A Quick Intro to Josh’s Project

2019: Transfer Learning Internship w/ Kelly Davis

2020: ASR + KWS for under-resourced languages

2019 research:

Background

Transfer Learning + DeepSpeech

This work is best explained in Chapter 8 of Josh’s dissertation:
http://jrmeyer.github.io/misc/MEYER_dissertation_2019.pdf

2019 research:

Background

Transfer Learning + DeepSpeech

Approach:

Use a trained LVCSR English model to bootstrap a new target language LVCSR model*

*an ASR equivalent of [Yosinski 2014]

2019 research:

Background

Transfer Learning + DeepSpeech

Research Questions:

(1) How deep should we slice the source model?

(2) Should we freeze or fine-tune transferred params?

2019 research:

Target Datasets

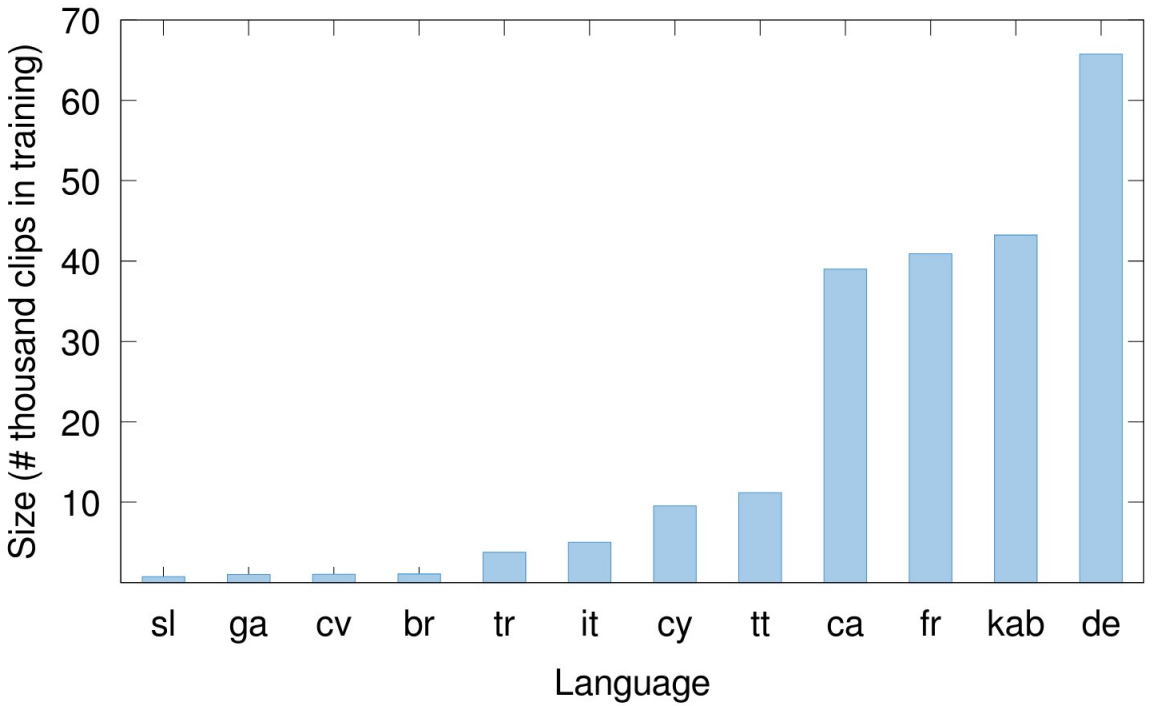


Figure 8-3: Number of audio clips per train split for each language. Audio clips for any individual speaker are only found in one split of the data (i.e. dev / test / train).

2019 research:

