NATIONAL UNIVERSITY OF
SCIENCES AND TECHNOLOGY

Real-time Document Localization in Natural Images by Recursive Application of a CNN

By Khurram Javed, Faisal Shafait

School of Electrical Engineering and Computer Science (SEECS)
National University of Sciences and Technology (NUST), Islamabad

Introduction

- TUKL-NUST R&D Center, National University of Sciences and Technology, Islamabad

Introduction

- TUKL-NUST R&D Center, National University of Sciences and Technology, Islamabad

- Smart capture, structured form processing, scene text detection and recognition, object detection, classification, and tracking.

Problem Statement

Camera Image

Problem Statement

Camera Image

Detected Boundaries

Problem Statement

Camera Image

Detected Boundaries

Extracted Document

Industry Solutions

Office Lens

Google Drive

CS

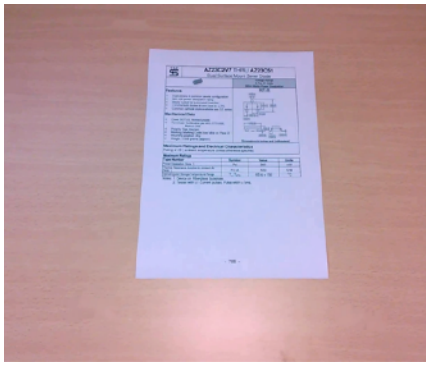
insiders
technologies

Academia Solutions

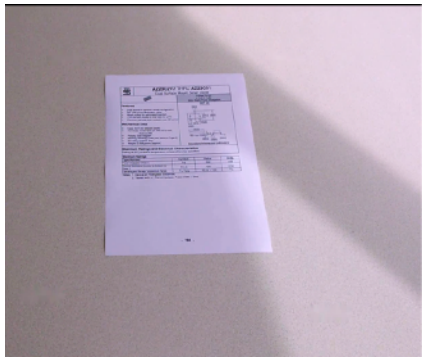
- ICDAR SmartDoc 2015 Challenge 1 : Smartphone document capture competition.
  - 5 backgrounds
  - 8 submitted methods

Literature Review

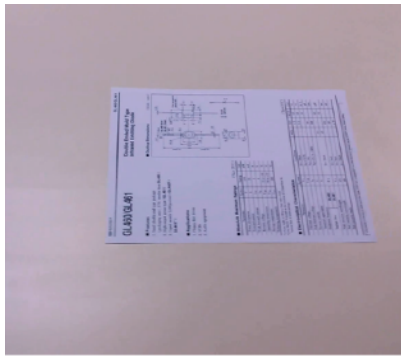
- ICDAR SmartDoc 2015 Challenge 1 : Smartphone document capture competition.
  - 5 backgrounds
  - 8 submitted methods



1



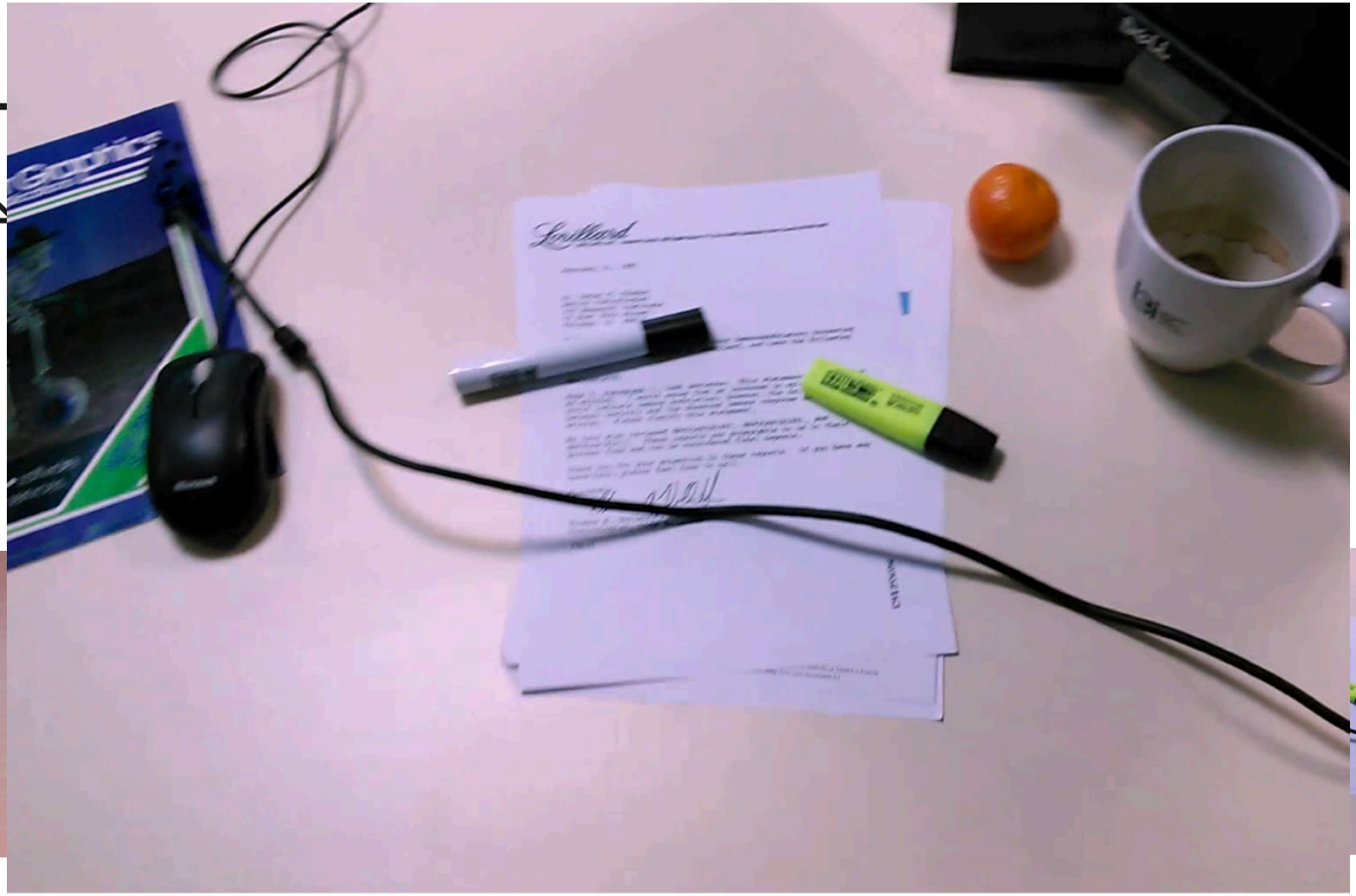
2



3

5

L



The complex background!

Competition Results

Method	Background 1	Background 2	Background 3	Background 4	Background 5
A2iA-1 [1]	0.9724	0.8006	0.9117	0.6352	
A2iA-2 [1]	0.9597	0.8063	0.9118	0.8264	
ISPL-CVML [1]	**0.9870**	0.9652	0.9846	0.9766	
LRDE [1]	0.9869	0.9775	0.9889	**0.9837**	
NetEase [1]	0.9624	0.9552	0.9621	0.9511	
SEECS-NUST [1]	0.8875	0.8264	0.7832	0.7811	
RPPDI-UPE [1]	0.8274	0.9104	0.9697	0.3649	
SmartEngines [1]	0.9885	**0.9833**	**0.9897**	0.9785	
L. R. S. Leal, et al [2]	0.9605	0.9444	0.9647	0.9300	

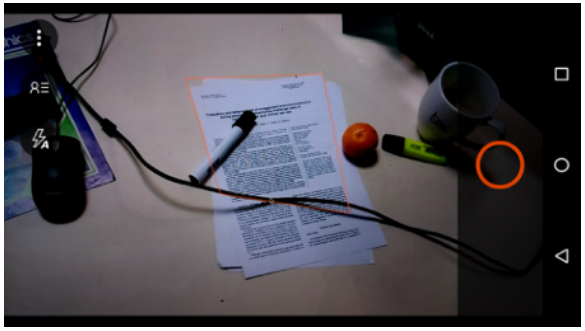
[1] JC. Burie, J. Chazalon, et al. ”ICDAR2015 competition on smartphone document capture and OCR (SmartDoc).”13th International Conference on Document Analysis and Recognition, IEEE, 2015.
[2] LRS Leal, BLD Bezerra. ”Smartphone camera document detection via Geodesic Object Proposals.” Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on. IEEE, 2016.

Competition Results

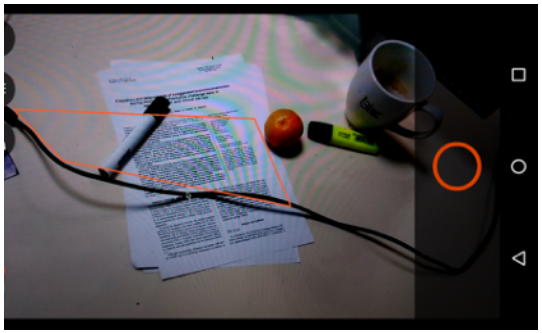
Method	Background 1	Background 2	Background 3	Background 4	Background 5
A2iA-1 [1]	0.9724	0.8006	0.9117	0.6352	0.1890
A2iA-2 [1]	0.9597	0.8063	0.9118	0.8264	0.1892
ISPL-CVML [1]	**0.9870**	0.9652	0.9846	0.9766	0.8555
LRDE [1]	0.9869	0.9775	0.9889	**0.9837**	**0.8613**
NetEase [1]	0.9624	0.9552	0.9621	0.9511	0.2218
SEECS-NUST [1]	0.8875	0.8264	0.7832	0.7811	0.0113
RPPDI-UPE [1]	0.8274	0.9104	0.9697	0.3649	0.2163
SmartEngines [1]	0.9885	**0.9833**	**0.9897**	0.9785	0.6884
L. R. S. Leal, et al [2]	0.9605	0.9444	0.9647	0.9300	0.4117

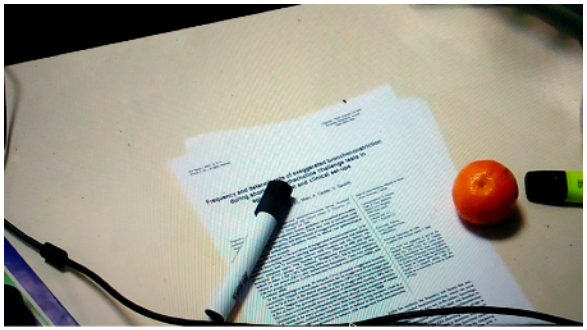
[1] JC. Burie, J. Chazalon, et al. ”ICDAR2015 competition on smartphone document capture and OCR (SmartDoc).”13th International Conference on Document Analysis and Recognition, IEEE, 2015.
[2] LRS Leal, BLD Bezerra. ”Smartphone camera document detection via Geodesic Object Proposals.” Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on. IEEE, 2016.

