

ON THE PERFORMANCE OF OPTICAL MUSIC RECOGNITION IN THE ABSENCE OF SPECIFIC TRAINING DATA

Juan C. Martinez-Sevilla¹, Adrian Rosello¹, David Rizo^{1,2} and Jorge Calvo-Zaragoza¹

¹ Pattern Recognition and Artificial Intelligence Group, University of Alicante, Spain

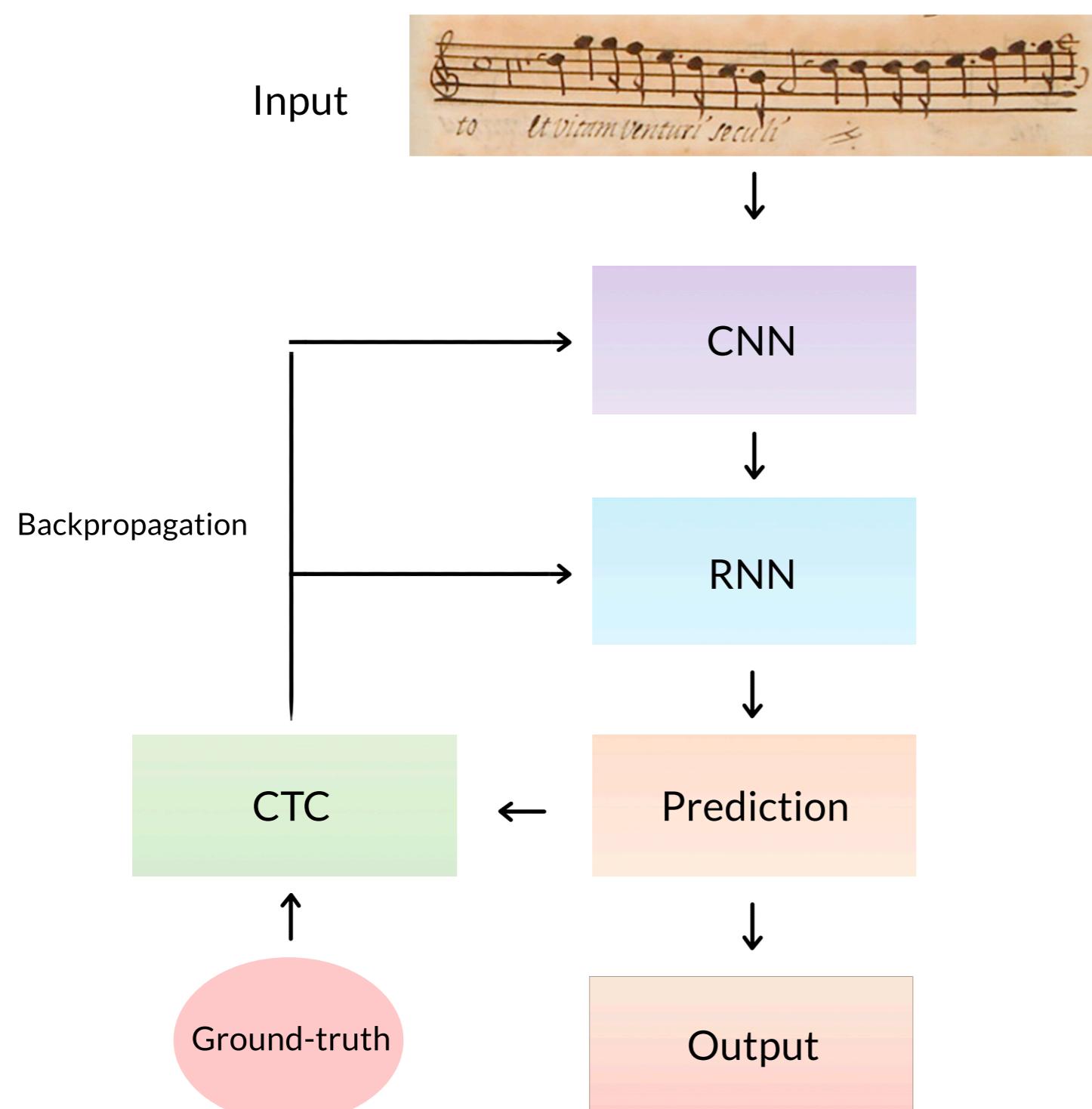
²Instituto Superior de Enseñanzas Artísticas de la Comunidad Valenciana, Spain

adrian.rosello@ua.es, {jcmartinez, drizo, jcalvo}@dlsi.ua.es

What to do when facing a new collection recognition process?

