INFORMED DEMOCRACY: VOTING-BASED NOVELTY DETECTION FOR ACTION RECOGNITION

Karlsruhe Institute of Technology

cv:hci

BMVC 2018 Newcastle upon Tyne

British Machine Vision Conference 2018

Alina Roitberg*, Ziad Al-Halah*, Rainer Stiefelhagen

alina.roitberg@kit.edu ziad.al-halah@kit.edu

Computer Vision for Human-Computer Interaction, Karlsruhe Institute of Technology, Germany

* denotes equal contribution

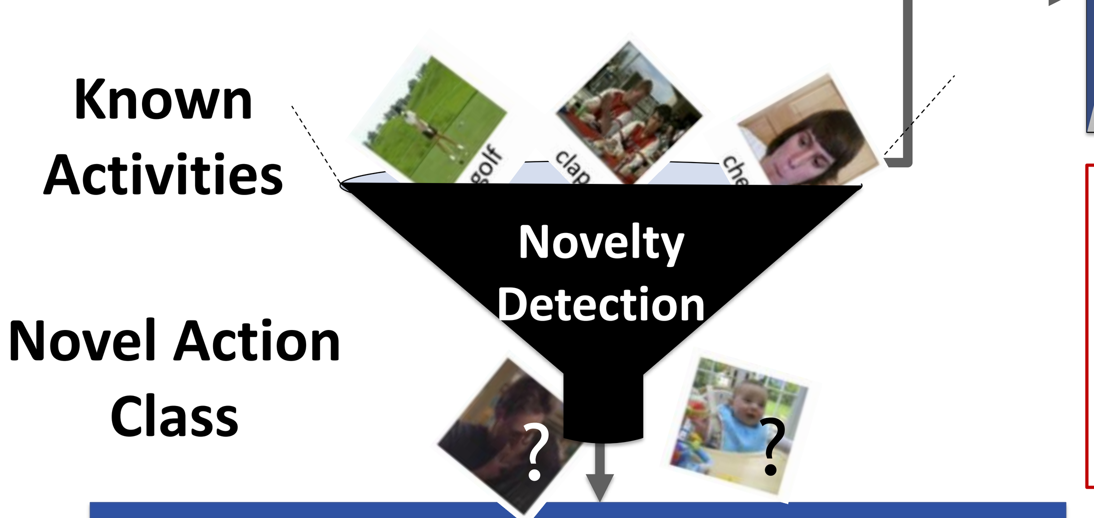
Motivation: Action Recognition in Open-Set Case

- **Goal**: Identifying actions not previously seen by the classifier (novelty detection)

Recognizing known actions

Supervised Action Recognition Models
Assume a static set of action categories

Closed-set case, cannot handle real-world scenario, where new actions can occur at any time



Generalizing to novel actions

Zero-Shot Action Recognition:
Using language models to recognize actions *without any training data*

Test set is limited to the unseen classes

Overview

Contributions

- **New model** for detecting previously **unseen action classes**
- Generic framework for zero-shot action recognition in ***generalized*** case (GZS)

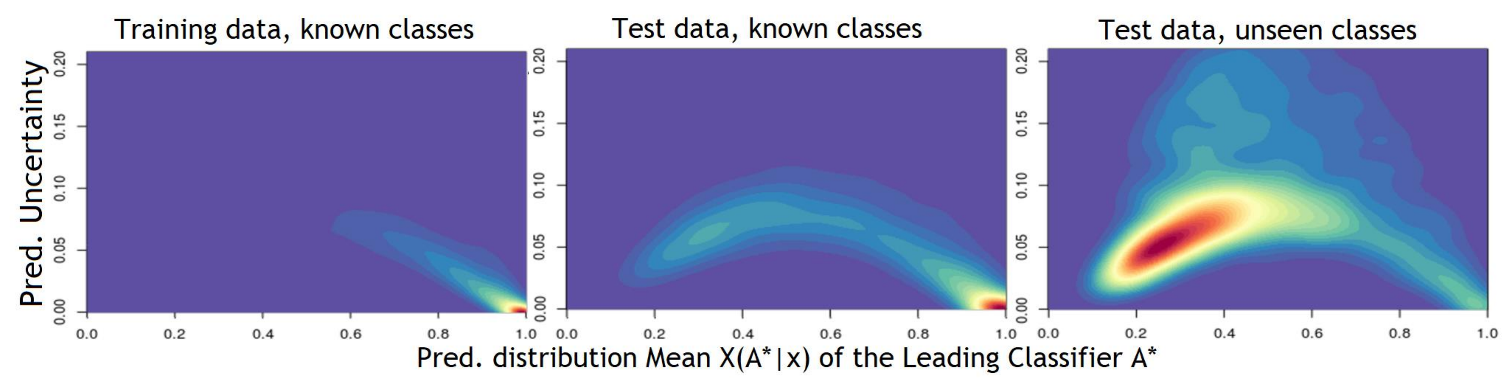
Main Idea

- Leverage the **predictive uncertainty** of the classifiers
- Two Concepts: the *Leader* and it's *Council*
- ***Leader***: the classifier with the highest confidence score (→ votes for the predicted "known" category)
- ***Council***: a selected ***subset*** of the classifiers validates the leader's decision
- **Informed Voting:** voting for novelty based on the classifiers uncertainty is **privileged to the council**

