Feature-Supervised Action Modality Transfer
Fida Mohammad Thoker, Cees G. M. Snoek

**Introduction**

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**

Cosine distance loss is minimized between action embeddings.

**Source**

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

**Target**

NTU-RGB-D 60, PKU-MMD

**Setup**

Ablation

Which Source?

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 Granularity?

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

Results

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

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