Università di Genova | MaLGa

Single view learning in action recognition

Gaurvi Goyal, Nicoletta Noceti, Francesca Odone

MaLGa – Machine Learning Genoa center, DIBRIS, Università di Genova

nicoletta.noceti@unige.it

UniGe | MaLGa

Motivations, context and goals

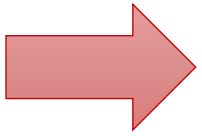
Cross-view action recognition

What do learnt features have to say?

- A main challenge in human action recognition is the fact that actions may look very different depending on the point of observation → Possible solutions: (very big) training sets incorporanting multiple views, use of additional information (as depth or 3D poses)

- We address the challenging problem of learning high level *view-tolerant representations* of an action from **a single view point**, through a *domain adaptation* procedure that *transfers information* from a generic action recognition network

Learning…

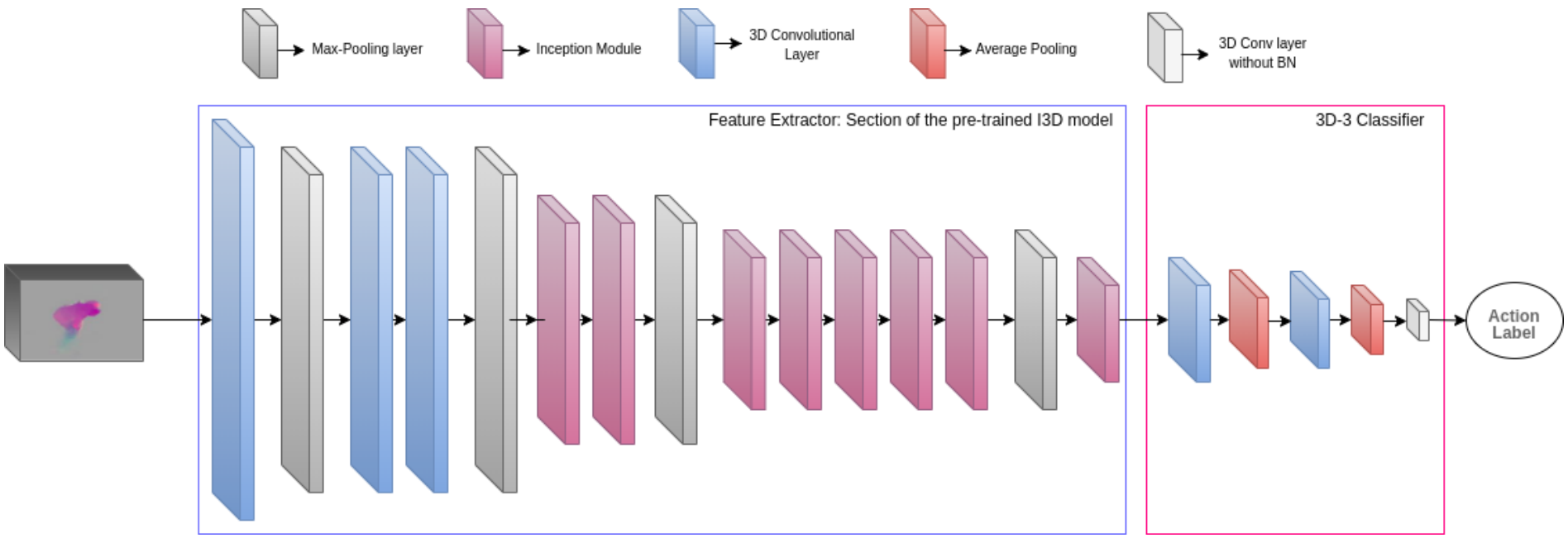


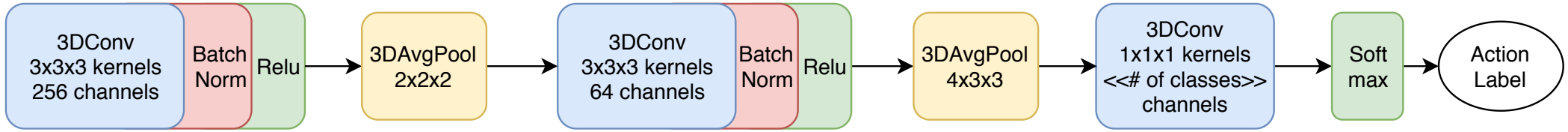
