Automate Data Entry With Computer Vision

New College of Florida - Nate Wagner

Survey Data

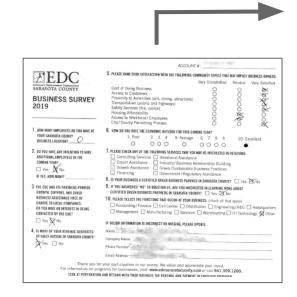
- Sent to 20,000+ Sarasota Businesses
- Received about 2,000 surveys
- Scanned as PDFs
- Goal: Build a model to read each survey and translate to CSV using computer vision techniques

	ACCOUNT	# :			
EDC BESSELE DEFECTION OF THE PROPERTY.	5. PLEASE RANK YOUR SATISFACTION WITH THE FOLLOWING CO	OMMUNITY TOPICS THAT Very Dissatisfied	MAY IMPACT I	BUSINESS OWNERS Very Satisfied	
SARASOTA COUNTY	Cost of Doing Business	0	0	8	
BUSINESS SURVEY 2019	Access to Customers Proximity to Amenities (arts, dining, attractions) Transportation (public and highways) Safety Services (fire, police) Housing Affordability Access to Workforce / Employees City / County Permitting Process	0000000	0000×	\$008 \$008	
1. HOW MANY EMPLOYEES DO YOU HAVE AT	6. HOW DO YOU RATE THE ECONOMIC OUTLOOK FOR THIS CON	WING YEAR?			
YOUR SARASOTA COUNTY	1: Poor 2 3 4 5: Average	6 7 8 9	10. 5	cellent	
BUSINESS LOCATION?	0 000 0	0000	10. 2	Cenent	
2. DO YOU HAVE ANY INTENTION TO HIRE ADDITIONAL EMPLOYEES IN THE COMING YEAR? Yes No IF YES, HOW MANY?	7. PLEASE CHECK ANY OF THE FOLLOWING SERVICES THAT YOU Consulting Services Workforce Assistance Export Assistance Industry/ Business R Growth Assistance Green/Sustainable B Financing Government/Regulat 8. IS YOUR BUSINESS A CERTIFIED GREEN BUSINESS PARTNE	e delationship Buildin Business Practices tory Assistance R IN SARASOTA COUNTY	g	B(No	
3. THE EDC AND ITS PARTNERS PROVIDE GROWTH, SUPPORT, AND OTHER BUSINESS ASSISTANCE FREE OF CHARGE TO LOCAL COMPANIES. DO YOU HAVE AN INTEREST IN BEING CONTACTED BY THE EDC?	9. IF YOU ANSWERED "NO" TO QUESTION #7, ARE YOU INTERE	ESTED IN LEARNING MO	RE ABOUT		
	CERTIFIED GREEN BUSINESS PARTMERS IN SARASOTA COUNTY? Yes No				
	 PLEASE SELECT THE FUNCTIONS THAT OCCUR AT YOUR BUSINESS: (check all that apply). 				
	☐ Accounting/Finance ☐ Call Center ☐ Distril☐ Management ☐ Manufacturing ☐ Services	bution	ing/R&D [] IT/Techno	Headquarters	
☐ Yes No	IF BELOW INFORMATION IS INCORRECT OR MISSING, PLEASE U	JPDATE.			
4. IS MOST OF YOUR REVENUE GENERATED BY SALES DUTSIDE OF SARASOTA COUNTY? Yes \(\subseteq \text{No} \)	Name				
	Company Name				
	Phone Number				

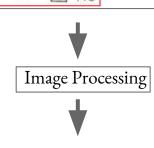
Thank you for your participation in our survey. We value and appreciate your input. For information on programs for businesses, visit: www.edcsarasotacounty.com or call 941.309.1200. TEAR AT PERFORATION AND RETURN WITH YOUR BUSINESS TAX RENEWAL AND PAYMENT IN PAYEL DES PROVINCE.

