Generative Adversarial Approach for HAR Zero-Shot Learning

Mat Ruiz
S’22 ECE 209AS Final Proj

Project Goals

- Zero-Shot Learning for HAR: train a HAR classifier to detect activities that are not present in an IMU training dataset
- Apply ZSL Generative approach to HAR, which has yet to be attempted

Background: Zero Shot Learning (ZSL)

Goal: Train a classifier that observes samples from classes not present in training, and be able to predict what class they belong to

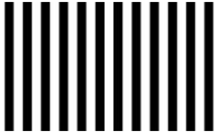
How? 1) Use auxiliary information from seen and unseen classes to create an embedding that encodes class relationships

2) Map training data to embedding space

Background: Zero Shot Learning (ZSL)

<u>Classic example:</u> Text (auxiliary space) for
Image (classification space)

CLASSIFICATION SPACE



AUXILIARY SPACE

“zebra” = “horse” + “w/b stripes”

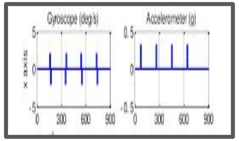
Zero Shot Learner

label = “zebra”

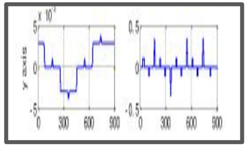
Background: Zero Shot Learning for HAR

Our use case: Video and/or text (auxiliary space) for
IMU (classification space)

CLASSIFICATION SPACE

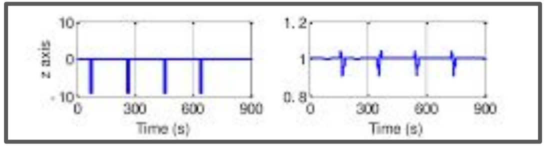


label = “sitting”



label = “looking up”

Train ZSL



Zero
Shot
Learner

classified label = “lying down”

AUXILIARY SPACE (text)

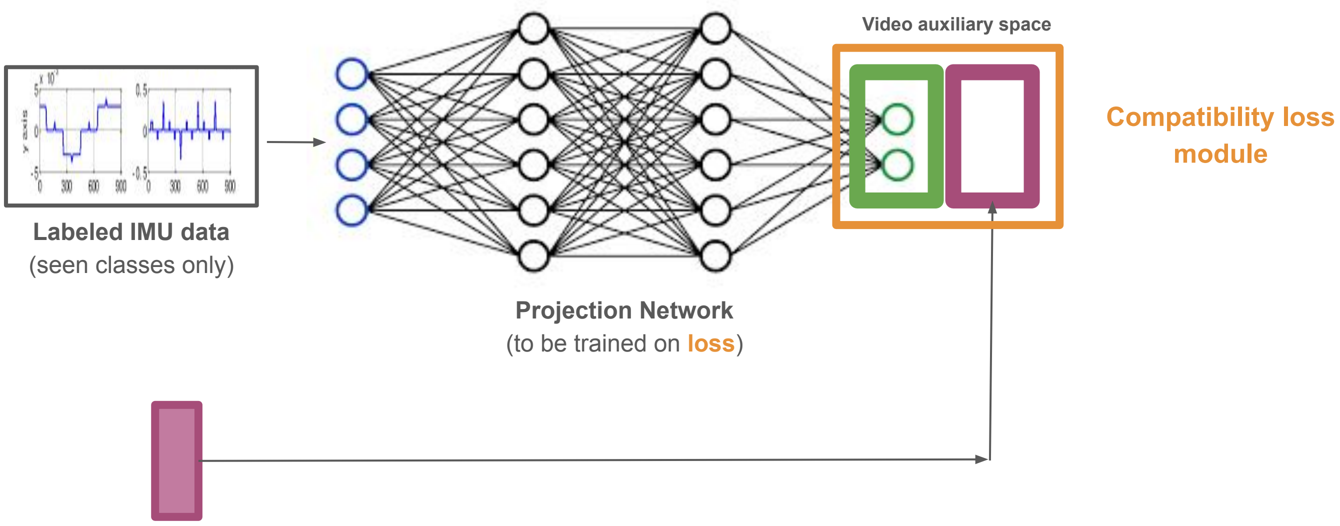
“lying down” =
“sitting” +
“looking up”

