Progress Report

08/05/2019
Mike Pack

Background

1. Page segmentation: Aims to identify image-like components—such as cartoons, illustrations, photographs, and maps—from Chronicling America corpus by using state-of-the-art deep learning model.
2. Metadata generation: Aims to build a metadata generator for image-like components.

Discussion of achievements

1. Page segmentation
   1.1. Using different training configuration, different models are trained using two datasets: (1) Beyond Words and (2) European historical NewsPaper.

Model	train/eval size	Classes	Weighted training	Pre-processing (Normalization)	Best Score	
					Accuracy	mIoU
BW_1500_v1	1226/306	0: Background 1: Editorial cartoon 2: Comics/cartoon 3: Illustration 4: Photograph 5: Map	No	No	0.87	0.24
BW_1500_v2			Yes [10;22;20;18;8;22]		0.88	0.26
ENP_500_v1	385/96	0: Background 1: Text 2: Figure 3: Separator 4: Table	Yes [5;10;40;10;35]	No	0.88	0.64
ENP_500_v2				Yes	0.89	0.64
ENP_500_v3			**No**	**No**	**0.91**	**0.69**
ENP_500_v4				**Yes**	**0.91**	**0.69**

*Accuracy: Pixel-wise accuracy.
*mIoU: Average intersection over union.
*Normalization: Zero mean unit variance

   1.2. Note that **models trained with ENP dataset** show better segmentation performance than with BW dataset in both accuracy and mIoU.
      1.2.1. High accuracy of BW_1500_v1 and v2 is not a surprising result. Since most of each image has only a few regions of interest, so from a model's point of view, assuming and predicting most of pixels to be a background is guaranteed to obtain high accuracy. The models' this sort of behavior can be verified based on their poor performance on mIoU metric, which quantifies the *percent overlap* between the target class and model's prediction.
      1.2.2. On contrast, high accuracy of from ENP_500_v1 to v5 is a meaningful result. Since text regions are included in the ground-truth, and thus a model's simple guessing that everything is background will get penalized. Also, we can see relatively high mIoU measures.
   1.3. Note that pre-processing does not play a significant role in improving segmentation performance.
   1.4. Note that weighted training causes a performance degradation.
   1.5. Post-processing has been implemented
      1.5.1. Eliminate small regions
      1.5.2. Draw bounding-box or polygon
   1.6. Actual testing on some Chronicling America images are shown in Figure 1 to 4.

2. Metadata-generation
    2.1. Approach 1: Google Cloud Platform (GCP) Vision API
        2.1.1. GCP Vision API provides a set of pre-trained machine learning models that can assign labels to images and quickly classify them into a number of predefined categories. For example, we can utilize their (1) object detection, (2) face recognition, (3) read printed and handwritten text, (4) similar image recommendation, or (5) basic image property generation (e.g., color space).
        2.1.2. For the visual demonstration, see Figure 5.
    2.2. Approach 2: Explore Automatic Image Annotation (AIA) research field and find the best model that fits our dataset.
        2.2.1. Since my work has mainly focused on page segmentation and GCP Vision API last week, I need more time to work on this.

Discussion of problems

1. Page segmentation
    1.1. As can be seen in Figure 2 to 4, there are some false-positive and false-negative results.
        1.1.1. We might improve the performance of our model with (1) more advanced data augmentation, (2) enlarged data set, (3) hyperparameter tuning, and (4) modifying architecture.
2. Meta-data generation
    2.1. Approach 1: GCP Vision API
        2.1.1. One thing in my mind is that the resultant metadata would be not that useful or end up with just entertaining result as my previous sentence generation idea. Since most of figures in newspapers are people, so most of the time, the GCP Vision API will label images to "person" or "people" as shown in Figure 5.
        2.1.2. For a large-scale data, there is a monthly usage.

