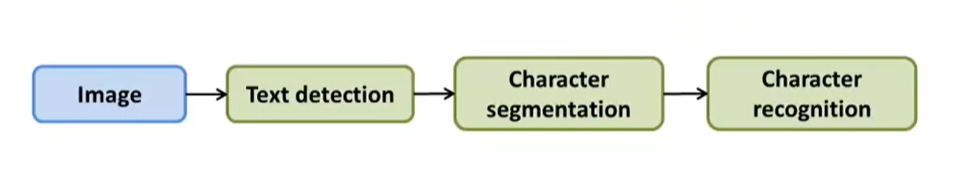
Week 11

**Pipeline:**

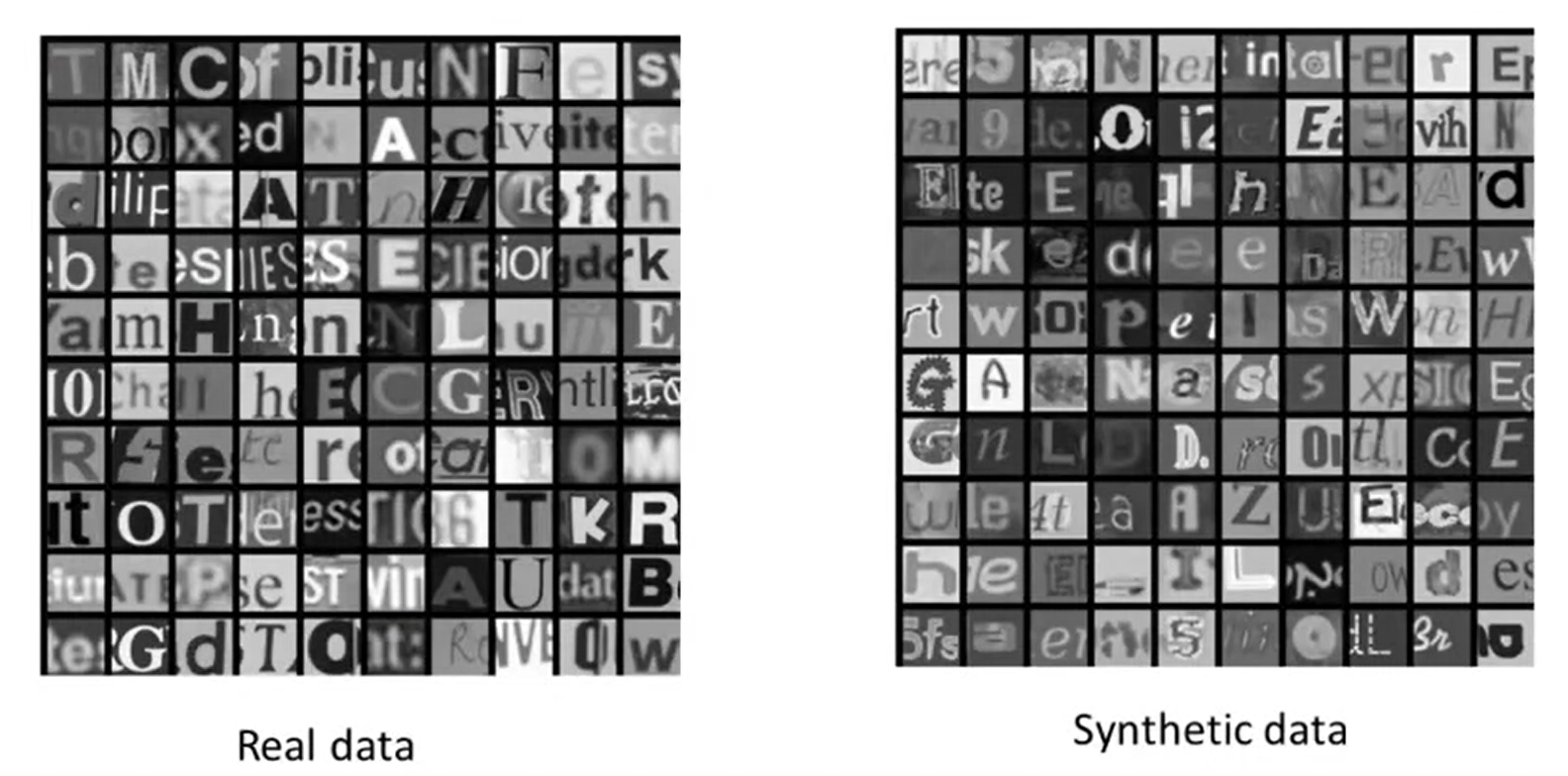
Example: Photo ORC pipeline



**Sliding Windows:**

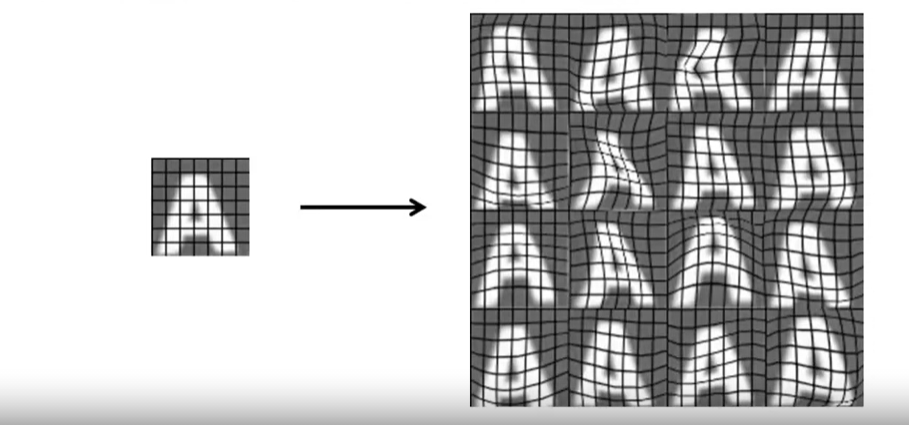
****

**Artificial Data:**

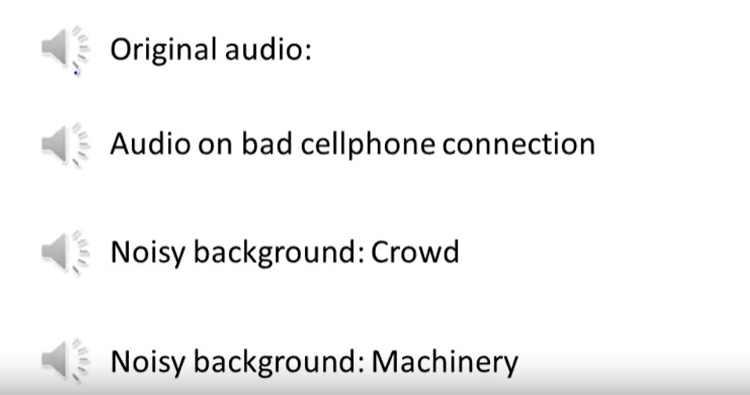
-synthetic data: (e.g: interner font vs random background)

-distortions:

(image)