First Attempt

- Built training set
 - Hand-entered 300 surveys
- First attempt was to crop each question, then train a CNN to classify as
 - o [yes, no, no-answer]
- Test accuracy on "Yes/No" style question was 95%
- Issue was on the other style questions



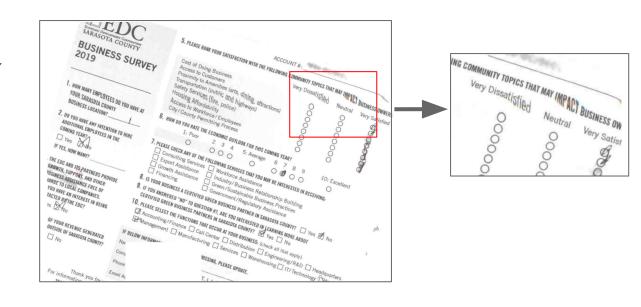
4. IS MOST OF YOUR REVENUE GENERATED BY SALES OUTSIDE OF SARASOTA COUNTY?





Issues With Scanning / Cropping

- When scanning 100s of surveys at a time, they don't always scan perfectly
- Cropping can be WAY off
- **Solution**: Deskew & Template Matching



OpenCV - Template Matching

Template Image





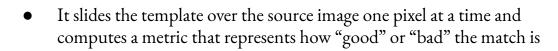
	0			
	ACCOUNT	#:	9.7	
	5. PLEASE RANK YOUR SATISFACTION WITH THE FOLLOWING CO	OMMUNITY TOPICS THAT	MAY IMPACT	BUSINESS OWNERS
SURVEY	Cost of Doing Business Access to Customers Proximity to Amenities (arts, dining, attractions) Transportation (public and highways) Safety Services (fire, police) Hassina Affordshilling	Very Dissatisfied		Very Satisfied

Source Image

BUSINESS SURVEY 2019	Across to Gatheries Plansifing to Annellise sint, diring, structure) Transportation (public and highways) Safety Services (the, galdes) Across to Washforst (Enginees) Across to Washforst (Enginees) City County Permitting Process
HOW MANY EMPLOYEES DO YOU HAVE AT YOUR SARASOTA COUNTY BUSINESS LOCATION?	6. NOW DO YOU RATE THE ECONOMIC GUTLOOK FOR THIS COMMING YEAR? 1: Poor 2 3 4 5: Average 6 7 8 9 10; Excellent
2. DO YOU HAVE ANY INTENTION TO HIRE ADDITIONAL EMPLOYEES IN THE COMING YEAR? Yes No IF YES, HOW MANY?	7. PLIASE CREEK AND OF HE FELLIPMING SWINGES THAT FIRS ME HE HET RETORATE IN RECEIVABLE Commuting Provinces Workfore Analisance Export Assistance Industry's Business Relationship Business Growth Assistance General Assistance Business Precisions Financing Government Propulsition Assistance Workfore W
3. THE COC AND ITS PARTNERS PROVIDE GROWTH, SUPPORT, AND DIMEN BUSINESS ASSISTANCE FREE OF CHARGE VO LOCAL COMPANIES. DO TOO HAVE AN INTEREST IN BEING CONTACTED BY THE EDC?	9. IF TO ASSENSE? OF TO DESCRIBE AP, OR TO DESCRIBE IN, LEARNING MODE LABOUT CONTROL OF THE MEMORISM PARKET OF TH
4. IS MOST OF YOUR REVENUE GENERATED BY SALES DUTSIDE OF SARASDIA COUNTY? Yes No	Name Compey Name Phone Nambor

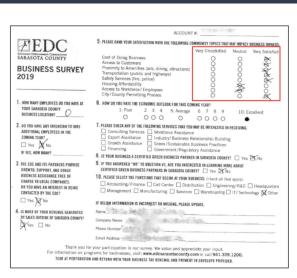
Thank you for your participation in our survey. We value and appreciate your input. For information on programs for businesses, visit: www.edcsarasotacounty.com or call 941.309.1200.

TEAR IF PERENBINA NA RETURN WITH TUR



Z EDO

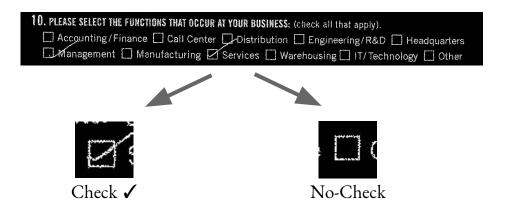
- (R) Result Matrix
- (x, y) Each Location
- Then for each location T over I, it computes and stores the result in the matrix R