Existing ZSL Approaches for HAR

- SVMs to predict new classes based on manually crafted binary auxiliary space
  - Cheng et al., 2013
  - Limitations: costly & non scalable to manually craft auxiliary space for new classes
- Word embeddings as an auxiliary space, and MLPs
  - Matsuki et al., 2019
  - Word embeddings not robust on creating relationships between certain activity classes
- Video-based auxiliary space
  - Tong and Ge et al., 2021
  - Learn to project seen IMU data onto video-based representations of classes
  - *… more info next slide*

Existing ZSL Approaches for HAR: Tong and Ge, 2021

ZSL Embeddings-based Framework

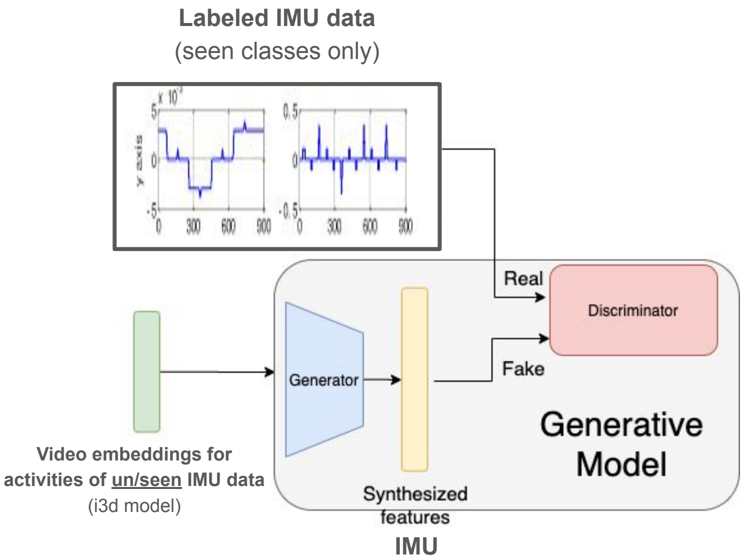


One limitation: training only utilizes video embeddings corresponding to seen IMU data classes

Importance of my Approach

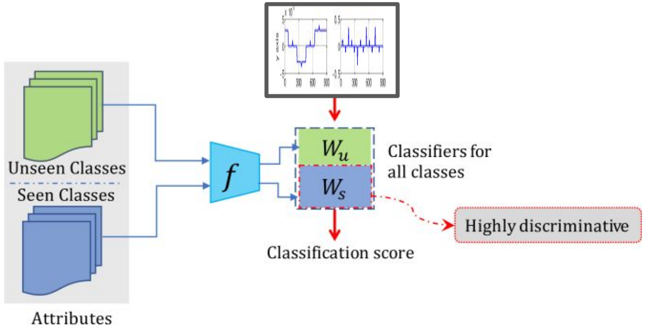
- Utilize auxiliary information of unseen classes during training
  - ones that aren't present in IMU dataset
- via <u>ZSL Generative based Framework:</u>
  - Proven better than ZSL Embeddings-based approach in Image-Text domains via [Zhu et al., 2018]

1) Train GAN



2) Generate IMU data of unseen & seen classes

- Generate IMU data based on I3D video attributes
  - Seen AND unseen classes
- Train classifier (KNN) on generated IMU data
- Use classifier on real-world IMU data

