Graph Distillation for Action Detection with Privileged Modalities

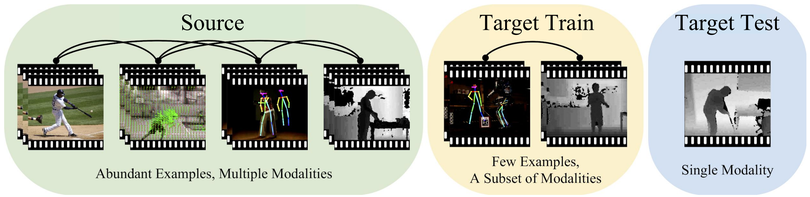
Zelun Luo[1,2], Jun-Ting Hsieh[1], Lu Jiang[2], Juan Carlos Niebles[1,2], Li Fei-Fei[1,2]
[1]Stanford University, [2]Google Inc.

Code available at
http://alan.vision/eccv18_graph/

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

Stanford | Google AI | Stanford Vision and Learning | Google Cloud

Introduction

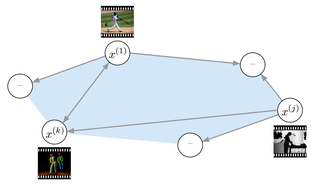


We systematically study the Learning Using Privileged Information (LUPI) problem in action detection and classification in which privileged modalities are only available

- In the source domain (but not target)
- During training (but not testing)

We propose the Graph Distillation method, which learns a dynamic distillation across multiple modalities based on

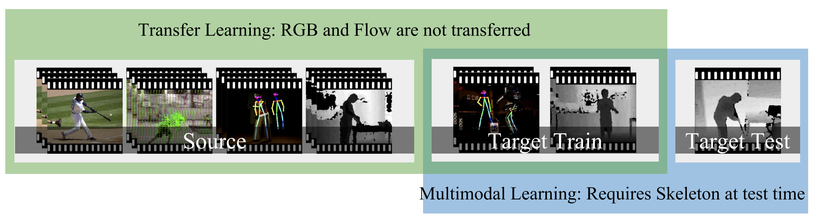
- Example-specific information
- Modality-specific prior



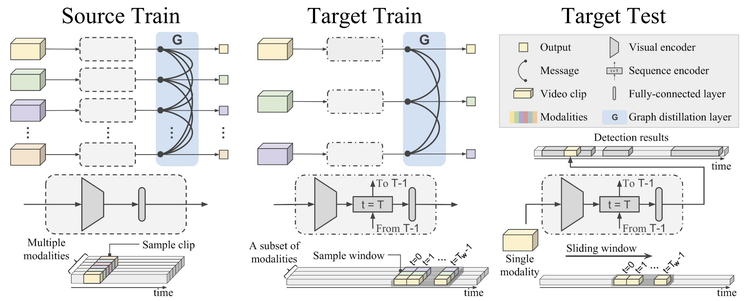
Motivation

In many real-world applications, only limited training data and partially observed modalities are available.

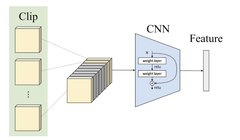
- Transfer learning does not take advantage of the extra modalities potentially available in the source domain.
- Multimodal learning focuses on a single domain or task and does not handle the modality discrepancy between training and testing.
- Existing LUPI and knowledge distillation approaches prespecify distillation directions and are suboptimal on multiple modalities.



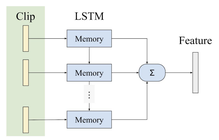
Overview



Visual Encoder for Images



Visual Encoder for Vectors



➔ Objective

$$\min \frac{1}{|\mathcal{D}_S|} \sum_{(\mathcal{X}_i, y_i) \in \mathcal{D}_S} \underbrace{\ell_c(f(\mathcal{X}_i), y_i)}_{\text{Task-Specific Loss}} + \underbrace{\ell_m(\mathcal{X}_i)}_{\text{Imitation Loss}}$$

$\mathcal{X}_i$: Input in all source modalities

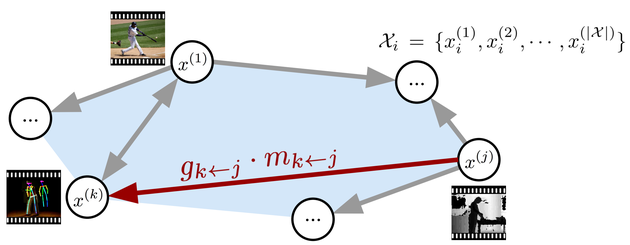
Graph Distillation

➔ Vectorization

$$\underbrace{g_{k \leftarrow j}}_{\text{Edge weight from modality j to k}} \cdot \underbrace{m_{k \leftarrow j}}_{\text{Message from modality j to k}} = \underbrace{\mathbf{G}_{jk}}_{\text{Distillation Graph}} \underbrace{\mathbf{M}_{jk}}_{\text{Message Matrix}}$$

