Training Results

Multi-Label Learning from Single Positive Labels

Results from Training

These results are from training $\mathcal{L}_{ROLE}$ on the *COCO* dataset. The training mode is *Linear Fixed Features* and the validation set variant is *Clean*.

Metric	Value (%)
Train MAP	68.99
Val MAP	66.78
Test MAP	66.36

Table 1: Mean Average Precision (Clean) at different stages.

Best epoch
Best epoch based on validation `map_clean`: **19**

Hyperparameters

- Batch size: 16
- Learning rate: 0.001

Results from Paper

		Linear				Fine-Tuned			
Loss	Labels Per Image	VOC12	COCO	NUS	CUB	VOC12	COCO	NUS	CUB
$\mathcal{L}_{\text{BCE}}$	All Pos. & All Neg.	86.7	70.0	50.7	29.1	89.1	75.8	52.6	32.1
$\mathcal{L}_{\text{BCE-LS}}$	All Pos. & All Neg.	87.6	70.2	51.7	29.3	90.0	76.8	53.5	32.6
$\mathcal{L}_{\text{IUN}}$	1 Pos. & All Neg.	86.4	67.0	49.0	19.4	87.1	70.5	46.9	21.3
$\mathcal{L}_{\text{IU}}$	1 Pos. & 1 Neg.	82.6	60.8	43.6	16.1	83.2	59.7	42.9	17.9
$\mathcal{L}_{\text{AN}}$	1 Pos. & 0 Neg.	84.2	62.3	46.2	17.2	85.1	64.1	42.0	19.1
$\mathcal{L}_{\text{AN-LS}}$	1 Pos. & 0 Neg.	85.3	64.8	48.5	15.4	86.7	66.9	44.9	17.9
$\mathcal{L}_{\text{WAN}}$	1 Pos. & 0 Neg.	84.1	63.1	45.8	17.9	86.5	64.8	46.3	**20.3**
$\mathcal{L}_{\text{EPR}}$	1 Pos. & 0 Neg.	83.8	62.6	46.4	**18.0**	85.5	63.3	46.0	20.0
$\mathcal{L}_{\text{ROLE}}$	1 Pos. & 0 Neg.	**86.5**	**66.3**	**49.5**	16.2	87.9	66.3	43.1	15.0
$\mathcal{L}_{\text{AN-LS}}$ +LinearInit.	1 Pos. & 0 Neg.	-	-	-	-	86.5	**69.2**	50.5	16.6
$\mathcal{L}_{\text{ROLE}}$ +LinearInit.	1 Pos. & 0 Neg.	-	-	-	-	**88.2**	69.0	**51.0**	16.8

Table 1. Multi-label test set mean average precision (MAP) for different multi-label losses on four different image classification datasets. We present results for two scenarios: (i) training a linear classifier on fixed features and (ii) fine-tuning the entire network end-to-end. In all cases the backbone network is an ImageNet pre-trained ResNet-50. All methods below the break use only one positive per image (*i.e.* 1 Pos. & 0 Neg.), while methods above the break use additional supervision. In each column we bold the best performing single positive method and underline the second-best. For each method and we select the hyperparameters that perform the best on the held-out validation set. For losses labeled with "LinearInit." we freeze the weights of the backbone network for the initial epochs of training and then fine-tune the entire network end-to-end for the remaining epochs. Note that this linear initialization phase is identical to the training protocol for the "Linear" results.

Loss	VOC12	COCO	NUS	CUB
$\mathcal{L}_{\text{AN}}$	85.8	63.8	49.3	16.8
$\mathcal{L}_{\text{AN-LS}}$	86.9	65.4	49.7	17.4
$\mathcal{L}_{\text{ROLE}}$	**90.3**	**69.5**	**56.0**	**19.6**

Table 2. Training set MAP for multi-label predictions evaluated with respect to the *full* ground truth labels. These values measure how well each method recovers the true training labels despite being trained with one positive label per image. Note that all results are for the linear case. Hyperparameters and stopping epoch are selected using the validation set as before.