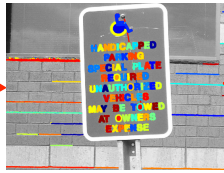


OCR



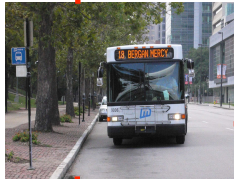
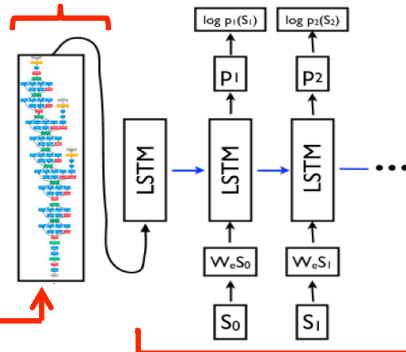
MSER



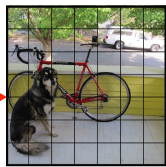
Remove non-text

Captioning

Encoder



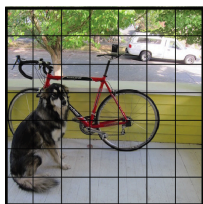
Object Detection



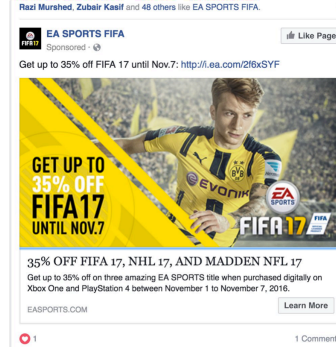
Grid on input image



Individual bounding box



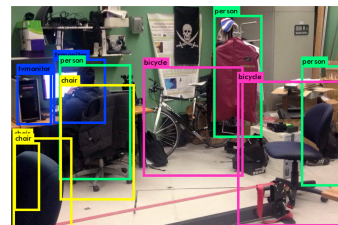
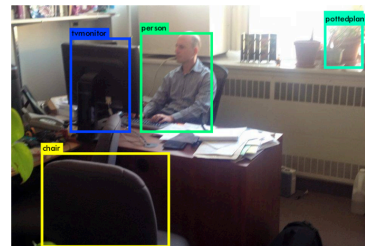
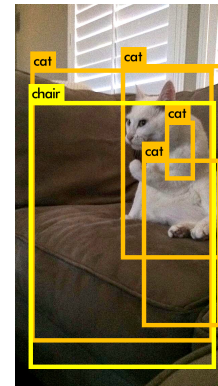
Grid on input image



web site, website, internet site, site (score = 0.98888)
jean, blue jean, denim (score = 0.00023)
comic book (score = 0.00017)
groenendael (score = 0.00011)
bannister, banister, balustrade, balusters, handrail (score = 0.00007)



Final Caption:
A white bus driving down a
street next to tall buildings
Text reads: MERCY BERGAN

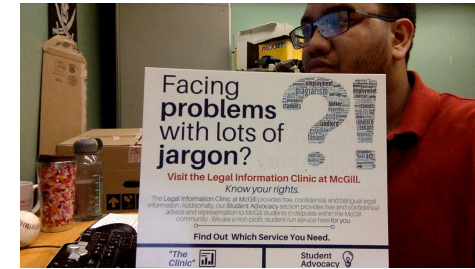
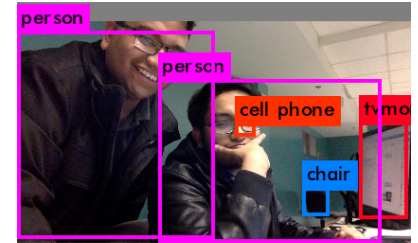


Examples

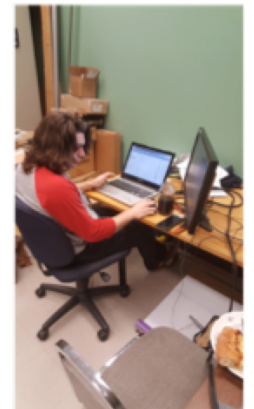


A street sign

Text reads: HAND\CAPPED'
PARK\NG PERMIT REQU\RED



A man holding a sign
Text reads: With lots of : jargon?
Visit the Legal Information
Clifiic at McGil Know your



A woman sitting in
front of a laptop
computer
Text Reads: none



A closeup of a clock on a table
Text reads: Roll Royce Trent
Powex



A man and a woman sitting on a
couch
Text Reads: none



IMAGE CAPTIONING

Sharhad Bashar
Professor: Dr. Jeremy Cooperstock
McGill University

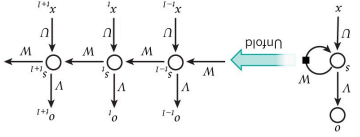
Faculty of Electrical and
Computer Engineering
ECSE 499 Honors Thesis

<https://github.com/SharhadBashar>



OCR

1. Maximally Stable External Regions (MSER)
 - Uniform Intensity
 - Surrounded by contrasting backgrounds
2. Remove non text region
3. Create bounding boxes
4. Combine bounding boxes
5. Apply OCR to get the text



Captioning

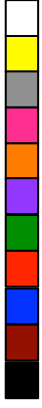
- Input Image I
Maximize P(S|I)
1. Encoder:
 - 48 Layer CNN
 - Input: Image
 - Output: Vector rep. of image
 2. Decoder:
 - RNN with LSTM blocks
 - Input: Vector rep. of image
 - Output: 3 captions with confidence

Object Detection

1. 32 Layer CNN
2. Break image into small squares
3. Generate Bounding boxes based on confidence
4. Combine bounding Boxes
5. Apply threshold

Color Detection

1. Break image into small squares
2. Avg. RGB value of each square
3. Use it to get a range of 1 most common colors:
4. Caption with highest count and confidence
5. score chosen



Cross Check

$\begin{bmatrix} caption[0] \\ caption[1] \\ caption[2] \end{bmatrix}$

$\begin{bmatrix} object[0] \\ object[1] \\ object[2] \\ \vdots \\ object[n] \end{bmatrix}$

Color Detection

