Feature-Supervised Action Modality Transfer

Fida Mohammad Thoker, Cees G. M. Snoek

UNIVERSITY OF AMSTERDAM

ICPR 2020
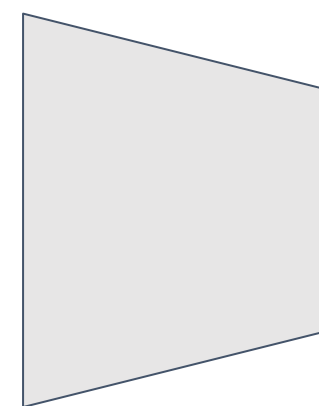
Introduction

Depth Maps or 3D-Skeletons   → Learn →   Action model
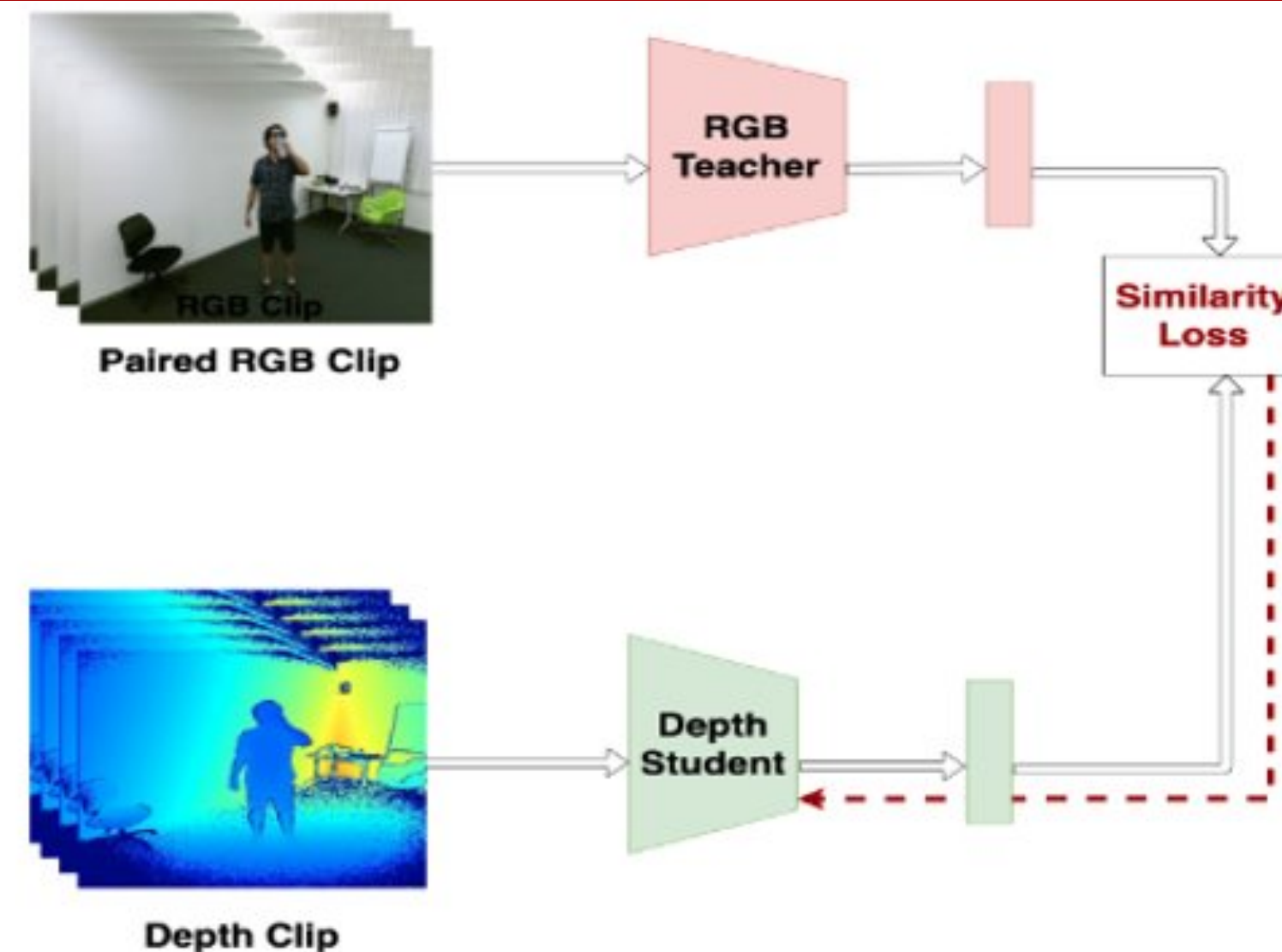
Non-RGB action classification/detection with limited labeled examples.

