school of computing, informatics, & decision systems engineering

ARIZONA STATE UNIVERSITY

ICPR 2020

Concept Embedding through Canonical Forms: A Case Study on Zero-Shot American Sign Language Recognition

- Azamat Kamzin akamzin@asu.edu
- Apurupa Amperayani vamperay@asu.edu
- Prasanth Sukhapalli psukhapa@asu.edu
- Ayan Banerjee abanerj3@asu.edu
- Sandeep K. S. Gupta sandeep.gupta@asu.edu

Arizona State University

IMPACT Lab: impact.asu.edu

25th International Conference on Pattern Recognition 2020, Milan, Italy

IMPACT Lab Research Overview

Safe, Secure and Intelligent **AI enabled** Cyber-Physical Systems

- Dr. Sandeep Gupta, director of school of computing (CIDSE), pervasive mobile computing
- Dr. Ayan Banerjee, assistant research professor, cyber physical systems (CPS)

- Lab members
  - Imane Lamrani, postdoc — **model mining and verification of cps**
  - Azamat Kamzin, PhD student — **zero shot learning, concept Learning**
  - Vinaya Chakati, PhD student — **grid computing**
  - Subhasish Das, PhD student — **model driven deep learning**
  - Bernard Nanganbonziza, PhD student — **mobile security**
  - Javad Sohankar, PhD student — **mobile security, brain mobile interface**
  - Sameena Hossain, PhD student — **education technology, accessible computing**

Motivation

- Gesture understanding requires a language model
- Advantages of Developing a Gesture Language Model
  - Language Translation
  - Gesture-based searching and mining
  - Automated Transcription of gestures
  - **Zero Shot Learning of Gestures – focus of the paper**
    - Recognize unseen gestures without access during training

Traditional Solutions

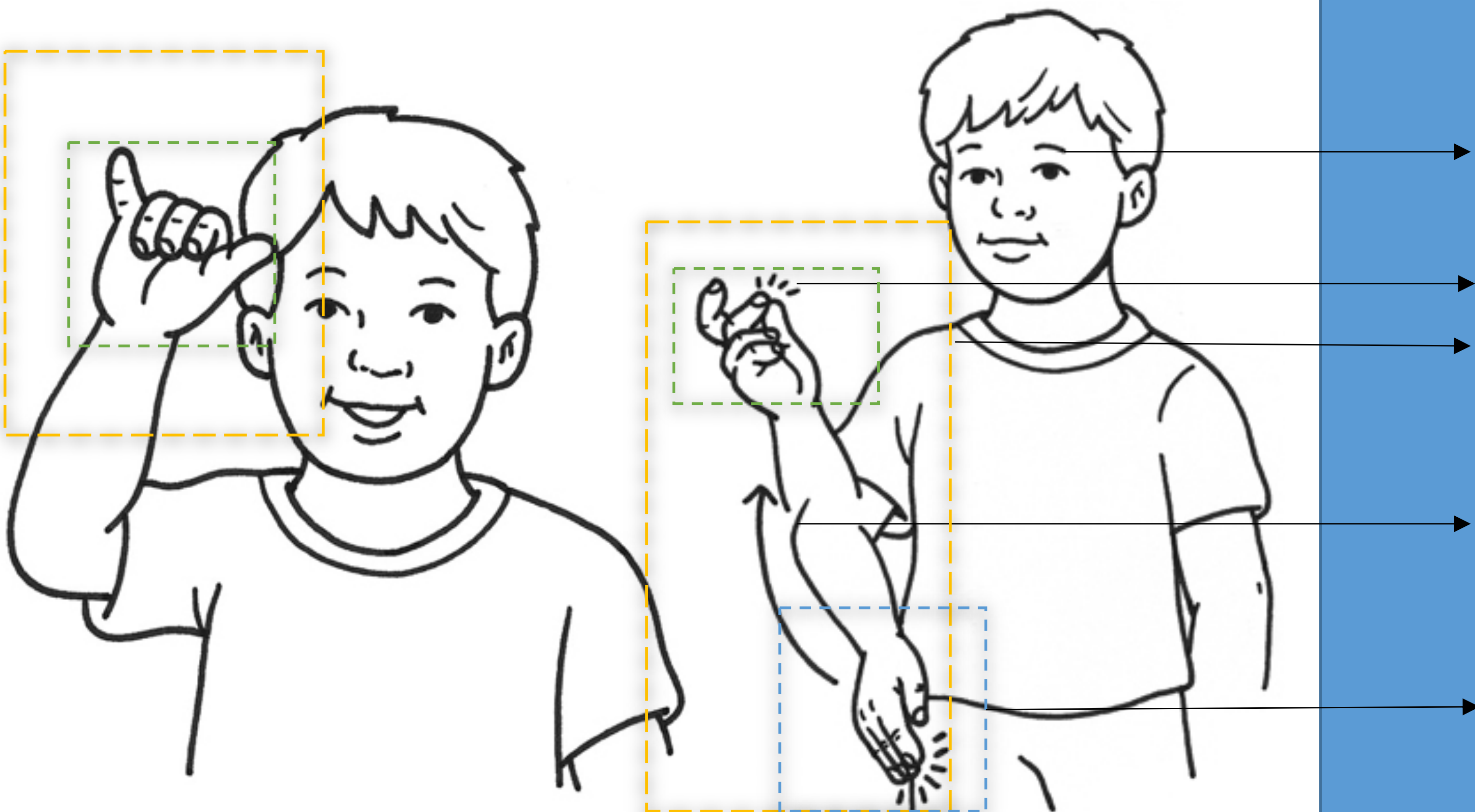
- Gesture recognition requires video classification
- Solution 1: Apply 3D-CNN or similar technique directly to video to predict gesture
  - No feature engineering
  - Problems:
    - Depends on signal features from examples
    - Requires large datasets
    - American Sign Language has limited dataset
- Solution 2: Adding high-level knowledge improve accuracy
  - Similar to Transfer learning, and its benefits
  - Problems:
    - No Transfer learning models for gestures