Industry Systems Results



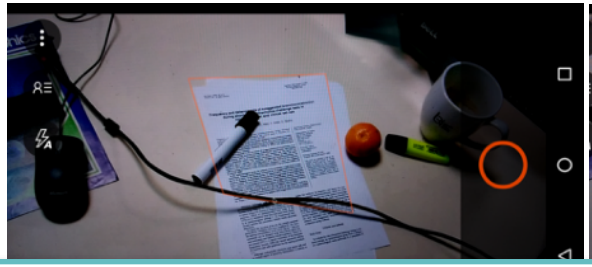
Microsoft Office Lens

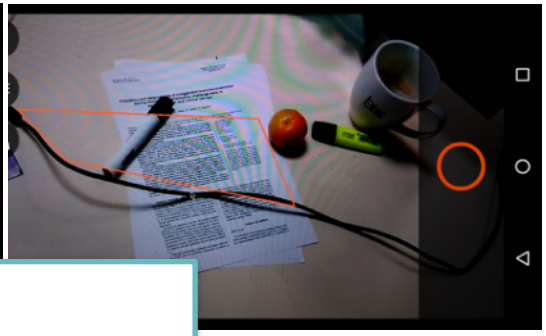


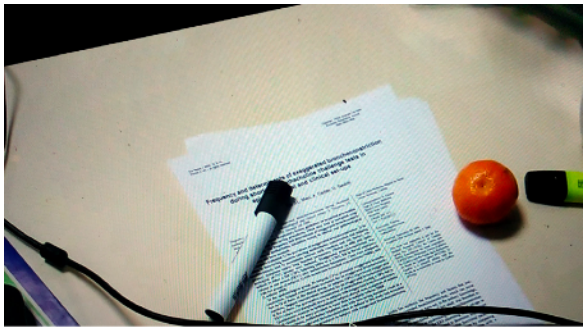


Google Drive

Industry Systems Results







Deep learning?

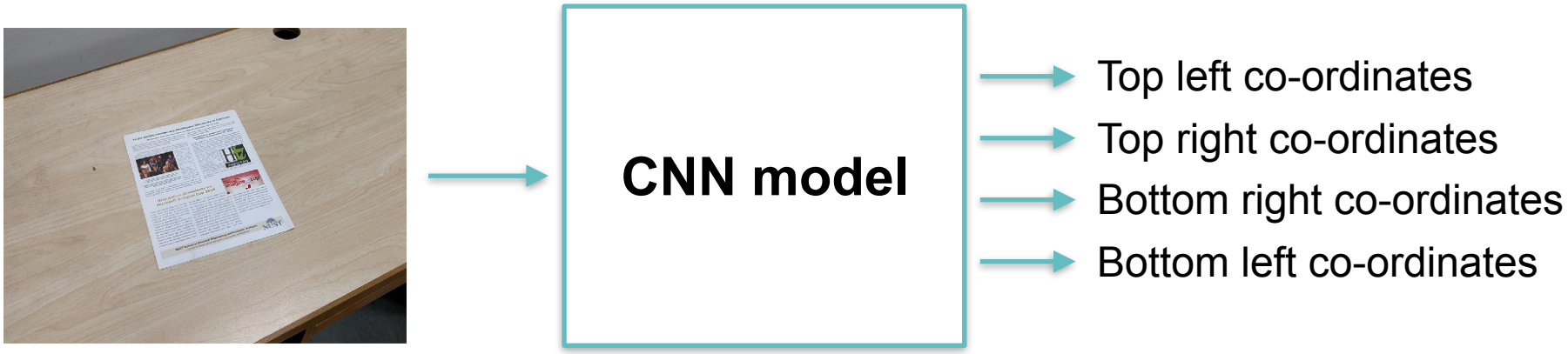
Google Drive

Approach

↘ **IDEA 1.0** : Regress co-ordinates of the document.

Approach

↘ **IDEA 1.0** : Regress co-ordinates of the document.



Idea 1.0

Expectation!

o-ordinates

co-ordinates

ght co-ordinates

eft co-ordinates

Idea 1.0

Prediction!

o-ordinates
co-ordinates
ght co-ordinates
eft co-ordinates

Why it doesn’t work!

Why it doesn't work!

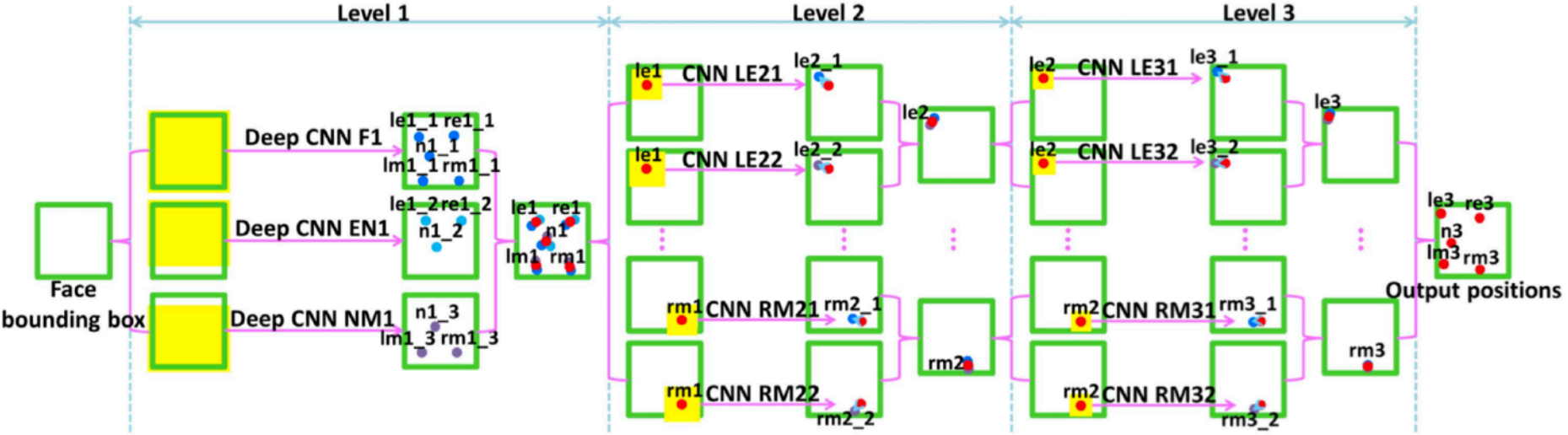
- Only relying on high level features.
  - Known problem for key-point regression [1].

[1] Y. Sun, X. Wang, et al. "Deep convolutional network cascade for facial point detection." Conference on Computer Vision and Pattern Recognition, 2013.

Why it doesn't work!

↘ Only relying on high level features.

↘ Known problem for key point regression [1]



Cascaded CNNs

[1] Y. Sun, X. Wang, et al. "Deep convolutional network cascade for facial point detection." Conference on Computer Vision and Pattern Recognition, 2013.

Why it doesn't work!

- Only relying on high level features.
  - Known problem for key-point regression [1].

[1] Y. Sun, X. Wang, et al. "Deep convolutional network cascade for facial point detection." Conference on Computer Vision and Pattern Recognition, 2013.

Why it doesn't work!

↘ Only relying on high level features.

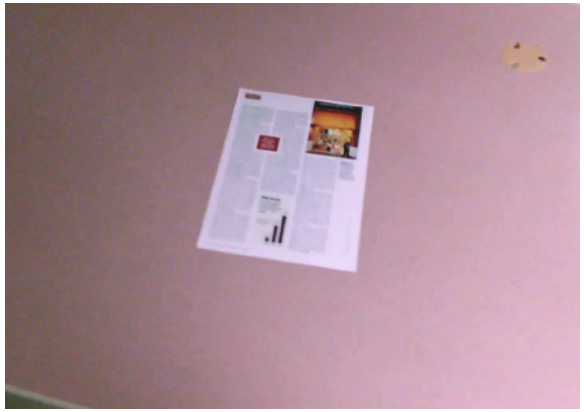
↘ Known problem for key-point regression [1].

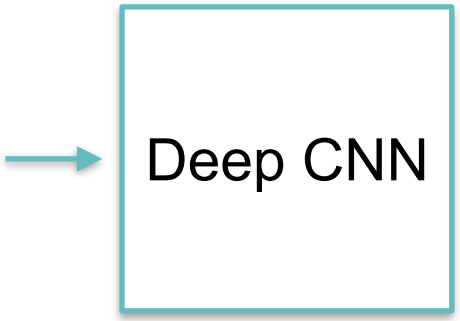
↘ We don't use the full resolution image.

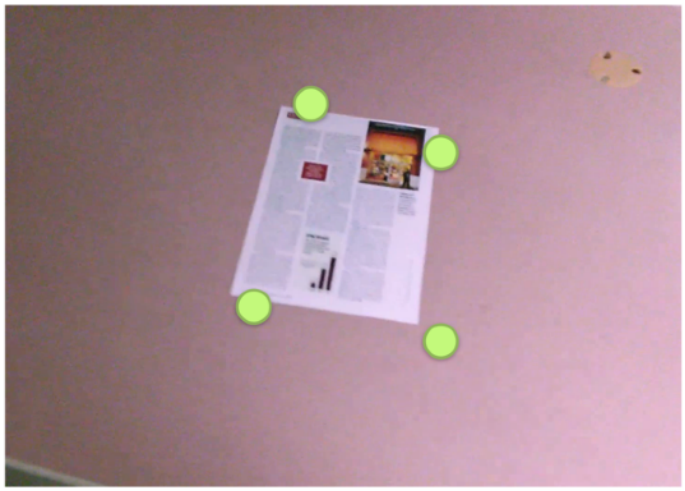
[1] Y. Sun, X. Wang, et al. "Deep convolutional network cascade for facial point detection." Conference on Computer Vision and Pattern Recognition, 2013.

