FIU-UM AT TRECVID2021

Weakly-Supervised Deep Neural Networks with Label Engineering for the DSDI task

Maria Presa-Reyes[1], Yudong Tao[2], Erik Coltey[1], Tianyi Wang[1], Rui Ma[2], Shu-Ching Chen[1], and Mei-Ling Shyu[2]

[1]Knight Foundation School of Computing and Information Sciences, Florida International University

[2]Department of Electrical and Computer Engineering, University of Miami

Presenter:
Maria Presa-Reyes
Email: mpres029@cs.fiu.edu

AGENDA

Submission Details

Introduction

Proposed Framework

Performance

Conclusion & Future Work

SUBMISSION DETAILS

Training Type: LADI-Based

Team ID: FIU-UM (Florida International University – University of Miami)

Year: 2021

Submission	Description
run1	fully automated feature score fusion through differential evolution
run2	mean aggregation of the predicted scores from the best performing models in the ensemble
run3	fully automated feature score fusion with z-score normalization and averaged z-scores

SUBMISSION DETAILS (CONT.)

Training Type: LADI + Others (O)

Team ID: FIU-UM (Florida International University – University of Miami)

Year: 2021

Submission	Description
run1	fully automated feature score fusion through differential evolution
run2	further enhancement of the feature score fusion by the removal of the less relevant feature scores as determined by the differential evolution
run3	fully automated feature score fusion with z-score normalization and averaged z-scores
run4	applied the team’s best performing model used to rank videos in TRECVID2020-DSDI [1] [2]

AGENDA

- Submission Details
- Introduction
- Proposed Framework
- Performance
- Conclusion & Future Work

INTRODUCTION

DSDI track: Pilot challenge introducing a new dataset (LADI) emphasizes disaster-related features, i.e., damage labels and scene descriptors

Testing dataset: 2,801 video shots with a total duration of around 6.7 hours

Features: 32 features divided by 5 broader categories (damage, environment, infrastructure, vehicles, and water)

Results: A maximum of 1000 possible shots for each of the 32 selected features

INTRODUCTION (CONT.)

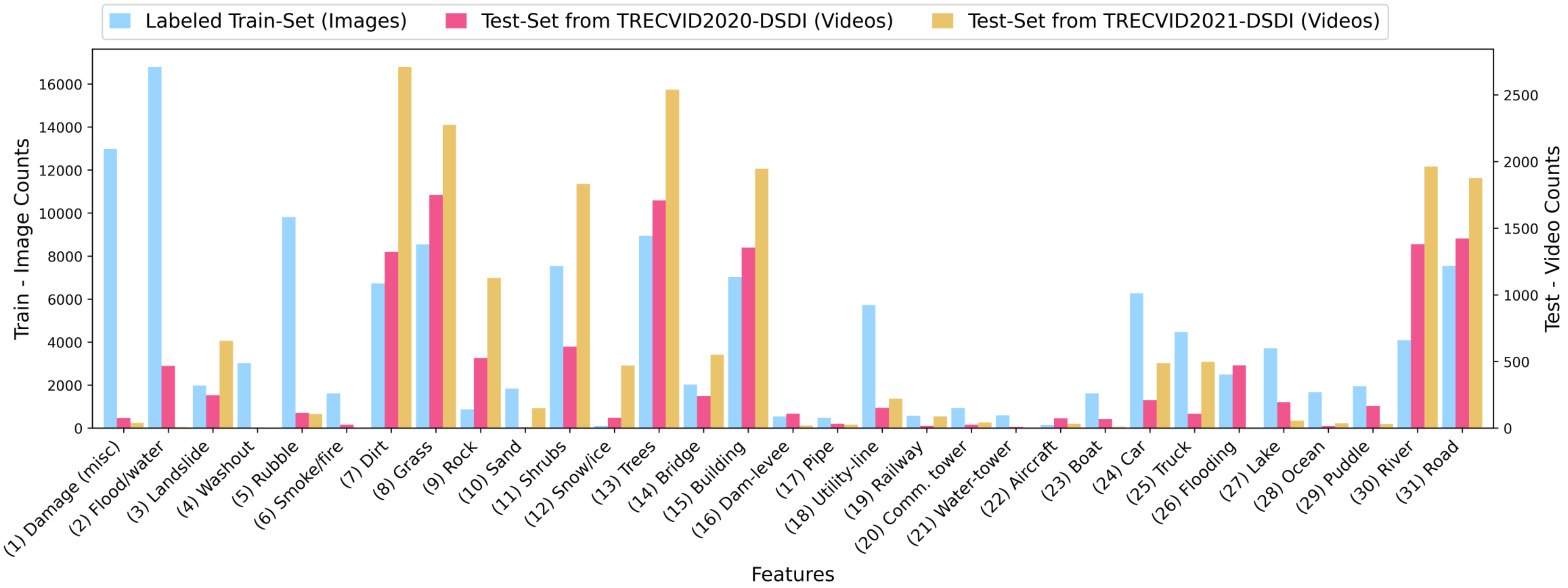
LADI (Low Altitude Disaster Imagery) dataset