Source Model

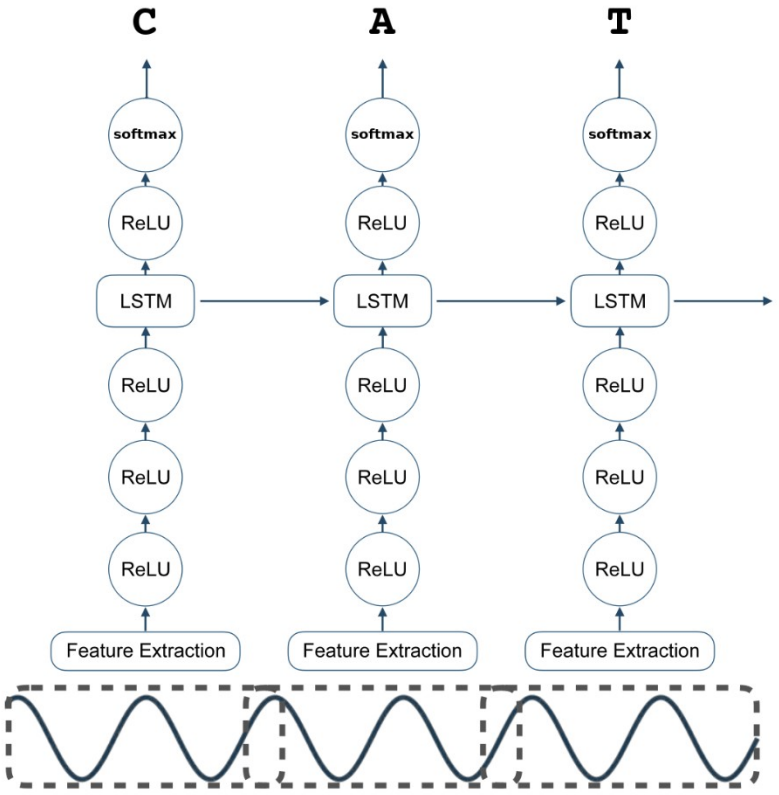


Figure 8-1: Architecture of Mozilla's DeepSpeech ASR system. A six-layer unidirectional CTC model, with one LSTM layer.

2019 research:

Transfer Approach

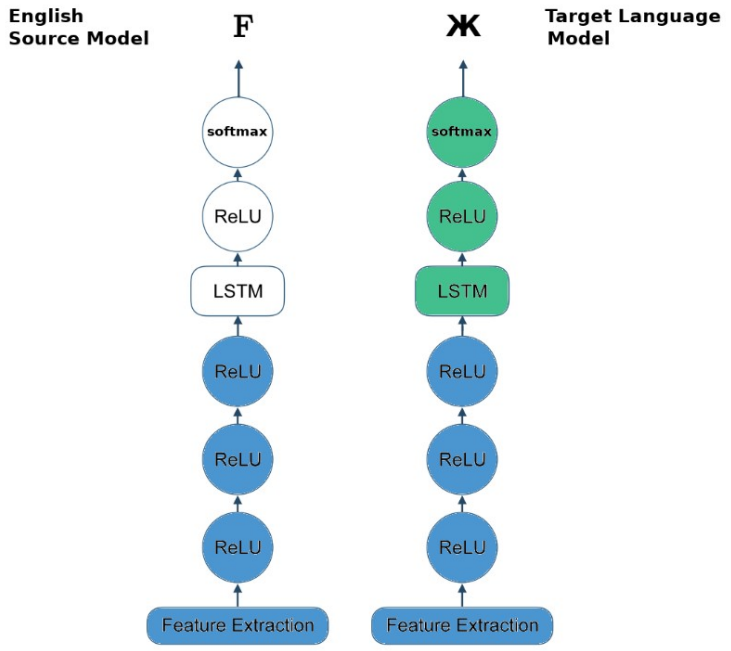


Figure 8-2: "Copy-Paste" Transfer Learning. In this example, a subset of model parameters trained on a language using the Latin alphabet is used to create a new model for a new language, written with the Cyrillic alphabet. Blue model parameters have been copied from the source model, and the green model parameters have been re-initialized from scratch.

2019 research:

Results per Language

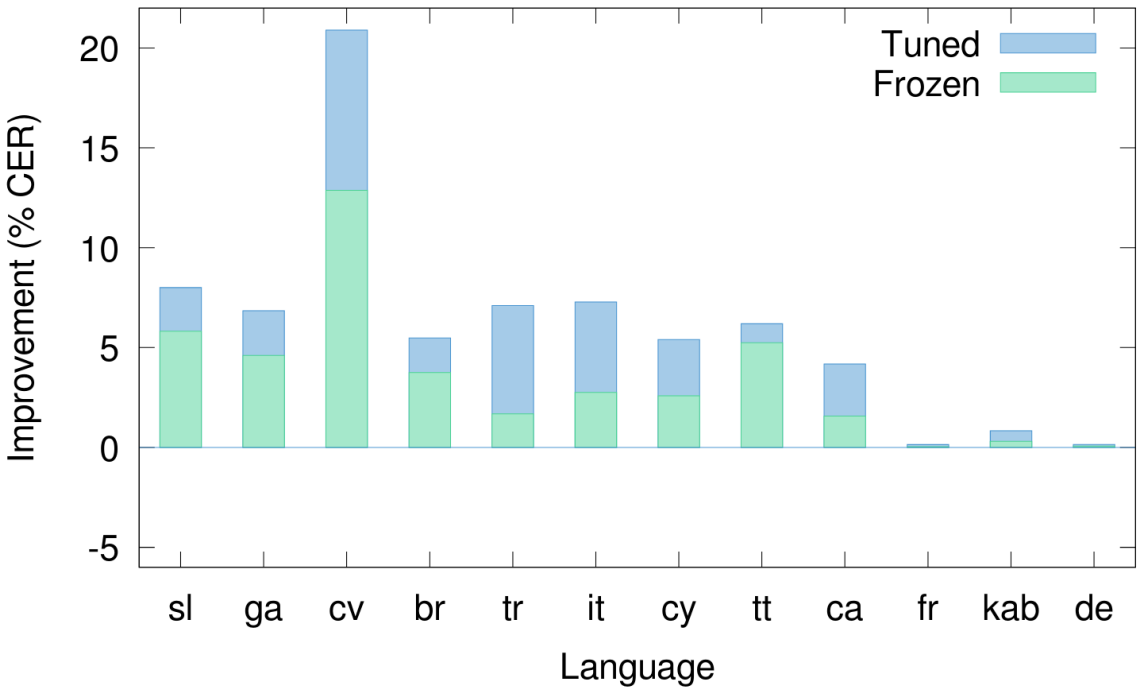
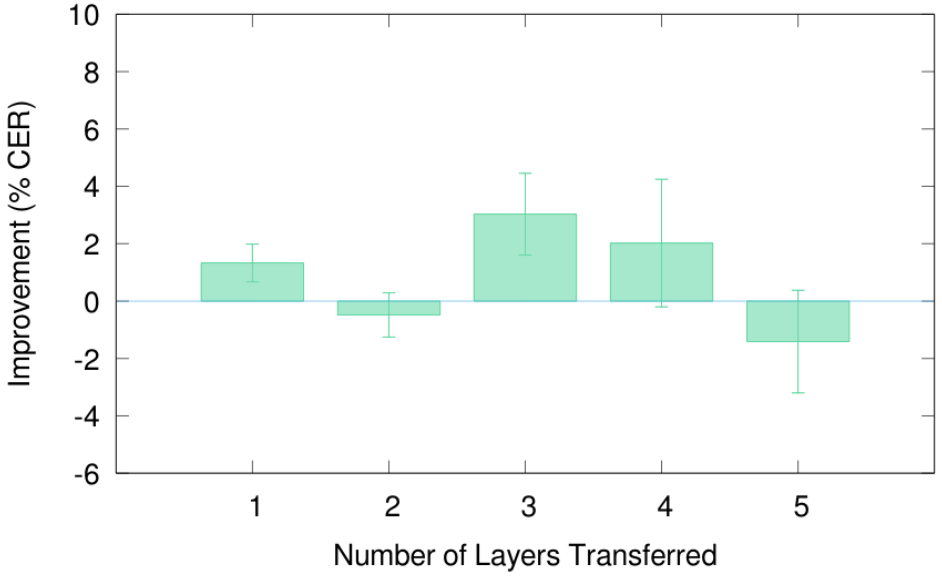