Idea 2.0

Idea 2.0



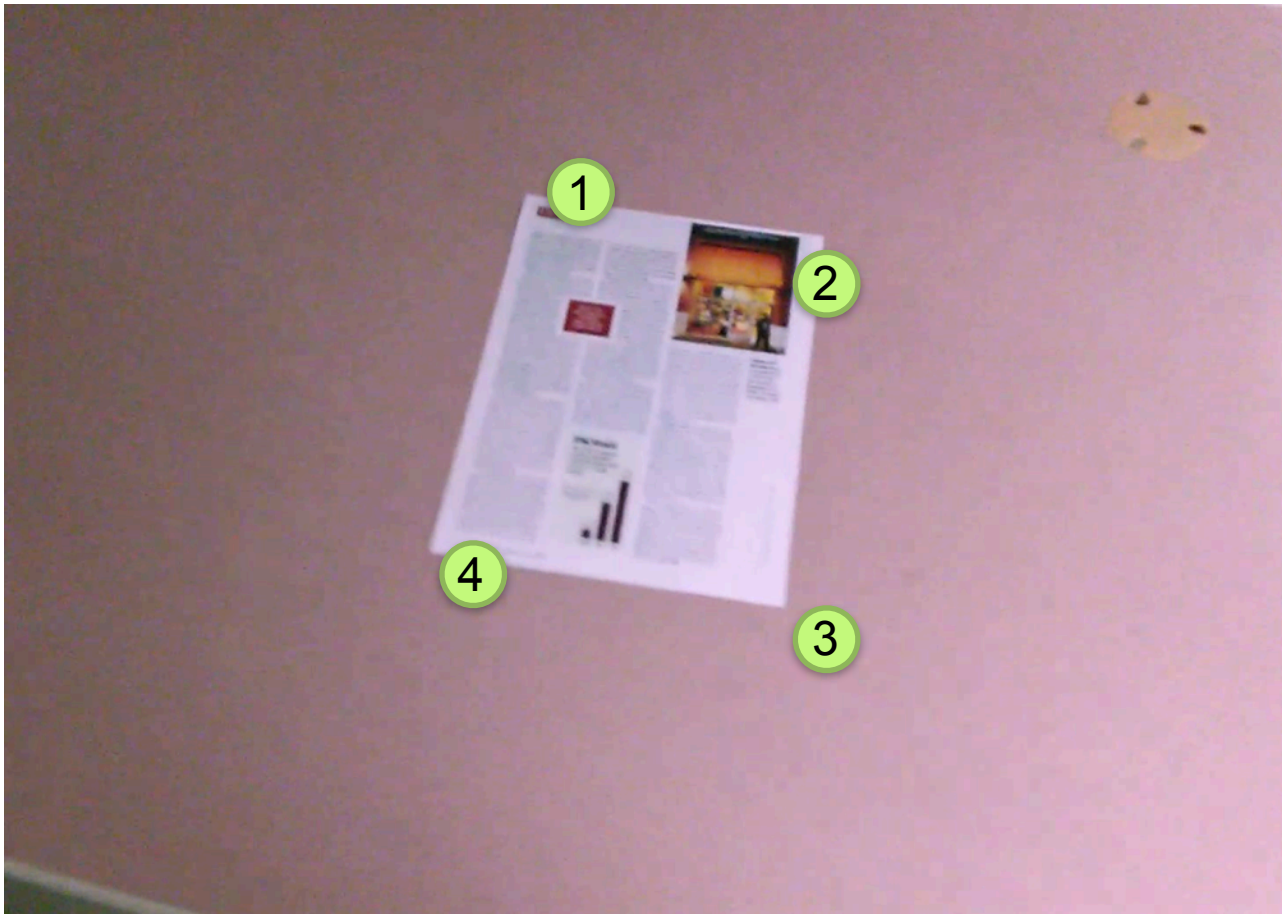




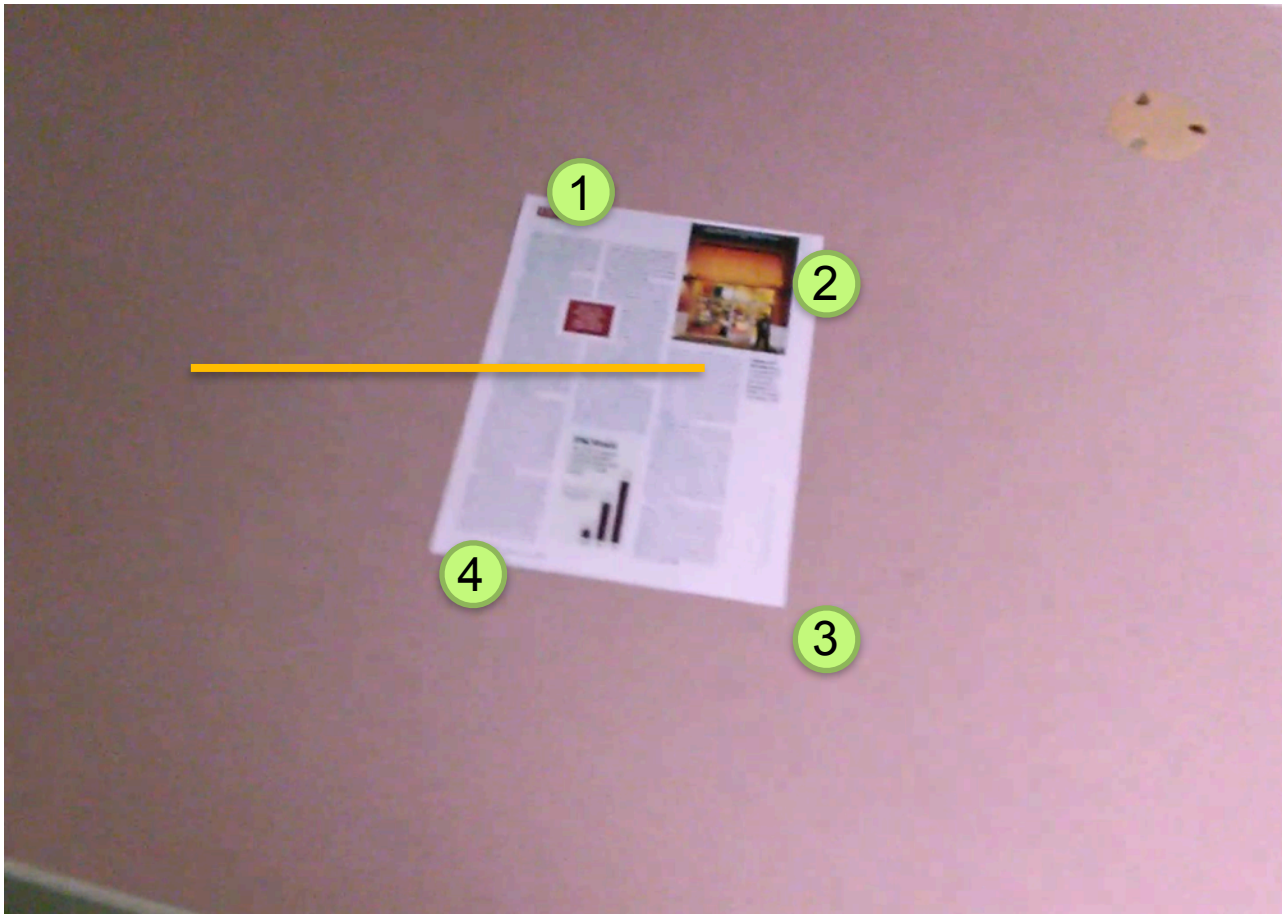
Input Image

Results visualized

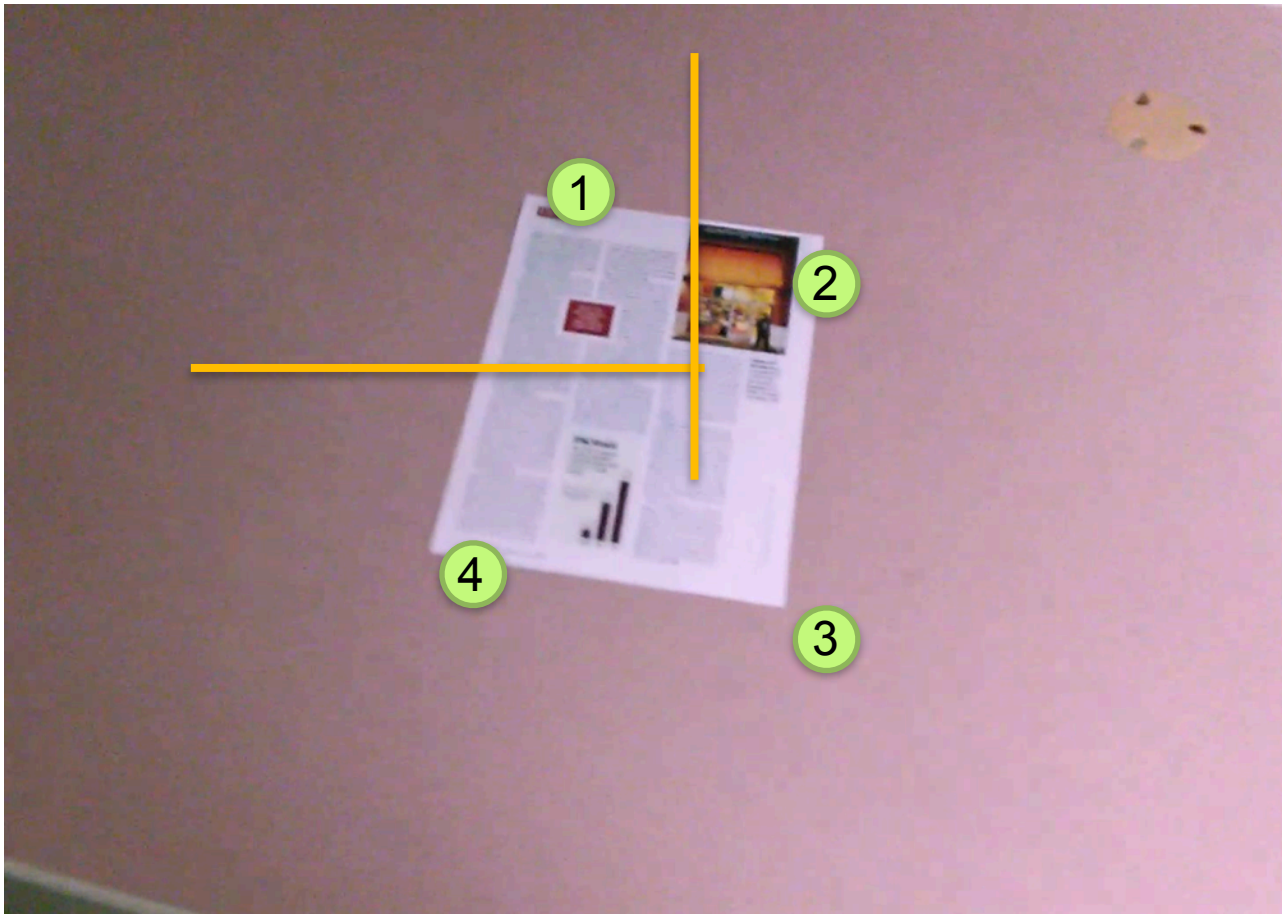
Idea 2.0



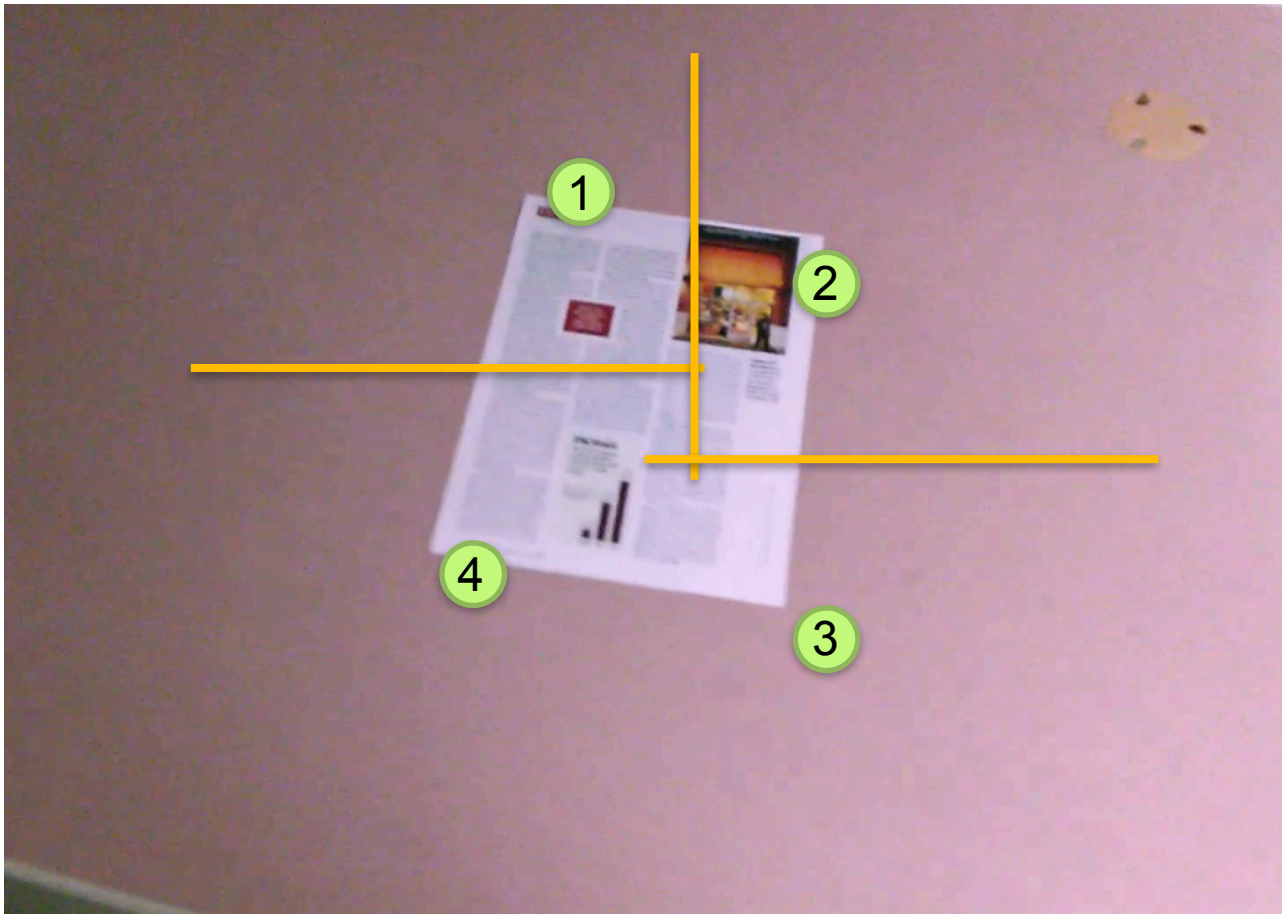
Idea 2.0



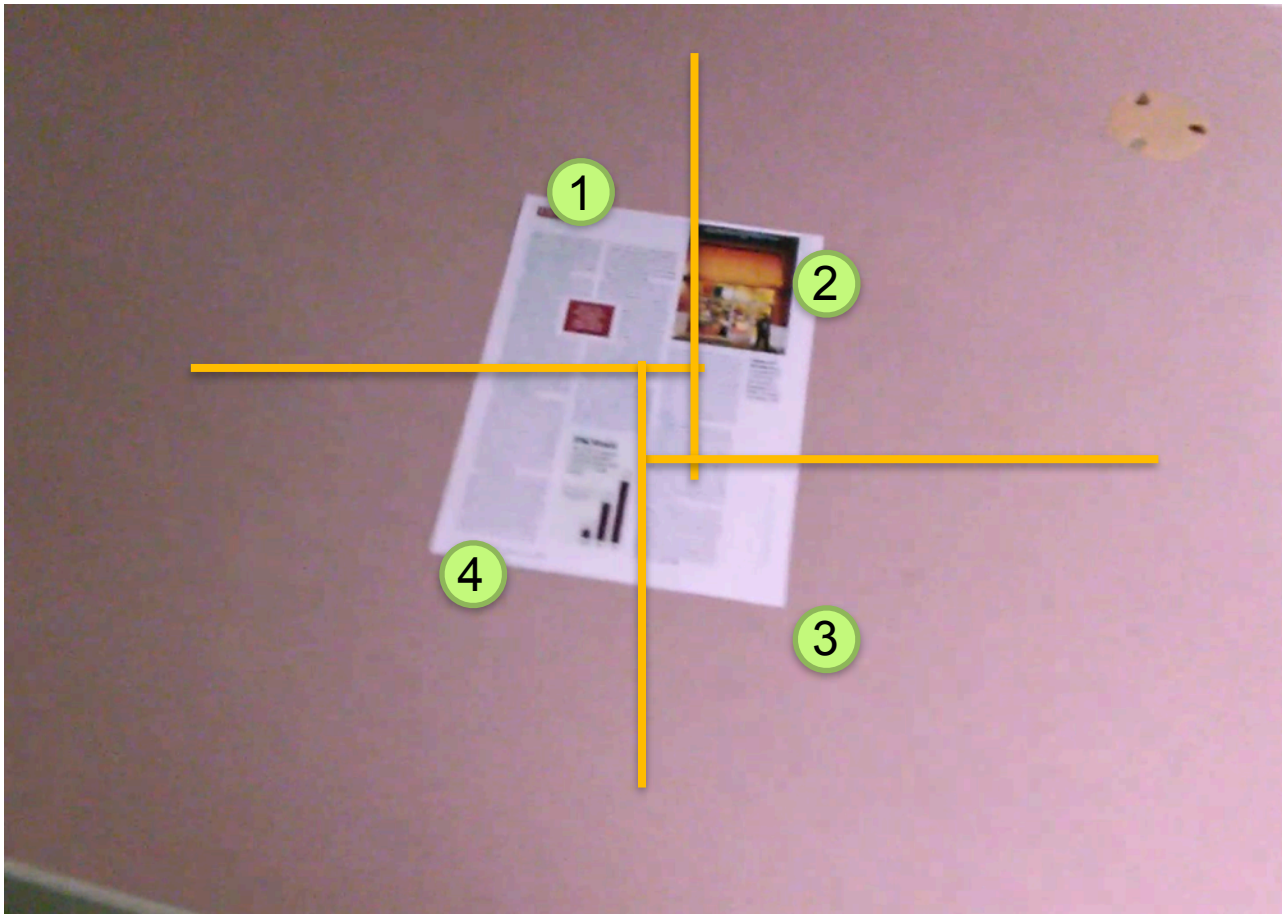
