Capstone Proposal

Udacity ML Engineering Nanodegree

Kenneth Preston

In this capstone project proposal I will outline the following project details:

- The project's domain background — the field of research where the project is derived;
- A problem statement — a problem being investigated for which a solution will be defined;
- datasets and inputs — data or inputs being used for the problem;
- A solution statement — the solution proposed for the problem given;
- A benchmark model — some simple or historical model or result to compare the defined solution to;
- A set of evaluation metrics — functional representations for how the solution can be measured;
- An outline of the project design — how the solution will be developed and results obtained.

1) The project's domain background

...The field of research where the project is derived

This project is derived from the marketing field. The technique I am using is OCR (i.e. computer vision and object detection) to read images of instore product sales tags and other promotional material.

There are many examples of academic papers that use machine learning to address this problem. Some include:

- 2016; Sonal Paliwal, Rajesh Shyam Singh and H.L. Mandoria; Text Localization and Extraction from Still Images using Fast Bounding Box Algorithm; Oriental Journal of computer science & Technology; https://pdfs.semanticscholar.org/b822/8fa26ebba5f5f45f980e3bd41ddc7eb647e4.pdf (https://pdfs.semanticscholar.org/b822/8fa26ebba5f5f45f980e3bd41ddc7eb647e4.pdf)
- 2017; Baoguang Shi, Xiang Bai, Serge Belongie; Detecting Oriented Text in Natural Images by Linking Segments; he IEEE Conference on Computer Vision and Pattern Recognition (CVPR); http://openaccess.thecvf.com/content_cvpr_2017/html/Shi_Detecting_Oriented_Text_CVPR_2017_paper.html (http://openaccess.thecvf.com/content_cvpr_2017/html/Shi_Detecting_Oriented_Text_CVPR_2017_paper.htm
- 2016; Darko Zelenika, Janez Povh, Bernard Zenko; Text Detection in Document Images by Machine Learning Algorithms; Advances in Intelligent Systems and Computing book series (AISC, volume 403); https://link.springer.com/chapter/10.1007/978-3-319-26227-7_16 (https://link.springer.com/chapter/10.1007/978-3-319-26227-7_16)

2) A problem statement

...a problem being investigated for which a solution will be defined

A company currently has thousands of instore images of product tags and promotional material. This project would be similar to this one I found online (https://super-geek-news.github.io/articles/415657/index.html (https://super-geek-news.github.io/articles/415657/index.html) & https://medium.com/capital-one-tech/learning-to-read-computer-vision-methods-for-extracting-text-from-images-2ffcdae11594 (https://medium.com/capital-one-tech/learning-to-read-computer-vision-methods-for-extracting-text-from-images-2ffcdae11594)). This data is currently underutilzied as it requires a human to review all the images. There is no process currently for transcribing these images into text. This project is looking to use a machine learning application to process these images and extract the text with a high level of accuracy.

This problem involves object detection and then finally a classification problem to identify each character.

3) Datasets and inputs

...data or inputs being used for the problem

I decided to use a training set that does not include any private data to avoid any legal or ethical complications. To complete this project and be able to release it publically, I will be using a public dataset. This dataset is DeTEXT: A Database for Extracting TEXT from biomedical literature figures. It is available here for download: http://prir.ustb.edu.cn/DeTEXT/ (http://prir.ustb.edu.cn/DeTEXT/)

Each sample has two files:

1. JPG file of the image
2. gt file of the text that appears on the image and the location of the text

The features for this dataset are the pixals in the image files (JPG). The target is the text at the locatons identified in the gt file.

- Train Set has a sample size of 100 images and corresponding gt file
- Validation Set has a sample size of 80 images and corresponding gt file (Orinal validation set is 100n, but splitting up to have a test and validation set)
- Testing Set has a sample size of 20 images and corresponding gt file

The dataset is slightly unbalanced because it has many character symbols. Characters all alphanumeric symbols and additional ones like dashes and brackets. So there are some characters that appear often, like 'a' and some that appear infrequently, like the '=' symbol.

4) A solution statement

...the solution proposed for the problem given

Using sagemaker's training resources I will design a custom deep learning ocr model to locate text in an image and return the text. This model will then be connected with an API and a website will be built where a image can be uploaded and the text results will be returned.

5) A benchmark model

...some simple or historical model or result to compare the defined solution to

I have previously attempted to use tesseract ocr alone to address this problem. The results were inaccurate and unreliable. I can compare to the tesseract model to this approach.

6) A set of evaluation metrics

...functional representations for how the solution can be measured

To evaluate this model I will use a F-Score because it is used in the image to text academic papers I reviewed (http://openaccess.thecvf.com/content_cvpr_2017/papers/Shi_Detecting_Oriented_Text_CVPR_2017_paper.pdf (http://openaccess.thecvf.com/content_cvpr_2017/papers/Shi_Detecting_Oriented_Text_CVPR_2017_paper.pdf)). It also seems like the right metric because it incorporates precision and recall. Precision is important because it is important because it looks at the the portion of true positives of all positives. It is important to have high percision in this instance to avoid faslely identifying characters. Recall (or sensitivity) on the other hand is the model's ability to correctly find all of the characters that exist in the document (the proportion of true positives and false negatives).

So in conclusion I will look at F-Score, Precision, and Recall when evaluating my model.

7) An outline of the project design

...how the solution will be developed and results obtained

Development:

I will being by uploading the data into an s3 bucket to use for training, validation, and testing. The splitting has already been done for the training set. The validation set will be split randomly of the 100 included in the validation set to end up with 80 in the validation set and 20 in the test set.

Although I could use the prepackaged textract api from amazon, I will instead build a custom approach. The approach I am going to use an object detection algorithm to detect text in images. I will consider either of these approaches outlined below:

- https://github.com/awslabs/amazon-sagemaker-examples/blob/master/introduction_to_amazon_algorithms/object_detection_pascalvoc_coco/object_detectior (https://github.com/awslabs/amazon-sagemaker-examples/blob/master/introduction_to_amazon_algorithms/object_detection_pascalvoc_coco/object_detectior
- https://github.com/awslabs/amazon-sagemaker-examples/blob/master/advanced_functionality/distributed_tensorflow_mask_rcnn/mask-rcnn-s3.ipynb (https://github.com/awslabs/amazon-sagemaker-examples/blob/master/advanced_functionality/distributed_tensorflow_mask_rcnn/mask-rcnn-s3.ipynb)

Then I will use an approach to convert the located text on the images into actual characters.

I will try some of these approaches to accomplish this:

- BeamSearch: https://github.com/githubharald/CTCWordBeamSearch (https://github.com/githubharald/CTCWordBeamSearch)
- Tesseract: https://pypi.org/project/pytesseract/ (https://pypi.org/project/pytesseract/)

Model Evaluation

Using the results from the model on F Score, Precision, and Recall. I will compare versus the simple Tesseract approach to benchmark the model's performance

Deployment:

Deployment will be carried out in the following way:

```
- Tested model will be deployed to an endpoint.
- API will be set up to connect with a website.
- Lambda function will be used to processs the image before going to the model
- Website ui will be designed to upload file from website to the api
```

User results:

The processed image with extracted text will be returned to the user through the website.