Signreco

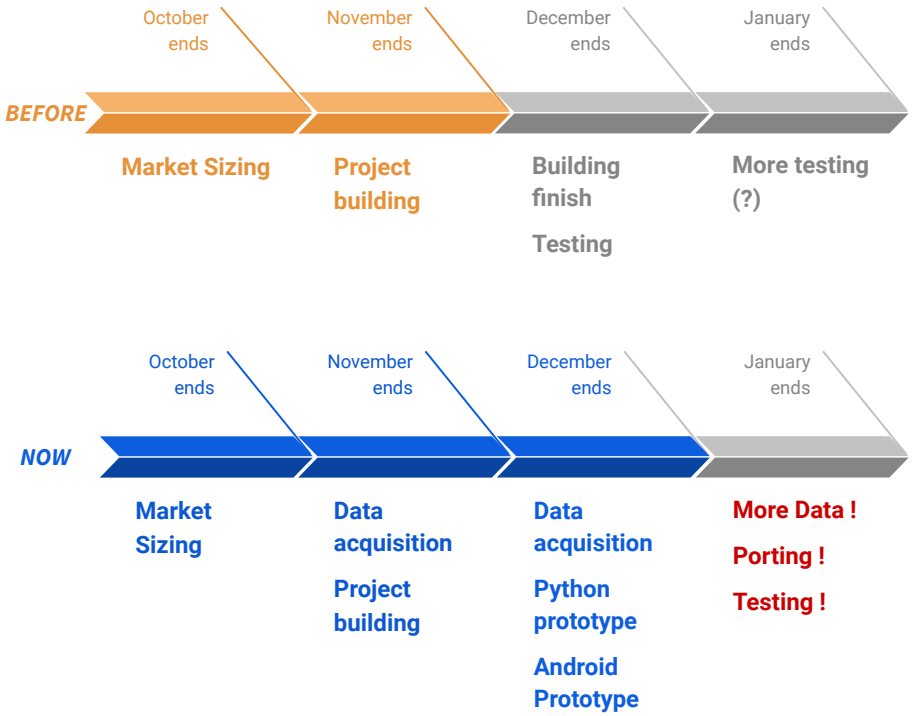
Raquel Pérez
raquel.leandra.perez@est.fib.upc.edu