Idea 2.0



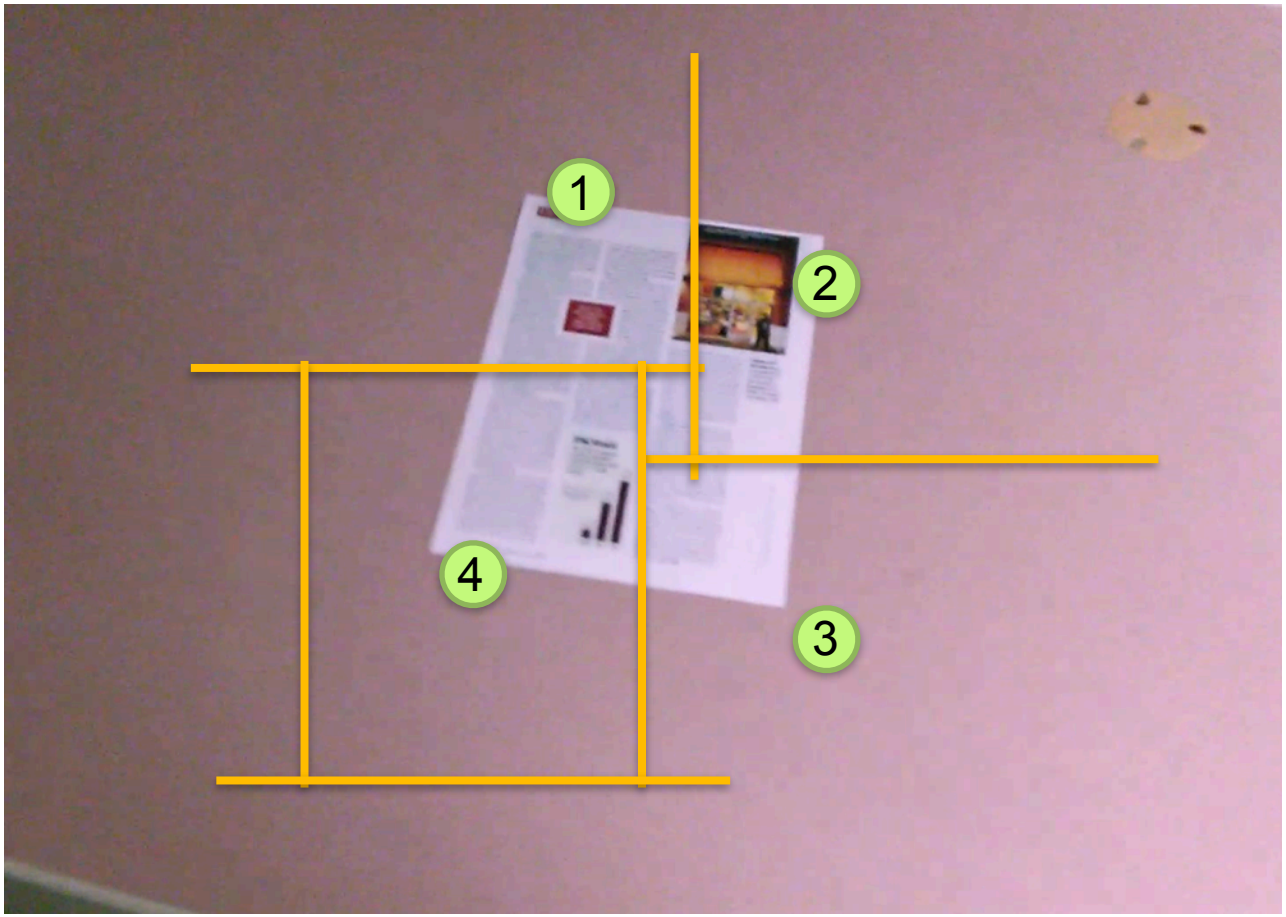
Idea 2.0



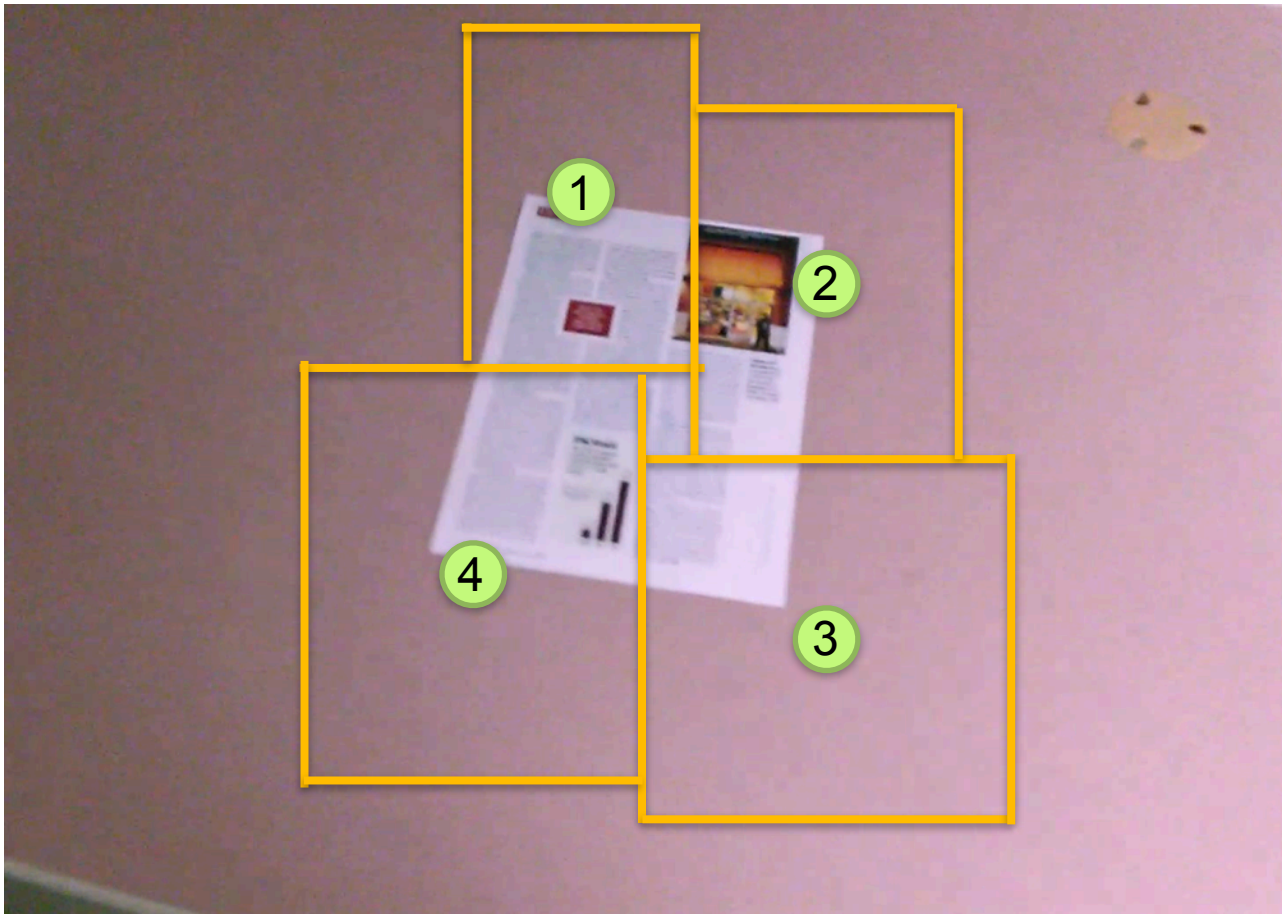
Idea 2.0



Idea 2.0

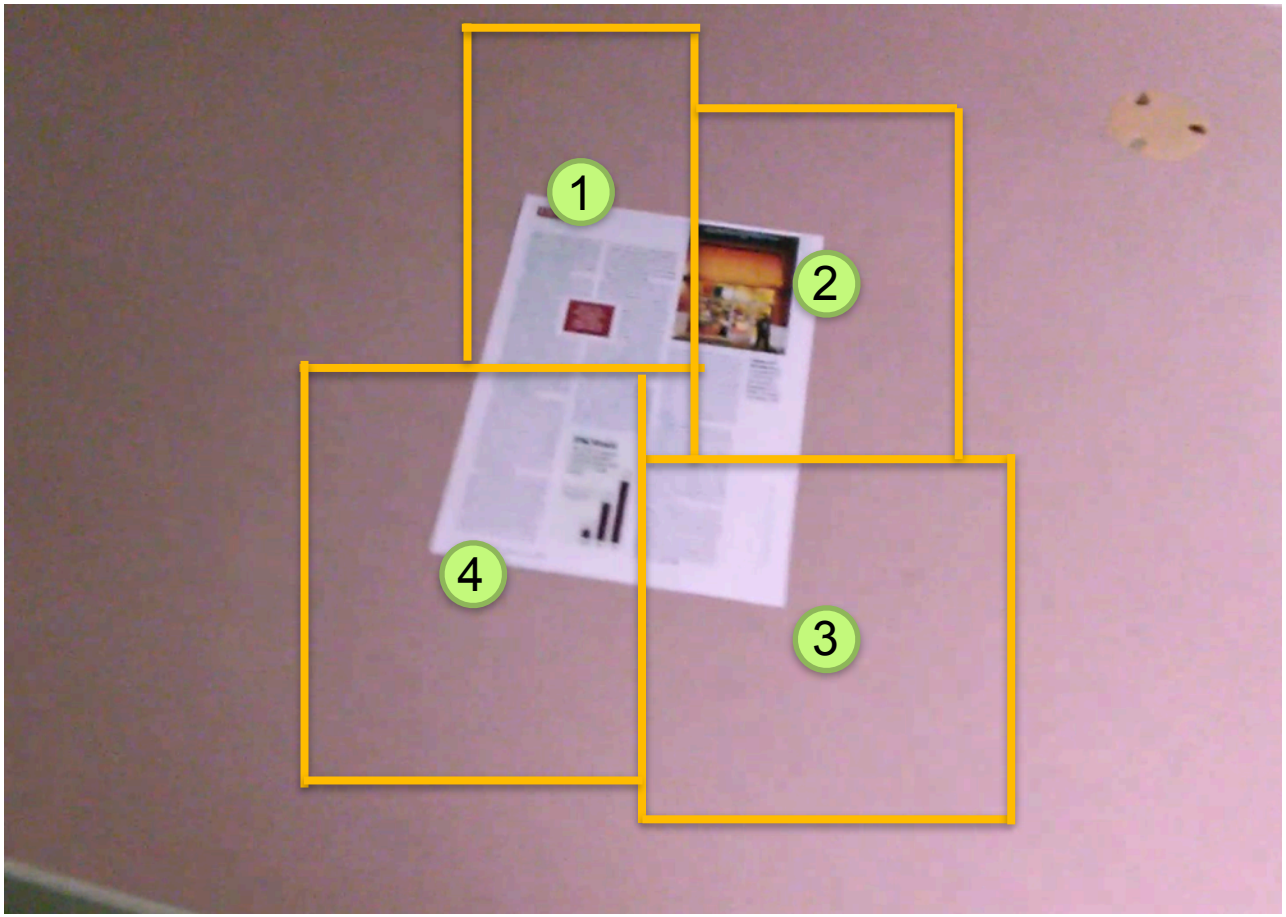


Idea 2.0

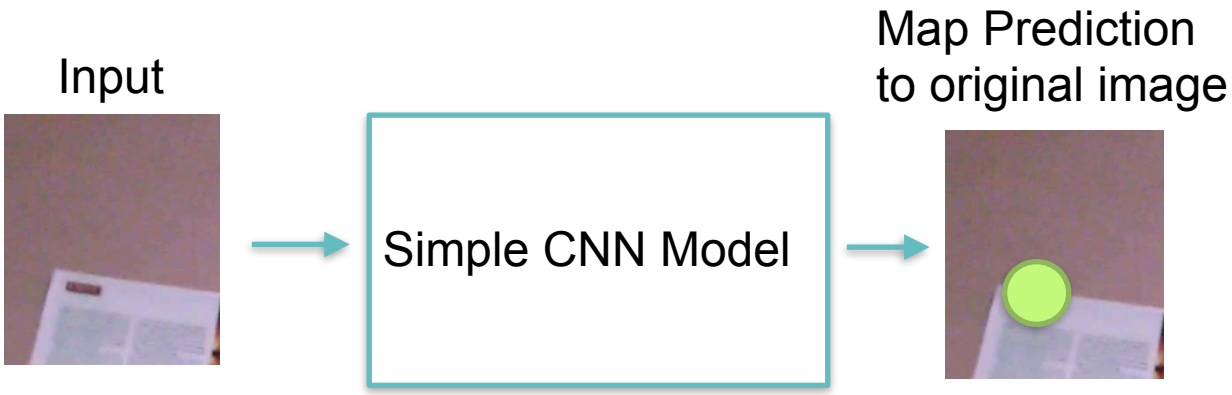


Idea 2.0

Regions normalized by document Size