- 500k total images (6-7% are manually annotated)
- <u>Metadata</u>: timestamp, location, camera details, altitude, etc.

Challenges unique to LADI

- Class Imbalance
- Limited/Noisy Labels
- View of objects/scene from low-altitudes

TARGET FEATURE DISTRIBUTION



AGENDA

- Submission Details
- Introduction
- Proposed Framework
- Performance
- Conclusion & Future Work

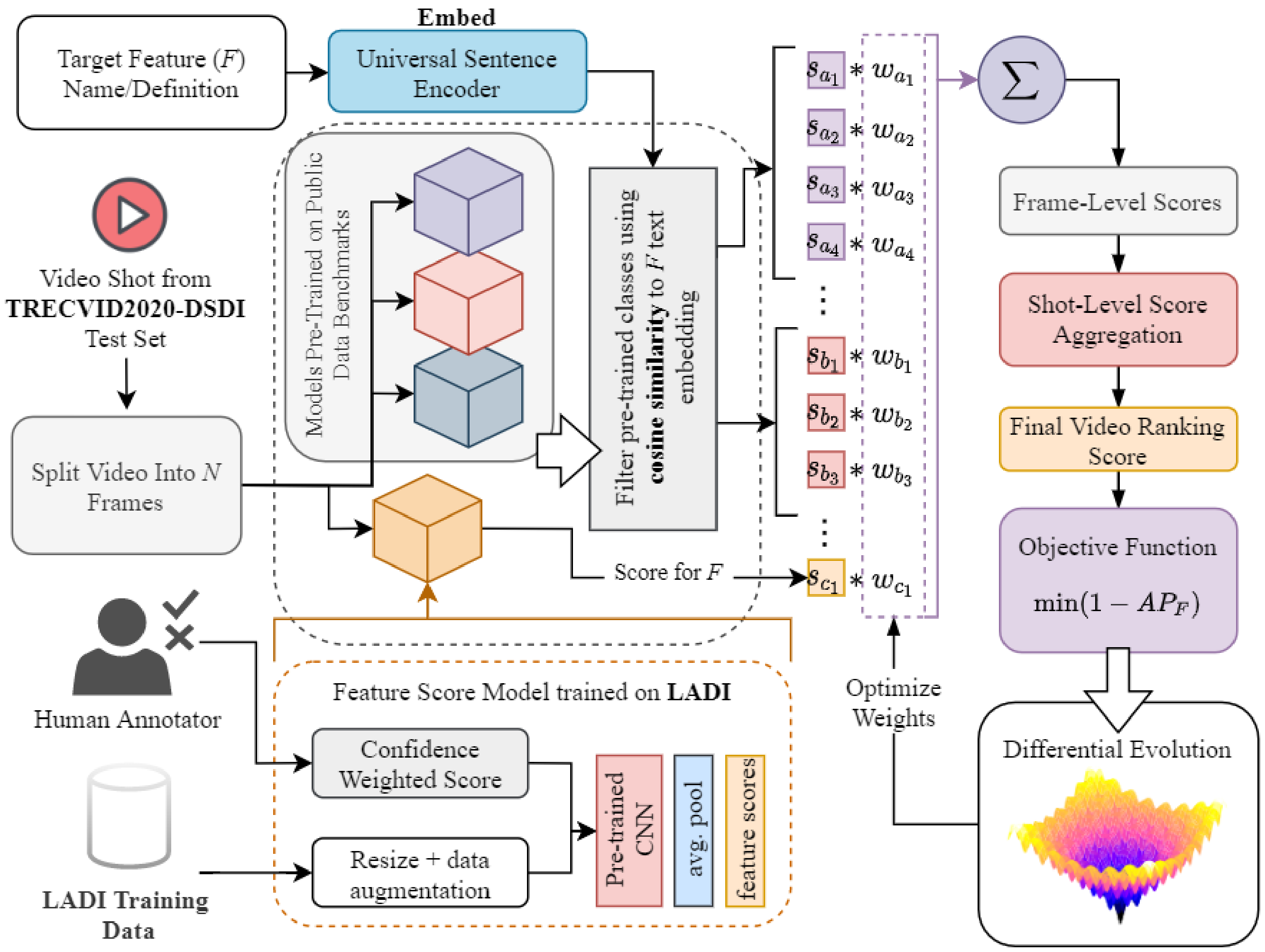
PROPOSED FRAMEWORK

(1) Feature Score Model trained on LADI using Confidence Learning approach