Jorge Sierra
jorge.sierra@est.fib.upc.edu

Gaspard Debussche
gaspard.debussche@est.fib.upc.edu

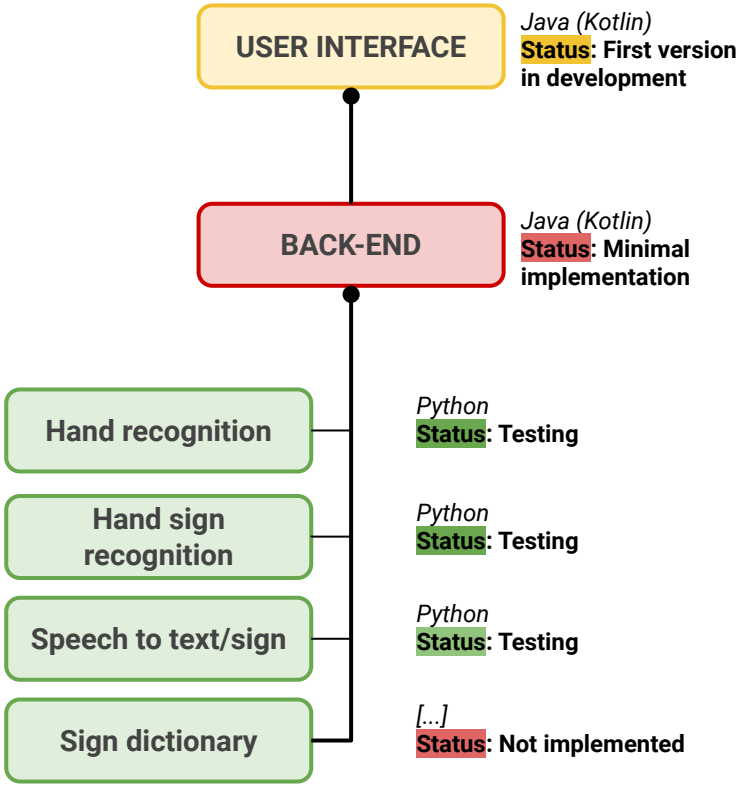
Schedule
Update

Signreco



App Modules

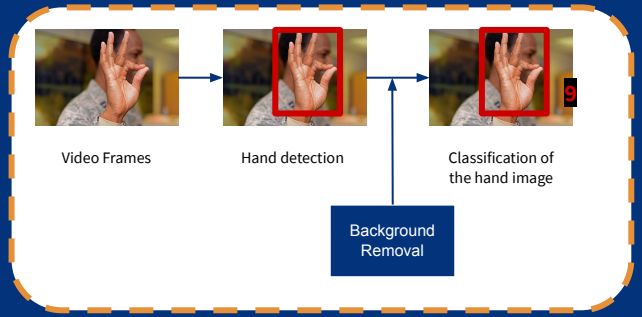
Building & Proto

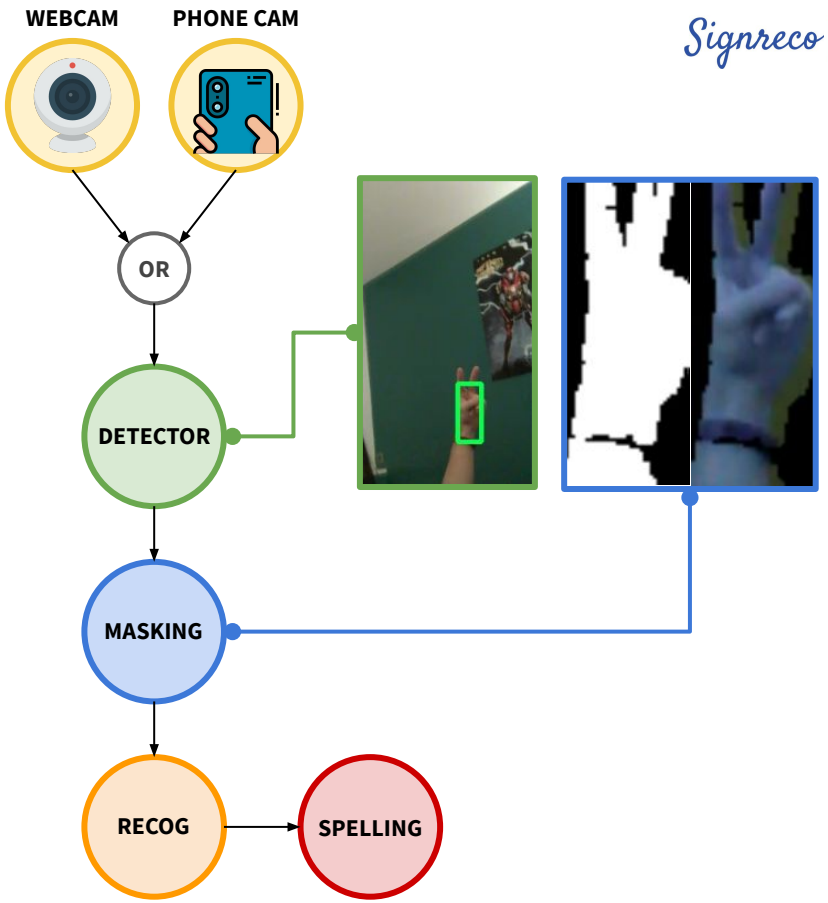


Architecture

Pipeline

Old architecture



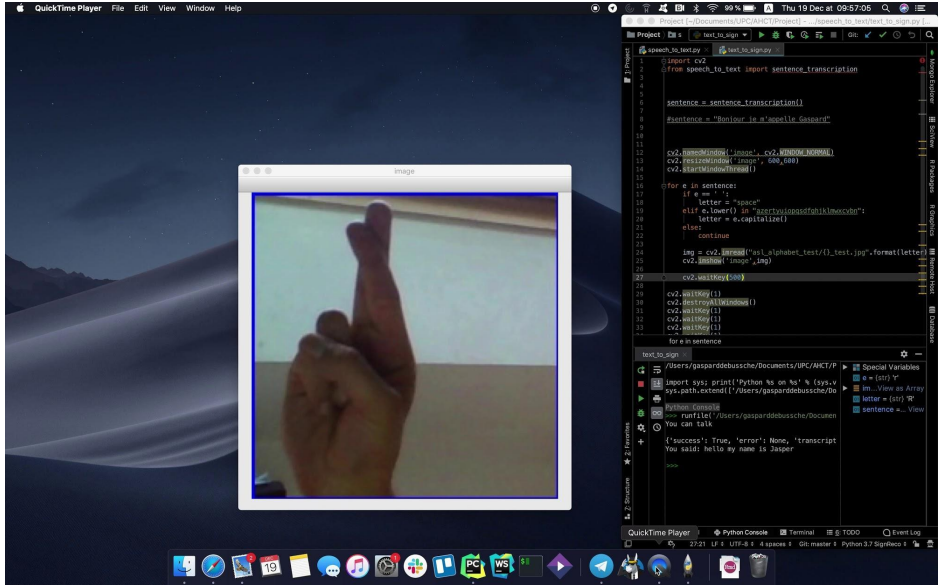


output: “2”

Speech to Sign

Testing

- Language Python
- Libraries
  - SpeechRecognition
  - Opencv-python



OpenCV

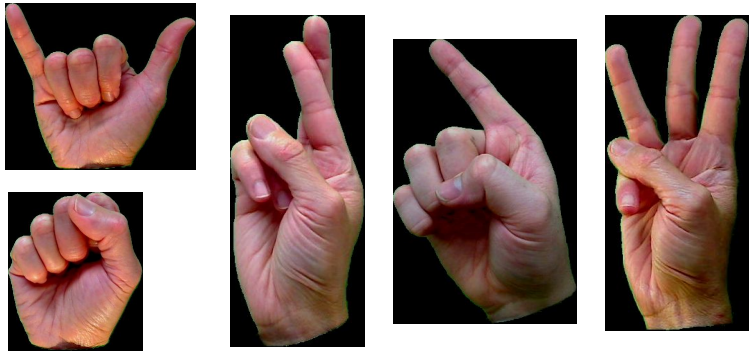
Sign Recognition

Dataset

Signreco

	Train	Val	Test	Total
%	82	10	8	100
#	2,123	258	208	2,589

- 36 clases
- Different sizes
- Different hand sizes
- Different illuminations
- Some sign language letters are very similar



Sign Recognition

CNN Architecture

- Medium size