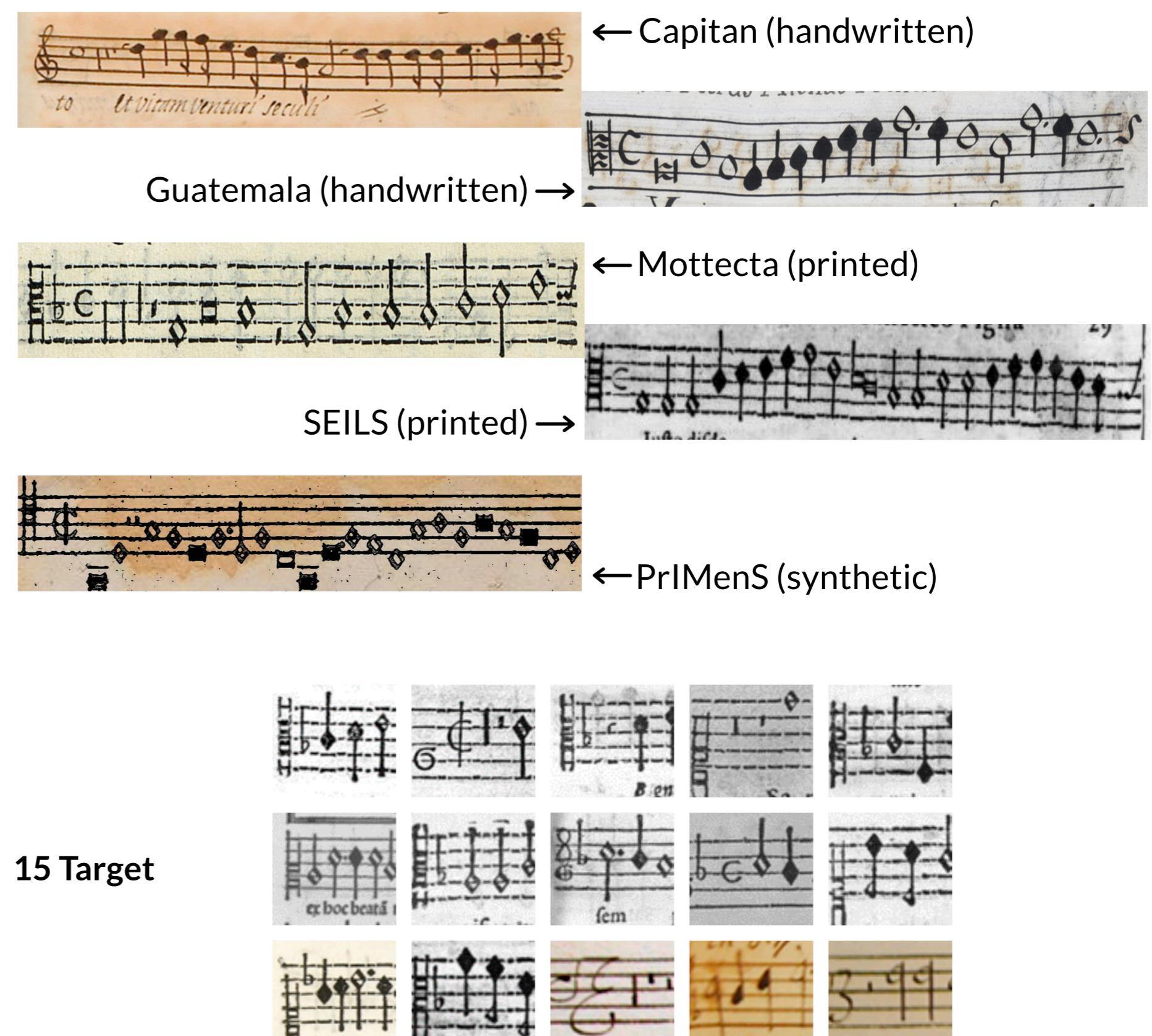
Learning framework

Convolutional Recurrent Neural Network (CRNN)