Recognizing…

UniGe | MaLGa

Our approach

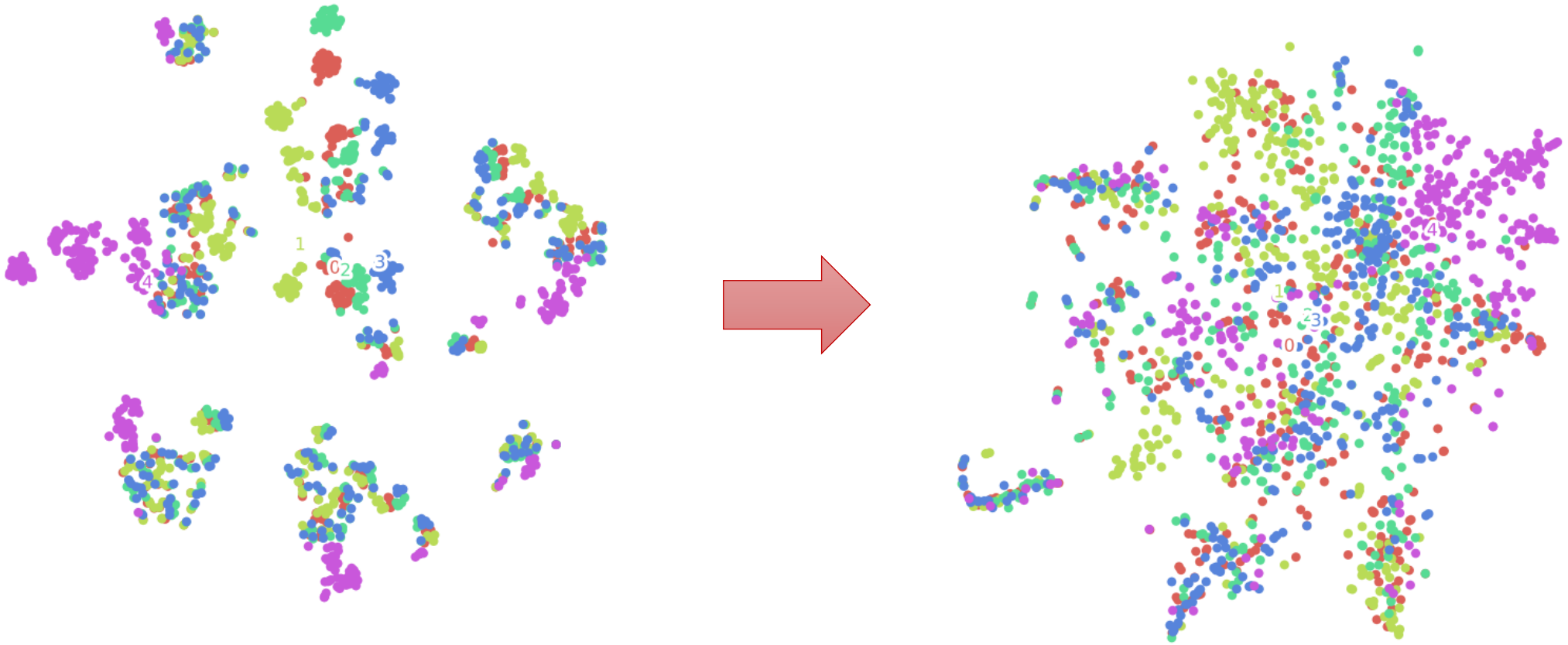
The architecture





J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in IEEE CVPR, 2017, pp. 4724–473

View-wise batch normalization



An analysis of the features in a space of low dimensionality (with TSNE) reveils that a covariate shift is present in the data

View-wise batch normalization

Given two batches $\mathbf{B}\,\mathbf{B}'$ with means $\mu_{\mathbf{B}}\,\mu_{\mathbf{B}'}$ and standard deviations $\sigma_{\mathbf{B}}\,\sigma_{\mathbf{B}'}$ we define a normalization function

$$\mathcal{V}(\mathbf{B}, \mu_{\mathbf{B}'}, \sigma_{\mathbf{B}'})$$

to normalize the first batch according to the second. Then

$$\forall \mathbf{B}_s \subseteq \mathbf{X}_s \quad : \quad \mathcal{V}(\mathbf{B}_s, \mu_{\mathbf{B}_s}, \sigma_{\mathbf{B}_s})$$

$$\forall \mathbf{B}_t \subseteq \mathbf{X}_t \quad : \quad \mathcal{V}(\mathbf{B}_t, \mu_{\mathbf{B}_t}, \sigma_{\mathbf{B}_t})$$

