Demo Project Presentation

PRESENTED TO

Professor Ouyang

PRESENTED BY

Nuanyu Cao

Content

Challenges Observed

- Unaccessible original image data

(Main challenge)

Processed to low resolution images

(Apply VLMsMMs)

Extracted to skeleton data

(Apply DL models)

- Insufficient original data

(Main challenge)

Apply defusion model

Challenges in details

- data preprocessing-- increase details
- data preprocessing-- from pose to continuous frames
- Model Accuracy
- Alignment with human perceprtion

Brief Summery

	Brief description	Work done
[1] Automated Analysis of Stereotypical Movements in Videos of Children With ASD	Own model of transferring video data(not provided) into skeleton datasets. A corresponding model to evaluate.	Replicated the method starting from the model evaluation part. Tried to use ST-Transformer but ended in overfitting due to the small dataset
[2] MMASD: A Multimodal Dataset for Autism Intervention Analysis	No model contained, just 3 types of datasets. 2d, 3d, skeleton datasets in “npz” form and continuous photos in low-resolution “png” form.	Transferred the skeleton datasets into pkl forms-- from scattered x,y,z data to continuous frame data. Compared the performance of 32, 64 frames-- chose 64 in the end.

Brief Summery

[1] Automated Analysis of Stereotypical Movements in Videos of Children With ASD

[2] MMASD: A Multimodal Dataset for Autism Intervention Analysis

Meanings

- Dataset 1's idea could be applied to evaluate whether a certain model's prediction fits the reality-- or human perception.
- Works done for the second dataset proved the reliability of using the ST-transformer model
- Combining both ideas could help evaluate the correctness and feasibility of defusion models

Works can be done in the future (chanllenges)

(Leftovers from the second dataset + some reserach)

- Try to preprocess dataset 2's "png" dataset in order to apply VLMs.
- A dataset named MMASD+ may be of help to the work.
- Train appropriate defusion models