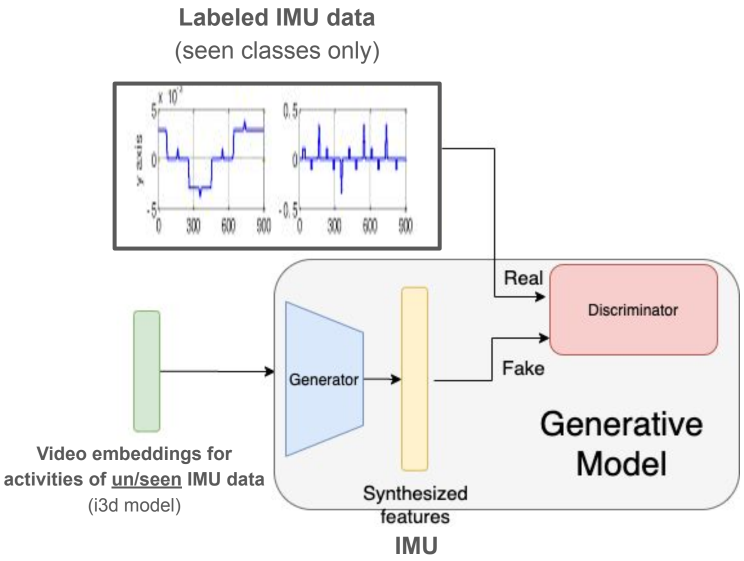


Data

- PAMAP for **IMU data**
  - 18 different activity classes
  - 3 IMU sensors: ankle, chest, wrist
    - 3 axial Accelerometer (XYZ), 3 Axial Gyroscope (XYZ)
    - For each sliding window, calculated **mean** and **std** for each instrumental axial
    - 36 features per window/instance
- I3D Video Activity Recognition Model for **video auxiliary space**
  - Gathered 10 videos for each of the 18 PAMAP classes
  - Passed videos through I3D, and created mean prototypes of each activity in video space
    - 400 video feature vector
    - Total of 18 vectors (one per each class)

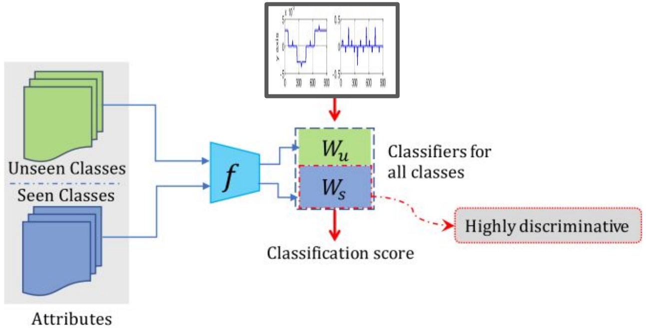
My Approach

1) Train GAN



2) Generate IMU data of unseen & seen classes

- Generate IMU data based on I3D video attributes
  - Seen AND unseen classes
- Train classifier (KNN) on generated IMU data
- Use classifier on real-world IMU data



Evaluation + Results

- Used k-fold evaluation
  - Divided 18 activity classes into 5 folds similar like Tong and Ge et al., 2021
    - Custom mix of “static” and “dynamic” activities in each fold
  - 4 folds used as “seen” data, 1 evaluation fold acts as “unseen” data
- Tong and Ge et al., 2021’s KFOLD acc: 56.4%
- My approach acc: 48.5%

Limitations

- Model didn’t perform well on eval folds that included “sitting”, “standing
    - Guess: poor video representations for these classes; classes not present in I3D model
- This approach of ZSL (GAN) takes longer to train, compared to Tong and Ge et al., 2021 method
    - Limited time for hyperparameter and model architecture selection/testing
- GAN model not as lightweight as Tong and Ge et al., 2021 MLP projection model

Future Explorations

- Experiment combining I3D video features with IMUTube data of videos, to create potentially more robust auxiliary space
- Explore other generative models other than GAN
- Does dataset classes affect learning, how much?
    - If class labels are very distinct, would a dataset with a lot of classes(100) degrade performance or make it easier for adding new classes?
- Add contrastive learning to end of pipeline to make classification on unseen classes more robust