$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x + x', y + y'))^{2}$$

Final Attempt

- 1.Deskew
- 2. Template Match Each Question
- 3. Crop
- 4. Trained CNN to classify each bubble as[check, no-check]
- Achieved test accuracy of +99% on all questions



Issues With Reproducibility

- EDC may want to change questions/survey design
- Template matching will no longer work without generating new templates / cropping dimensions
- Solution: Object Detection

Template Image



Source Image

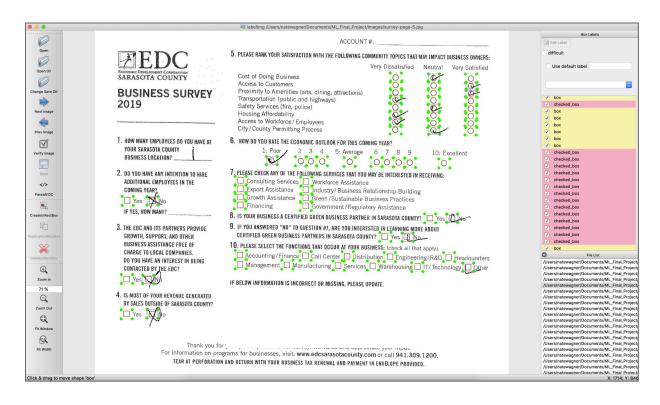


6. DID YOU RECEIVE COVID-19 RELIEF FUNDS?

Yes No

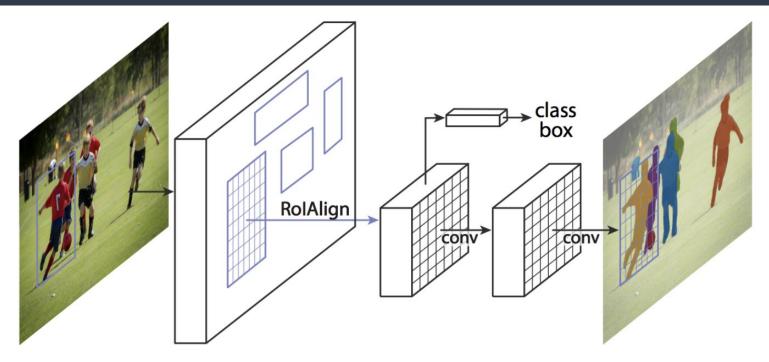
TILDO	5. PLEASE RANK YOUR SETISFACTION WITH THE FOLLOWING COMMUNITY TOPICS THAT MAY IMPACT BUSINESS OWNER			
TEDU	Very Dissatisfied Neutral Very Satisfied			
SARASOTA COUNTY	Cost of Doing Business			
BUSINESS SURVEY 2019	Access to Customers Provining to Amenities (aris, dining, attractores) Variance and the Amenities (aris, dining, attractores) Variance and the Amenities (aris, dining) Variance and the Amenities (aris, dining) Access to Workforce / Employees City / County Premitting Proces			
1. HOW MANY EMPLOYEES DO YOU HAVE AT	6. HOW DO YOU RATE THE ECONOMIC OUTLOOK FOR THIS COMING YEAR?			
YOUR SARASOTA COUNTY BUSINESS LOCATION?	1: Poor 2 3 4 5: Average 6 7 8 9 10: Excellent			
2. DO YOU HAVE ANY INTENTION TO HIRE ADDITIONAL EMPLOYEES IN THE COMING YEAR? YES NO NO IF YES, HOW MAKY?	7. PLEASE CREEK AND ON THE PLEASEMENTS SERVICES THAN 100 MAN FM KINESETED IN PECETIVING. Communiting Reviewing Westprices Assistance Export Assistance Industry's Business Relationship Busining Growth Assistance General Sustainable Businings Practices Pleasecing Government Pleasecing of Services Britancing Government Pleasecing of Services Government Pleasecing of Services			
3. THE EDC AND ITS PARTNERS PROVIDE GROWTH, SUPPORT, AND OTHER	9. IF YOU ANSWERED "NO" TO QUESTION 97, ARE YOU INTERESTED IN LEARNING MORE ABOUT CERTIFIED GREEN DUSINESS PARTMERS IN SARASDIA COUNTY?			
BUSINESS ASSISTANCE FREE OF CHARGE TO LOCAL COMPANIES.	10. PLEASE SELECT THE FUNCTIONS THAT OCCUR AT YOUR BUSINESS: (check all that apply)			
CHARGE TO LOCAL COMPANIES. DO YOU HAVE AN INTEREST IN BEING CONTACTED BY THE EDC?	□ Accounting/Finance □ Call Center □ Distribution □ Engineering/R&D □ Headquarter □ Management □ Manufacturing □ Services □ Warehousing □ IT/Technology ⋈ Other			
☐ Yes X No	IF BELOW INFORMATION IS INCORRECT OR MISSING, PLEASE UPDATE.			
4. IS MOST OF YOUR REVENUE GENERATED BY SALES OUTSIDE OF SARASOTA COUNTY? YOU NO	Name			
	Compety Name			
	Phone Number			
	Email Address			
For information on pro	r your participation in our survey. We value and appreciate your input, grams for businesses, visits www.edcsarasobacounty.com or call 94.1.309,1200, 1.488 EFERS WITH YOUR SUSPECTS AT RETURN LAW SPERMED IN STRUCT IN STRUCT AND OF PROMISES.			