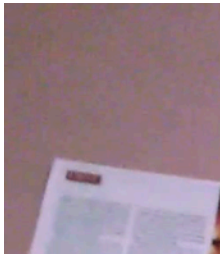


Idea 2.0 : Recursive Refinement



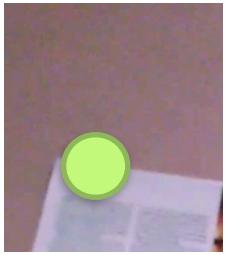
Idea 2.0

Input

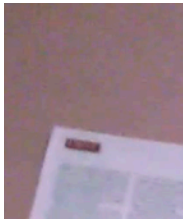


Simple CNN Model

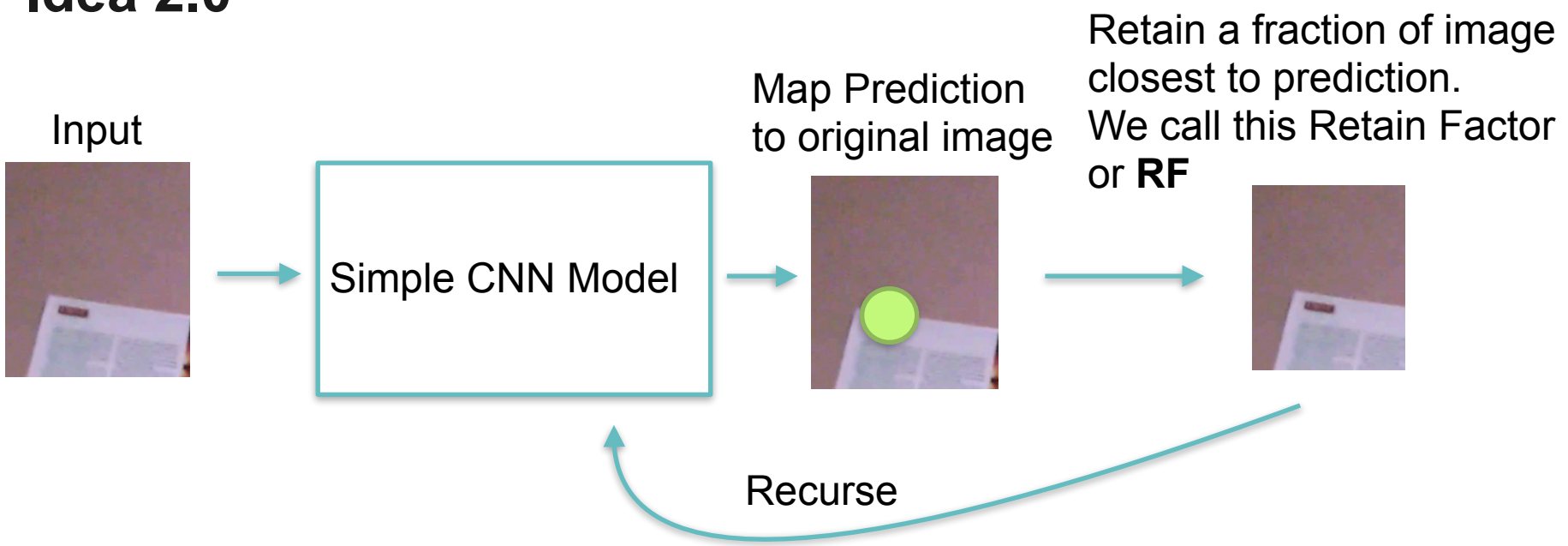
Map Prediction
to original image



