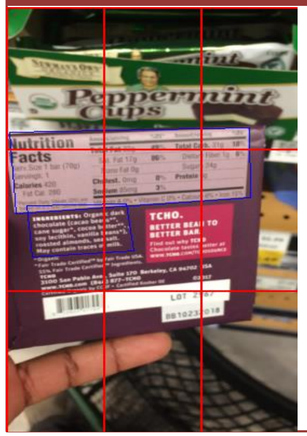
Spotting & Transcribing Structured Nutrition Information From Product Images.

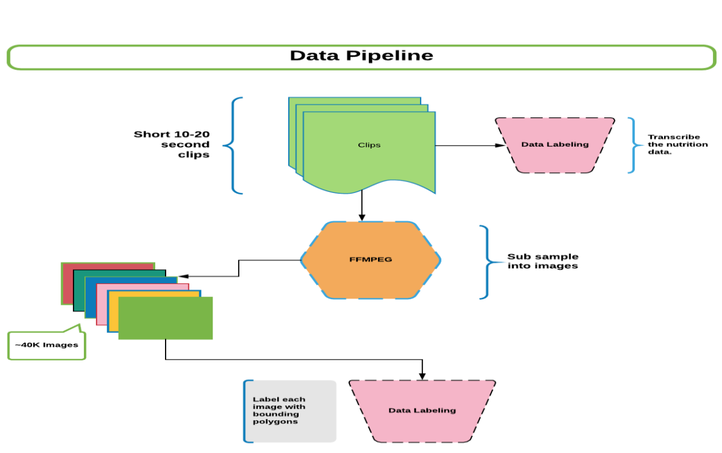
Okonda Joseph L.

OVERVIEW



- Nutrition Information encoded in nutrition fact-box and List of ingredients is hard to reason about.
- Computer Program would do a much better job of objectively assessing a product's healthfulness.
- Need to develop a method for machines to ingest the encoded information.

DATA & PRE-PROCESSING



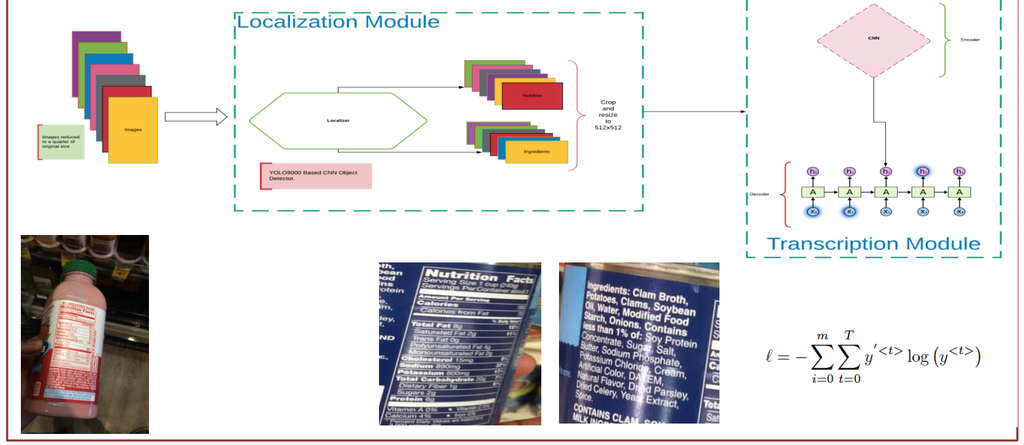
References:

[1] M. Jaderberg, A. Vedaldi, and A. Zisserman. Deep features for text spotting. In ECCV, 2014.

[2] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6517–6525, 2017.

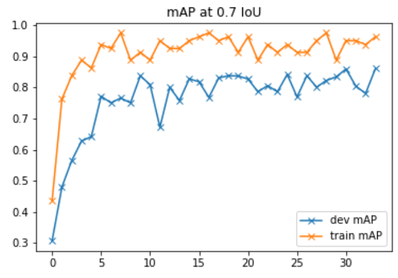
[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation. In ACL, 2002.