➔ Imitation Loss

$$\ell_m = (\mathbf{G} \odot \mathbf{M})^T \mathbf{1}$$



➔ Message

$$\mathbf{M}_{jk} = \lambda_1 \ell_{logit}(x_i^{(j)}, x_i^{(k)}) + \lambda_2 \ell_{repr}(x_i^{(j)}, x_i^{(k)})$$

➔ Distillation Graph (Adjacency Matrix)

$$\mathbf{G} = \underbrace{\mathbf{L}}_{\text{Normalized Learned Graph}} + \underbrace{\mathbf{P}}_{\text{Normalized Prior Graph}}$$

Learned Graph L

$$z_i^{(k)} = W_1 \phi_k^{l-1}(x_i^{(k)}) + W_2 \phi_k^l(x_i^{(k)})$$
$$\mathbf{L}_{jk} = e_{k \leftarrow j} = W_3[z_i^{(j)} \| z_i^{(k)}]$$

Prior Graph P

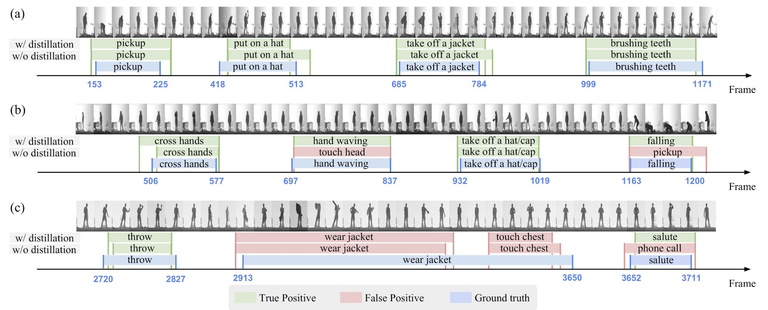
$$\mathbf{P} = \mathbf{c}\mathbf{1}^T$$

c: Cross validation performance

Results

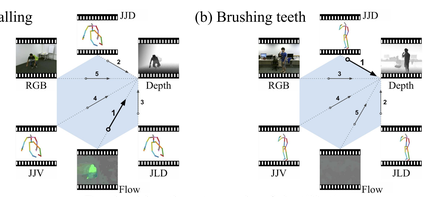
Action Detection on PKU-MMD

Method	Test Modality	mAP @ tIoU thresholds (θ) 0.1	0.3	0.5
Deep RGB (DR) [28]	RGB	0.507	0.323	0.147
Qin and Shelton [43]	RGB	0.650	0.510	0.294
Deep Optical Flow (DOF) [28]	F	0.626	0.402	0.168
Raw Skeleton (RS) [28]	S	0.479	0.325	0.130
Convolution Skeleton (CS) [28]	S	0.493	0.318	0.121
Wang and Wang [55]	S	0.842	-	0.743
RS+DR+DOF [28]	RGB+F+S	0.647	0.476	0.199
CS+DR+DOF [28]	RGB+F+S	0.649	0.471	0.199
Ours (w/o | w/ transfer)	RGB	0.824 | 0.880	0.813 | 0.868	0.743 | 0.801
Ours (w/o | w/ transfer)	D	0.823 | 0.872	0.817 | 0.860	0.752 | 0.792
Ours (w/o | w/ transfer)	F	0.790 | 0.826	0.783 | 0.814	0.708 | 0.747
Ours (w/o | w/ transfer)	S	0.836 | 0.857	0.823 | 0.846	0.764 | 0.784
Ours (w/ transfer)	RGB+D+F+S	**0.903**	**0.895**	**0.833**



Action Classification on NTU RGB+D

Method	Test Modality	mAP
Shahroudy [47]	RGB+D	0.749
Liu [29]	RGB+D	0.775
Liu [32]	S	0.800
Ding [9]	S	0.823
Li [24]	S	0.829
Ours	RGB	**0.895**
Ours	D	0.875
Ours	F	0.857
Ours	S	0.837



Comparison with Baseline Methods

Method (mini-NTU / RGB)	mAP
Empty graph	0.464
Multi-task	0.456
Cross-distillation	0.503
Knowledge distillation	0.524
Distillation graph	**0.619**

Efficacy of Distillation Graph

Graph	mini-NTU mAP / RGB	mini-PKU mAP @ 0.5 / D
Empty graph	0.464	0.501
Uniform graph	0.537	0.513
Prior graph	0.571	0.515
Distillation graph	**0.619**	**0.559**

Efficacy of Privileged Information

	Method (mini-PKU / D)	mAP @ tIoU thresholds (θ) 0.1	0.3	0.5
1	trg only	0.248	0.235	0.200
2	src + trg	0.583	0.567	0.501
3	src w/ PIs + trg	0.625	0.610	0.533
4	src + trg w/ PIs	0.626	0.615	0.559
5	src w/ PIs + trg w/ PIs	0.642	0.629	0.562
6	src w/ PIs + trg	0.625	0.610	0.533
7	src w/ PIs + trg w/ 1 PI	0.632	0.615	0.549
8	src w/ PIs + trg w/ 2 PIs	0.636	0.624	0.557
9	src w/ PIs + trg w/ all PIs	0.642	0.629	0.562