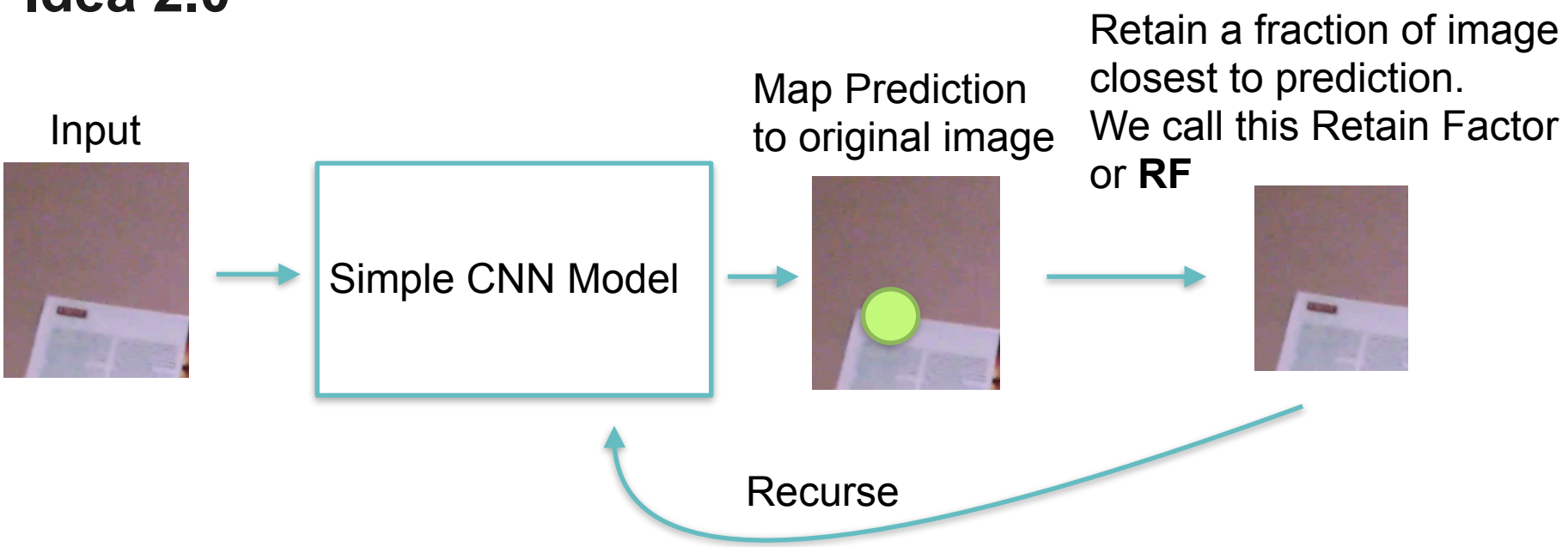
Retain a part of the image
closest to prediction by
a factor called Retain Factor
or **RF**



Idea 2.0


Input
Simple CNN Model
Map Prediction
to original image
Retain a fraction of image
closest to prediction.
We call this Retain Factor
or **RF**
Recurse