Proposed Method

Measuring Classifier Uncertainty

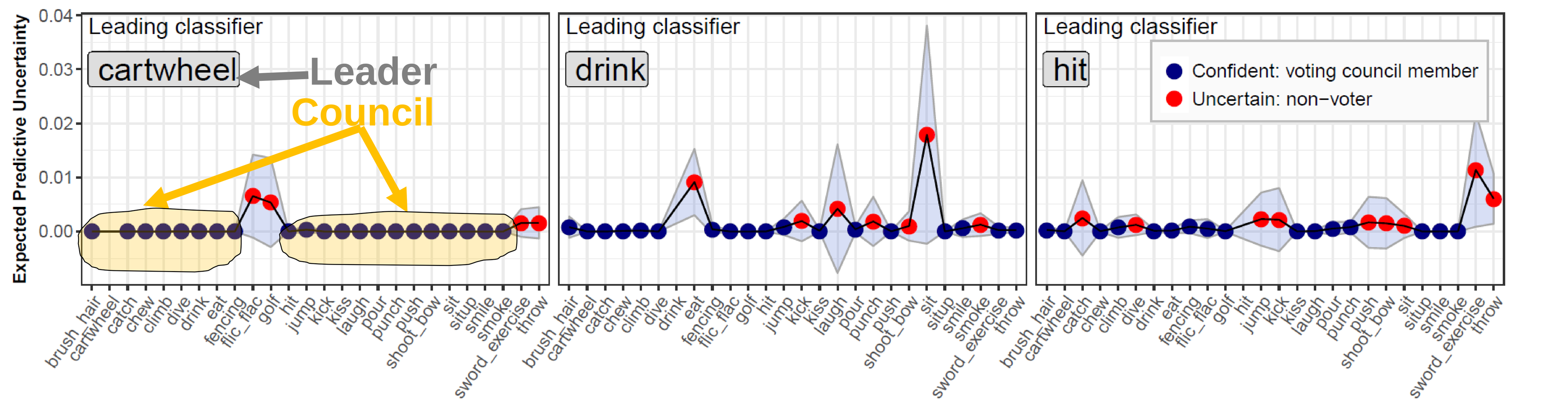
- Monte-Carlo Dropout for approximation of Bayesian Neural Network uncertainty [Gal et.al, 2016]
- Mean over M stochastic forward passes $E(A_i|\boldsymbol{x})$ instead of deterministic Softmax estimates
- Uncertainty $\mathrm{U}(A_i|\boldsymbol{x})$ is measured by the output's variance



Informed Voting for Novelty

- For the leader A^* and it's council C_{A^*}, compute novelty score $\nu(\mathbf{x})$ based on the council uncertainties $\nu(\mathbf{x}) = \frac{\sum_{A_i \in C_{A^*}} U(A_i|x)}{|C_{A^*}|}$

Council members for three different Leaders



Selecting the Leader and it's Council

- Select the **Leader**: $A^* = argmax_{A_{k \in A}} E(A_i|\boldsymbol{x})$
- Select the **Council** C_A* based the uncertainty statistics of the classifiers for the current leader A^* on a held-out set: $Var(A_j|A^*) = \frac{1}{N}\sum_{n=1}^{N}\left(U(A_j|\boldsymbol{x_n}) - E[U(A_j|\boldsymbol{x})]\right)$
- Select if $Var(A_j|A^*) < c$, where c is the *credibility constant*

Voting Scheme Variants

- **Informed Democracy:** voting is privileged to the council
- **Uninformed Democracy:** all classifiers voting
- **Dictator:** Leader's uncertainty

Deep Architecture

- Inflated 3D CNN (I3D) architecture as backbone [Carreira et al., 2017]
- Two FC-layers with MC-Dropout
- Sample output for M=100 forward passes at test-time

