Supplementary material of "Asymmetric metric learning for knowledge transfer"

Mateusz Budnik Yannis Avrithis
Inria, Univ Rennes, CNRS, IRISA

A. More results

Complete contrastive–regression ablation Here, we present the full version of the results of the ablation from Table 2, for all four student-teacher combinations. Apart from mAP, we also report mP@10. Table 5 and Table 6 present the symmetric and asymmetric testing results, respectively. All results agree with the results of Table 2. For symmetric testing, contrastive loss with a single positive and no negatives is again the worst. The addition of the anchor as a positive for itself as well as the negatives improve the results substantially. Contr$^+$, which uses both, performs best in most cases with the exception of VGG16→EfficientNet. For asymmetric testing, regression is the best. The inclusion of the anchor as positive for itself gives better results than without it.

Complete results including mP@10 Table 7 supplements Table 3 by adding mP@10 scores for all the symmetric testing experiments. Similarly, Table 8 adds mP@10 results to all asymmetric testing experiments. Overall, the conclusions drawn based on mAP apply to mP@10 too.

Experiments on $\mathcal{R}$1M distractors Table 9 and Table 10 report symmetric and asymmetric testing results on both $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k with the addition of $\mathcal{R}$1M distractors. The structure of the tables mirrors exactly that of Table 7 and Table 8, which includes both the mAP and mP@10 metrics. This is far more challenging than the standard setting. Therefore, results are lower across the board. Besides this observation, the general conclusions from the previous experiments still apply here, with the gain of our approach being even more pronounced.

In *symmetric testing*, student models trained with Contr$^+$ and contrastive give the best results, often surpassing the performance of the teacher model. For ResNet101 teacher in particular, EfficientNet student outperforms the teacher in all cases, with a gain of up to 3.3% mAP for $\mathcal{R}$Oxf+$\mathcal{R}$1M, while MobileNetV2 is on par or outperforms the teacher in certain cases, with a gain of up to 2.1% mAP for $\mathcal{R}$Oxf+$\mathcal{R}$1M. In *asymmetric testing*, models trained with regression have the highest performance, followed by Contr$^+$. However, the gap in performance compared with symmetric testing is even greater in the presence of $\mathcal{R}$1M.

Student	d	Teacher	Lab	Loss	Self	Pos	Neg	Mining	Medium				Hard			
									$\mathcal{R}$Oxford5k		$\mathcal{R}$Paris6k		$\mathcal{R}$Oxford5k		$\mathcal{R}$Paris6k	
									mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
MobileNetV2	512	VGG16	✓	Contr (4)		✓	✓	hard	**57.3**	77.1	67.1	95.7	31.1	47.3	41.3	80.4
			✓	Contr (4)	✓	✓	✓	hard	**57.3**	78.4	**68.4**	96.1	**31.5**	46.9	**42.2**	78.9
			✓	Contr (4)		✓		hard	55.9	79.2	66.7	95.0	31.1	44.0	40.6	78.9
			✓	Contr (4)	✓	✓		hard	55.5	76.1	67.0	96.0	30.4	44.1	40.9	81.4
				Reg (7)	✓			–	53.3	75.1	67.5	95.6	28.9	43.6	40.9	81.3
	2048	ResNet101	✓	Contr (4)		✓	✓	hard	60.8	81.7	72.1	97.3	36.1	50.4	47.6	85.1
			✓	Contr (4)	✓	✓	✓	hard	**63.2**	84.4	**75.0**	98.0	**37.9**	52.1	**52.0**	87.3
			✓	Contr (4)		✓		hard	51.8	72.5	67.6	96.0	27.6	38.1	41.3	80.0
			✓	Contr (4)	✓	✓		hard	60.6	80.0	74.1	97.0	35.7	49.4	50.9	85.6
				Reg (7)	✓			–	59.8	80.3	73.1	96.9	35.7	49.4	49.5	84.7
EfficientNet-B3	512	VGG16	✓	Contr (4)		✓	✓	hard	56.8	75.7	**70.4**	96.3	31.2	43.9	**45.4**	81.7
			✓	Contr (4)	✓	✓	✓	hard	56.9	75.6	69.0	96.0	31.1	46.7	43.5	80.9
			✓	Contr (4)		✓		hard	56.1	77.0	69.3	96.4	30.1	42.1	44.7	78.4
			✓	Contr (4)	✓	✓		hard	**57.6**	78.3	69.9	96.9	**31.4**	46.7	44.9	82.6
				Reg (7)	✓			–	55.0	75.0	69.4	96.6	27.1	42.3	44.5	80.4
	2048	ResNet101	✓	Contr (4)		✓	✓	hard	66.3	85.3	**77.4**	98.4	41.3	58.9	**55.5**	88.3
			✓	Contr (4)	✓	✓	✓	hard	**66.8**	84.7	77.1	98.6	**42.5**	58.7	**55.5**	87.9
			✓	Contr (4)		✓		hard	61.7	81.7	74.3	97.1	36.1	51.7	51.6	85.9
			✓	Contr (4)	✓	✓		hard	63.8	83.1	75.9	98.3	40.1	54.3	54.4	87.1
				Reg (7)	✓			–	64.9	83.7	74.4	97.7	40.5	55.9	52.4	87.1