- Not pre-trained
- Dropout of 0.2 to avoid overfitting
- Inspired on [2]

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 254, 254, 64)	1792
max_pooling2d_1 (MaxPooling2	(None, 127, 127, 64)	0
conv2d_2 (Conv2D)	(None, 125, 125, 16)	9232
max_pooling2d_2 (MaxPooling2	(None, 62, 62, 16)	0
conv2d_3 (Conv2D)	(None, 60, 60, 16)	2320
max_pooling2d_3 (MaxPooling2	(None, 30, 30, 16)	0
conv2d_4 (Conv2D)	(None, 28, 28, 8)	1160
max_pooling2d_4 (MaxPooling2	(None, 14, 14, 8)	0
flatten_1 (Flatten)	(None, 1568)	0
dropout_1 (Dropout)	(None, 1568)	0
dense_1 (Dense)	(None, 256)	401664
dropout_2 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 54)	13878
dense_3 (Dense)	(None, 36)	1980

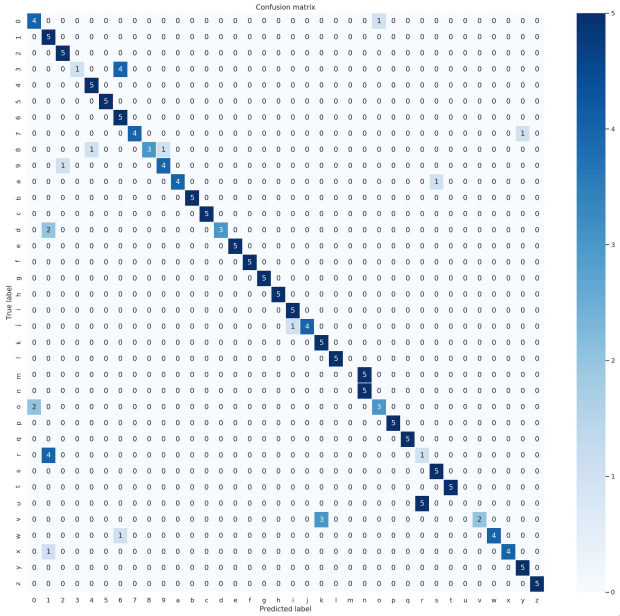
Total params: 432,026
Trainable params: 432,026
Non-trainable params: 0