(2) Match Target Feature to pre-trained classes, including:

- Incidents Dataset
- MS COCO
- Places365
- ImageNet21k

(3) Optimize the weighted Frame-Level Scores aggregation using **Differential Evolution**

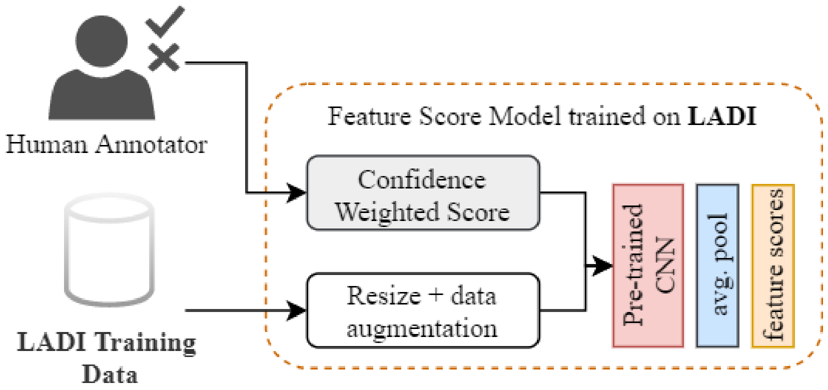


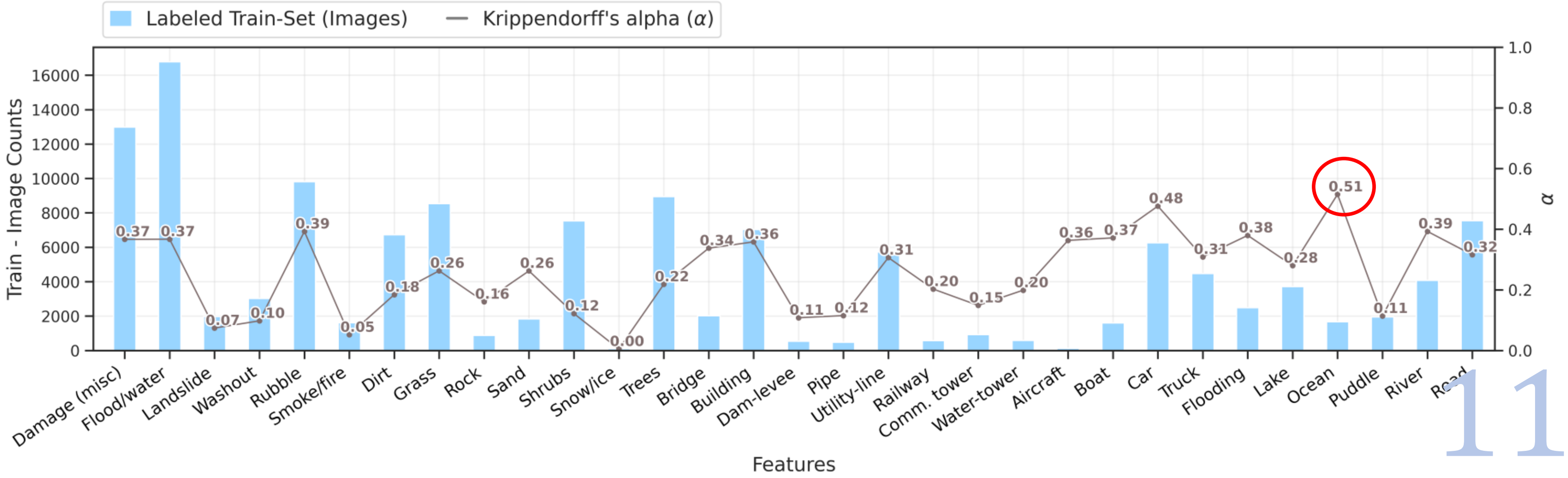
CONFIDENT LEARNING

Objective: Denoise the training labeled-set to improve reliability

- What is the extent of the mislabels?
- Krippendorff's Alpha [5] Reliability Estimate or Inter-Rater Agreement
- Example:

1	0	…
NA	0	…
…	…	…





CONFIDENT LEARNING (CONT.)