Table 5. *Complete contrastive–regression ablation: symmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52]. Lab: using labels in student model training. Pos, Neg: Using positives, negatives. Self: Using anchor (by teacher) as positive for itself (by student). Using asymmetric similarity (3) at training in all cases. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases.

Student	d	Teacher	Lab	Loss	Self	Pos	Neg	Mining	Medium				Hard			
									$\mathcal{R}$Oxford5k		$\mathcal{R}$Paris6k		$\mathcal{R}$Oxford5k		$\mathcal{R}$Paris6k	
									mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
MobileNetV2	512	VGG16	✓	Contr (4)		✓	✓	hard	38.3	53.7	49.8	84.4	18.4	32.8	23.8	55.7
			✓	Contr (4)	✓	✓	✓	hard	42.9	59.1	55.9	88.4	22.6	35.2	31.4	66.3
			✓	Contr (4)		✓		hard	34.1	48.9	47.3	82.0	17.0	25.6	24.5	53.4
			✓	Contr (4)	✓	✓		hard	38.2	52.0	52.2	86.0	15.3	26.2	28.9	64.1
				Reg (7)	✓			–	**48.0**	64.3	**57.9**	90.7	**26.5**	37.9	**32.6**	67.1
	2048	ResNet101	✓	Contr (4)		✓	✓	hard	32.3	49.7	51.5	83.3	9.6	18.3	28.2	62.4
			✓	Contr (4)	✓	✓	✓	hard	47.1	65.4	61.5	92.6	21.8	33.1	37.7	74.1
			✓	Contr (4)		✓		hard	27.3	38.4	47.7	80.9	8.4	15.3	24.3	50.6
			✓	Contr (4)	✓	✓		hard	40.5	58.2	55.8	87.6	17.4	26.3	29.9	63.4
				Reg (7)	✓			–	**49.2**	67.9	**65.0**	92.6	**23.3**	36.9	**40.7**	72.1
EfficientNet-B3	512	VGG16	✓	Contr (4)		✓	✓	hard	43.8	74.7	24.9	39.3	23.0	51.3	6.1	15.6
			✓	Contr (4)	✓	✓	✓	hard	44.7	61.5	58.0	93.3	23.9	37.9	32.4	69.1
			✓	Contr (4)		✓		hard	32.4	45.4	47.8	84.4	14.1	22.0	25.8	56.3
			✓	Contr (4)	✓	✓		hard	41.6	57.5	53.9	90.1	20.3	30.6	30.2	64.0
				Reg (7)	✓			–	**49.4**	70.0	**58.2**	92.4	**26.0**	39.6	**33.0**	70.6
	2048	ResNet101	✓	Contr (4)		✓	✓	hard	37.4	56.8	57.4	90.4	10.9	24.6	33.7	65.9
			✓	Contr (4)	✓	✓	✓	hard	45.2	67.2	63.7	92.1	19.6	35.5	40.9	73.6
			✓	Contr (4)		✓		hard	30.8	44.5	51.2	83.7	10.2	16.1	27.8	57.0
			✓	Contr (4)	✓	✓		hard	40.1	56.7	59.1	91.1	14.6	24.3	35.0	71.0
				Reg (7)	✓			–	**52.9**	71.8	**65.2**	93.3	**27.8**	41.5	**42.4**	71.9

Table 6. *Complete contrastive–regression ablation: asymmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52]. Lab: using labels in student model training. Pos, Neg: Using positives, negatives. Self: Using anchor (by teacher) as positive for itself (by student). Using asymmetric similarity (3) at training in all cases. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases.