Sign Recognition

CNN Results (1)

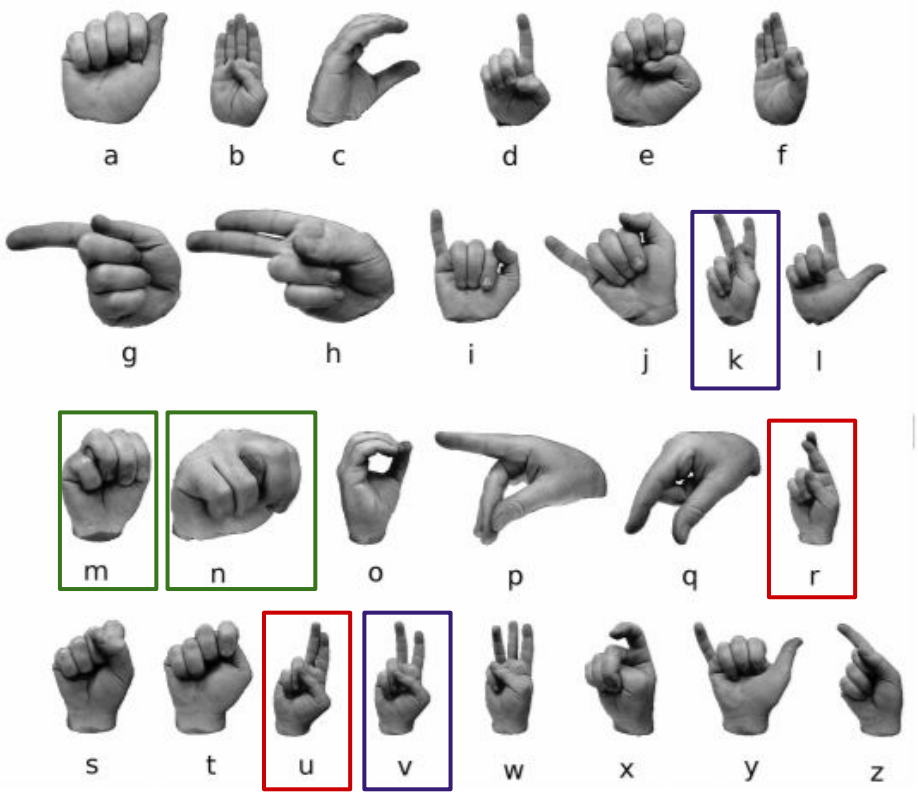
Averages	Precision	Recall	F1-score
Macro	0.82	0.81	0.79
Weighted	0.82	0.81	0.79

Accuracy: 0.81



Sign Recognition

CNN Results (2)



Hand Recognition

SSD Network

- Faster than R-CNN for object detection
- Real time networks
- COCO - dataset
- Pretrained network with transfer learning

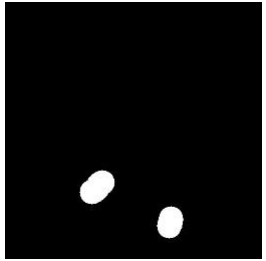
https://github.com/victordibia/handtracking

System	VOC2007 test *mAP*	FPS (Titan X)	Number of Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	~6000	~1000 x 600
YOLO (customized)	63.4	45	98	448 x 448
SSD300* (VGG16)	77.2	46	8732	300 x 300
SSD512* (VGG16)	**79.8**	19	24564	512 x 512

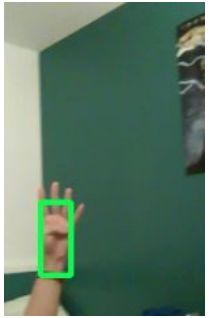
Hand Recognition

Dataset, what didn't work

- Custom network + Labelling data manually
- 500 images labelled
- COCO Dataset (Labelled as "person")