Discussion of work that lies ahead

1. Segmentation
    ~~1.1. Training model with Beyond Words dataset to address data imbalance problem~~
    ~~1.2. Training model with enlarged ENP dataset~~
    1.3. Data augmentation
    1.4. Hyperparameter tuning
2. Meta-data generation
    2.1. Explore techniques to generate meta-data relevant to image quality
    ~~2.2. Explore techniques to generate meta-data relevant to image context~~
    2.3. Explore state-of-the-art methods in AIA field

Figures

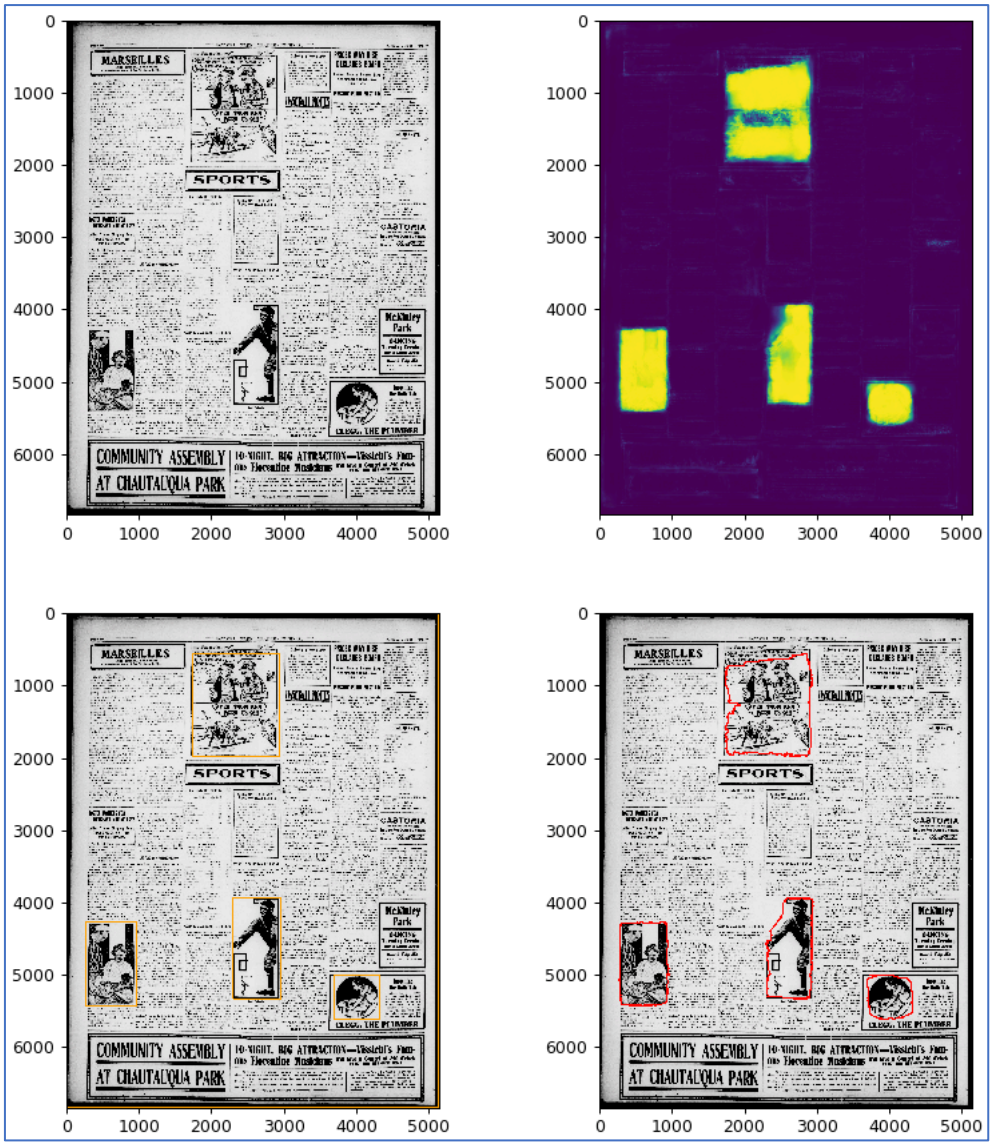

Figure 1. Segmentation result of ENP_500_v4 on Chronicling America image (sn92053240-19190805.jpg). Clockwise from top-left: (1) Input, (2) probability map for figure class, (3) detected figures in polygon, and (4) detected figures in bounding-box. In the probability map, pixels with higher probability to belong to figure class are shown with brighter color.

