Project Beta Progress Report

Yuan Wang (Aria), Cindy Mo, Yucheng Dong (Steve), Rishi Sinha, Raj Agrawal

November 12, 2015

Background

Data Description

- high-resolution fMRI dataset
- 20 participants
- audio description in German with start and end times
- dsnum: ds000113

Background (Cont)

The Data

- 2 hour in total for 1 subject
- 8 runs (7 runs for later training and 1 run for validation)
- 15 mins each run
- Format: 4D volumetric images (160x160x36)in NIfTI format
- TR: 2s
- Sequence: T2*-weighted gradient-echo EPI sequence (1.4 mm isotropic voxel size).

Background (Cont)

The Data

- IMPORTANT NOTES: These images have partial brain coverage — centered on the auditory cortices in both brain hemispheres and include frontal and posterior portions of the brain. There is no coverage for the upper portion of the brain (e.g. large parts of motor and somato-sensory cortices)."
- Subject we downloaded: 004, (014, 015)

Completed Steps

Stimuli Preprocessing

- Translated German audio description into English
- Removed English stopwords (common words that hold no meaning) from audio descriptions

Completed Steps (Cont)

Voxelwise Modeling - WordNet

- Tagged audio descriptions with WordNet labels and generated a “word to WordNet” dictionary
- Grouped audio stimuli according TRs
- Generated a time (TRs) by features (WordNet tags) design matrix

