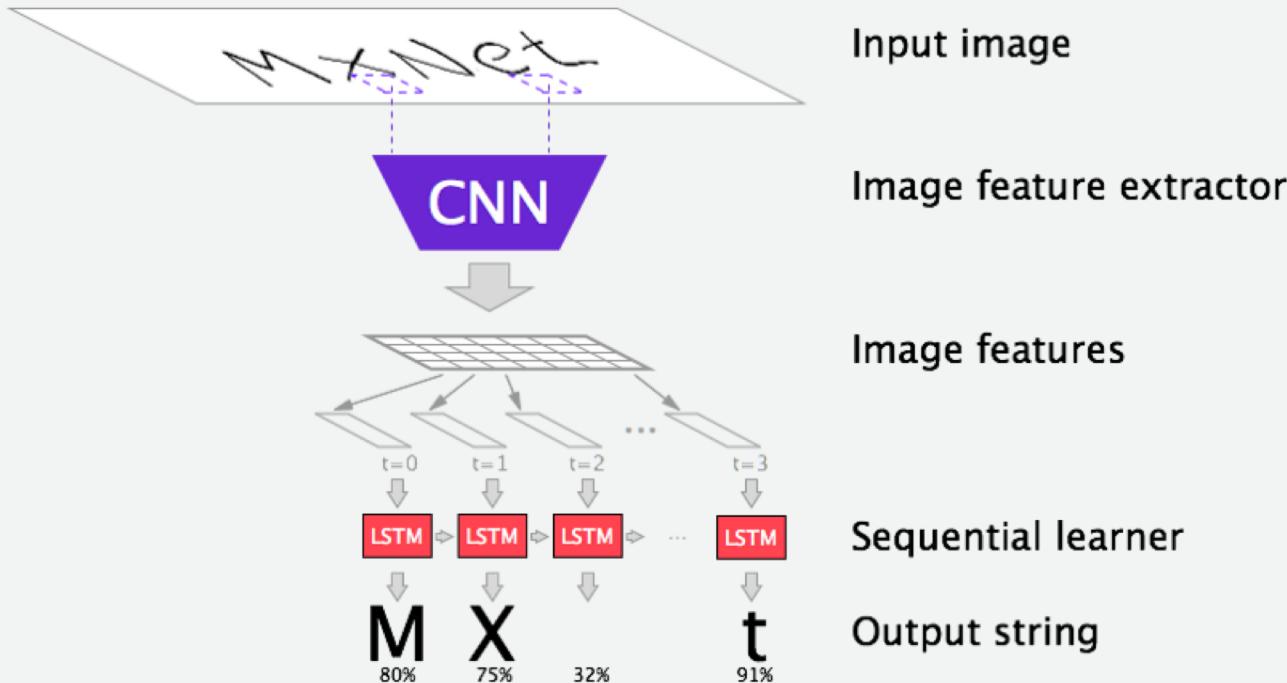
Output: Probabilities of each character for each time step.

i.e. an array of shape (sequence_length × vocab. of characters)

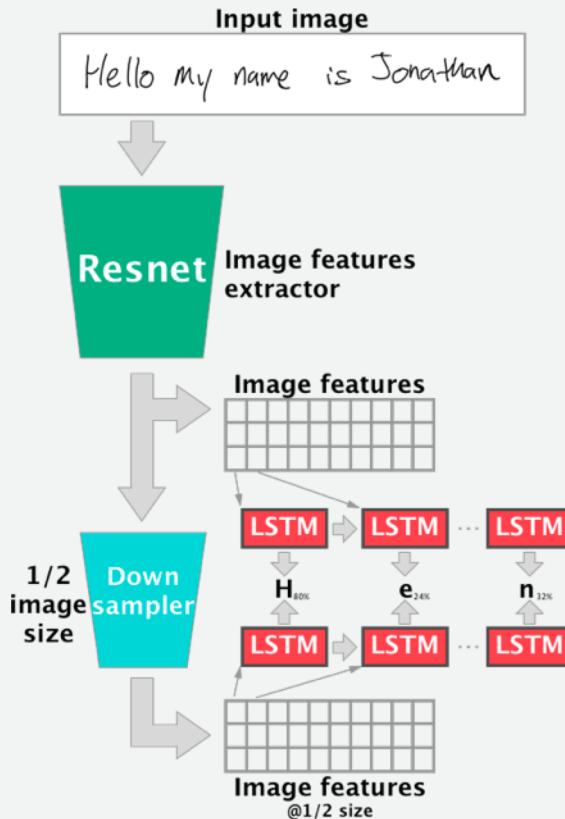
Inference Output: String of recognized text