(1) Calculate the soft-labels weighted on majority-vote

- Assumption: each image may have multiple annotations from multiple workers
- Compute an initial soft-label based on the majority vote

(2) Calculate Confident Scores

- Randomly split the data into 5 folds
- Generate out-of-sample predicted probabilities
- Focus on label quality rather than quantity

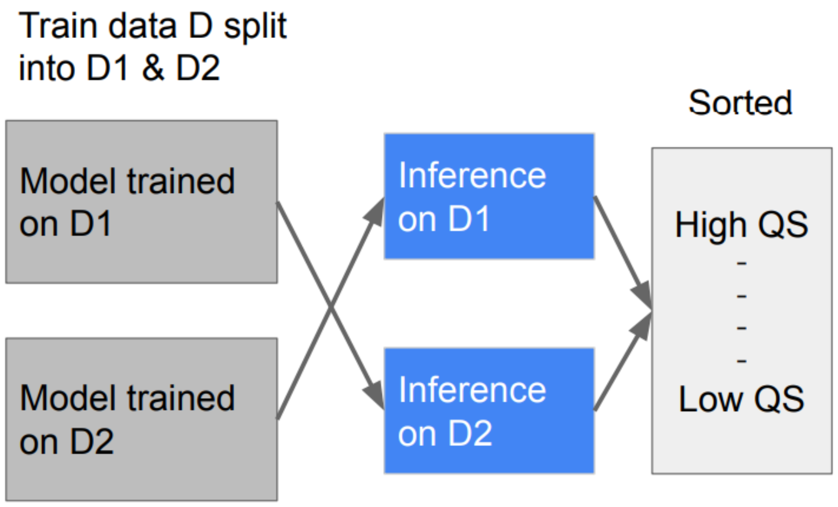


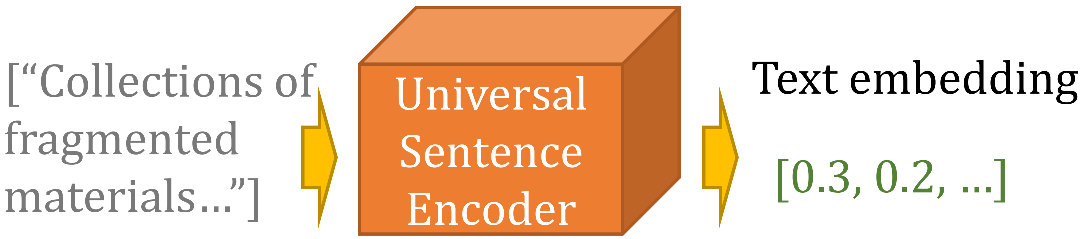
Figure 1: Schematic of stratified noisy cross-validation (SNCV).

Figure retrieved from reference [6]