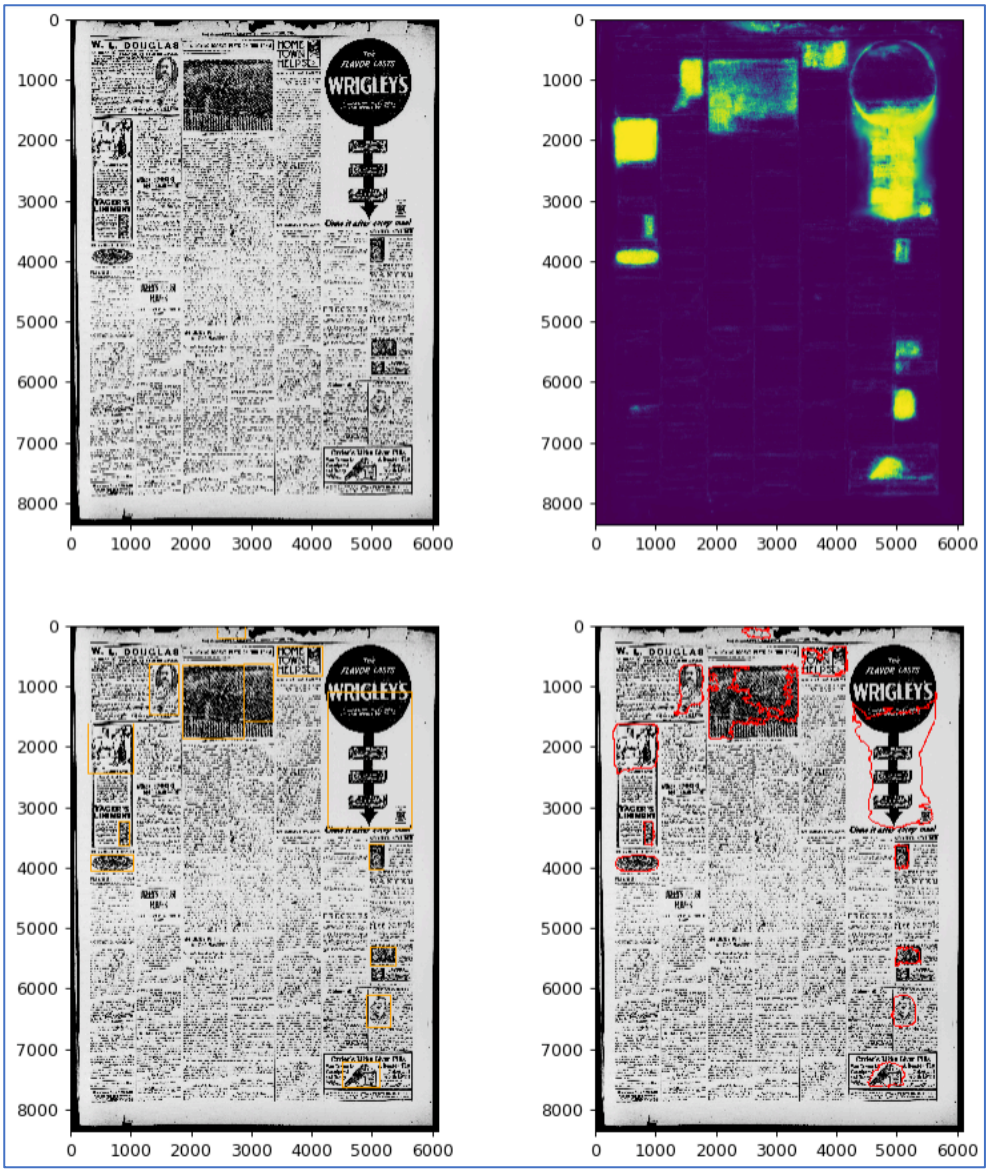

Figure 2. Segmentation result of ENP_500_v4 on Chronicling America image (ndnp-jpeg-surrogates_deu_descendo_ver01_data_sn84026820_00271765095_1917050501_0153.jpg). Clockwise from top-left: (1) Input, (2) probability map for figure class, (3) detected figures in polygon, and (4) detected figures in bounding-box. In the probability map, pixels with higher probability to belong to figure class are shown with brighter color.