Detection: Using CNN-BiLSTM



Detection: Adding down-sampling



Detection: With CTC loss

Problem:

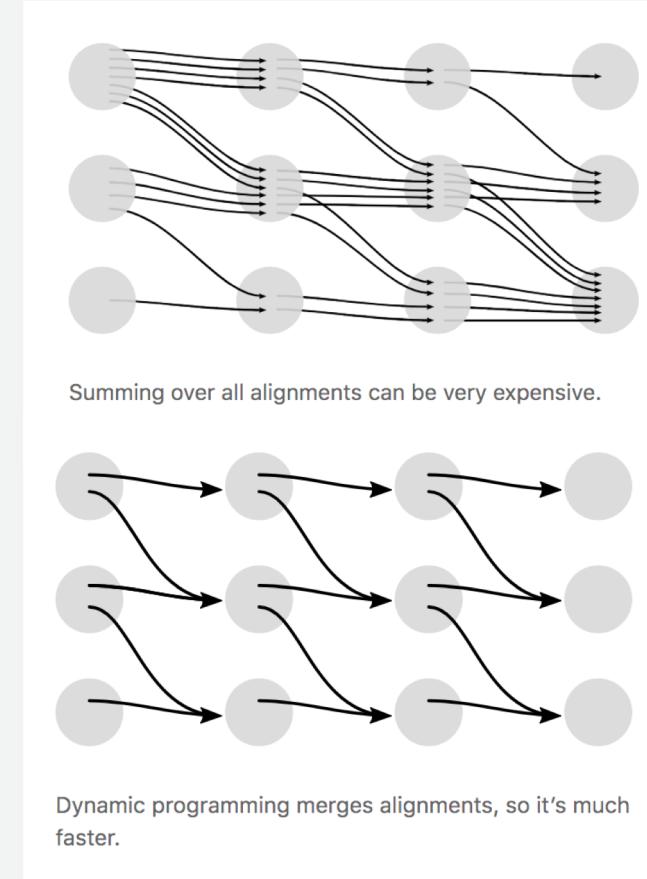
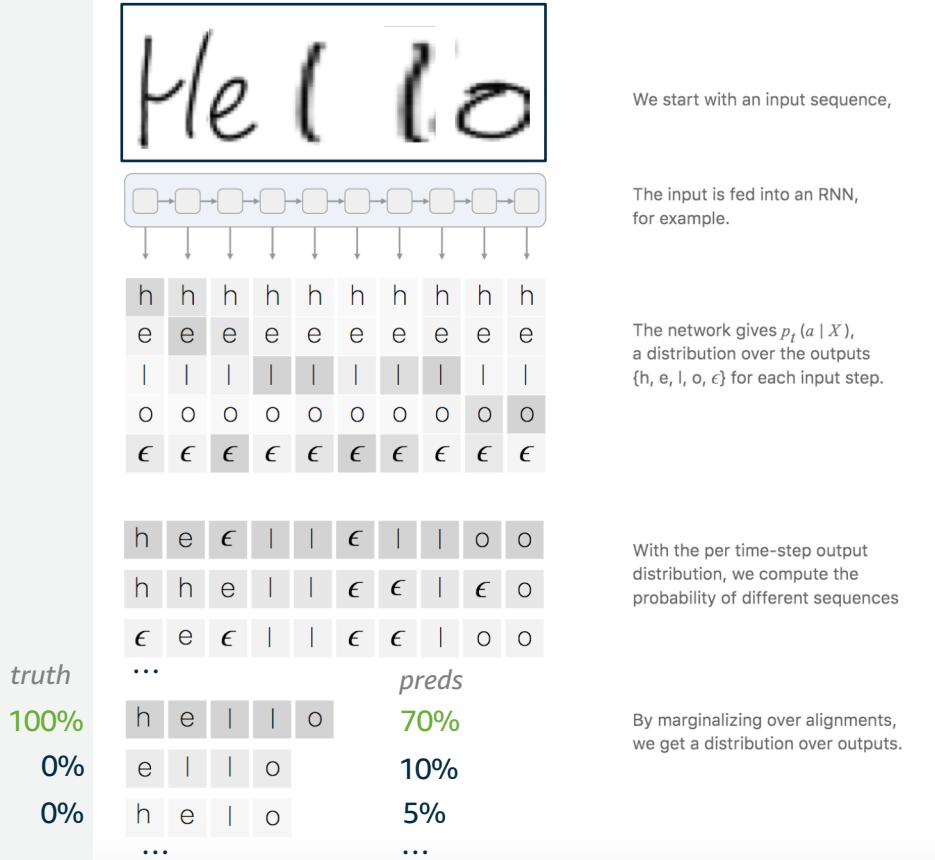
- Must have label for each time step.
- Or must model alignment.
- Must compress final sequence.

