TOWARDS A FAIR EVALUATION OF ZERO-SHOT ACTION RECOGNITION USING EXTERNAL DATA

Karlsruhe Institute of Technology

cv:hci

ECCV 2018 European Conference on Computer Vision 8 – 14 September 2018 | Munich, Germany

European Conference on Computer Vision 2018

SiVL

Alina Roitberg, Manuel Martinez, Monica Haurilet, Rainer Stiefelhagen

Computer Vision for Human-Computer Interaction (cv:hci)

Karlsruhe Institute of Technology, Germany

cvhci.anthropomatik.kit.edu

Introduction: Zero-Shot Action Recognition

- **Task:** classifying actions without any training data (**unseen target classes**)
- **How?** By linking visual and semantic features through **seen source classes**

Zero-Shot Learning premise: source and target classes are disjoint!

Contributions & Summary

Motivation

- Recent work shows extraordinary results when using external data sources for zero-shot action recognition
- **Problem**: in a cross-dataset setup source and target categories are often not disjoint

Contributions

- We show that external sources often **have actions excessively similar to the target classes**, strongly influencing the performance and **violating the ZSL premise**
- We propose an **evaluation procedure** that enables fair use of external data for zero-shot action recognition
- Side-contribution: we propose the **hybrid evaluation regime**, which uses the available training data of the source domain **and** the large-scale external datasets

Fair transfer of foreign categories

Evaluation regimes for ZSL

Intra-dataset: same origin of training- and test data

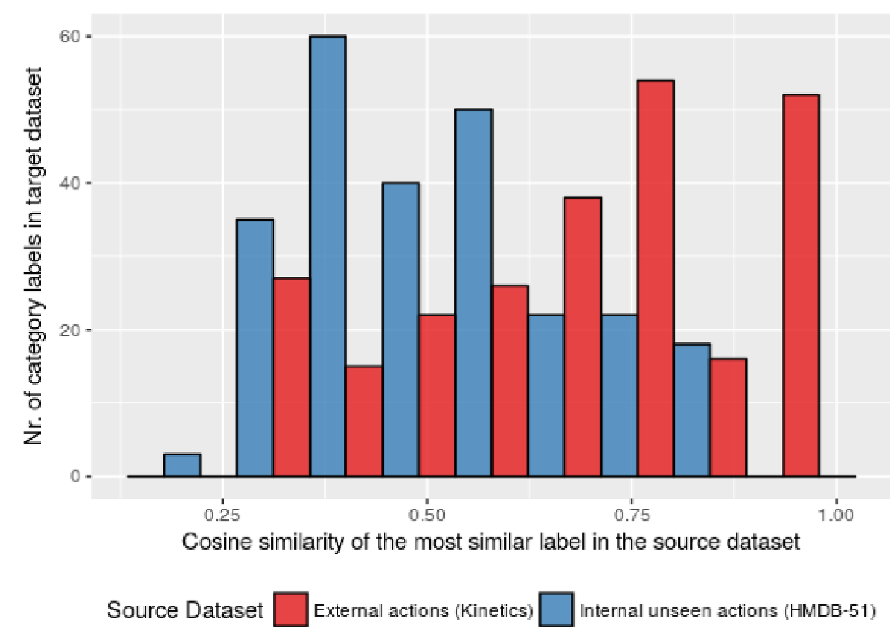
Supervised Action Recognition (AR)	▪ Classifying the already known categories ▪ $T \subset S$	standard approaches
Zero-Shot AR Intra-dataset	▪ Source from the **same domain**: $S = S_{native}$ ▪ $T \cap S_{native} = \emptyset$ → **ZSL premise satisfied** ☺	standard approaches