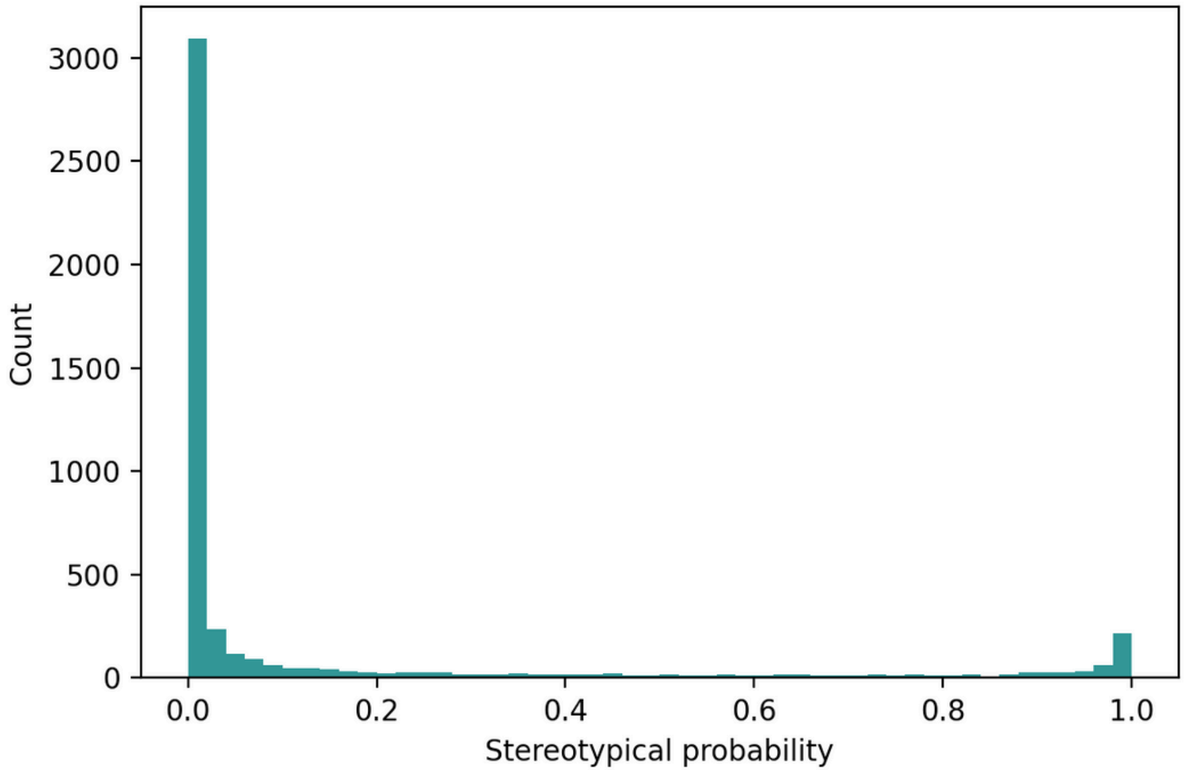
First Dataset

The model named ASDPose in the paper, essentially PoseC3D within the MMAction2 framework, is a binary classifier trained on skeleton datasets, which takes skeleton sequence clips as input to output SMM (Stereotypical Motor Movement) probabilities (0–1) for identifying whether there is SMM.

Before this process, the model uses “OpenPose” to transform video data into skeleton data, which is given in the dataset. (But without the original set I did not replicate this part).

In this way, the model will be able to diagnose the Children’s ASD problem.

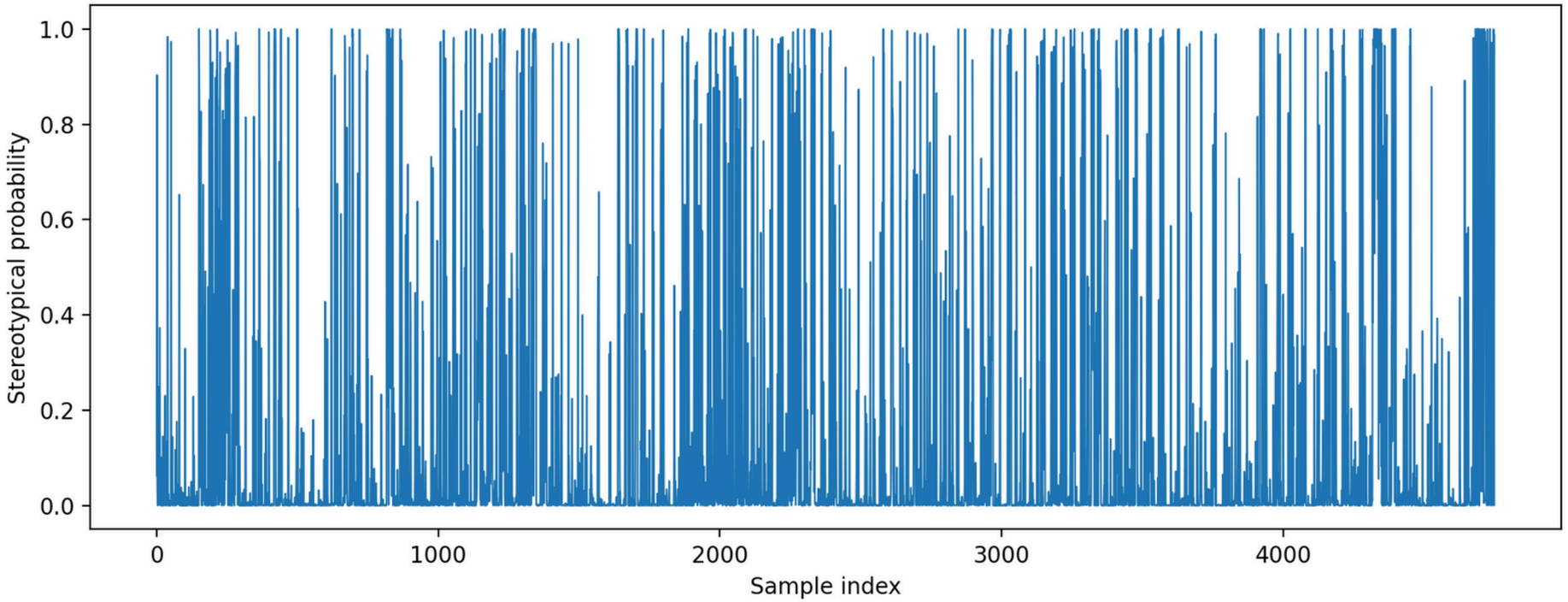
First Dataset



- around 10%-20% abnormal activities (in this graph, SMM stereotypical).

