# Feature-Supervised Action Modality Transfer

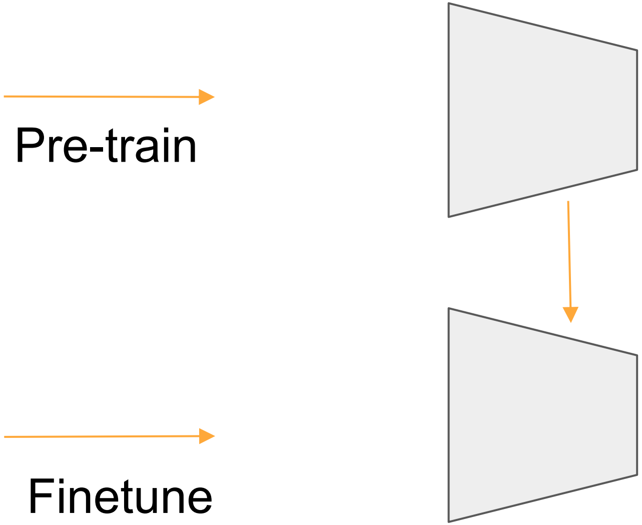
Fida Mohammad Thoker, Cees Snoek

University of Amsterdam

# Action classification for small datasets

Transfer Learning

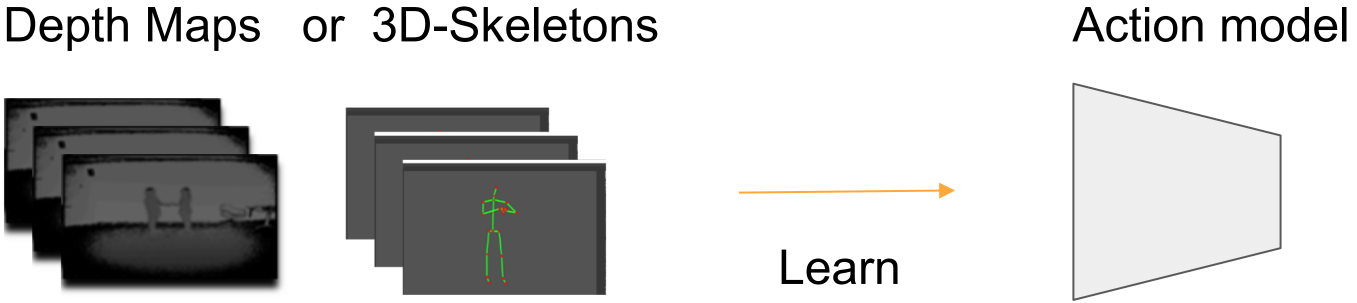
Large source dataset



Small target dataset

***Problem: pre-training datasets for non-RGB modalities unavailable.***

# Problem Statement



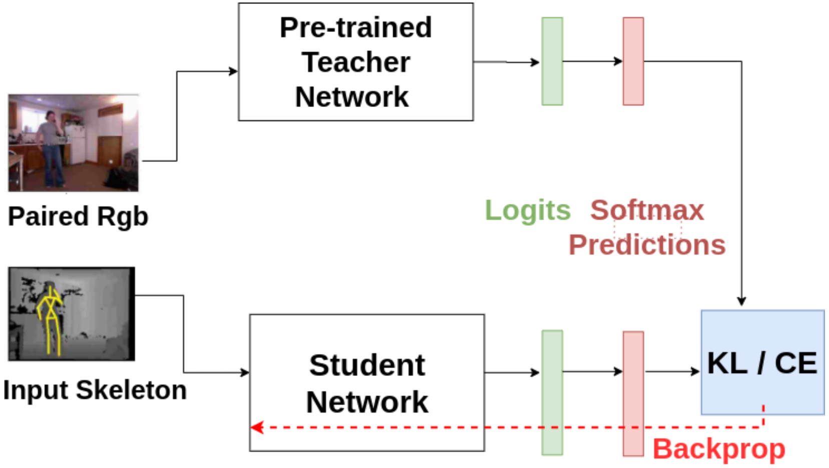
Depth Maps or 3D-Skeletons

Learn

Action model

***Can we transfer RGB action information to non-RGB modalities?***

# Standard Knowledge Distillation (Hinton et al. NeurIPS wshop, 2015)



Transfer class-level information from the pretrained RGB teacher.