Finding Boxes in an Image



- Annotated with Labelimg
 - Python program used to annotate images
- Annotated 300 survey images
- Goal: Use a Mask R-CNN to locate boxes and checked-boxes

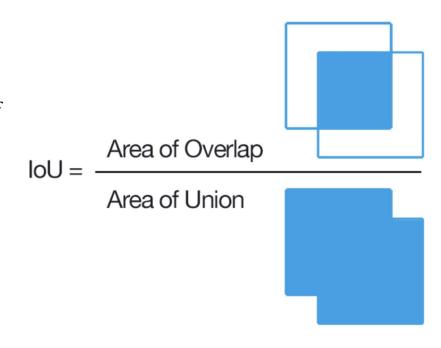
Architecture - Mask R CNN



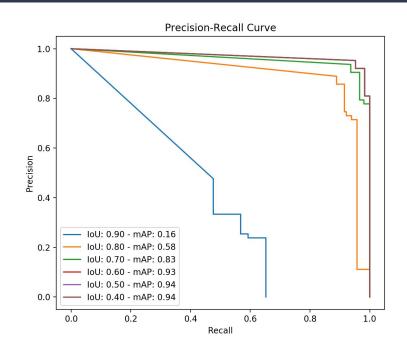
Mask R-CNN

Evaluate Model - mAP (Mean Average Precision)

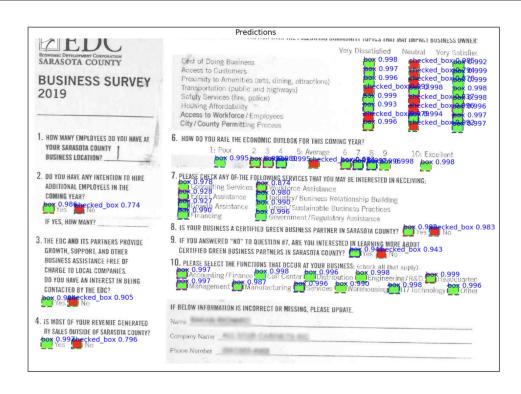
- Precision = TP / TP + FP
- Recall = TP / TP + FN
- IoU (intersection over union) is used to determine if a predicted bounding box is TP, FP or FN
- We define a TP, FP, or FN as
 - \circ TP if IoU > 50%
 - FP if IoU < 50% or duplicated bounding box
 - FN if no detection at all,
 IoU >50% but wrong classification
- The mAP is calculated as the area under precision-recall curve averaged over all images