- Each frame’s probability detection squeezes to either 0.0 or 1.0

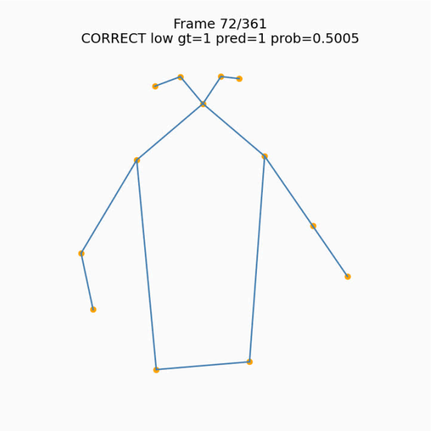
First Dataset

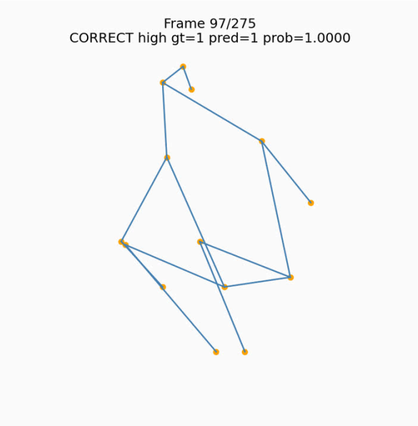


- No obvious patterns – randomly scattered from human's understanding.

First Dataset

I also did some image (only including skeleton data) rendering in gif





First Dataset

Transformer model overfit

```
Using device: cuda
Start Training...
Training: 100%|██████████| 1006/1006 [00:10<00:00, 94.89it/s, acc=0.775, loss=0.319]
Epoch [1/10] Loss: 0.4933 | Train Acc: 0.7755
Training: 100%|██████████| 1006/1006 [00:10<00:00, 98.96it/s, acc=0.776, loss=0.524]
Epoch [2/10] Loss: 0.4887 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 99.05it/s, acc=0.776, loss=0.457]
Epoch [3/10] Loss: 0.5013 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 98.09it/s, acc=0.776, loss=0.384]
Epoch [4/10] Loss: 0.4913 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:09<00:00, 100.79it/s, acc=0.776, loss=0.428]
Epoch [5/10] Loss: 0.4899 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 97.40it/s, acc=0.776, loss=0.619]
Epoch [6/10] Loss: 0.5235 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 98.11it/s, acc=0.776, loss=0.479]
Epoch [7/10] Loss: 0.5330 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 98.77it/s, acc=0.776, loss=0.675]
Epoch [8/10] Loss: 0.5326 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:09<00:00, 100.98it/s, acc=0.776, loss=0.709]
Epoch [9/10] Loss: 0.5328 | Train Acc: 0.7759
Training: 100%|██████████| 1006/1006 [00:10<00:00, 99.21it/s, acc=0.776, loss=0.268]
Epoch [10/10] Loss: 0.5325 | Train Acc: 0.7759
```

[ ]:

Second Dataset-- Data Pre-processing

- Core Objectives：
  - Prepare data for 2D/3D skeleton action classification: fix raw data issues (inconsistent formats, mixed modalities) and screen valid data to boost training efficiency.
- Key Methods：
  - Standard environment (PyTorch, etc.) and parameters (64 frames, 71 joints); verify paths.
  - Map theme to labels, and keep samples with both 2D/3D data.
  - Use custom tools to preprocess 2D/3D separately (unify sequences, normalize 3D, filter anomalies).
  - Export pure 2D/3D data; build PyTorch datasets (8:2 train/test split) and DataLoaders.
- Achievements:
- Valid samples for 13 Themes + clear label system.
- Standardized datasets (2D: 32/64×71×2; 3D: 32/64×71×3).
- Independent 2D/3D train/test sets + model-ready DataLoaders.
- No repeated processing; real-time memory cleaning for stability.