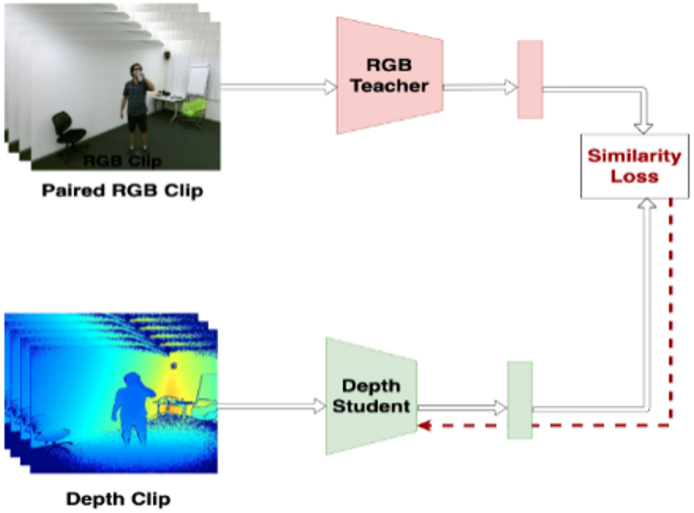
Match student softmax predictions with those of the teacher.

***Requires teacher to be pre-trained for same action classes as the student.***

# Our Proposal: Feature-Supervised Action Modality Transfer

1. Match action embeddings of modality pairs via **feature-level supervision**.

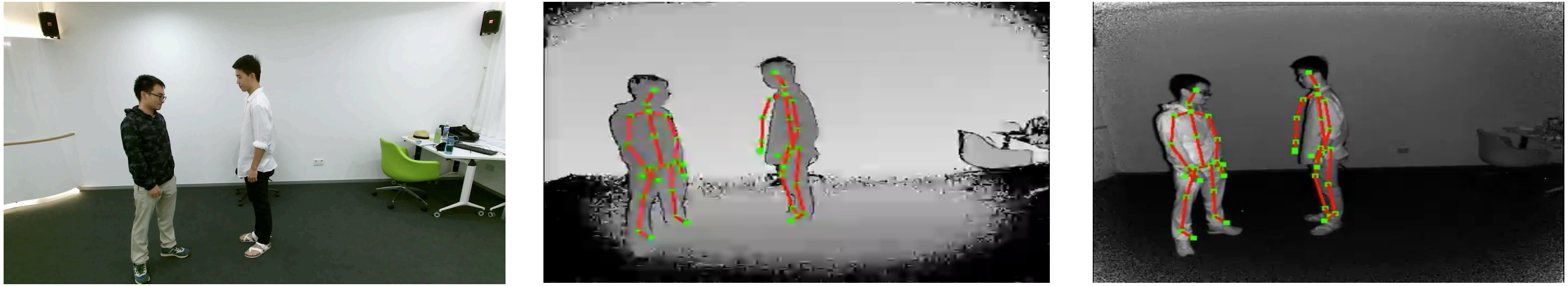
2. Finetune for new action classes on a small labeled **non-RGB** dataset.



***RGB teacher trained on a source dataset with non-overlapping action classes.***

# Multi-modal paired video data

A sample action scene captured in multiple modalities (Handshaking).



RGB

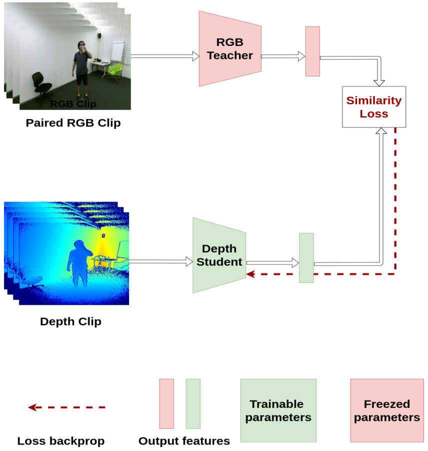
Depth

Infrared

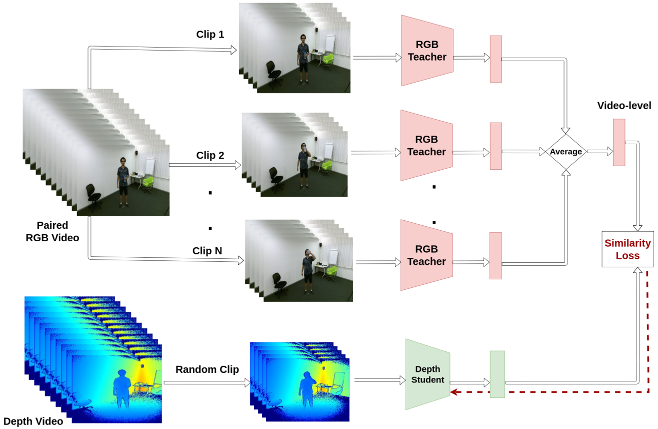
***Transfer knowledge from pre-trained RGB models via unlabelled modality pairs.***