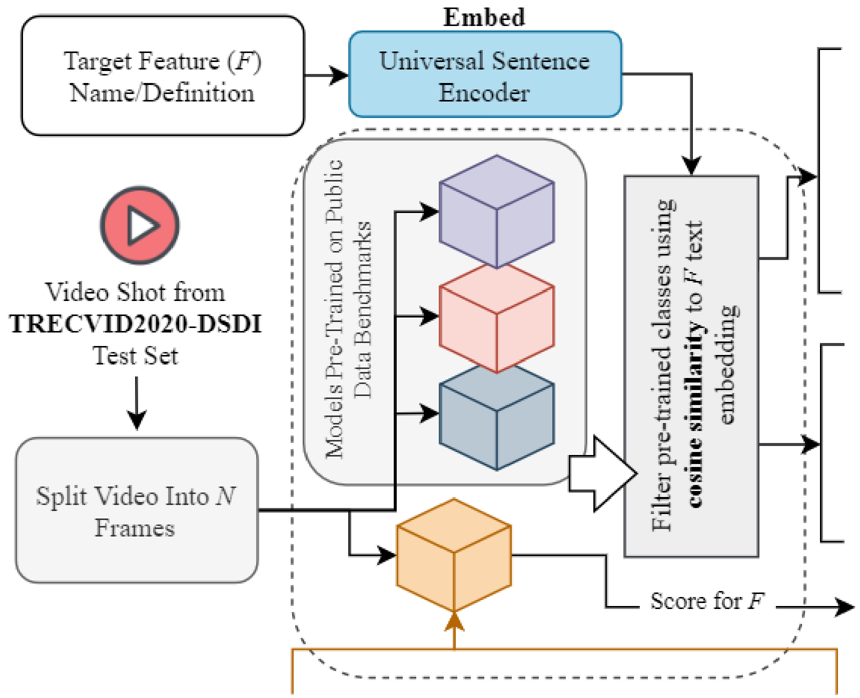
FEATURE MATCHING

Objective: Filter the classes to select the most relevant to the Target Feature (*F*)

(1) Embed the text from name/definition of *F*



(2) Select most relevant classes to *F* based on the **cosine similarity** of the embeddings



LADI:rubble, damaged, heavy rainfall

LADI:utility line, utility room, power line

LADI:road, railroad track, desert road, street, forest road highway, field road

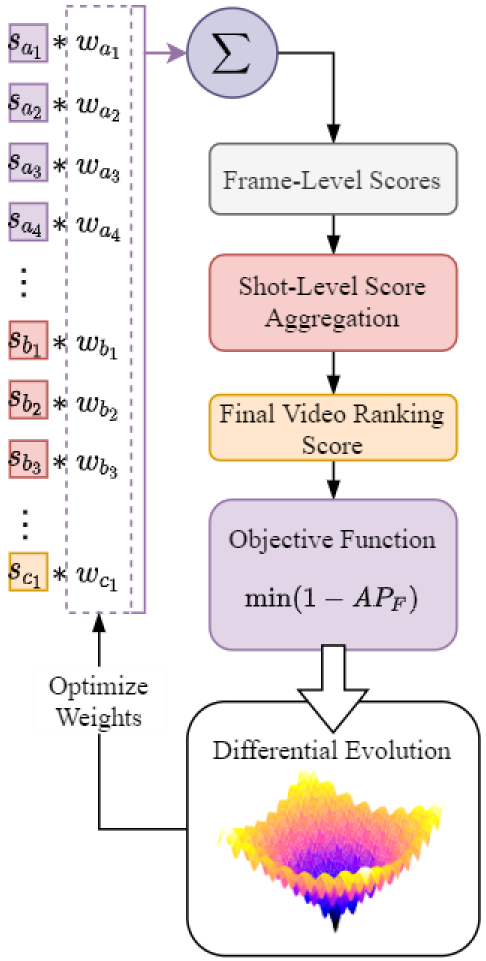
FEATURE FUSION

Objective: Find the most optimal aggregation of the SoftMax scores to generate the Frame-Level Score

Differential Evolution

- Population-based metaheuristic search method

- Final score is generated from the optimum combination of multiple scores



LADI:rubble, damaged, **heavy rainfall**

LADI:utility line, **utility room**, **power line**

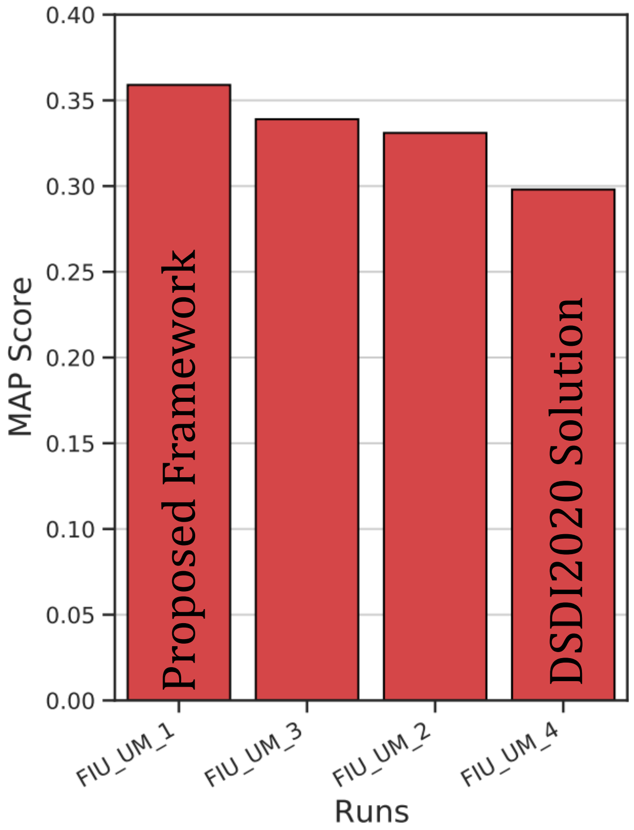
LADI:road, railroad track, desert road, **street**, forest road, highway, **field road**