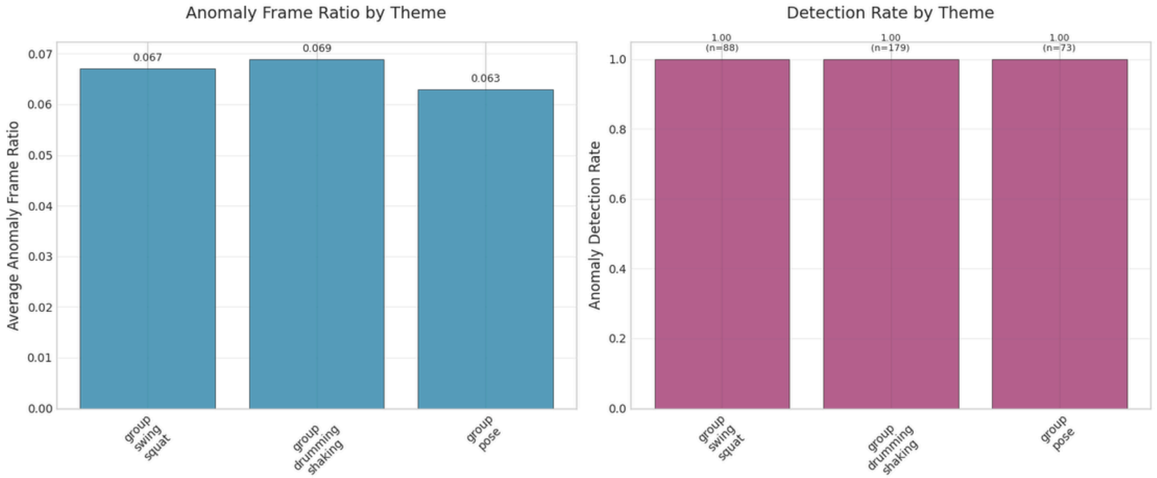
Second Dataset-- ST-Transformer training

- Objective： Develop a time-series anomaly detection model (e.g., for abnormal limb movement recognition) by training on fully mixed datasets (with stratified cross-validation for unbiased evaluation), and compare performance with Fine-tuning.

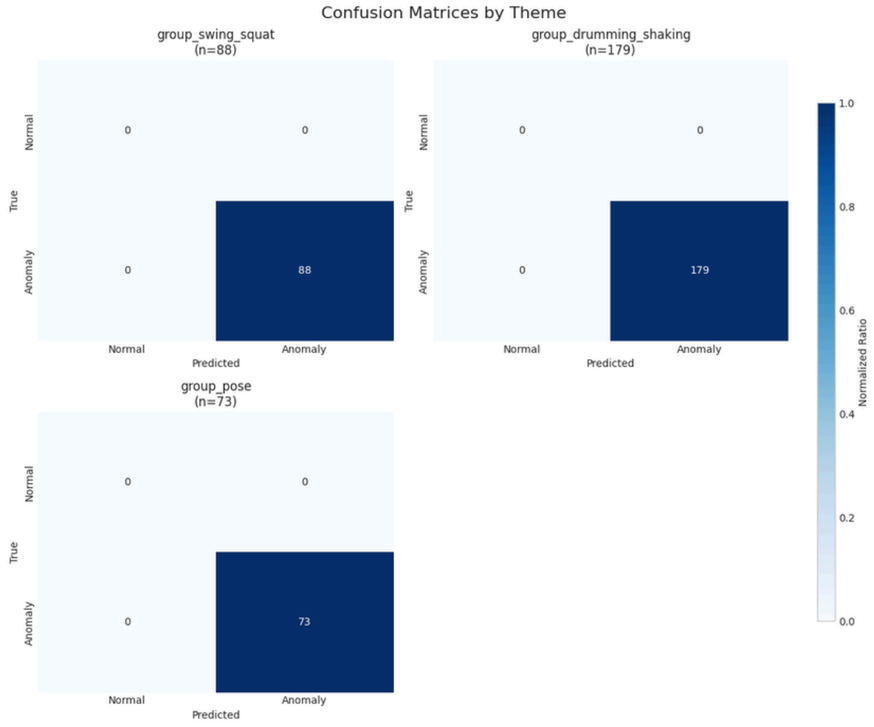
Stage	Core Actions
Data Preprocessing	Merge multi-source time-series data → shuffle → standardize features → split features/labels (0=normal, 1=abnormal)
Cross-Validation	5-fold stratified sampling (preserve normal/abnormal ratio) → train time-series Transformer per fold → track loss/accuracy/anomaly ratio
Final Training	Train on full data with optimized hyperparameters → dynamic learning rate → save model + meta-info