Experiments

Novelty Detection

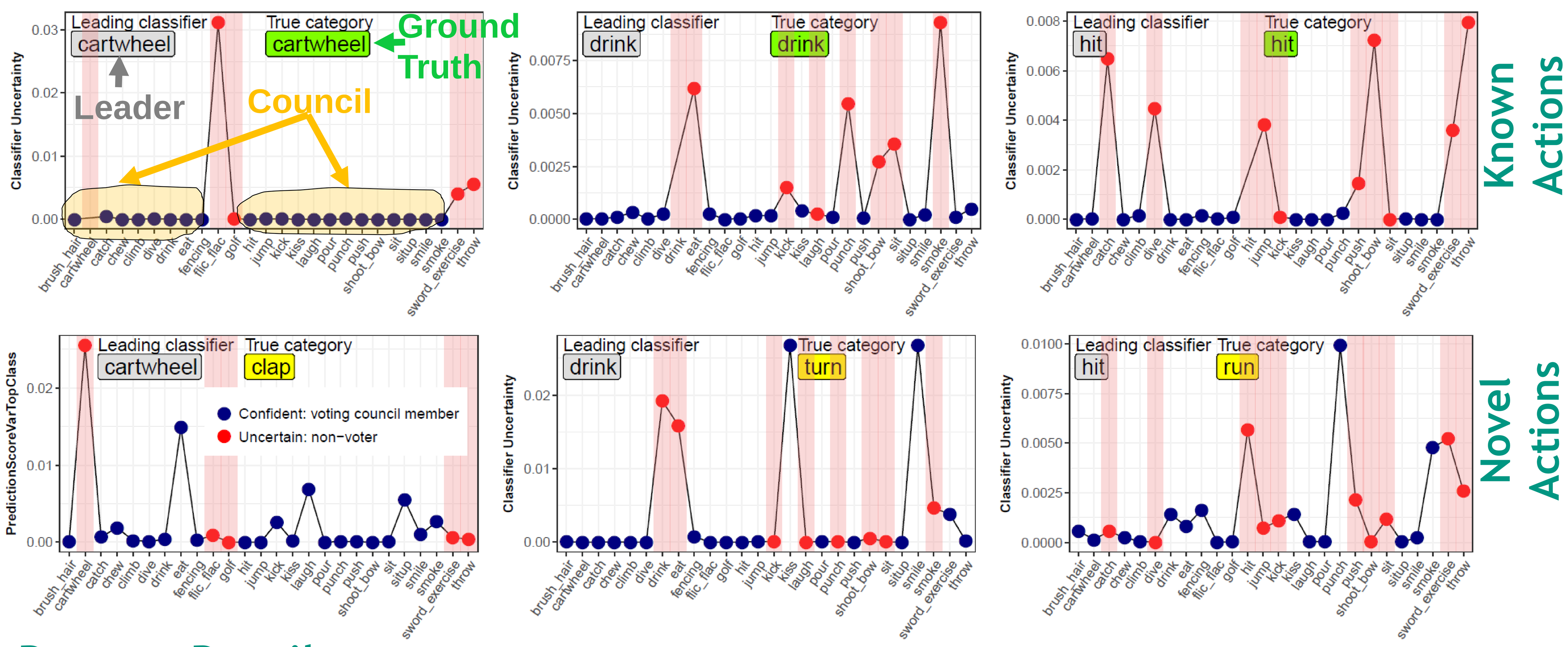
Novelty Detection Model	HMDB-51 ROC AUC %	HMDB-51 PR AUC %	UCF-101 ROC AUC %	UCF-101 PR AUC %
Baseline Models				
One-class SVM	54.1 (±3.0)	77.9 (±4.0)	53.6 (±2.0)	78.6 (±2.4)
GMM	56.8 (±4.2)	78.4 (±3.6)	59.2 (±4.2)	79.5 (±2.2)
Conventional NN Conf.	67.6 (±3.3)	84.2 (±3.0)	84.2 (±1.9)	93.9 (±0.7)
Our Proposed Model based on Bayesian Uncertainty				
Dictator	71.8 (±1.8)	86.8 (±2.5)	91.4 (±2.3)	96.7 (±1.0)
Uninformed Democracy	73.8 (±1.7)	87.8 (±2.3)	92.1 (±1.8)	97.2 (±0.7)
Informed Democracy	**75.3 (±2.7)**	**88.7 (±2.3)**	**92.9 (±1.7)**	**97.5 (±0.6)**

Generalized Zero-Shot Action Recognition

- ZSL Methods: **ConSE** and **Devise**
- Test has seen (S) and unseen (U) classes (GZS)
- Pure ZSL methods fail due to seen-classes-bias
- Our novelty detection leads to a clear improvement in GZS

Zero-Shot Method	HMDB-51 U→U+S	HMDB-51 U+S→U+S	UCF-101 U→U+S	UCF-101 U+S→U+S
ConSe	0.0	0.0	0.1	0.1
Devise	0.3	0.5	0.8	1.6
ConSe + Novelty Detection				
OC SVM	11.0	17.4	10.3	16.6
GMM	13.3	19.9	9.3	16.0
NN Conf.	11.0	18.6	12.2	20.9
ID (ours)	**13.7**	**22.3**	**13.6**	**23.4**
Devise + Novelty Detection				
OC SVM	8.9	14.7	8.7	14.3
GMM	10.6	16.7	7.3	12.9
NN Conf.	8.7	15.1	10.1	17.7
ID (ours)	**10.7**	**18.2**	**11.0**	**19.5**

Examples of Informed Voting



Dataset Details

- HMDB-51 and UCF-101 datasets for action recognition
- Ten splits into seen/unseen categories (26/25 for HMDB-51 and 51/50 for UCF-101).
- Set containing the seen classes is split into training (70%) and testing (30%)
- Baseline models trained on I3D model features (last avg. pooling layer)
- Dataset splits will be provided at cvhci.anthropomatik.kit.edu/~aroitberg/novelty_detection_action_recognition