Design Matrix

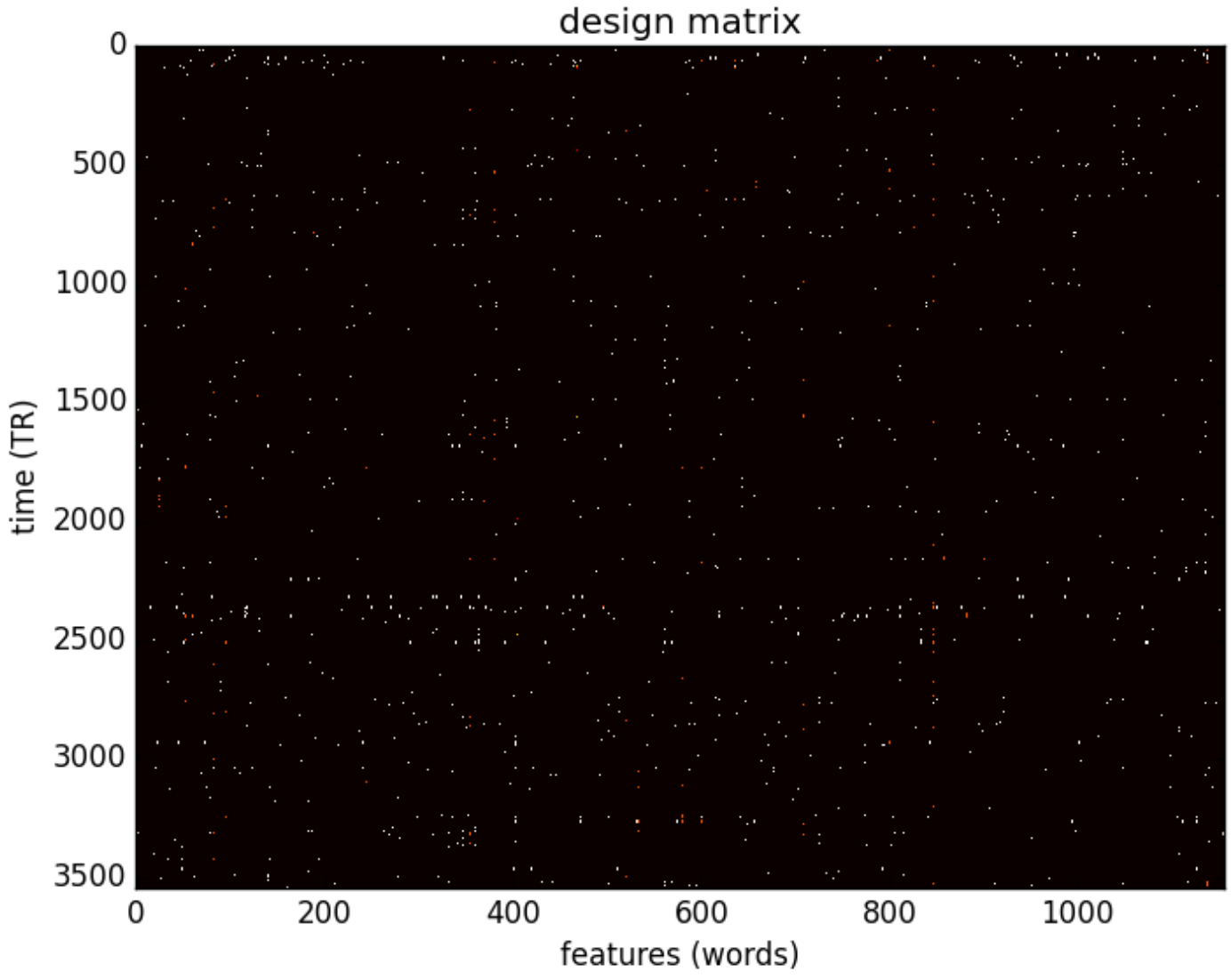

Figure 1:Design Matrix

Completed Steps (Cont)

Scene modeling

- Created task-courses and scene categories for each run
- Extracted scenes from description and splitted scenes according to runs