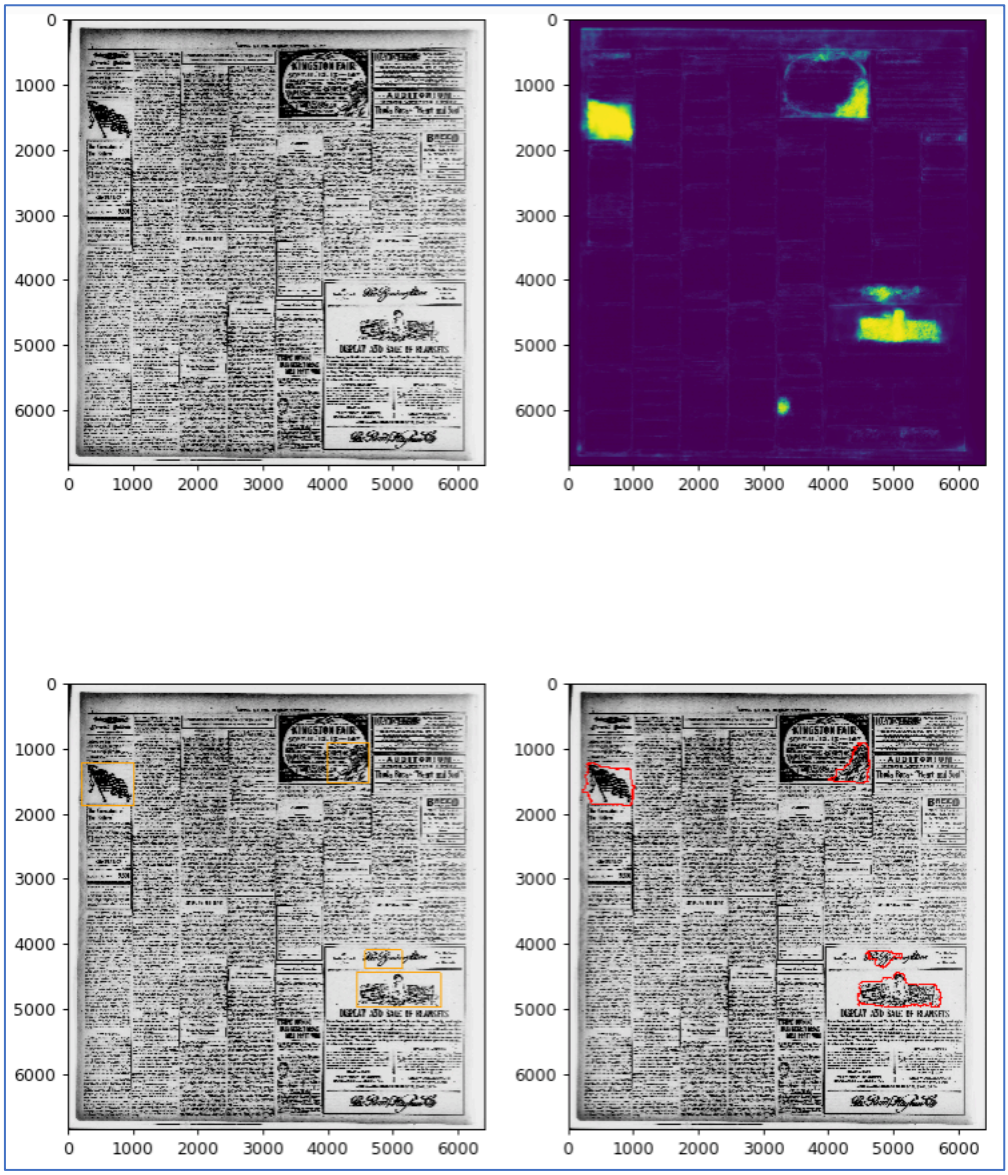

Figure 3. Segmentation result of ENP_500_v4 on Chronicling America image (ndnp-jpeg-surrogates_ct_berlin_ver01_data_sn82014086_00295866135_1917091301_0116.jpg). Clockwise from top-left: (1) Input, (2) probability map for figure class, (3) detected figures in polygon, and (4) detected figures in bounding-box. In the probability map, pixels with higher probability to belong to figure class are shown with brighter color.