# Three Knowledge Transfer Granularities

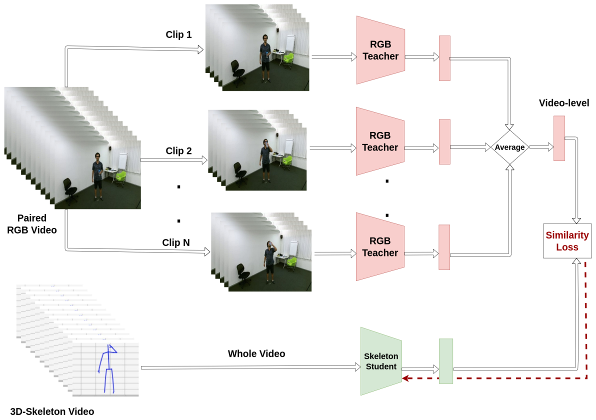
## Clip-to-Clip



## Video-to-Clip



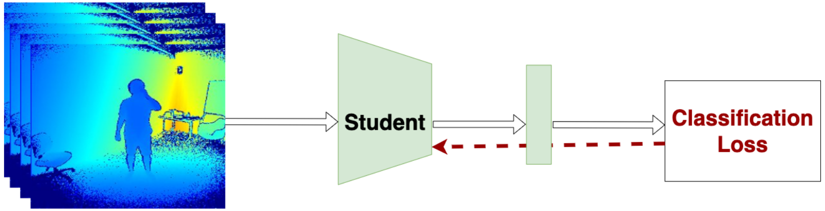
## Video-to-Video



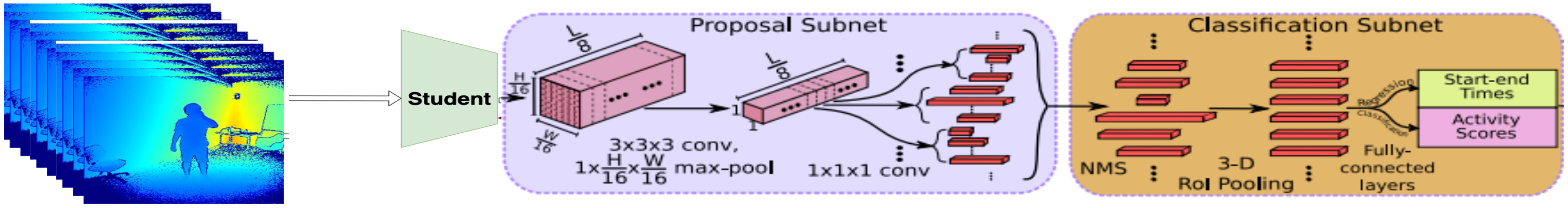
***Cosine distance loss is minimized between action embeddings.***

# Fine-tuning with non-RGB examples

## Task I: Action Classification



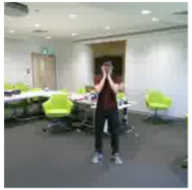
## Task II: Action Detection (Xu et al. ICCV 2017)