All scenes

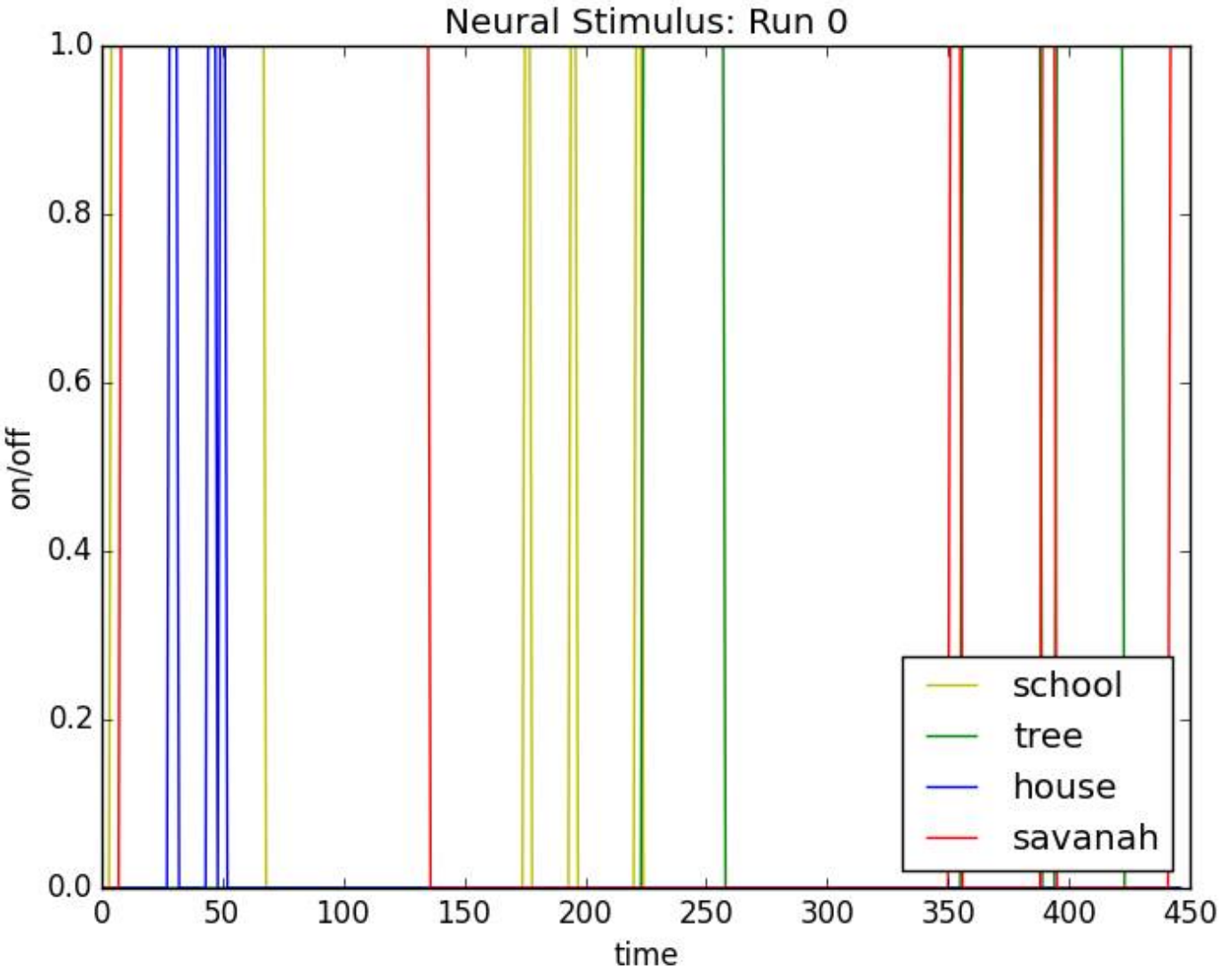

Figure 2:Scenes Conditions

fMRI Data Preprocessing

- Inspected the data by generating different kinds of summary statistics
- Saved outlier data indices