Cross-dataset: utilize large-scale external data sources

Zero-Shot AR Cross-dataset (Zhu et. al, 2018)	▪ Source from a different domain: $S = S_{ext}$ ▪ Boost in accuracy ▪ $T \cap S_{ext} \neq \emptyset$. ***ZSL* premise not given by default** ☹ **Our corrective protocol eliminates too unfamiliar concepts → ZSL premise given**	novel approaches
Zero-Shot AR Hybrid (ours)	▪ Source from native and external domains: $S = S_{ext} \cup S_{native}$ ▪ Boost in accuracy, **lower-bounded by the intra- and cross-dataset regimes** ▪ Same evaluation issues as in cross-dataset **Our corrective protocol eliminates too unfamiliar concepts → ZSL premise given**	novel approaches

Problem with the source-target synonyms

- A dataset does not contain the same action class twice
- External datasets intersect with datasets for zero-shot AR!
- Example: ***brushing hair*** in ActivityNet, Kinetics and HMDB51 (***brush hair***)
- Specializations: ***drinking beer*** vs. ***drinking***
  → **Getting rid of the direct matches is not enough!**



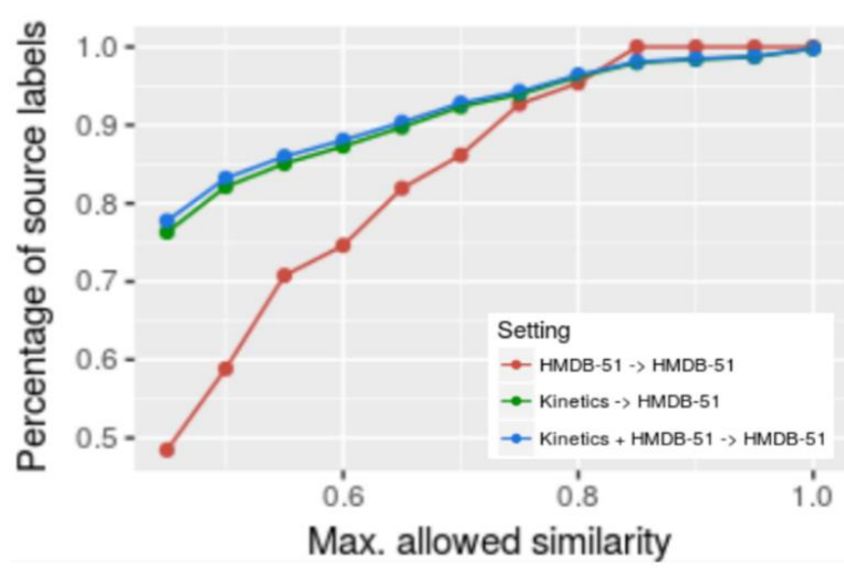
Proposed corrective protocol for fair cross-dataset transfer

1) Calculate the maximum intra-dataset similarity as our threshold s_{th} :

$$s_{th} = \max_{a_k \in S_{intra},\, t_m \in T} s\left(\omega(a_k), \omega(t_m)\right)$$