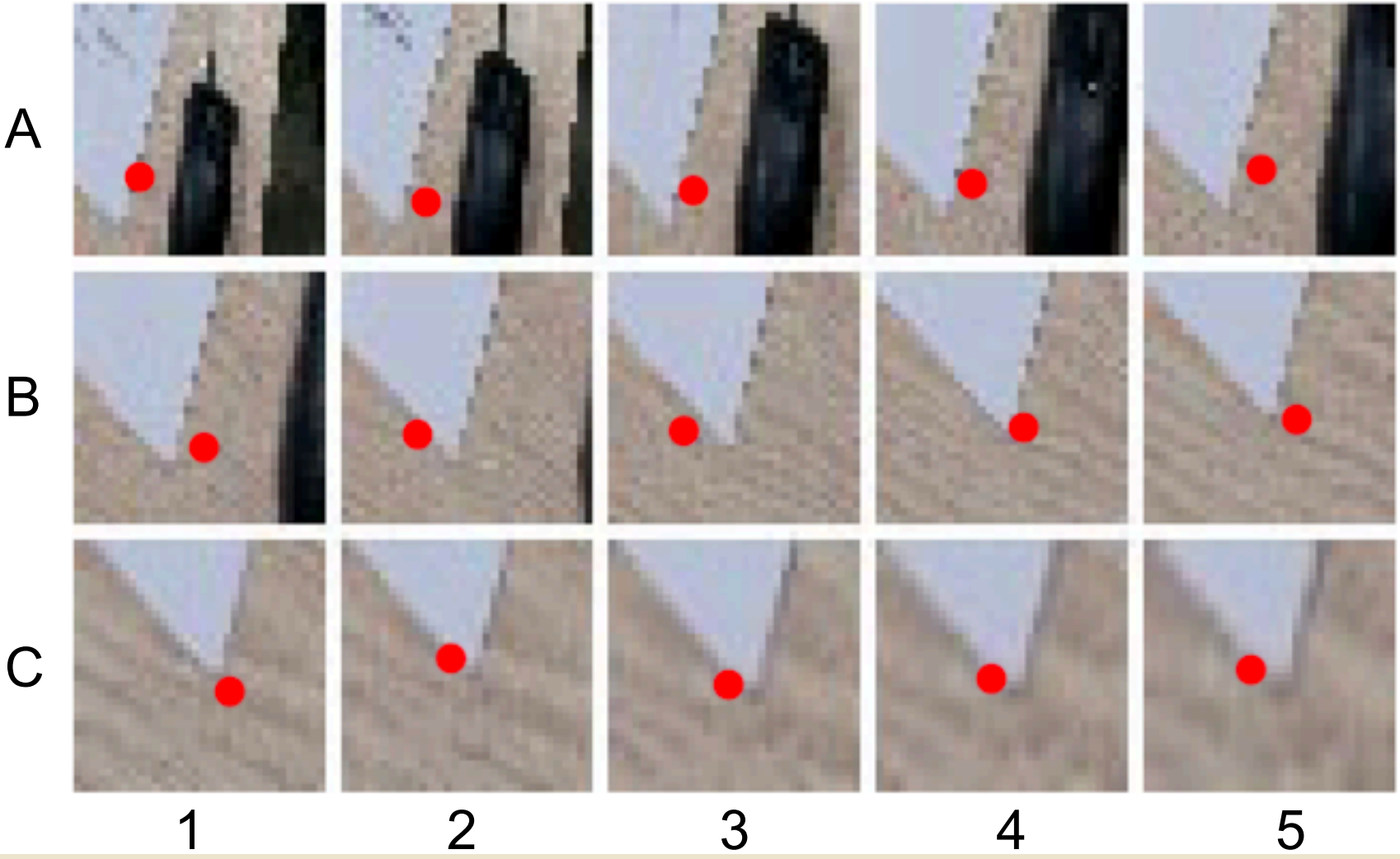
Idea 2.0



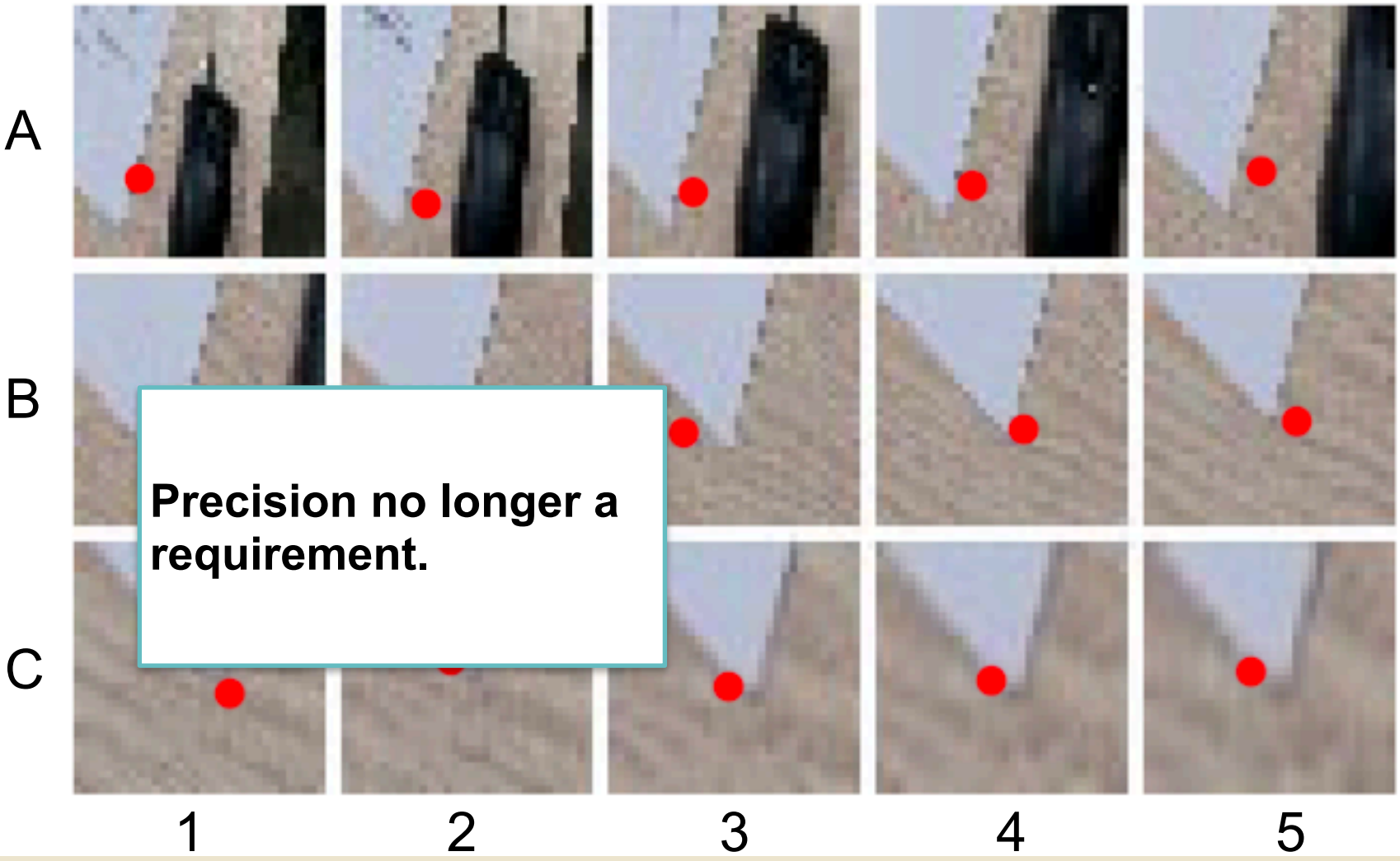
Stopping Criteria

$$(H \times RF^n, W \times RF^n) < (10 \times 10)$$

Recursive Refinement

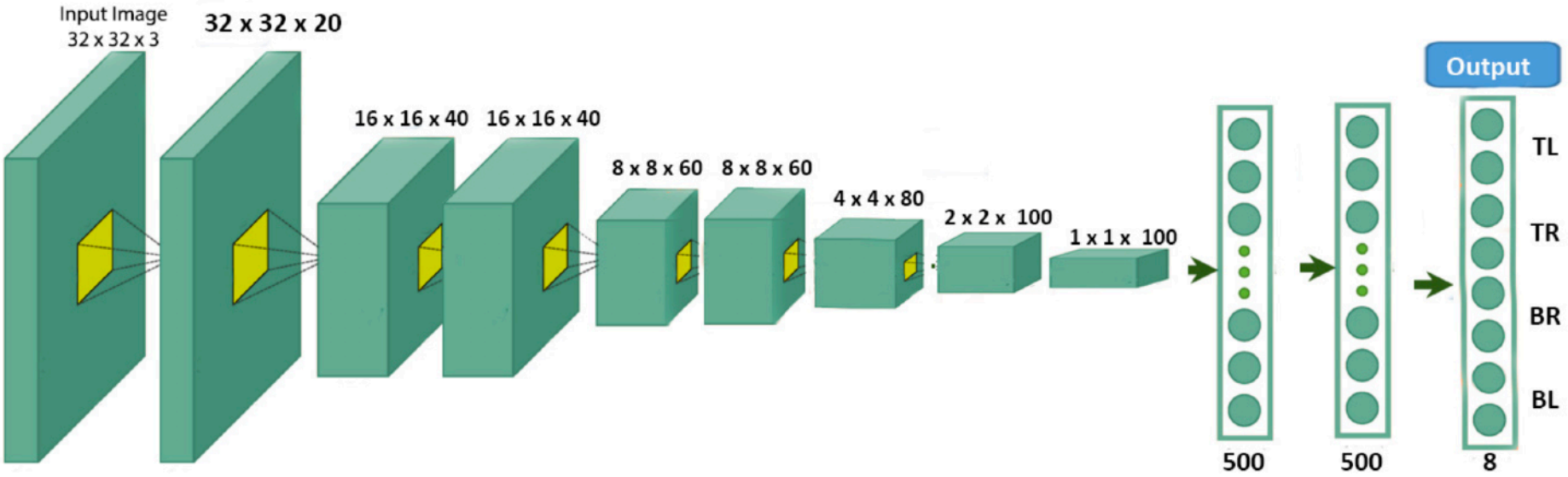


Recursive Refinement



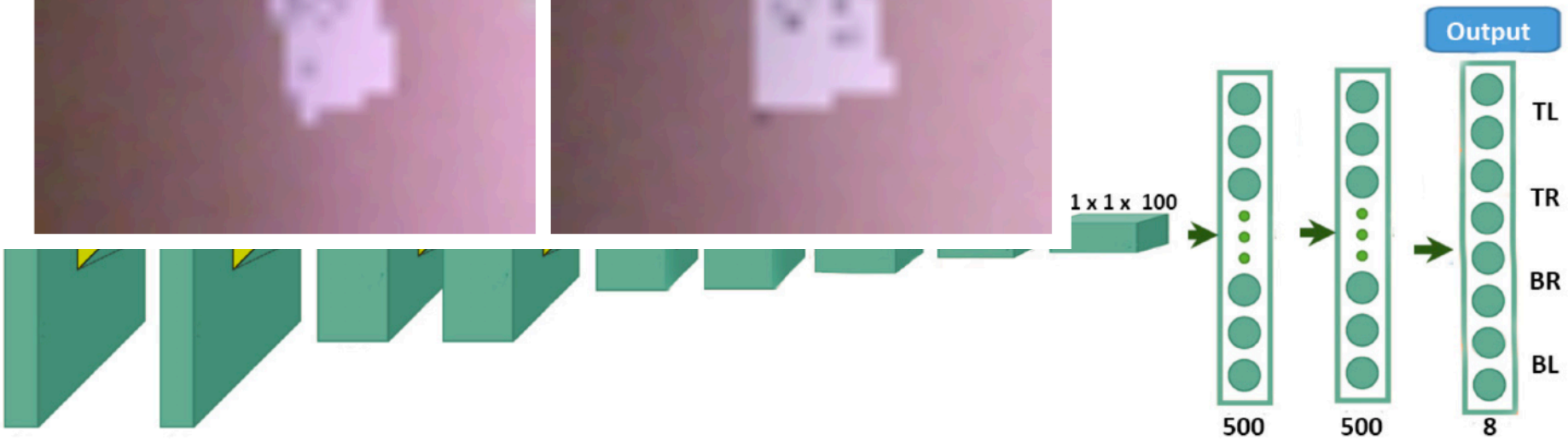
Model Details

↘ Architecture similar to Alex-Net for 4 points regression.



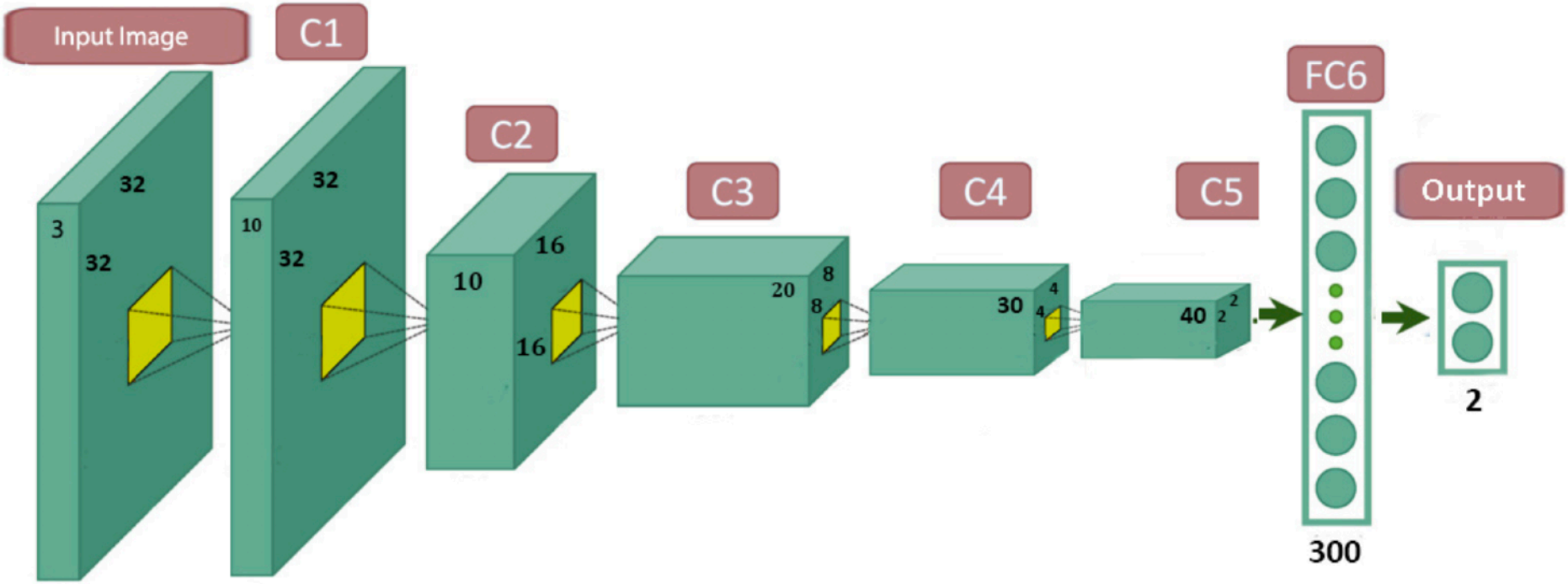
Information in 32 x 32 Image

nts regression.