Standard deviation

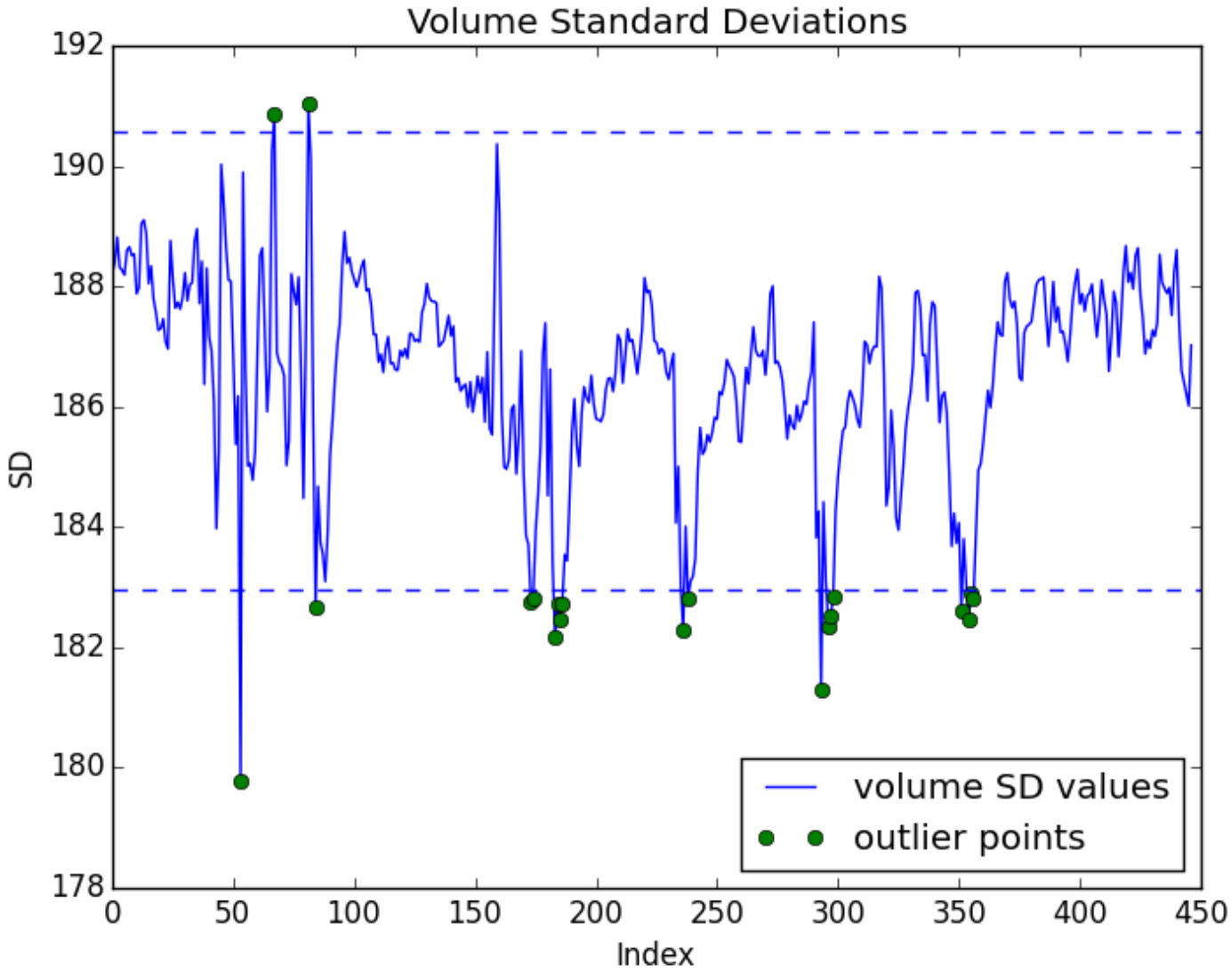

Figure 3:Standard Deviation

RMS difference

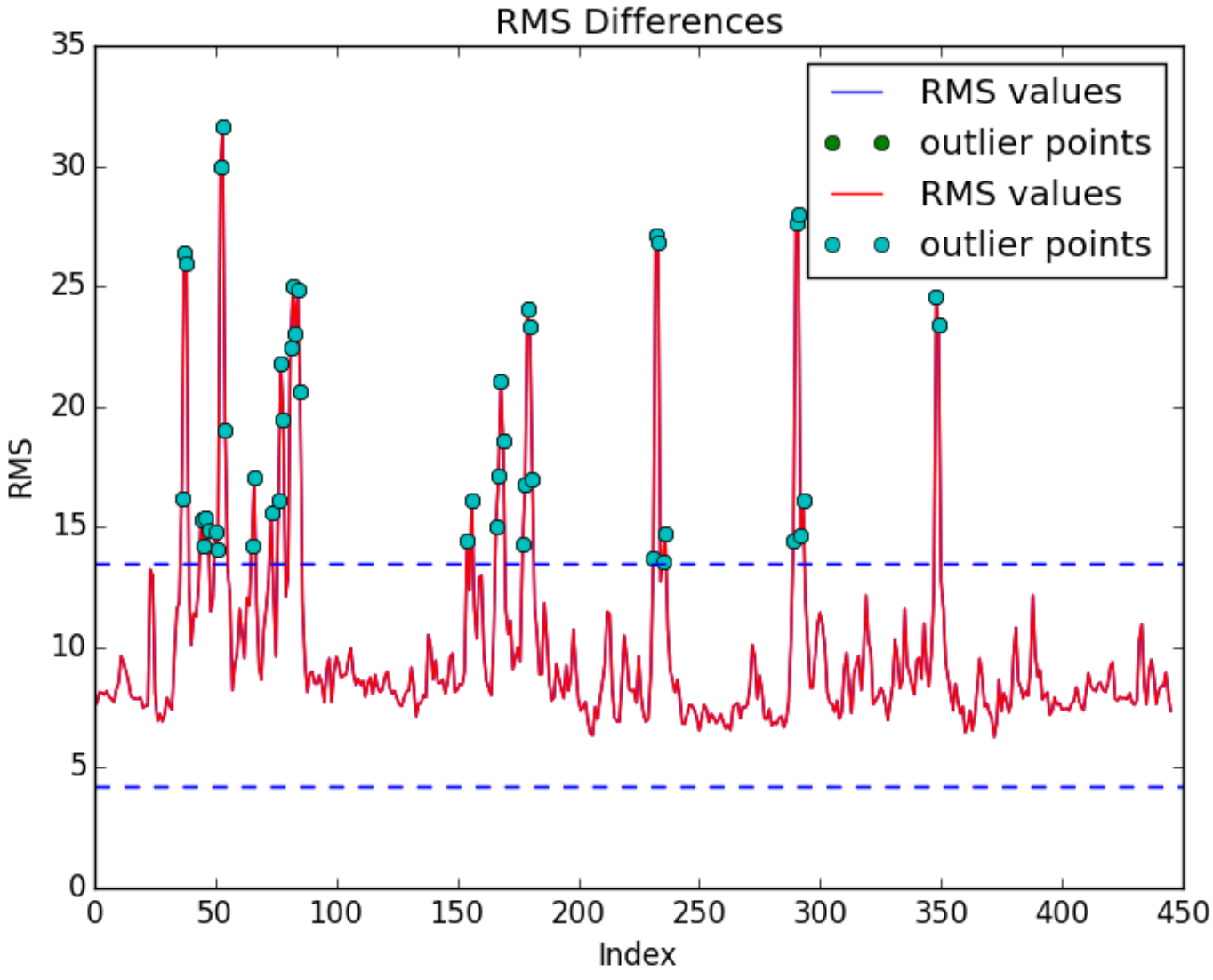

Figure 4:RMS difference

Completed Steps (Cont)