***Finetune pre-trained student with the task specific non-RGB labelled examples.***

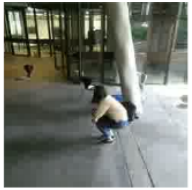
# Expiremental Setup

**Source** (NTU RGB+D 120 minus 60, Kinetics-400)



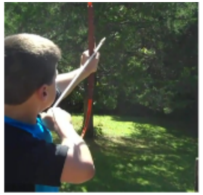
Apply Cream

Exchange Things



Squat Down

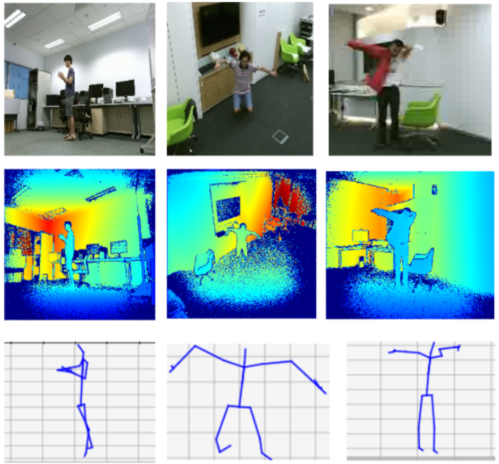
Zumba



Archery

Dribbling basketball

**Target** (NTU RGB+D 60, PKU-MMD)



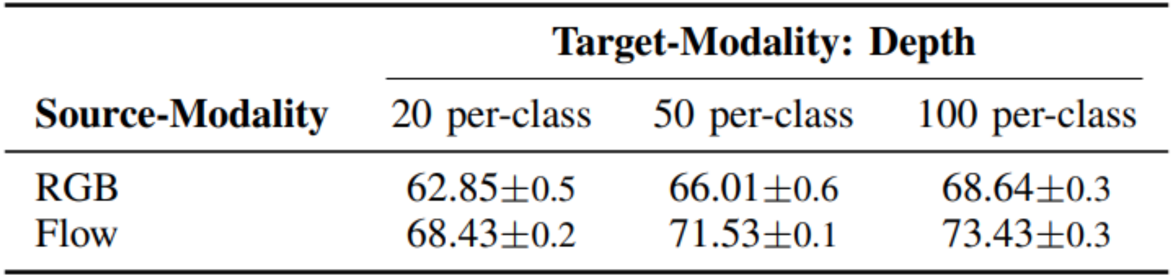
Clapping Falling Down Wear Jacket

*Pretrain teacher on RGB/Flow modality of the source dataset.*

*Transfer via unlabeled modality pairs of NTU RGB+D 60 training set.*

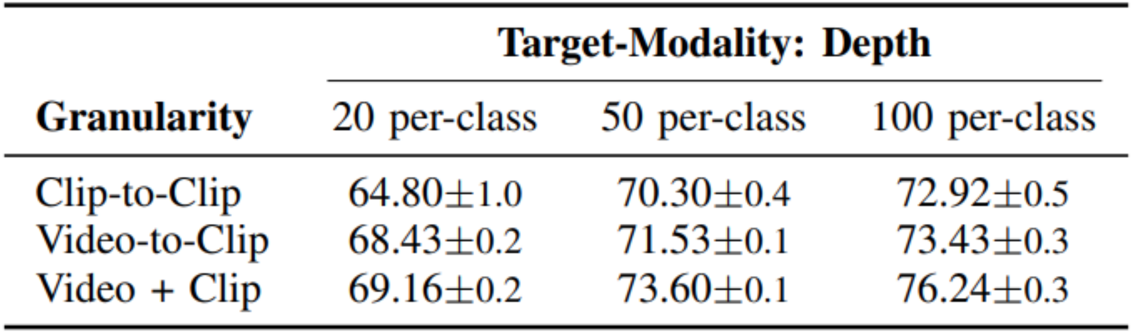
*Finetune with labeled examples from NTU RGB+D 60 / PKUMMD training set.*

# Ablation Studies



| | **Target-Modality: Depth** | | |
|---|---|---|---|
| **Source-Modality** | 20 per-class | 50 per-class | 100 per-class |
| RGB | 62.85±0.5 | 66.01±0.6 | 68.64±0.3 |
| Flow | 68.43±0.2 | 71.53±0.1 | 73.43±0.3 |

Which Source Modality?



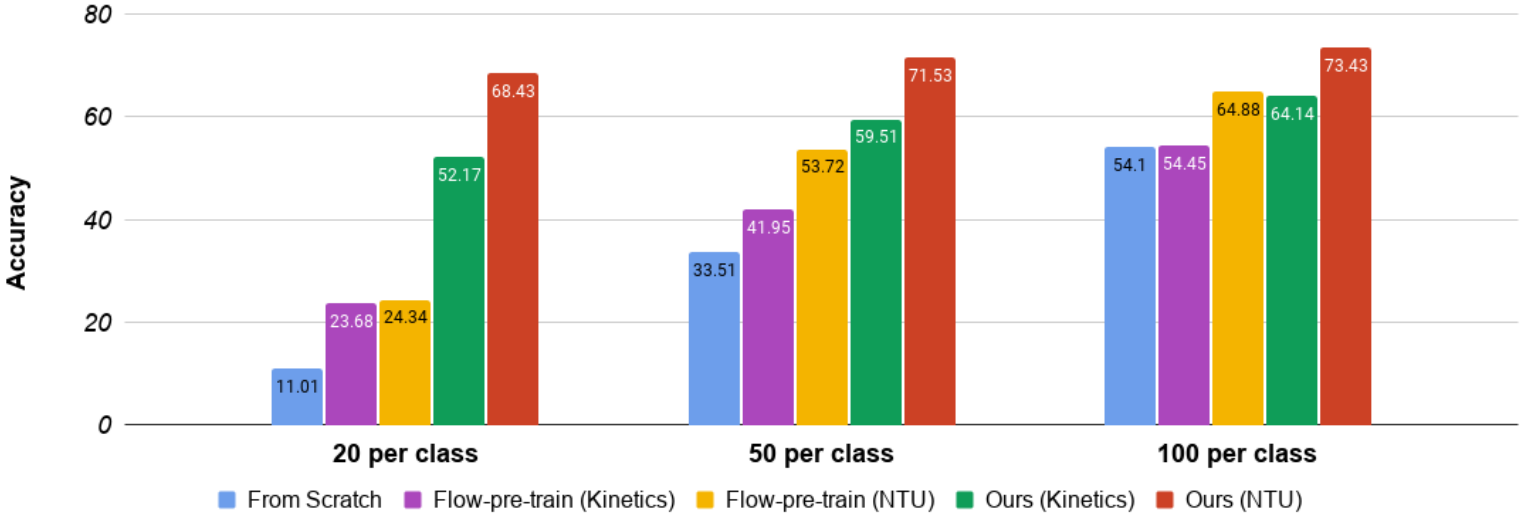
| | **Target-Modality: Depth** | | |
|---|---|---|---|
| **Granularity** | 20 per-class | 50 per-class | 100 per-class |
| Clip-to-Clip | 64.80±1.0 | 70.30±0.4 | 72.92±0.5 |
| Video-to-Clip | 68.43±0.2 | 71.53±0.1 | 73.43±0.3 |
| Video + Clip | 69.16±0.2 | 73.60±0.1 | 76.24±0.3 |