Model Details

↘ Shallow, 5 layer network for recursive refinement.



Performance Analysis

- Intel i5-4200U 1.6 Ghz CPU 8 GB ram.
- In-efficient implementation.
- 1920 x 1080 images.

Retain Factor	Time in ms
0.85	320
0.75	210
0.65	150
0.60	130
0.50	100

Run-time complexity : $\sqrt{N}$

Results

Method	Background 1	Background 2	Background 3	Background 4	Background 5
A2iA-1 [1]	0.9724	0.8006	0.9117	0.6352	0.1890
A2iA-2 [1]	0.9597	0.8063	0.9118	0.8264	0.1892
ISPL-CVML [1]	**0.9870**	0.9652	0.9846	0.9766	0.8555
LRDE [1]	0.9869	0.9775	0.9889	**0.9837**	**0.8613**
NetEase [1]	0.9624	0.9552	0.9621	0.9511	0.2218
SEECS-NUST [1]	0.8875	0.8264	0.7832	0.7811	0.0113
RPPDI-UPE [1]	0.8274	0.9104	0.9697	0.3649	0.2163
SmartEngines [1]	0.9885	**0.9833**	**0.9897**	0.9785	0.6884
L. R. S. Leal, et al [2]	0.9605	0.9444	0.9647	0.9300	0.4117
SEECS-NUST-2					

[1] JC. Burie, J. Chazalon, et al. ”ICDAR2015 competition on smartphone document capture and OCR (SmartDoc).”13th International Conference on Document Analysis and Recognition, IEEE, 2015.
[2] LRS Leal, BLD Bezerra. ”Smartphone camera document detection via Geodesic Object Proposals.” Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on. IEEE, 2016.

Experiments done with RF = 0.85

Results

Method	Background 1	Background 2	Background 3	Background 4	Background 5
A2iA-1 [1]	0.9724	0.8006	0.9117	0.6352	0.1890
A2iA-2 [1]	0.9597	0.8063	0.9118	0.8264	0.1892
ISPL-CVML [1]	**0.9870**	0.9652	0.9846	0.9766	0.8555
LRDE [1]	0.9869	0.9775	0.9889	**0.9837**	**0.8613**
NetEase [1]	0.9624	0.9552	0.9621	0.9511	0.2218
SEECS-NUST [1]	0.8875	0.8264	0.7832	0.7811	0.0113
RPPDI-UPE [1]	0.8274	0.9104	0.9697	0.3649	0.2163
SmartEngines [1]	0.9885	**0.9833**	**0.9897**	0.9785	0.6884
L. R. S. Leal, et al [2]	0.9605	0.9444	0.9647	0.9300	0.4117
SEECS-NUST-2	**0.9832**	**0.9724**	**0.9830**	**0.9695**	

[1] JC. Burie, J. Chazalon, et al. ”ICDAR2015 competition on smartphone document capture and OCR (SmartDoc).”13th International Conference on Document Analysis and Recognition, IEEE, 2015.
[2] LRS Leal, BLD Bezerra. ”Smartphone camera document detection via Geodesic Object Proposals.” Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on. IEEE, 2016.

Experiments done with RF = 0.85

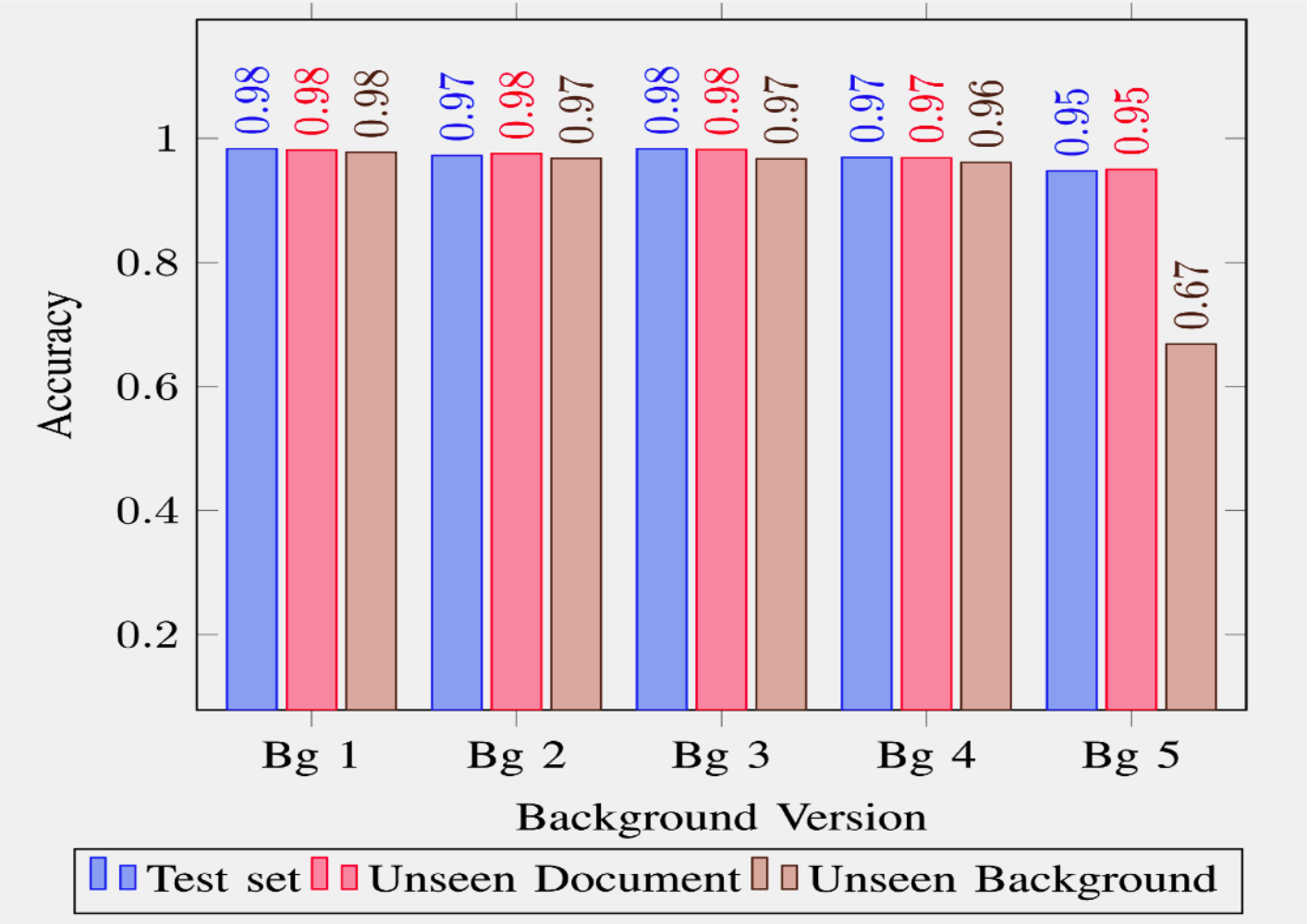
Results

Method	Background 1	Background 2	Background 3	Background 4	Background 5
A2iA-1 [1]	0.9724	0.8006	0.9117	0.6352	0.1890
A2iA-2 [1]	0.9597	0.8063	0.9118	0.8264	0.1892
ISPL-CVML [1]	**0.9870**	0.9652	0.9846	0.9766	0.8555
LRDE [1]	0.9869	0.9775	0.9889	**0.9837**	**0.8613**
NetEase [1]	0.9624	0.9552	0.9621	0.9511	0.2218
SEECS-NUST [1]	0.8875	0.8264	0.7832	0.7811	0.0113
RPPDI-UPE [1]	0.8274	0.9104	0.9697	0.3649	0.2163
SmartEngines [1]	0.9885	**0.9833**	**0.9897**	0.9785	0.6884
L. R. S. Leal, et al [2]	0.9605	0.9444	0.9647	0.9300	0.4117
SEECS-NUST-2	**0.9832**	**0.9724**	**0.9830**	**0.9695**	**0.9478**

[1] JC. Burie, J. Chazalon, et al. ”ICDAR2015 competition on smartphone document capture and OCR (SmartDoc).”13th International Conference on Document Analysis and Recognition, IEEE, 2015.
[2] LRS Leal, BLD Bezerra. ”Smartphone camera document detection via Geodesic Object Proposals.” Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on. IEEE, 2016.

Experiments done with RF = 0.85

Generalization Results



Performance Analysis

- Intel i5-4200U 1.6 Ghz CPU 8 GB ram.
- In-efficient implementation
- Run-time complexity where N is no of pixels: $\sqrt{N}$

Retain Factor	Overall Accuracy	Time in ms
0.85	0.9743	320
0.75	0.9701	210
0.65	0.9617	150
0.60	0.9604	130
0.50	0.9513	100

Experiments done with 1920 x 1080 images

Future Directions

↘ Finding the ideal model configuration.

↘ Training on better representative data.

↘ Using a single end-to-end model.

↘ Code : **https://github.com/Khurramjaved96/Recursive-CNNs**

Last note!

↘ **Special thanks to ICDAR for Student's Travel Award!**