Given a pre-trained RGB action model, we aim to transfer action knowledge to non-RGB action modalities like depth maps, 3D-skeletons, etc.
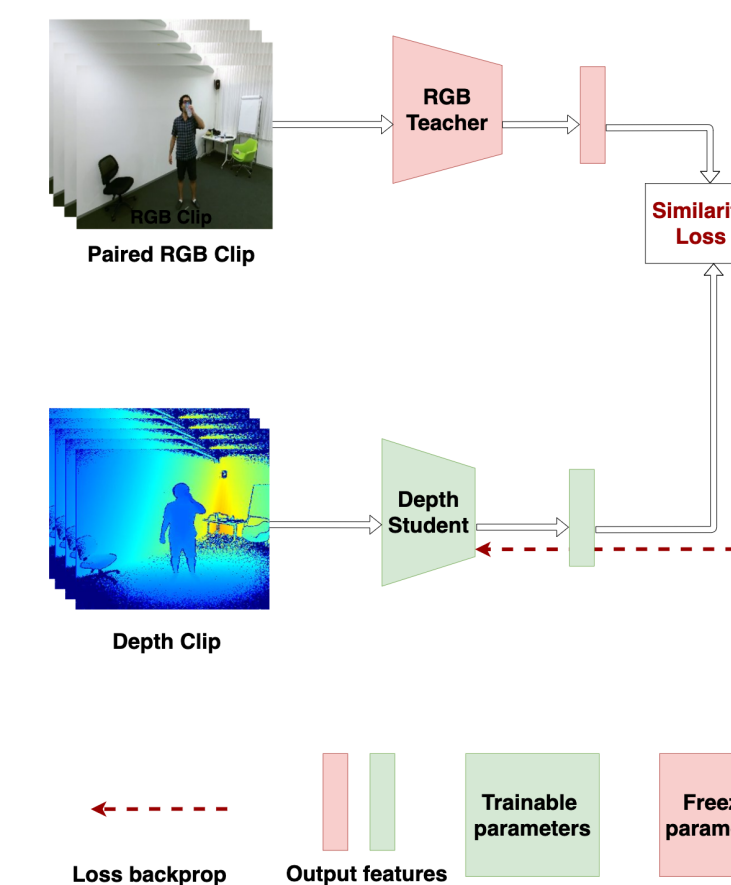
Approach