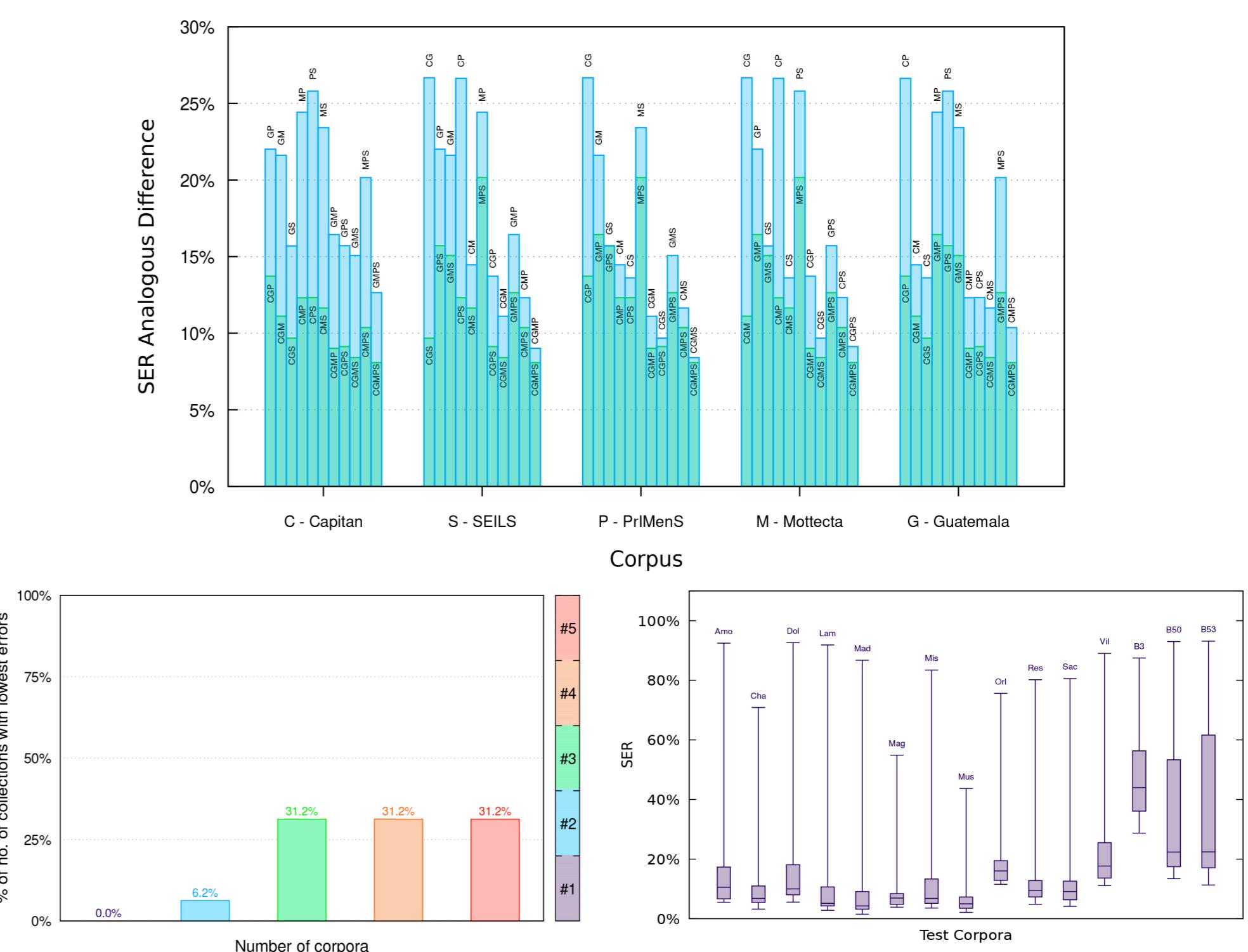
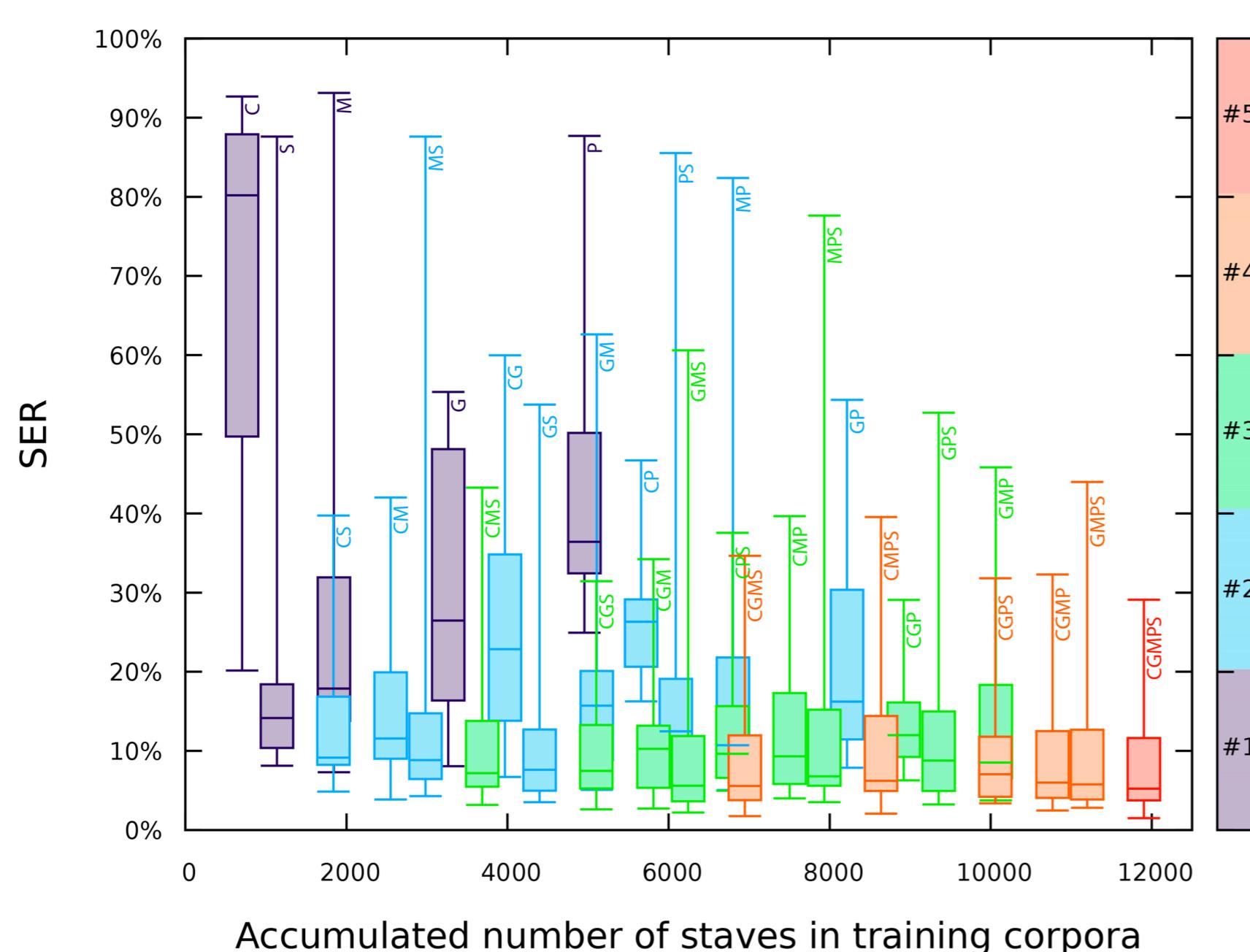
Datasets (mensural notation)

Training



Results

Performance measured using *Symbol Error Rate* (SER). Computed as the average number of elementary editing operations (insertions, deletions, or substitutions) required to convert the predicted sequence into the reference.



Lessons learned

Which is the best choice to transcribe a new collection?

One must use all the available training corpora even if some of them are quite different from the target collection.

How important is it to be aware of the collection to transcribe for selecting the right corpora to train the model?

It is indeed relevant, and depending on the difficulty (for example, whether or not it is handwritten) the differences in performance can be very varied.

Is it better to have fewer collections with a high number of samples or more collections with fewer samples each?

It is preferable to have more variability even at the cost of a smaller sample set.

Does the introduction of a synthetic corpus improve the performance?

Yes, the introduction of a reliable synthetic collection adds size and variability to the training data, enabling better performance rates.