Where $\mathbf{X}_s$ us the training set from the **source** view, while $\mathbf{X}_t$ is the test set from the **target** view

UniGe | MaLGa

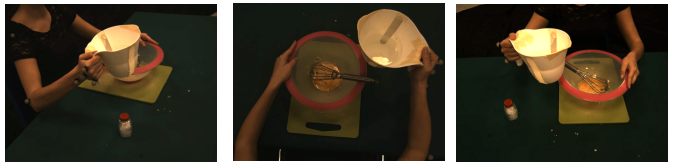
Experiments

Single view recognition

NTU

Source|Target(°)	0 | 45	0 | 90	45 | 0	45 | 90	90 | 0	90 | 45	Mean
I3D + SLP	58.94	48.61	58.39	57.92	50.08	59.41	55.56
I3D + Inc.	67.19	54.87	69.00	68.81	55.10	66.94	63.65
I3D + CV-3	**74.08**	**57.94**	**76.78**	**75.82**	**60.27**	**75.39**	**70.05**

MoCA

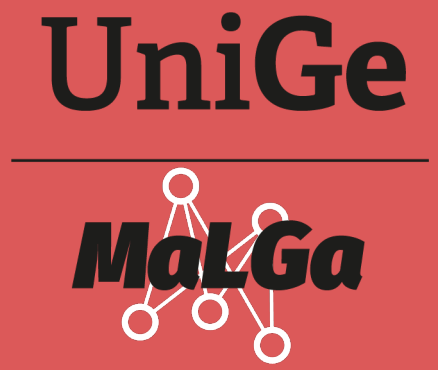


Source|Target	0|1	0|2	1|0	1|2	2|0	2|1
I3D + SLP	47.38	68.33	47.38	32.86	66.27	34.84
I3D + Inc.	50.63	64.84	33.10	36.35	61.67	54.92
I3D + CV-3	**54.84**	**79.52**	**65.71**	**69.52**	**88.20**	**61.83**

IXMAS

With regular batch norm. → 45.1%
With view-wise batch norm. → 78.4%

S|T	0|1	0|2	0|3	0|4	1|0	1|2	1|3	1|4	2|0	2|1	2|3	2|4	3|0	3|1	3|2	3|4	4|0	4|1	4|2	4|3	Mean
DT [18]	93.9	64.2	81.8	27.6	87.6	66.4	75.2	22.4	70.0	83.0	73.9	53.3	75.5	77.0	67.0	34.8	42.1	25.8	63.3	48.8	61.7
Hank. [23]	83.7	59.2	57.4	33.6	84.3	61.6	62.8	26.9	62.5	65.2	72.0	60.1	57.1	61.5	71.0	31.2	39.6	32.8	68.1	37.4	56.4
DVV [30]	72.4	13.3	53.0	28.8	64.9	27.9	53.6	21.8	36.4	40.6	41.8	37.3	58.2	58.5	24.2	22.4	30.6	24.9	27.9	24.6	38.2
CVP [29]	78.5	19.5	60.4	33.4	67.9	29.8	55.5	27.0	41.0	44.9	47.0	41.0	64.3	62.2	24.3	26.1	34.9	28.2	29.8	27.6	42.2
I3D + SVM-Lin	94.2	84.5	70.0	43.3	94.2	83.6	65.5	47.3	82.7	77.0	90.0	60.0	35.5	25.8	76.7	26.1	43.6	38.2	69.4	53.0	63.0
I3D + SVM-RBF	91.8	81.2	71.8	42.7	87.0	72.4	47.3	38.5	78.8	66.7	77.6	54.5	38.3	33.9	71.2	33.3	43.9	40.0	61.2	47.9	59.0
I3D + SLP	84.4	80.3	79.2	48.6	87.4	77.6	72.1	47.0	79.6	78.6	83.0	65.9	72.1	72.0	98.3	45.0	**53.3**	**56.6**	69.3	53.5	69.4
I3D + Inc.	88.6	76.7	84.0	44.4	87.4	70.6	79.4	42.5	78.9	77.9	80.8	61.2	86.0	85.4	77.3	45.4	48.6	49.1	57.9	46.7	68.5
I3D + **CV-3**	**97.1**	**92.7**	**94.6**	**50.3**	**95.4**	**85.6**	**92.8**	**47.5**	**88.9**	**86.8**	**95.3**	**77.5**	**91.6**	**90.5**	**94.9**	**54.4**	49.4	52.3	**73.7**	**56.5**	**78.4**


UniGe
MaLGa