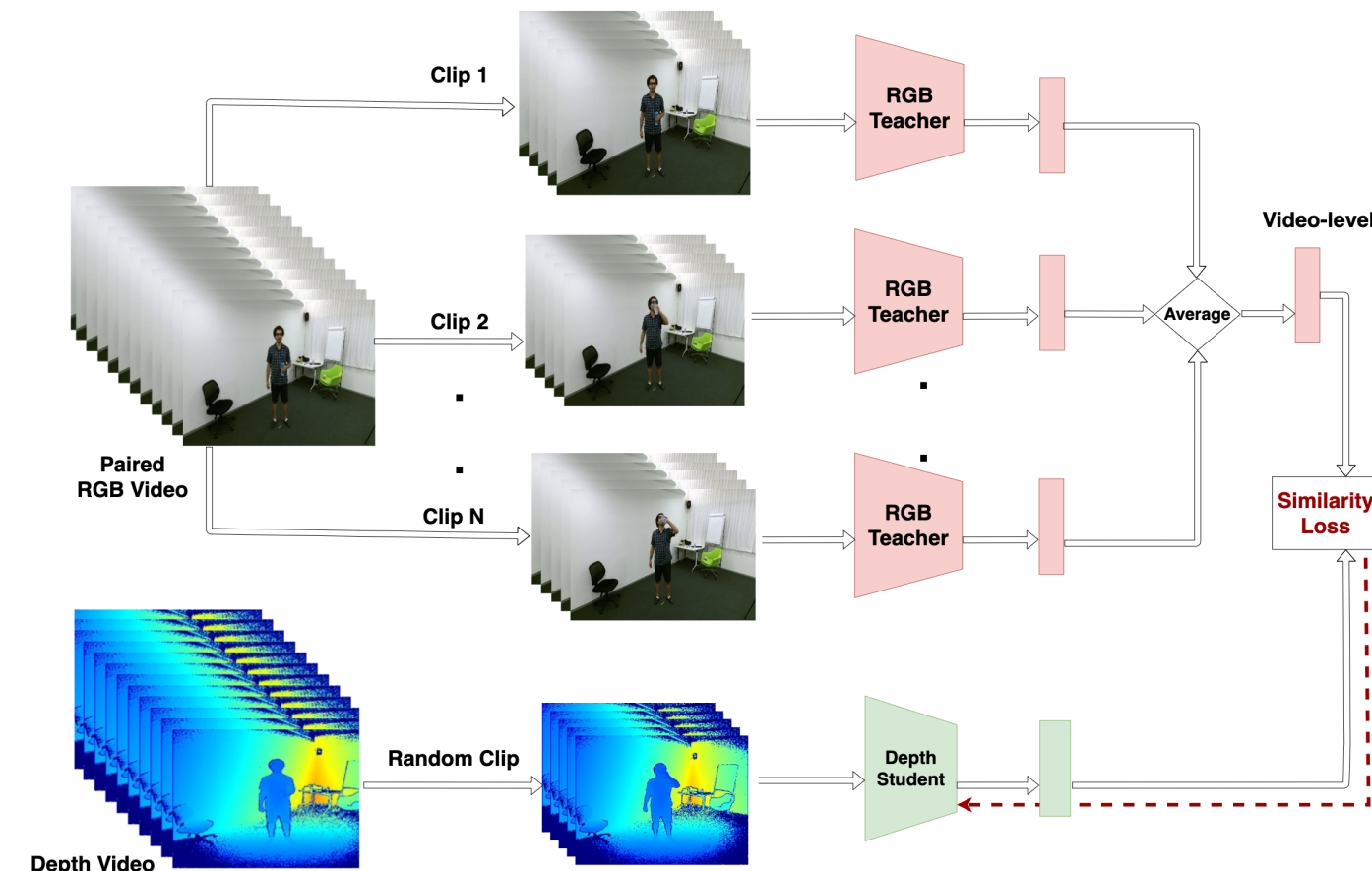*RGB teacher trained on a source dataset with non-overlapping action classes.*