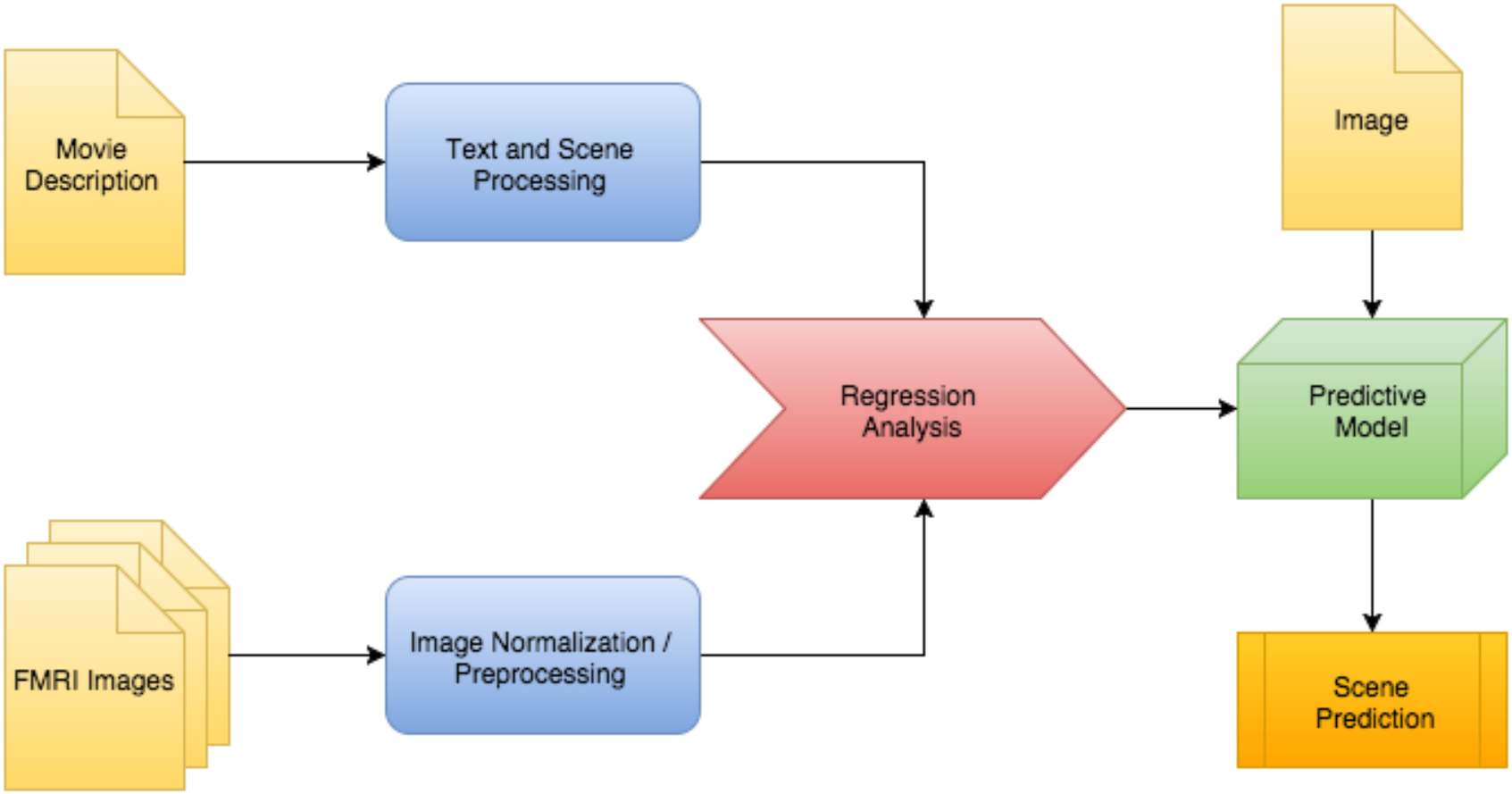

Figure 5:Flow Chart

Two Kinds of Modeling

- Voxelwise Modeling of audio description using Ridge Regression Model
  - train models with BOLD signal on audio description
  - Goal 1: predict voxelwise BOLD response based on description
  - Goal 2: predict words based on BOLD response
- Linear Modeling of brain activity on Scenes
  - Goal: predict the scene category based on brain activity
- Hypothesis Testing:
  - Parametric: t-test, z-test
  - Non-Parametric: permutation test, sign-test
- Checking Robustness, assesing model peformance (e.g. AIC), resampling

Future Plan

- Accomplish two modeling goals
- Checking out-of-sample peformance
- Adding more tests, docstrings, and improving readability

Problems

- Learning Github flow
- Learn to "unpush" something
- Understanding fMRI data
- Reproducibility and handling such large datasets
- Data manipulation to fit the question we are trying to answer

Process

Issues as Team

- Finding right time to meet
- Varying degrees of statistical and coding knowledge

Feedback

Useful:

- Code exercises and solutions
- Github

Could Improve:

- Lack of notes
- More emphasis on fMRI data

Areas of Review:

- ANOVA
- Regression modeling