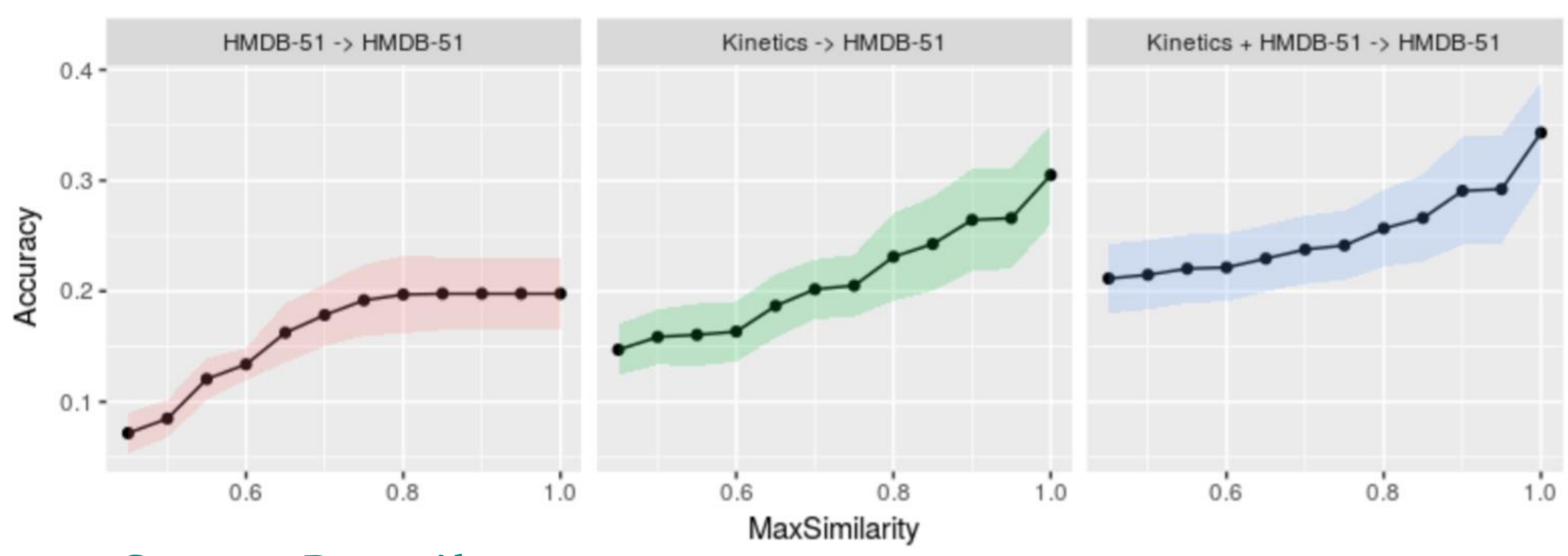
2) Purge the source category, if the label is too similar:

$$\forall t_m \in T, s\left(\omega(a_k), \omega(t_m)\right) \leq s_{th}$$



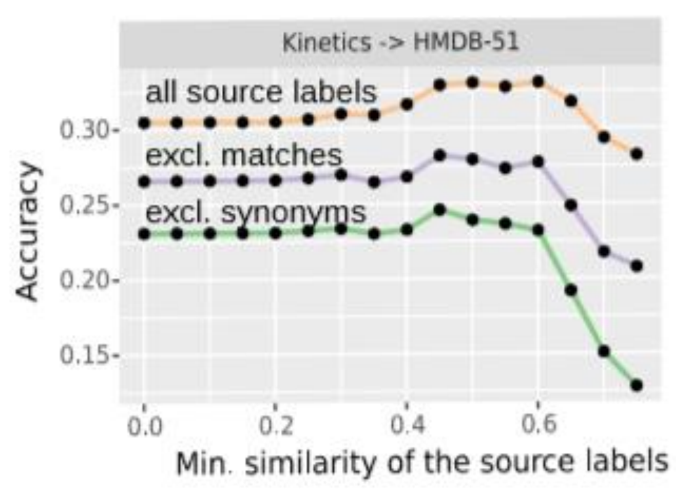
Experiments

Similar source and target classes influence the accuracy



Exclusion protocol	Source	# source labels	Accuracy	ZSL premise
n. a.	HMDB-51	26	19.92 (±3.3)	✓
Use all source labels	Kinetics	400	30.72 (±4.4)	–
	Kinetics+ HMDB-51	426	**34.77 (±4.5)**	–
Exclude exact labels	Kinetics	≈394.8	26.6 (±4.6)	–
	Kinetics+ HMDB-51	≈420.8	**29.22 (±4.9)**	–
Exclude similar labels **(ours)**	Kinetics	≈384.7	23.1 (±3.9)	✓
	Kinetics+ HMDB-51	≈410.7	**25.67 (±3.5)**	✓

Eliminating too unfamiliar concepts



An additional **lower bound** on the source similarity improves the performance

Setup Details

- ZSL Method: Convex Combination of Semantic Embeddings (**ConSE**)
- Language Model: **word2vec**, visual model: **I3D**
- Ten random splits into seen/unseen categories (26/25) for **HMDB-51**.
- **Kinetics** as external source (400 activity classes)