Match action embeddings of unlabeled modality pairs via **feature-level supervision**.
Finetune on a small labeled **non-RGB** dataset for action classification or detection.
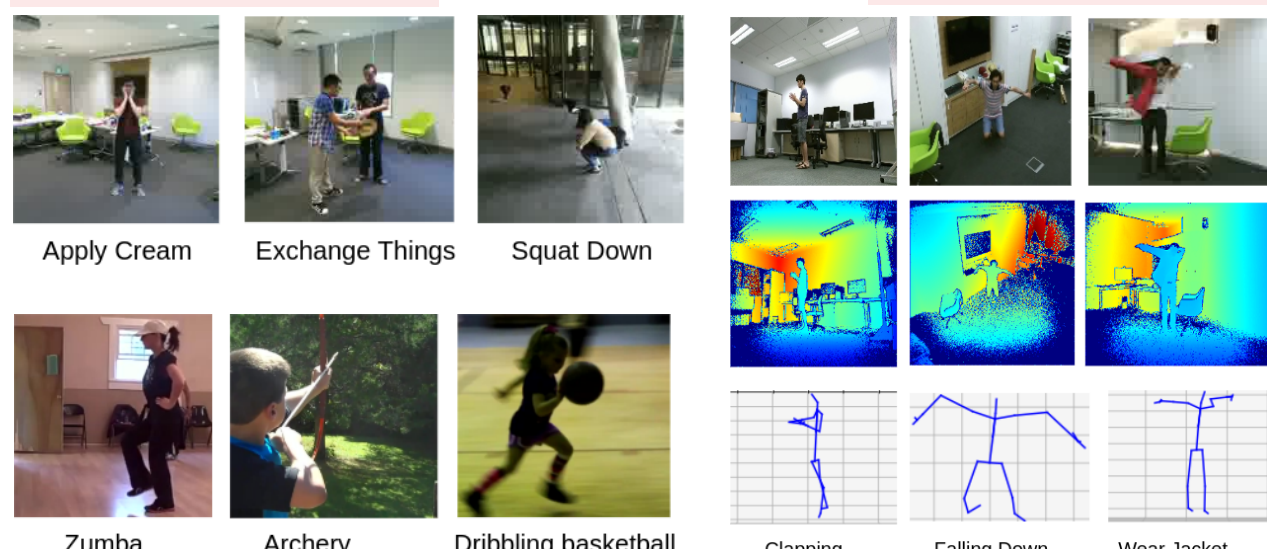
Transfer Granularities

Clip-to-Clip

Video-to-clip

Cosine distance loss is minimized between action embeddings.

Setup

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

Target
NTU-RGB+D 60, PKU-MMD

Apply Cream Exchange Things Squat Down
Zumba Archery Dribbling basketball
Clapping Falling Down Wear Jacket

Pretrain teacher on RGB/Flow of the source.

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

Finetune with small number of labeled examples from NTU RGB+D 60 / PKUMMD training set.

Ablation

Which Source ?

Source-Modality	Target-Modality: Depth		
	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 Granularity ?

Granularity	Target-Modality: Depth		
	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