Which Granularity?

***Optical-flow with video+clip granularity provides best feature-level supervision.***

# Results

Action classification from depth maps for NTU RGB+D 60 dataset

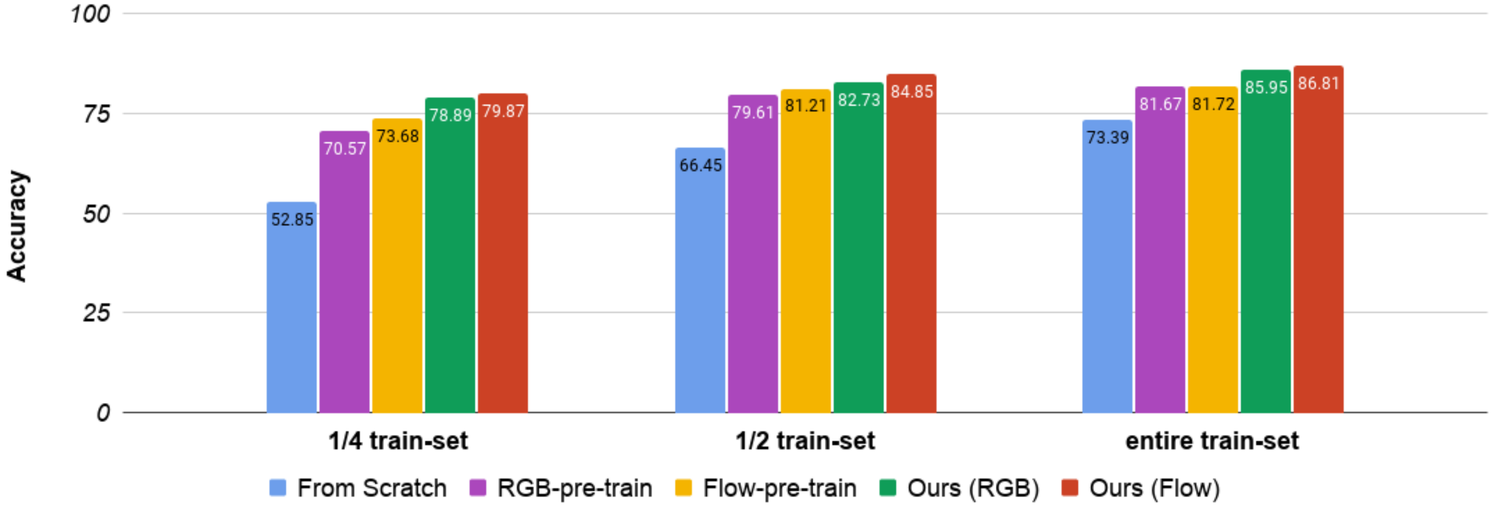


***Considerable improvement over training from scratch and simple pretraining.***

***Source dataset with a similar domain provides better action transfer features.***

# Results

Action detection from depth maps  for  PKU-MMD dataset



*Transfer results for 3D-skeleton action classification in paper.*

***Our method generalizes for temporal action detection as well.***

# Conclusion

***RGB action datasets act as pre-training source for non-RGB modalities.***

***Optical-flow from a similar domain provides best feature-supervision.***

***Boost non-RGB action classification and detection when labels are scarce.***

Contact: fmthoker@gmail.com