Concept: High-level knowledge

- Concept Definition
  - Attributes of examples with following properties
    - Common across examples of different classes
    - Each example can be uniquely represented in terms of concepts
    - Examples can be represented as a Spatio-Temporal sequence of concepts
    - Allows *soft matching*
- *Solution Approach:*
  - *A gesture parser that splits a gesture video into concepts following a grammar*
  - *Utilize transfer learning models for each concept*
- *Challenge:*
  - *Define concepts such that transfer learning models are available*
  - *Develop a grammar for language model for gestures*

American Sign Language

Concepts



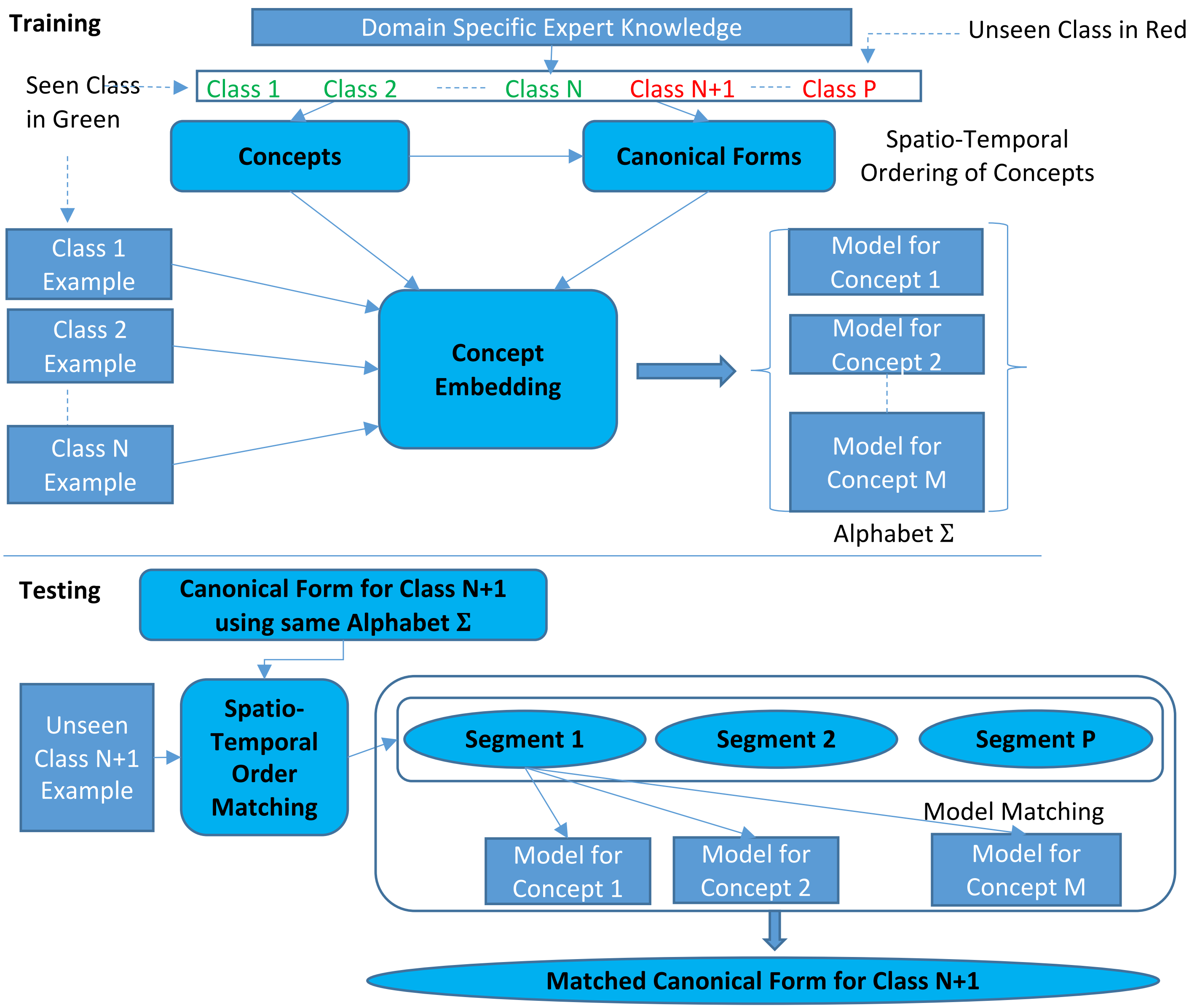
Context Free Grammar

- Canonical form of gesture representation

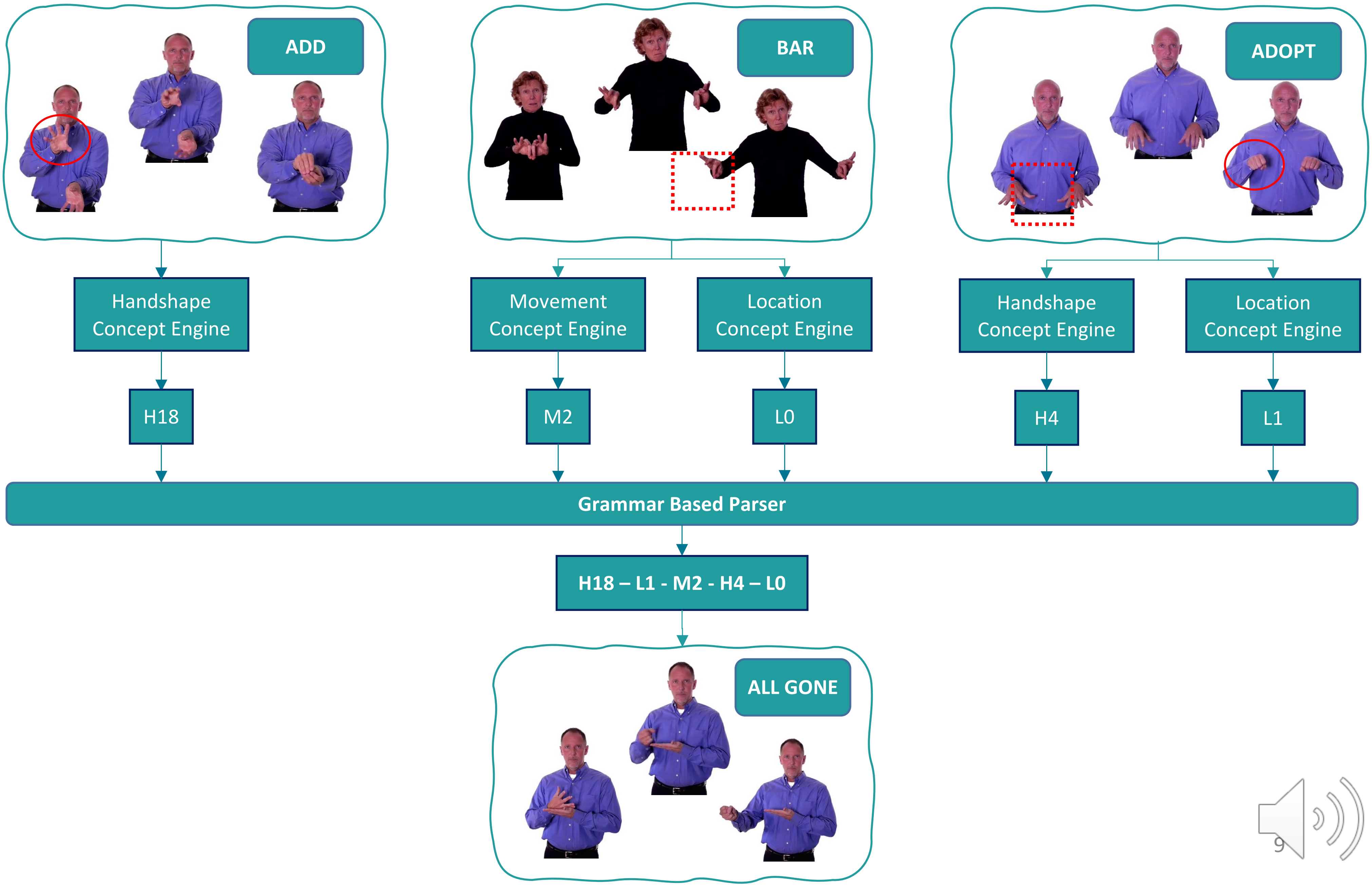
$$
\begin{array}{lll}
Hand & \rightarrow \Sigma_H & \longrightarrow \textbf{Handshape Alphabet} \\
Mov & \rightarrow \Sigma_M & \longrightarrow \textbf{Movement Alphabet} \\
Loc & \rightarrow \Sigma_L & \longrightarrow \textbf{Location Alphabet}
\end{array}
$$

$$
