**Speech to Image**

Deployment in Heroku

MELD Dataset

Text to Image

Speech to Text (Speech Recognition)

**Libraries**

PyTorch

**STEPS TO FOLLOW**

Reference:

<https://github.com/madewithml/e2e-ml-app-tensorflow/blob/master/notebook.ipynb>

1. Check-in file into Github
2. Download MELD dataset : Store in Floyd hub
3. Speech to text :
   1. Start with state of art model – 2020 best modes (Papers with code)
      1. Try hugging face models
   2. Model should have training, testing and inference code
   3. Test the models using inference code
   4. Choose the best model . Apply for MELD dataset and test
   5. Work on Colab
4. Coding improvements
   1. Create function and classes
   2. Organize code using cookiecutter : <https://github.com/madewithml/boilerplate>
   3. Log key events – Preprocessing, training, testing performance : <https://docs.python.org/2/howto/logging.html>
   4. Test code using pytest : <https://docs.pytest.org/en/stable/>
5. Performance tracking
   1. Use Weights and Biases
   2. Create dashboard pages
   3. Create W&B reports , which documents performance improvements from simple to more complex models
   4. Use W&B tool – Sweeps for hyperparameter tuning
   5. Use W&B tool- Artifacts for data/pipeline/model management
6. Robust prediction peipeline
   1. Output is not softmax scores – instead, we need to get the top prediction classes
   2. Consider equal class probabilities while training – stratify
   3. Observe training and testing input data – datatypes/other – If using tensorflow use tensor---
7. Wrap model as API
   1. Use Flask/Django/FastAPI . Try using FastAPI fro ease and performance boost
   2. Create Docker image using Dockerfile
8. Front end
   1. Use streamlit to create interactive UI
   2. Use Heroku service for deployment

**Testing Pre-trained model**

**Reference** : Pose detector : <https://github.com/lxy5513/hrnet>

For an identified model,we can import the mdl(weights) and dataset from git using the commands mentioned in the reference:

* Clone the complete folder from git
* Install all the lib required and make file
* Download the pretrained mdl file from google drive
* Download pretrained(yolo) model from server location
* Apply the models on dataset

**MELD Dataset overview :**

**Refer : MELD\_researchpaper.doc**

**What can we do with the dataset?**

1. **Audio clip 🡪 utterance(text)** 
   1. **MELD dataset can be used**
2. **Text -> image** 
   1. **COCO dataset can be used**
3. **GANs**

**REFERENCE MODELS**

Speech to text models

<https://madewithml.com/search-results/?tags=tts>