STUDENT	d	TEACHER	LAB	LOSS	MINING	ASYM	MEDIUM				HARD			
							$\mathcal{R}$Oxf		$\mathcal{R}$Par		$\mathcal{R}$Oxf		$\mathcal{R}$Par	
							mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
VGG16	512		✓	Contr (4)	hard		60.9	81.9	69.3	97.4	32.9	50.9	44.2	83.1
ResNet101	2048		✓	Contr (4)	hard		65.4	85.7	76.7	98.4	40.1	56.6	55.2	87.7
MobileNetV2	512		✓	Contr (4)	hard		53.6	75.8	66.4	96.7	28.8	42.9	39.7	79.0
	2048		✓	Contr (4)	hard		56.1	79.0	68.5	98.1	30.3	46.0	42.0	82.6
EfficientNet-B3	512		✓	Contr (4)	hard		53.8	76.6	70.9	96.6	26.2	42.3	46.0	83.7
	2048		✓	Contr (4)	hard		59.6	86.1	75.1	95.1	33.3	46.0	51.9	87.6
MobileNetV2	512	VGG16	✓	Contr$^+$ (10)	hard	✓	**57.3**	78.4	**68.4**	96.1	**31.5**	46.9	**42.2**	78.9
			✓	Contr (4)	hard	✓	**57.3**	77.1	67.1	95.7	31.1	47.3	41.3	80.4
			✓	Triplet (5)	hard	✓	37.0	55.2	62.7	94.4	11.6	23.0	36.4	73.7
			✓	MS (6)	hard	✓	36.8	55.2	62.8	94.4	11.5	22.2	36.5	75.0
				Reg (7)	–	✓	53.3	75.1	67.5	95.6	28.9	43.6	40.9	81.3
				RKD (8)	random		46.2	68.1	64.3	94.7	21.8	32.8	37.6	72.3
				DR (9)	random		45.2	66.5	60.6	92.1	24.6	34.9	33.1	74.1
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	**63.2**	84.4	**75.0**	98.0	**37.9**	52.1	**52.0**	87.3
			✓	Contr (4)	hard	✓	60.8	81.7	72.1	97.3	36.1	50.4	47.6	85.1
			✓	Triplet (5)	hard	✓	45.5	66.1	68.0	96.1	19.6	33.5	43.4	80.6
			✓	MS (6)	hard	✓	44.5	65.4	68.1	96.1	17.9	32.1	43.2	80.1
				Reg (7)	–	✓	59.8	80.3	73.1	96.9	35.7	49.4	49.5	84.7
				RKD (8)	random		56.1	79.3	69.8	96.3	31.8	46.0	44.2	82.3
				DR (9)	random		43.4	65.6	59.3	93.4	20.8	31.8	31.6	69.0
EfficientNet-B3	512	VGG16	✓	Contr$^+$ (10)	hard	✓	**56.9**	75.6	69.0	96.0	31.1	46.7	43.5	80.9
			✓	Contr (4)	hard	✓	56.8	75.7	**70.4**	96.3	**31.2**	43.9	**45.4**	81.7
			✓	Triplet (5)	hard	✓	33.7	48.5	64.6	94.4	8.0	20.1	40.3	76.1
			✓	MS (6)	hard	✓	33.9	49.5	64.9	94.4	8.1	20.4	40.6	76.9
				Reg (7)	–	✓	55.0	75.0	69.4	96.6	27.1	42.3	44.5	80.4
				RKD (8)	random		51.6	71.4	67.6	95.3	26.2	38.5	41.7	81.1
				DR (9)	random		52.4	72.1	65.2	95.4	26.5	38.1	37.2	73.7
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	**66.8**	84.7	77.1	98.6	**42.5**	58.7	**55.5**	87.9
			✓	Contr (4)	hard	✓	66.3	85.3	**77.4**	98.4	41.3	58.9	**55.5**	88.3
			✓	Triplet (5)	hard	✓	39.5	57.3	69.4	95.9	11.6	24.3	45.8	81.1
			✓	MS (6)	hard	✓	39.9	57.4	69.7	95.7	11.7	24.2	46.2	81.4
				Reg (7)	–	✓	64.9	83.7	74.4	97.7	40.5	55.9	52.4	87.1
				RKD (8)	random		56.3	75.8	73.0	98.4	30.5	46.4	50.4	82.3
				DR (9)	random		52.2	72.1	66.3	95.4	27.3	39.9	40.1	79.0

Table 7. *Symmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52]. LAB: using labels in student model training. ASYM: Using asymmetric similarity (3) at training. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases.