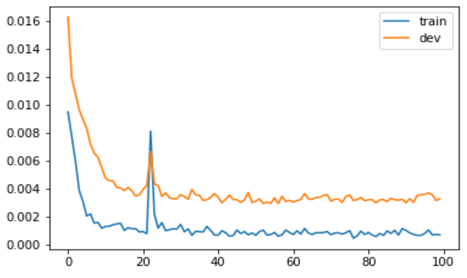
MODEL



$$\ell = -\sum_{i=0}^{m}\sum_{t=0}^{T} y'^{<t>} \log\left(y^{<t>}\right)$$

RESULTS: LOCALIZATION





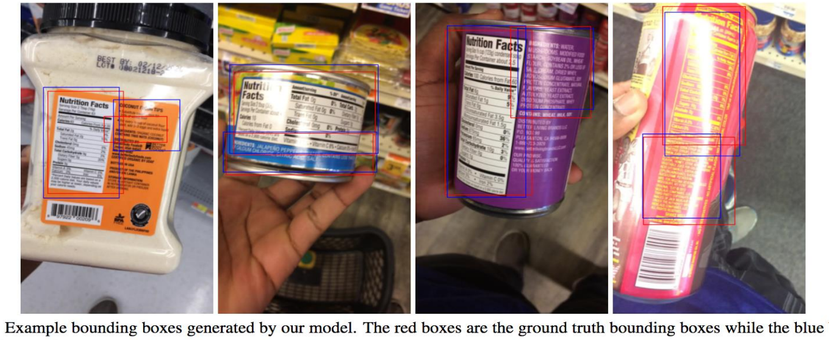
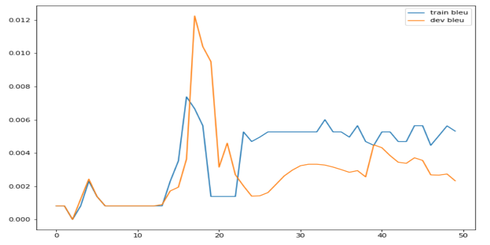


Figure 6. Example bounding boxes generated by our model. The red boxes are the ground truth bounding boxes while the blue boxes are predicted by our detector.

RESULTS: TRANSCRIPTION



- Fidelity of Transcriptions no so good.
- Model has hard time conditioning on image.
- Very Low BLEU Score.

```
Truth
---
<start> {
  num_servings : 2 8 serving_size : 4 0 g
  calories_per_serving : 1 6 0
  calories_from_fat : 4 0
  total_fat : 4 . 5 g sat_fat : 1 . 5 g
  trans_fat : 0 g poly_fat : 1 g
  mono_fat : 2 g cholesterol : 0 mg
  sodium : 4 1 0 mg potassium : 6 0 mg
  total_carb : 2 5 g dietary_fiber : 1 g
  total_sugar : 1 g added_sugar : n/a
  protein : 3 g
} <end>
Prediction
---
<start> {
  num_servings : 4 serving_size : 1 4 0 g
  calories_per_serving : 1 4 0
  calories_from_fat : 4 5
  total_fat : 4 . 5 g sat_fat : 0 . 5 g
  trans_fat : 0 g poly_fat : n/a
  mono_fat : n/a
  cholesterol : 0 mg
  sodium : 1 0 0 mg potassium : n/a
  total_carb : 2 4 g dietary_fiber : 2 g
  total_sugar : 1 3 g added_sugar : n/a
  protein : 1 4 g
} <end>
```

DISCUSSION & FUTURE

- One way to solve the decoder's conditioning problem would be to introduce an Attention module.
- Motion Blur in the training images is also a problem. We could solve this by having an explicit de-blurring module or by changing the data collection procedure.
- Localizer is main bottleneck because need to label a large number of images. Replacing it with a module that does not require bounding boxes would allow us to use more unique images.