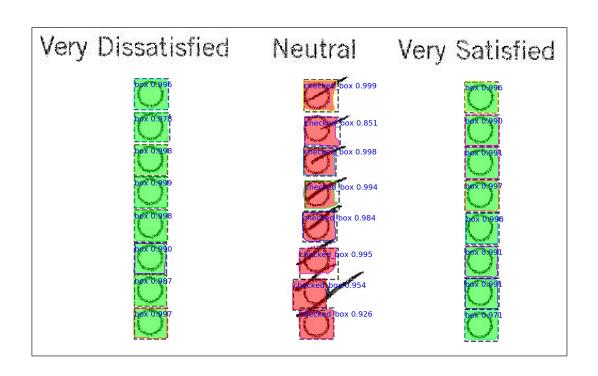
Results



• IoU: 0.50 - mAP: 0.94

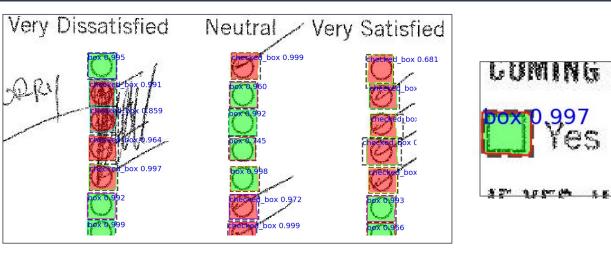


Results

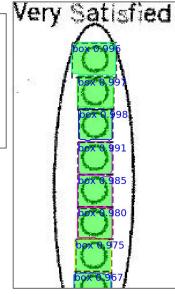


. DO YOU HAVE ANY INTENTION TO HIRE
ADDITIONAL EMPLOYEES IN THE
COMING YEAR?
Yes pt ked box 0.962
IF YES, HOW MANY?
a supramonant in
. THE EDC AND ITS PARTNERS PROVIDE
GROWTH, SUPPORT, AND OTHER
BUSINESS ASSISTANCE FREE OF
CHARGE TO LOCAL COMPANIES.
DO YOU HAVE AN INTEREST IN BEING
CONTACTED BY THE EDC?
998 No 0.998
. IS MOST OF YOUR REVENUE GENERATED
BY SALES OUTSIDE OF SARASOTA COUNTY?
997 Yes No.995

Results - How Can Data Be Bad?



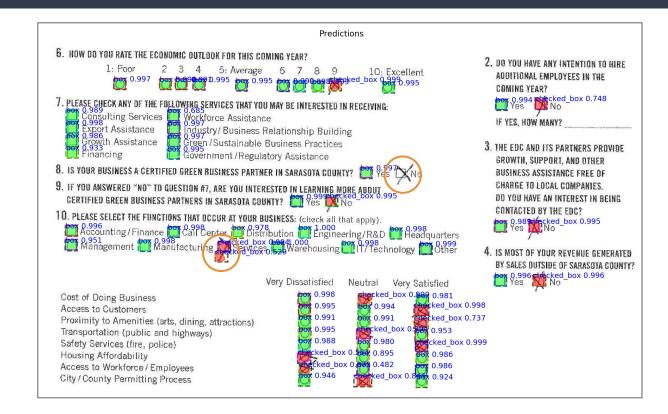




1. HOW MANY EMPLOYEES DO YOU HAVE AT YOUR SARASOTA COUNTY BUSINESS LOCATION?

What Happens When We Rearrange The Questions?

- Simulate the creation of a new survey
- Didn't do too bad, but still not perfect
- To improve results we would just need more training data (annotated images)



Model Output

• Predicted Class

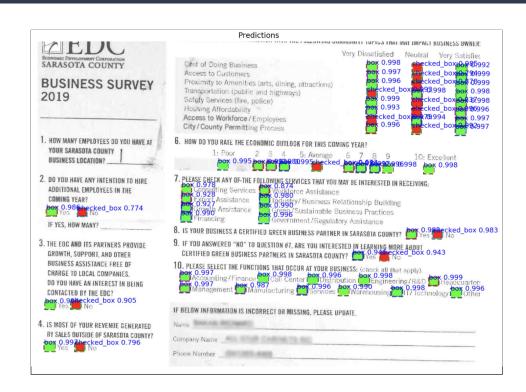
Probability

Bounding Boxes

```
[652, 1242, 682, 1277],
 [703, 1033, 734, 1069],
 [ 136, 667, 165, 703],
  [ 247, 107, 276, 142],
 [137, 759, 167, 794],
 [249, 377, 278, 411],
  [ 364, 886, 395, 921],
  [ 136, 710, 165, 746],
  [ 283, 377, 312, 411],
 [ 135, 558, 164, 591],
  [858, 726, 890, 762],
  [857, 882, 882, 912]
```

Extracting Data - Still In Progress

- Sort the bounding boxes from top to bottom, and maintain order of predicted classes
- Then take the first three bounding boxes and sort them from left to right
- Now we have an ordering, and can figure out which box the person checked
- Continue this procedure with the subsequent groups of boxes



Questions?