Solution:

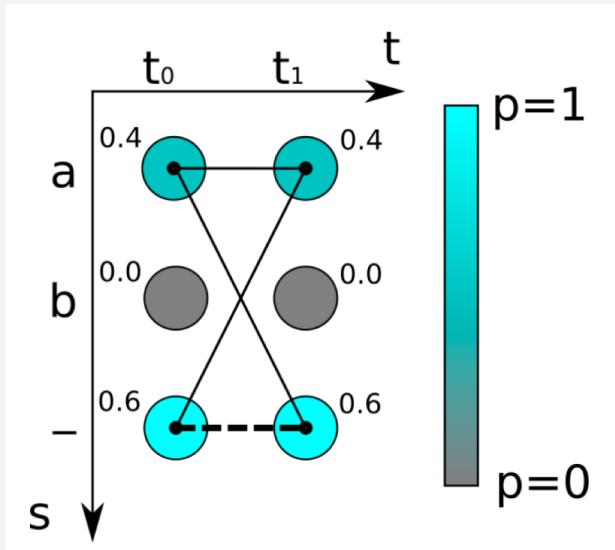
- Connectionist Temporal Classification loss.
- Defines a collapsing function.
- Assumes monotonic and many-to-one.



`mxnet.ndarray.contrib.CTCLoss`



Inference: Best Path vs Correct decoding



Best Path = $(-, -)$

Best Path Output = $C(-, -) = ""$