AGENDA

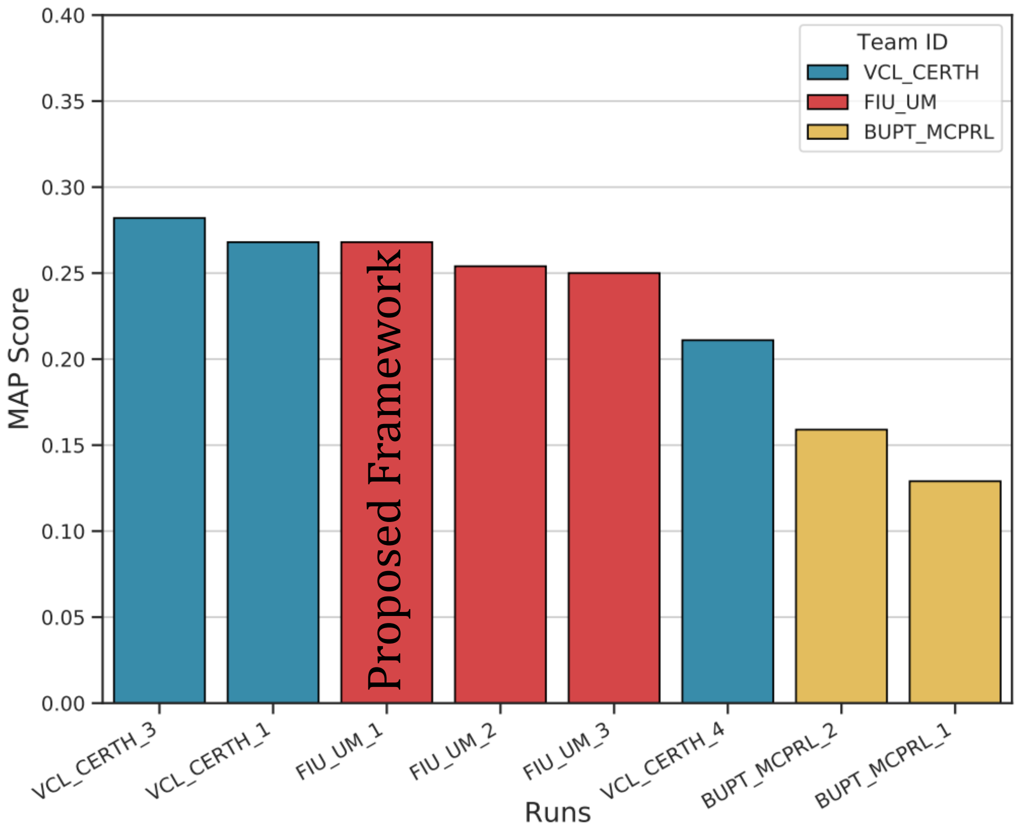
- Submission Details
- Introduction
- Proposed Framework
- Performance
- Conclusion & Future Work

PERFORMANCE: OVERALL MAP SCORE

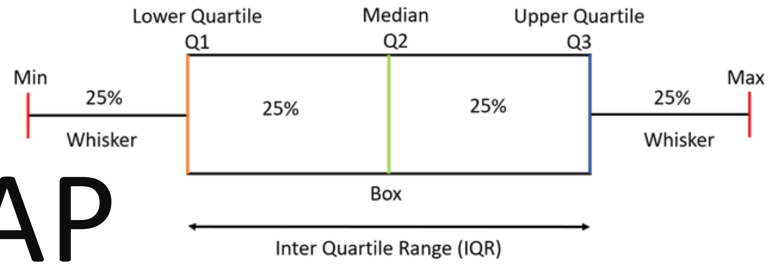
LADI + Others (O)