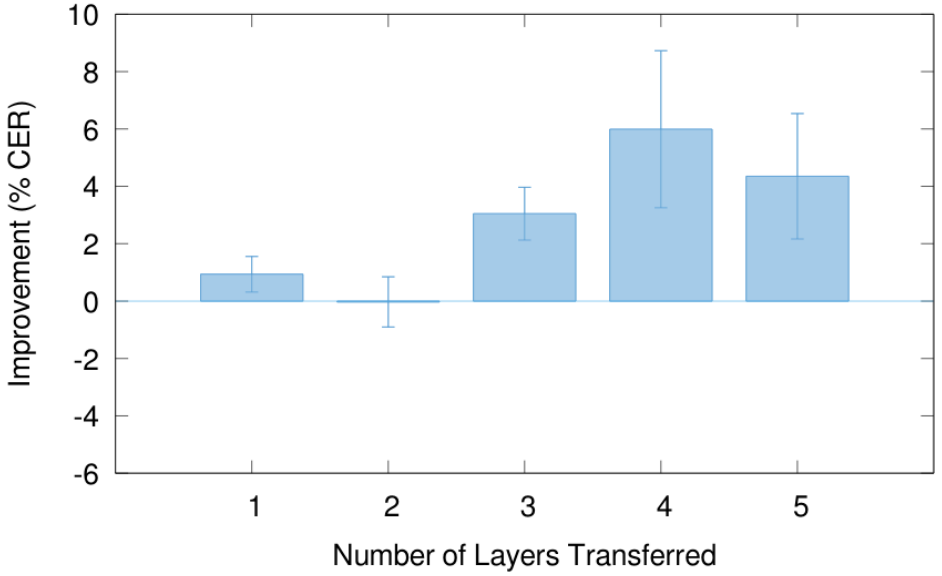
Figure 8-5: Largest improvement from Transfer Learning relative to the baseline, for each language. Languages are ordered from left → right in ascending order of size of training dataset. These improvements represent the bolded values in Table 8.2 and Table 8.3.

2019 Research:

Results averaged over Languages



(a) Frozen Transfer



(b) Fine-Tuned Transfer

Figure 8-4: Mean and standard deviation for CER improvement for different # layers transferred from English.

2019 Research:

Conclusions

Keep the LSTM layer, but drop everything after

+

always fine-tune

2020: Josh’s Project Goals

(1) Train useable models for East African languages

(2) Research best practices for training small vocab models

(3) Communicate best practices for training small vocab models

2020: Project Deliverables

(1) LVCSR Kinyarwandan model

(2) KWS Luganda model

(3) models benchmarked on digits/YES/NO for Common Voice languages

(4) playbook for training new models with DeepSpeech

Contact Info

joshua.richard.meyer@gmail.com

@_josh_meyer_

https://chat.mozilla.org/#/room/#machinelearning:mozilla.org
(username == josh)