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

Results

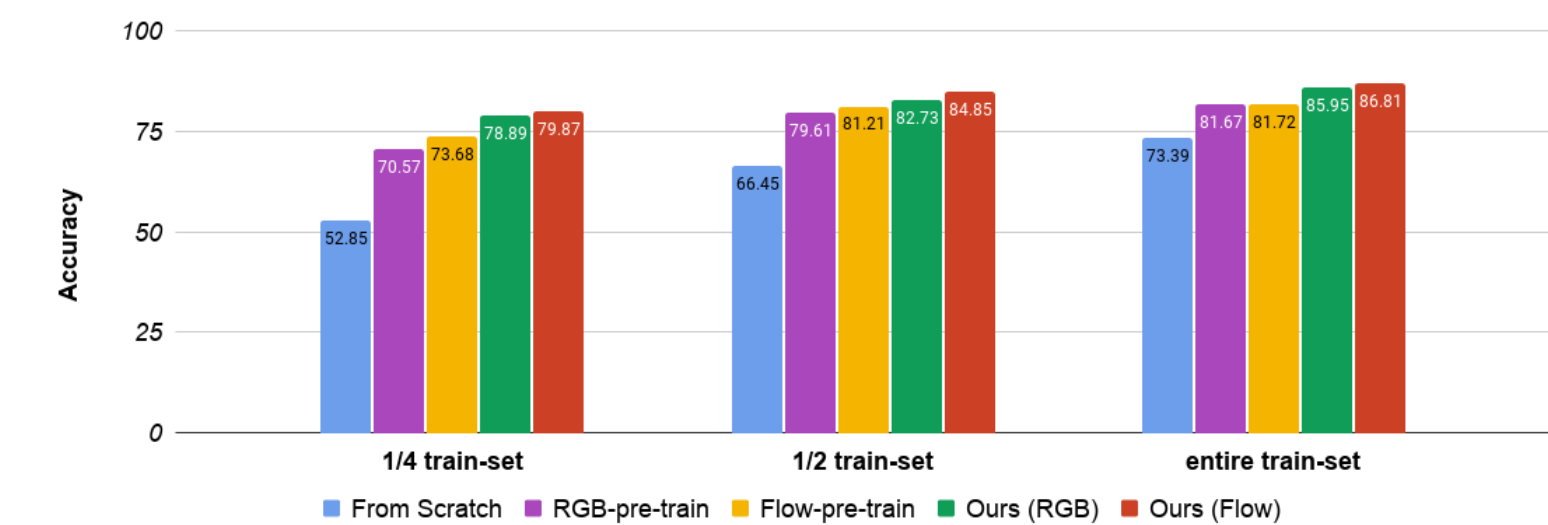*Transfer results for 3D-skeleton action classification in paper.*
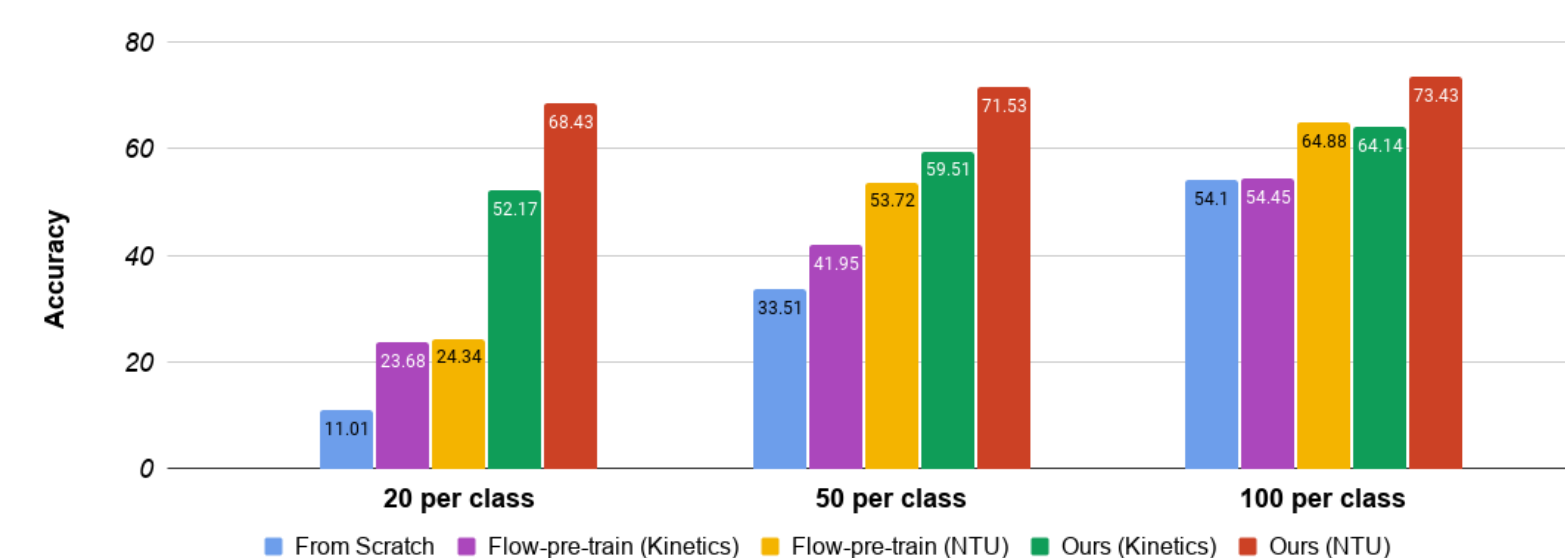
Action classification from depth maps for NTU RGB+D 60 dataset

Action detection from depth maps for PKU-MMD dataset

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

Considerable improvement over training from scratch and simple pretraining.

Optical-flow source from a similar domain provides better action transfer features.

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

Contact

fmthoker@gmail.com