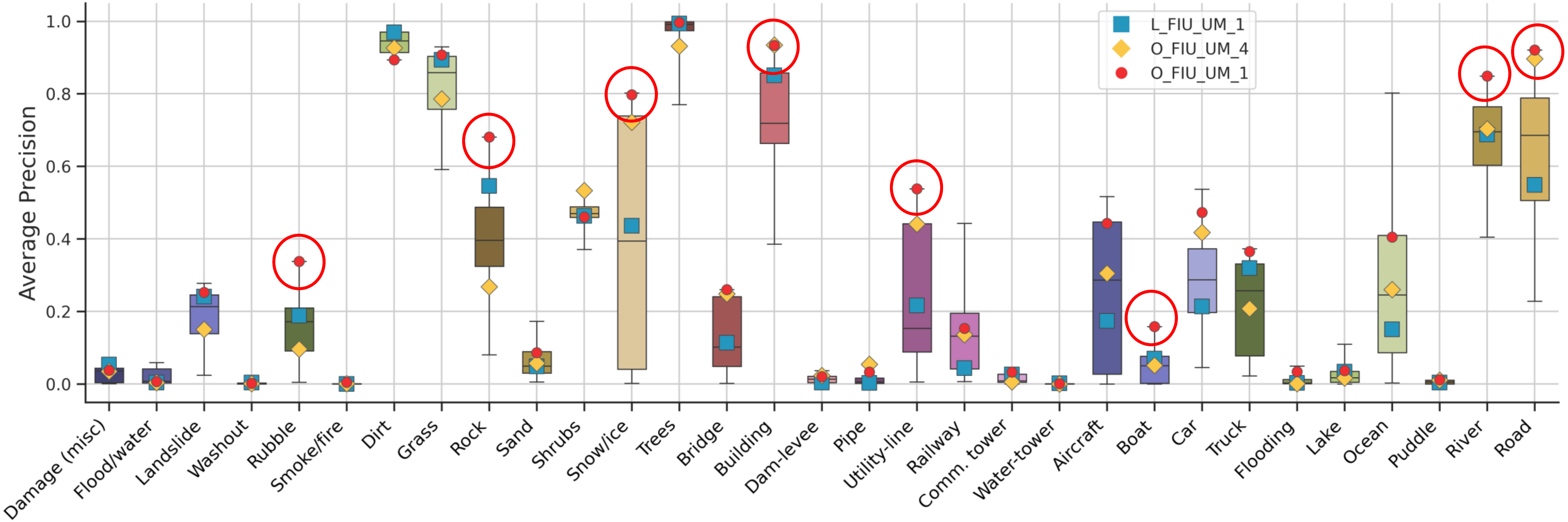


LADI-based (L)



PERFORMANCE: FEATURE AP





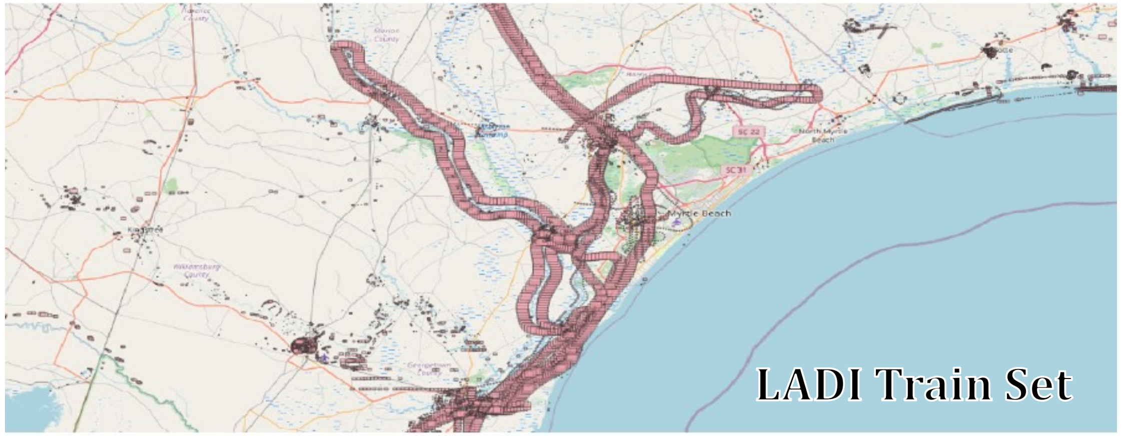
- Boxplot distribution for the precision score of a feature is compared across <u>all</u> submissions
- L_FIU_UM_1: TRECVID-DSDI 2021 Solution trained on LADI-based
- O_FIU_UM_4: TRECVID-DSDI 2020 Solution trained on LADI + Others
- O_FIU_UM_1: TRECVID-DSDI 2021 Solution trained on LADI + Others