- Key Advantages
  - Full-data utilization + stratified validation (no evaluation bias)
  - Time-series optimized model + anomaly detection-specific loss/metrics
  - Modular design (easily adaptable to custom models/hyperparameters)

Second Dataset-- ST-Transformer training



All data is to be detected as normal.
Frames are abnormal of small ratio.



Second Dataset-- ST-Transformer training

```
group_swing_squat 微调完成!
最优验证损失: 0.007192 | 最优验证准确率: 0.9329
最终训练准确率: 0.9125 | 最终验证准确率: 0.9329
模型保存路径: st_transformer_action_groups_ft/st_transformer_group_swing_squat_finetune_best.pth

group_drumming_shaking 微调完成!
最优验证损失: 0.013368 | 最优验证准确率: 0.9340
最终训练准确率: 0.9071 | 最终验证准确率: 0.9310
模型保存路径: st_transformer_action_groups_ft/st_transformer_group_drumming_shaking_finetune_best.pth

group_pose 微调完成!
最优验证损失: 0.016033 | 最优验证准确率: 0.9375
最终训练准确率: 0.9126 | 最终验证准确率: 0.9371
模型保存路径: st_transformer_action_groups_ft/st_transformer_group_pose_finetune_best.pth
```

Results are better after fine-tuning

```
预训练完成! 最优验证损失: 0.012108 | 最优验证准确率: 0.9261
```

Furture Works

Addressing Core Field Challenges

- Target Issues:
- (1) High dependence on natural data vs. scarce available data and public datasets;
- (2) Image data blurring in public datasets due to privacy protection.
- Synergistic Approach: Combine the diffusion model (for data augmentation) and optimised visual model (for low-frame-rate data) to alleviate data scarcity and poor visual quality, forming an integrated solution for the field’s practical pain points.

Furture Works

1. Development of Diffusion Model with Dual-Dataset Validation

- Current Foundation: Completed skeleton data processing for two datasets; summarised 3 evaluation criteria (derived from the first dataset's test validation) to verify if model predictions align with human cognition (successfully applied to the second dataset's ST-Transformer performance testing).
- Objective: Develop a diffusion model targeting the first dataset to address the issue of insufficient experimental data (a key challenge in the field: high reliance on natural data but limited data volume and public datasets).
- Validation Mechanism: Use the 3 human cognition-aligned evaluation criteria (from the first dataset) to validate the diffusion model's outputs, ensuring consistency between the new model and existing valid standards.

Furture Works

2. Optimisation of Visual Model for MMASD Dataset (Second Dataset)

- Current Challenge: The existing visual model for the second dataset (MMASD) shows unsatisfactory performance, partly due to low frame rates of the data.
- Reference Insight: Identified the MMASD+ dataset (with similar data acquisition methods to MMASD but non-public access), which proposes trainable models for visual tasks.
- Implementation Direction: Research and adapt the visual model training strategies and preprocessing techniques from MMASD+ to the second dataset, focusing on mitigating inaccuracies caused by low frame rates.

Thank you