\textbf{Temporal Sequencing} \left\{
\begin{array}{ll}
GE & \rightarrow GE_{Left} GE_{Right} \\
GE_X & \rightarrow Hand | \epsilon, \text{ where } X \in \{Right, Left\} \\
GE_X & \rightarrow Hand\ Loc \\
GE & \rightarrow Hand\ Loc\ Mov\ Hand\ Loc
\end{array}
\right.
$$

Concept Embedding



Example

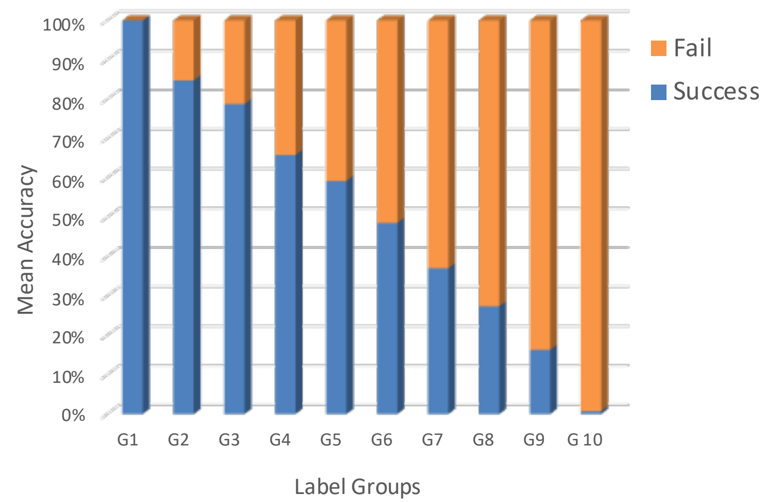

ADD
BAR
ADOPT
Handshape Concept Engine
Movement Concept Engine
Location Concept Engine
Handshape Concept Engine
Location Concept Engine
H18
M2
L0
H4
L1
Grammar Based Parser
H18 – L1 - M2 - H4 – L0
ALL GONE