$$P("") = P(-, -) = 0.36$$

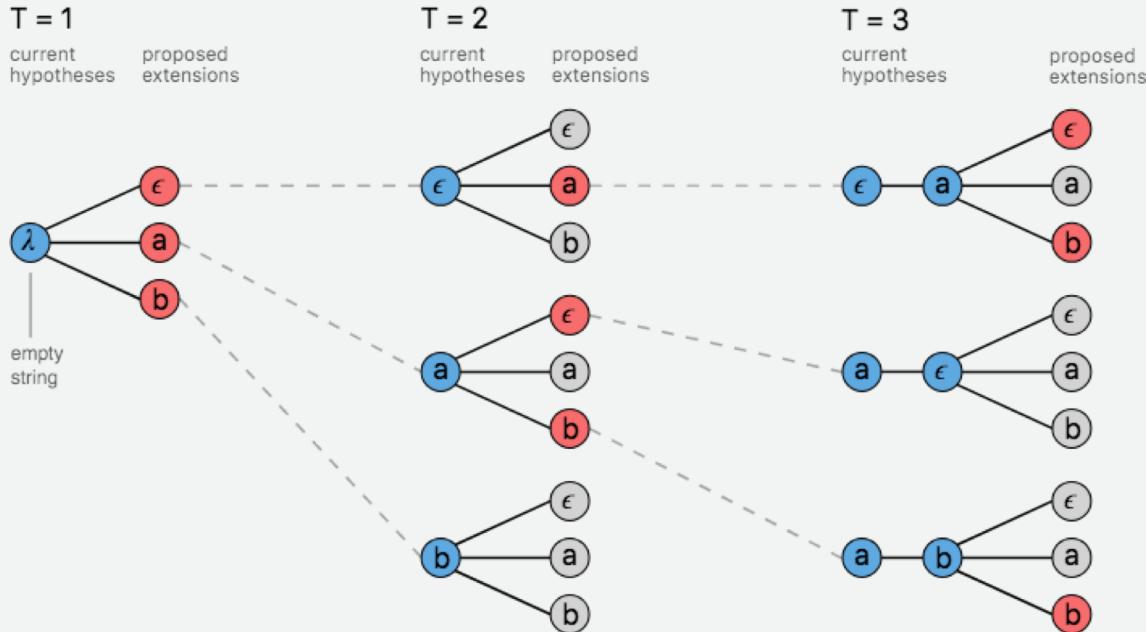
$$\begin{aligned} P("a") &= P(" - a") + P("a - ") + P("aa") \\ &= 0.6 * 0.4 + 0.4 * 0.6 + 0.4 * 0.4 \\ &= 0.64 \end{aligned}$$

Correct Output = "a"

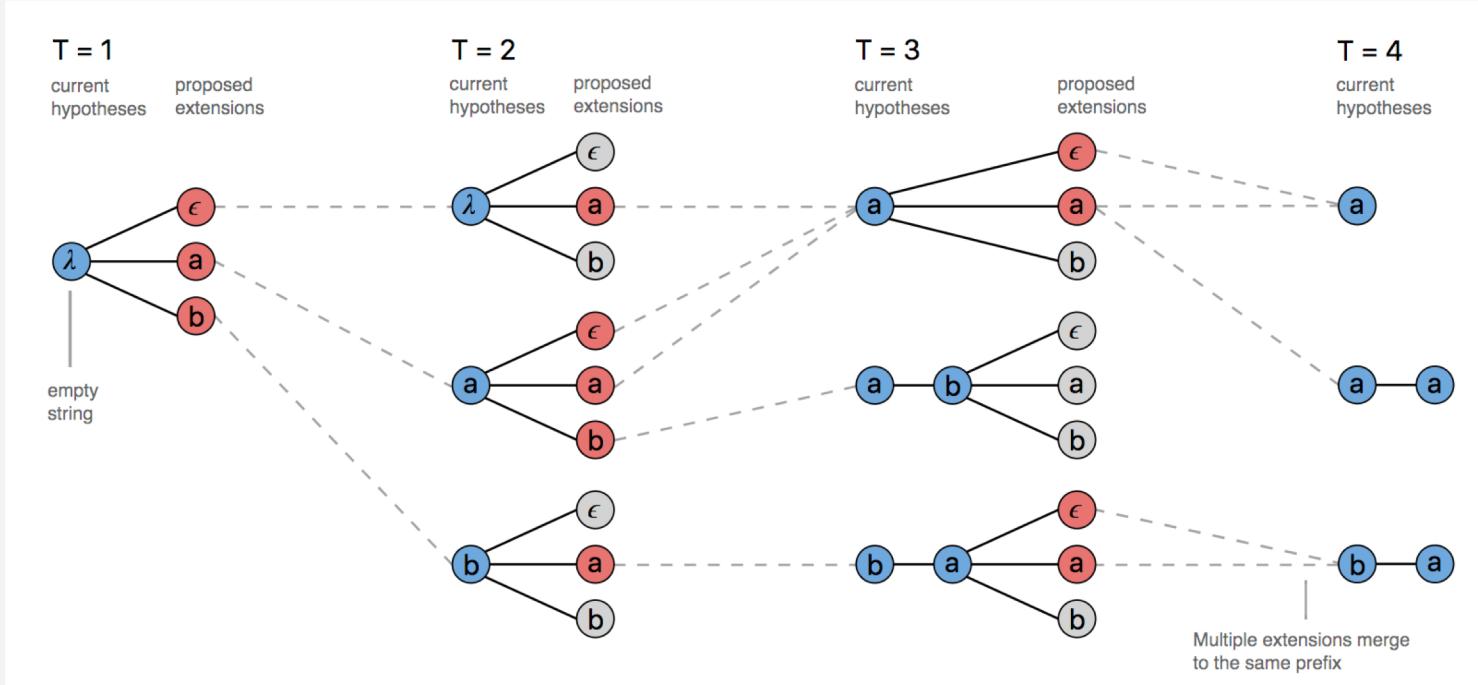
Can't check all paths though.