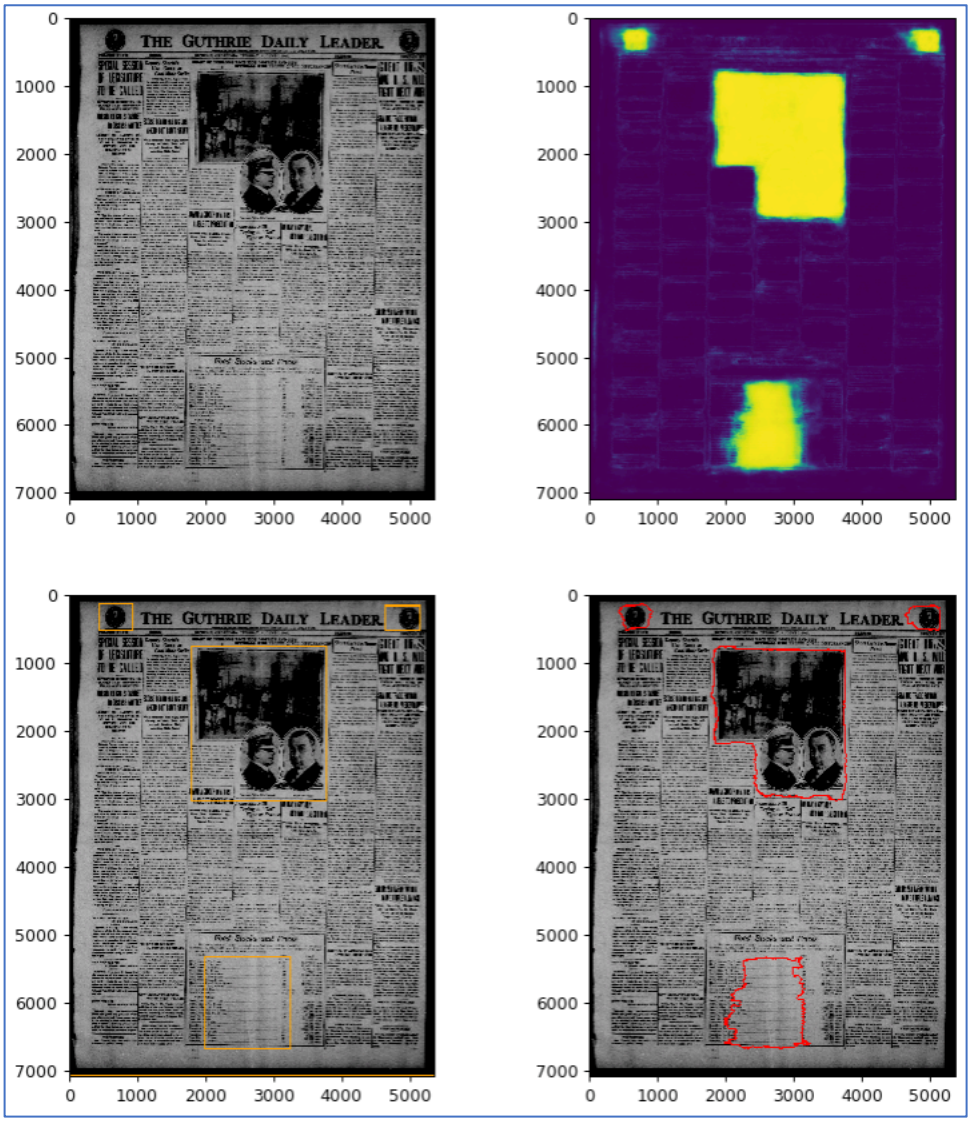

Figure 4. Segmentation result of ENP_500_v4 on Chronicling America image (sn86063952-19190805.jpg). Clockwise from top-left: (1) Input, (2) probability map for figure class, (3) detected figures in polygon, and (4) detected figures in bounding-box. In the probability map, pixels with higher probability to belong to figure class are shown with brighter color.