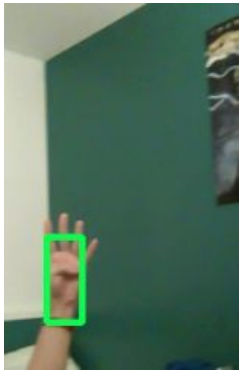
- Integration with Sign Recognition needs improvement



Hand Recognition

Finder

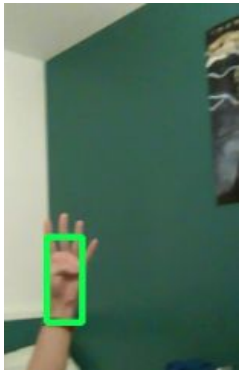
- SSD Box detector
- Discard boxes with minimal normalized area
- If no boxes are found, pick last found box within time
- Thresholding score for best boxes
- Pick box above threshold with best manhattan distance
- If distance above threshold, pick from all boxes
- Check for predominant colors (KMeans) for hand like
- Assert average last scores above threshold



Hand Recognition

Masking

- Expand ROI by factor (150%)
- Transform ROI to HSV, pick colors in range
  - `H:  5°, S: 19%, V: 19%`
  - `H: 50°, S: 58%, V: 58%`
- Open morphology (Erode + Dilate)
- Dilate morphology
- Find contours, pick closest with manhattan distance
- Fill contour



Integration

- Tests performed in single threaded program 3.40GHz
  - Sample phone [200€]: **Snapdragon octa-core 2.0GHz**
- Using CPU only. Expected to run faster on mobile GPU
- Speeds:
  - `Frame reading (Webcam): 25ms`
  - `Pre-processing: < 0ms`
  - `SSD Detector: 30ms`
  - `Post-processing: 10ms`
  - `Classifier: 10ms`
  - `Speller (expected): < 5ms`
  - `Total processing: 55ms (18 FPS)`
  - `Total: 85ms (11 FPS)`

Difficulties found during the project

Expected

- **Accuracy drop with real images:**
  - Real images even without background are different and contain a lot of noise
- **Compatibility problems when integrating the CNN model on the app**
  - The model was easier to integrate than we expected

Unexpected

- **Lack of time to test the model with real images**
  - Dependency of the interface to try the model with real data
- **The voice to text part was more difficult than we expected**
- **Less time than expected to work on the project**

Future Work: Sign Recognition

Signreco

Basic Future Work

- Fully integrate the keras model into the android interface
- Perform proper tests with real images
  - Fix any performance problem that might appear

Extended Future Work

- Add compatibility to whole words and movement
  - **Working with signs for words**:
    - **Reason**: Sign language is word oriented
    - **How**: New dataset, new model, same interface
- Spell-check the whole words to detect possible mismatches of the model
- **Include other language than English**
  - **Reason**: Scaling
  - **How**: New dataset, new model, change in the interface

References

[1] Barczak, A. L. C., Reyes, N. H., Abastillas, M., Piccio, A., & Susnjak, T. (2011). A new 2D static hand gesture colour image dataset for ASL gestures.

[2] https://edu.authorcafe.com/academies/6813/sign-language-recognition

[3] https://github.com/victordibia/handtracking

[4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu: "SSD: Single Shot MultiBox Detector", 2015; http://arxiv.org/abs/1512.02325 arXiv:1512.02325. DOI: https://dx.doi.org/10.1007/978-3-319-46448-0_2 10.1007/978-3-319-46448-0_2