Evaluation Datasets

- IMPACT Lab dataset:
  - Using Learn2Sign mobile application
  - 23 gestures from 130 learners with 3 repetitions
  - Varying light conditions, distance to the camera, recording pose
  - Used as training set
- ASLTEXT dataset
  - subset of ASL Lexicon Video Dataset from Boston University
  - 250 unique gestures. 1598 videos out of which we utilize 1200 videos of 190 gestures not in the IMPACT dataset.
  - Used as test set

Evaluation on ASLTEXT dataset

Groups	Labels
G1	AHEAD,AVERAGE,BOY,CAN,EMBARRASS,EMPHASIZE,FAMILY,FREE,FRIDAY,GHOST,HOW-MANYORMANY,INTRODUCE,MACHINE,MATCH,PASS,SET-UP
G2	AFRAID,AVOIDORFALL-BEHIND,MAD,PROCEED,LIVE,SAUSAGEORHOT-DOG,BANANA,CHAINOROLYMPICS,CHASE,COAT,EARTH,FAR,FENCE,FREEZE,LUNGS,TAKE-UP
G3	ACT,APPLE,BICYCLE,BOSS,BUT,COMB,DESTROY,DRESSORCLOTHES,FOLLOW,MEAT,MEET,METAL,RUN-OUT,DISCONNECT,CAR,DEAF
G4	ANY,CENTER,COUNTRY,CRUEL,EVERYDAY,FINALLY,GREEN,HELLO,BLAME,OVERORAFTER
G5	ASSOCIATION,COME-ON,COOPERATEORUNITE,GOVERNMENT,GRAB-CHANCE,GRASS,HOSPITAL,MAKE,MORNING,MOST,ONE-MONTH,SKIN,STRONG,DEPOSIT,LETTERORMAIL,MESSED-UP,COURT
G6	APPOINTMENT,ARRIVE,COLLECT,DECIDE,DRY,ENGAGEMENT,EXACT,FOOTBALL,GAMBLE,HALLOWEEN,LIPOR MOUTH,PRICE,SHAPEORSTATUE,INCLUDEORINVOLVE,DISAPPOINT,DRUNK,MERGEORMAINSTREAM
G7	BREAD,COUGH,COURSE,CRUSH,DISAPPEAR,EXPENSIVE,GASORGAS-UP,GIRL,IDEA,INSULT,INSURANCEORINFECTION,LIBRARY,MAGAZINE,ONE,WHERE,BRAVEORRECOVER,BAD,BO RE,BREAK-DOWN,CHERISH,DIVORCE,FORGET,FRIEND,GONE,GROW,LEFT,MOSQUITO,PROTEST
G8	BAR,HEAD-COLD,HELMET,ILLEGAL,COLD,GOAL
G9	ALONE,BAWL-OUT,BLACK,EXPLAIN,HARD,NOT-MIND,CANNOT,EAST,GRANDFATHER,GRANDMOTHER,HEAD,HEAVY,PAINT,WORK-OUT,AGAIN,FLY-BY-PLANE,MISSORASSUME,NICEORCLEAN,SHAME,ARTORDESIGN,A-LOT,CONFLICTORINTERSECTION
G10	ANSWER,EXPERT,CANCELORCRITICIZE,ACCEPT,ADVISEORINFLUENCE,AUTUMN,BEAUTIFUL,BLUE,CALL-BY-PHONE,CELEBRATE,DARK,DIRTY,DISMISS,DOWN,EAT,EXPERIENCE,EXPERIMENT,FED-UPORFULL,FULL,GENERAL,GENERATION,GET-UP,GRADUATE,HAPPEN,HAVE,HIT,HOME,INFORM,INJECT,LEARN,LESS-THAN,LIE,LINE,MEMBER,MONDAY,NAB,PULL,REALLY,SAME-OLD,SILLY,TO-FOOL,TRASHORBAG



- Overall normalized accuracy of 66% out of 1200 videos for ASLTEXT
- Closest state of the art using 3D-CNN reports 51.4%
  - While utilizing part of ASLTEXT as training set

Conclusion

- Defined canonical form representation of gestures to use for Zero-Shot Learning
- Surprisingly robust to changes in location, new users, settings, camera positions
- Developed an ensemble system that recognize novel unseen gestures

Thank you for your time and consideration.