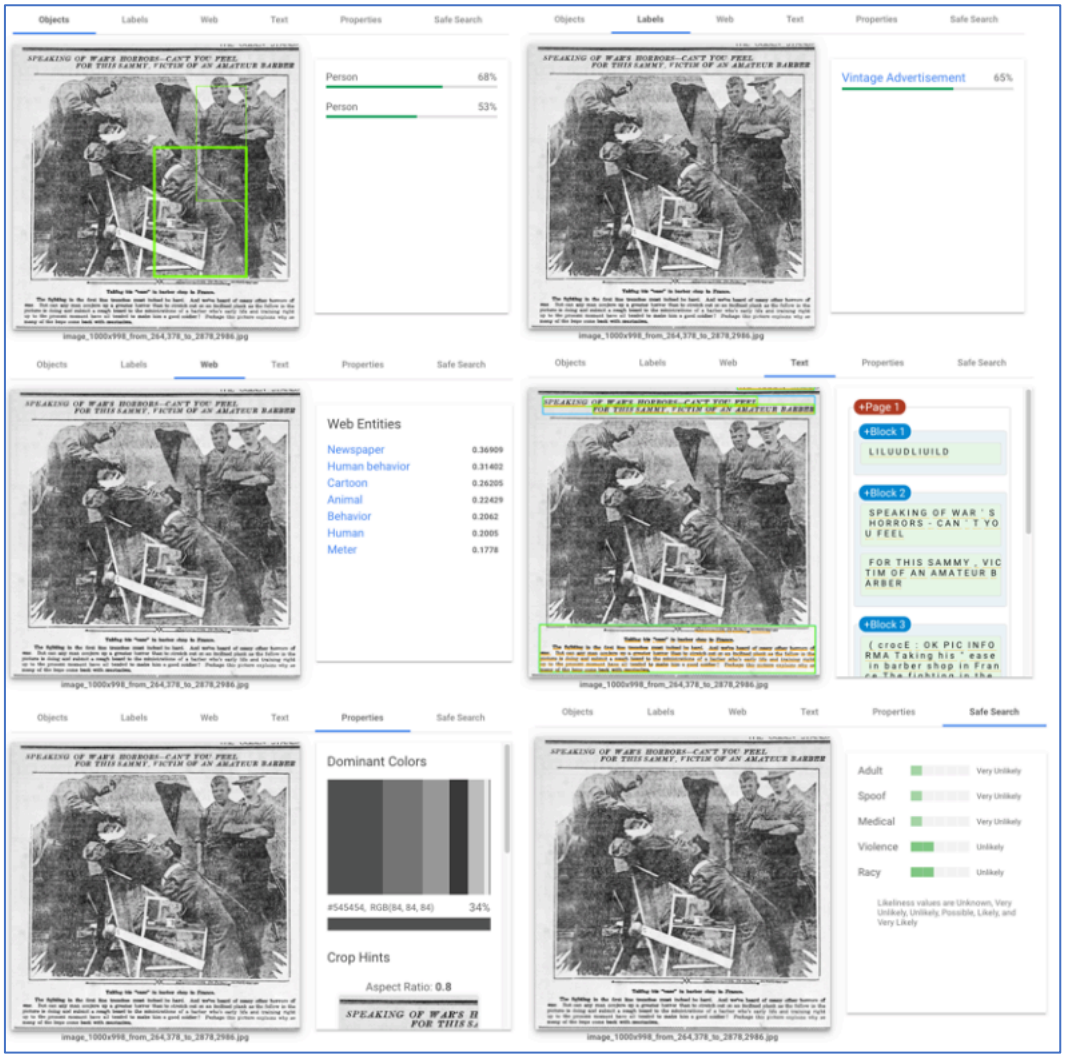

Figure 5. GCP Vision API demonstration on test image from Beyond Words (http://beyondwords.labs.loc.gov/#/view/5aa47eff639da00001002159).