Signreco

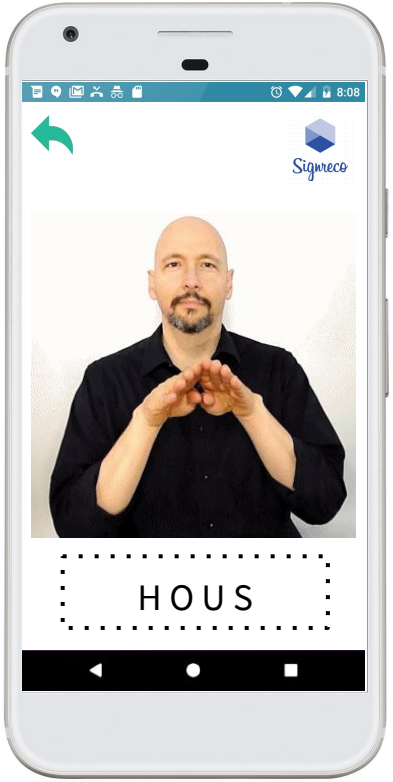
THANK YOU

Raquel Pérez
raquel.leandra.perez@est.fib.upc.edu

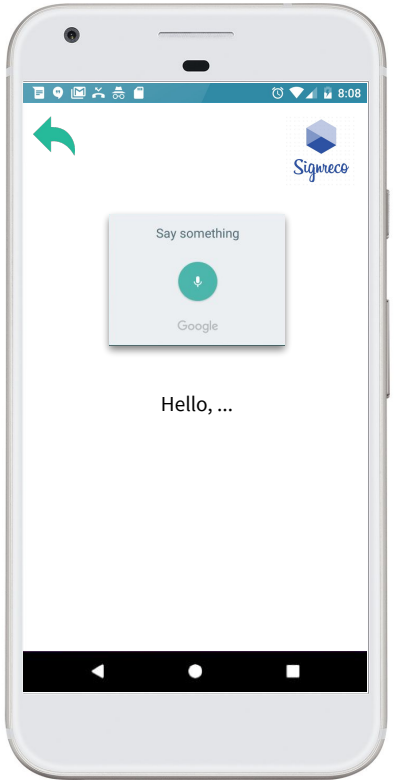
Jorge Sierra
jorge.sierra@est.fib.upc.edu

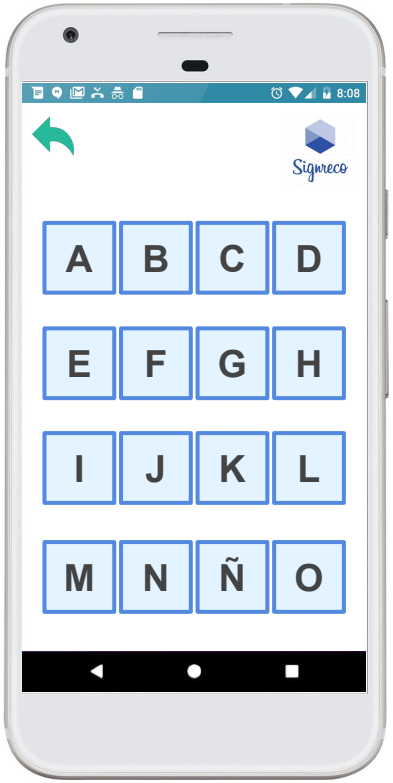
Gaspard Debussche
gaspard.debussche@est.fib.upc.edu

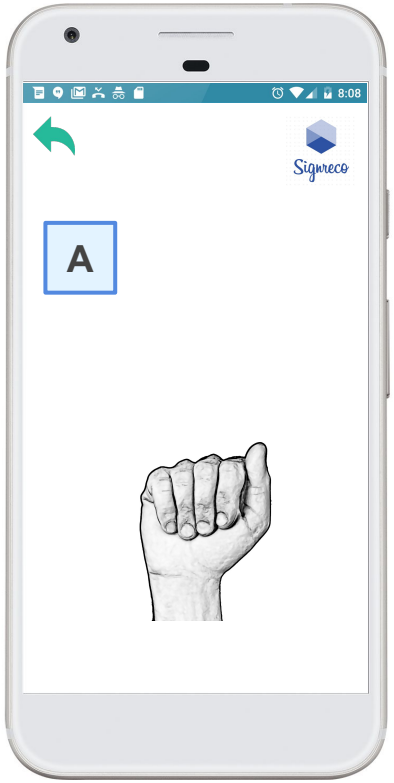
Mockup



Mockup


8:08
Signreco
Say something
Google
Hello, ...


8:08
Signreco
A B C D
E F G H
I J K L
M N Ñ O


8:08
Signreco
A

SSD Architecture