PERFORMANCE: QUALITATIVE RESULTS

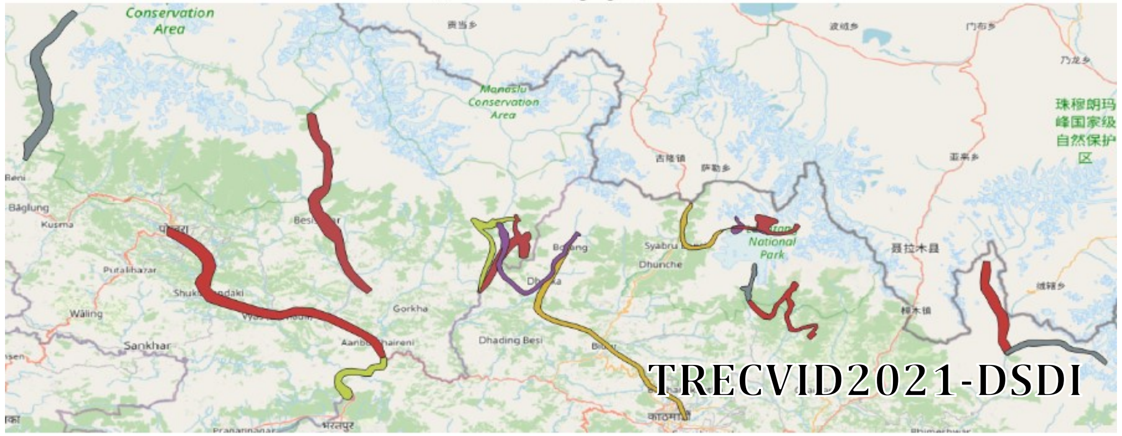
AGENDA

- Submission Details
- Introduction
- Proposed Framework
- Performance
- Conclusion & Future Work

CONCLUSION AND FUTURE WORK



(a) LADI Training Spatial Area



(b) LADI Testing Spatial Area

Conclusion

- We trained a model using a Confident Learning approach to handle the noisy labels
- Fused the scores generated by models pre-trained on datasets such as Incidents Dataset and ImageNet-21K
- Presented the best way to combine the predicted scores from multiple models using Differential Evolution
- Our technique saves a significant amount of time and resources while achieving great results

Future Work

- Consider the sequence information of the image
- Exploit the spatial knowledge from the test video

REFERENCES

[1] M. Presa-Reyes, Y. Tao, S. -C. Chen and M.-L. Shyu, "Deep Learning with Weak Supervision for Disaster Scene Description in Low-Altitude Imagery," in IEEE Transactions on Geoscience and Remote Sensing, doi:10.1109/TGRS.2021.3129443.

[2] M. Presa-Reyes, Y. Tao, S.-C. Chen, and M.-L. Shyu, “Florida international university-university of miami TRECVID2020 DSDI track,” in TRECVID. NIST, USA, 2020.

[3] J. Liu, D. Strohschein, S. Samsi, and A. Weinert, “Large scale organization and inference of an imagery dataset for public safety,” in IEEE High Performance Extreme Computing Conference, Sep. 2019, pp. 1–6.

[4] Northcutt, Curtis G., Lu Jiang and Isaac L. Chuang. “Confident Learning: Estimating Uncertainty in Dataset Labels.” J. Artif. Intell. Res. 70 (2021): 1373-1411.

[5] Krippendorff, K. Computing Krippendorff's Alpha-Reliability, 2011. Retrieved from https://repository.upenn.edu/asc_papers/43

[6] Hsu, Joy, Sonia Phene, Akinori Mitani, Jieying Luo, Naama Hammel, Jonathan Krause and Rory Sayres. “Improving Medical Annotation Quality to Decrease Labeling Burden Using Stratified Noisy Cross-Validation.” ArXiv abs/2009.10858 (2020): n. pag.

[7] Das, Swagatam and Ponnuthurai Nagaratnam Suganthan. “Differential Evolution: A Survey of the State-of-the-Art.” IEEE Transactions on Evolutionary Computation 15 (2011): 4-31.

THANK YOU!

Questions?

Presenter:
Maria Presa-Reyes
Email: mpres029@cs.fiu.edu