*alphabet_length^{sequence_length} paths.
i.e. 81^{30} paths!*

Inference: With Beam Search



Inference: With CTC Beam Search



Inference: Adding Lexicon Search

```
○○○

1 arg_max_string = get_arg_max_output(network_output)
2 cont = Contractions('word2vec model')
3 decontracted_text = list(cont.expand_texts([arg_max_string])[0])
4 words = tokenize(decontracted_text)
5 output_words = []
6 for word in words:
7     suggested_words = suggest_words(word)
8     dists = []
9     for suggested_word in suggested_words:
10         dists.append(edit_distance(suggested_word, word))
11     output_words.append(suggested_words[arg_min(dists)])
12 output_string = detokenize(output_words)
13 contracted_text = list(cont.contract_texts([output_string])[0])
```

Inference: Adding Language Model

```
○○○  
1 # using pretrained language model  
2  
3 lm_model, vocab = gluonnlp.model.get_model('awd_lstm_lm_1150', pretrained=True)  
4 hidden = lm_model.begin_state(batch_size=1, func=mx.nd.zeros)  
5  
6 tokens = tokenize(line_string)  
7 token_ids = [vocab[token] for token in tokens]  
8 data = mx.nd.array(token_ids).expand_dims(1)  
9  
10 output, hidden = lm_model(data, hidden)
```

And evaluate 'perplexity' of each candidate and take the one with lowest value as final prediction.

Inference: Results

a) Greedy algorithm outputs	b) Lexicon search outputs	c) Beam + lexicon search and language model output
got a these tovely things'; she woved a got all these tovely things'-she woved a	got a these lovely things'; she waved a got all these lovely things'-she waved a	got all these lovelythings'- she waved a got all these tovely things'-she waved a
self as a stanger in these panks self as a stranger in these pants	salt as a stranger in these pranks salt as a stranger in these pants	self as a stranger in these pars self as a stranger in these pants
the rannd-abont, which was pop the unanot - abont, which was fant	the rannd-abont, which was pop the unanot - abont, which was fant	the rounel abront, which mas p the unanot - abont, which was fant
has come forhim to he taken seriously has come for him to be taken seriously	has come forum to he taken seriously has come for him to be taken seriously	has come for him to be taken seriously has come for him to be taken seriously

Mean Character Error Rate (CER)

Greedy = 21.170.

Greedy + Lexicon Search: 21.152.

Beam Search + Lexicon Search: CER = 21.058.

Thanks!

And don't forget to check out:

<https://medium.com/apache-mxnet>

https://github.com/ThomasDelteil/Gluon_OCR_LSTM_CTC