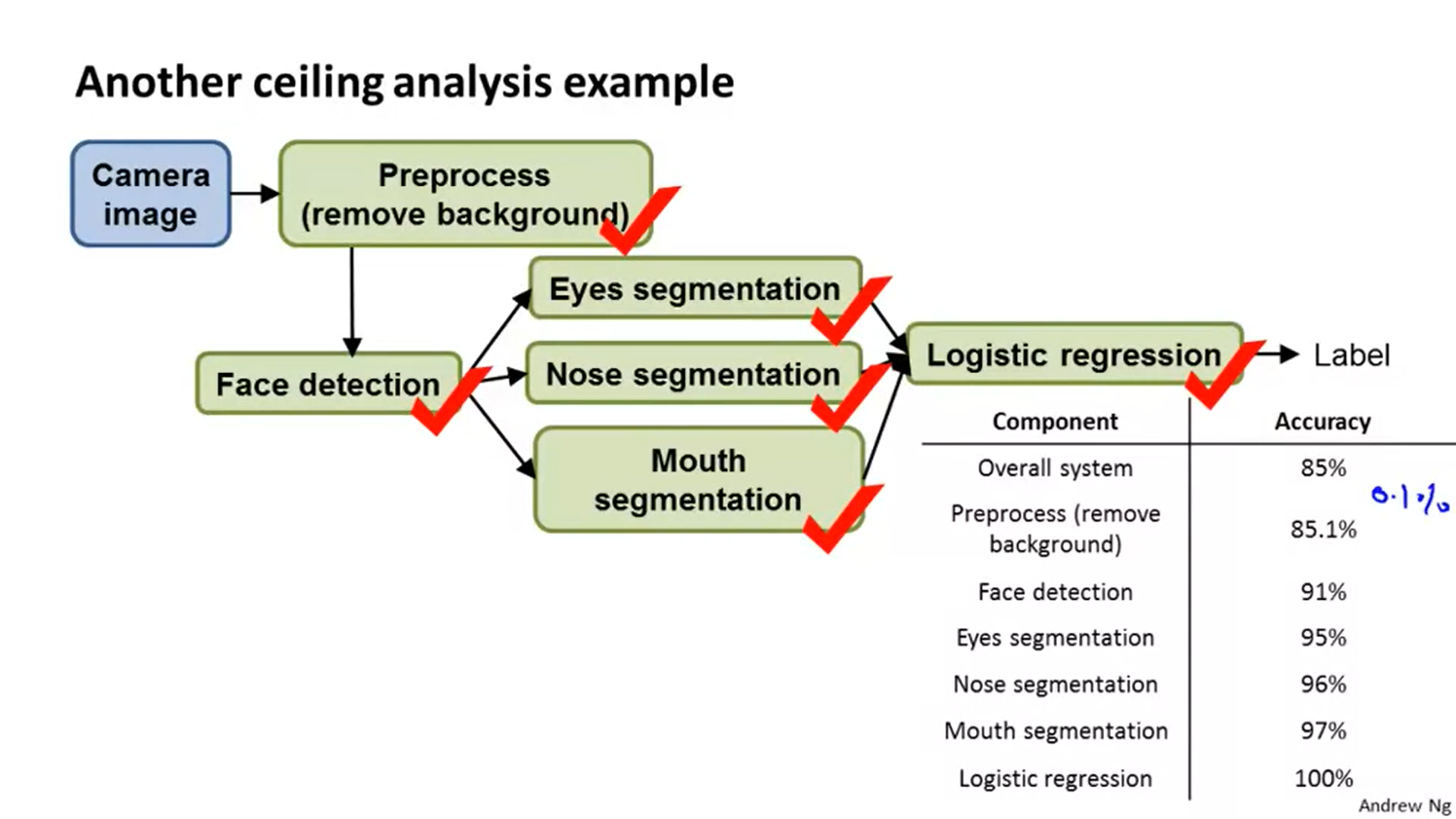
(or audio)

****

**Discussion on getting more data:**

* Make sure you have a low bias classifier. If not, keep increase the number of features/ number of hidden units in neural network until you have a low bias classifier
* 10x data?:
  + Artificial data synthesis
  + Collect/label it yourself
  + “Crowd source” (E.g. Amazon Mechanical Turk)

**Ceiling Analysis:**



-85% means the current overall system has 85% accuracy.

-if we change the output of the preprocess component with the ground truth label, overall system accuracy up from 85% to 85.1% (increase 0.1%)

-if we change the output of the face detection with the ground truth label, overall system accuracy up from 85.1% (accuracy with 100% accuracy preprocess) to 91%.

-and so on if we have ground truth label for logistic regression, overall system has 100% accuracy