STUDENT	d	TEACHER	LAB	LOSS	MINING	ASYM	MEDIUM				HARD			
							$\mathcal{R}$Oxf		$\mathcal{R}$Par		$\mathcal{R}$Oxf		$\mathcal{R}$Par	
							mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
VGG16	512		✓	Contr (4)	hard		60.9	81.9	69.3	97.4	32.9	50.9	44.2	83.1
ResNet101	2048		✓	Contr (4)	hard		65.4	85.7	76.7	98.4	40.1	56.6	55.2	87.7
MobileNetV2	512		✓	Contr (4)	hard		53.6	75.8	66.4	96.7	28.8	42.9	39.7	79.0
	2048		✓	Contr (4)	hard		56.1	79.0	68.5	98.1	30.3	46.0	42.0	82.6
EfficientNet-B3	512		✓	Contr (4)	hard		53.8	76.6	70.9	96.6	26.2	42.3	46.0	83.7
	2048		✓	Contr (4)	hard		59.6	86.1	75.1	95.1	33.3	46.0	51.9	87.6
MobileNetV2	512	VGG16	✓	Contr$^+$ (10)	hard	✓	42.9	59.1	55.9	88.4	22.6	35.2	31.4	66.3
			✓	Contr (4)	hard	✓	38.3	53.7	49.8	84.4	18.4	32.8	23.8	55.7
			✓	Triplet (5)	hard	✓	1.8	0.0	4.3	1.3	0.7	0.0	2.8	1.4
			✓	MS (6)	hard	✓	1.9	0.0	4.3	1.6	0.8	0.0	2.7	1.6
				Reg (7)	–	✓	**48.0**	64.3	**57.9**	90.7	**26.5**	37.9	**32.6**	67.1
				RKD (8)	random		2.0	0.0	4.1	1.0	0.8	0.0	2.6	1.0
				DR (9)	random		1.7	0.0	3.8	0.3	0.7	0.0	2.4	0.3
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	47.1	65.4	61.5	92.6	21.8	33.1	37.7	74.1
			✓	Contr (4)	hard	✓	32.3	49.7	51.5	83.3	9.6	18.3	28.2	62.4
			✓	Triplet (5)	hard	✓	1.3	0.0	3.7	1.4	0.7	0.0	2.4	1.4
			✓	MS (6)	hard	✓	1.4	0.3	3.6	1.0	0.7	0.3	2.3	0.9
				Reg (7)	–	✓	**49.2**	67.9	**65.0**	92.6	**23.3**	36.9	**40.7**	72.1
				RKD (8)	random		1.6	1.3	4.1	2.3	0.8	1.1	2.5	1.6
				DR (9)	random		1.5	0.4	3.7	3.7	0.6	0.3	2.3	2.4
EfficientNet-B3	512	VGG16	✓	Contr$^+$ (10)	hard	✓	44.7	61.5	58.0	93.3	23.9	37.9	32.4	69.1
			✓	Contr (4)	hard	✓	43.8	74.7	24.9	39.3	23.0	51.3	6.1	15.6
			✓	Triplet (5)	hard	✓	1.4	0.0	4.0	0.0	0.6	0.0	2.5	0.0
			✓	MS (6)	hard	✓	1.4	0.0	3.9	0.0	0.6	0.0	2.5	0.0
				Reg (7)	–	✓	**49.4**	70.0	**58.2**	92.4	**26.0**	39.6	**33.0**	70.6
				RKD (8)	random		1.3	0.0	3.8	0.7	0.6	0.0	2.5	0.3
				DR (9)	random		1.4	0.0	3.8	1.3	0.6	0.0	2.5	1.0
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	45.2	67.2	63.7	92.1	19.6	35.5	40.9	73.6
			✓	Contr (4)	hard	✓	37.4	56.8	57.4	90.4	10.9	24.6	33.7	65.9
			✓	Triplet (5)	hard	✓	1.5	0.7	4.0	1.6	0.7	0.7	2.5	0.9
			✓	MS (6)	hard	✓	1.5	0.7	4.0	1.4	0.7	0.7	2.4	1.0
				Reg (7)	–	✓	**52.9**	71.8	**65.2**	93.3	**27.8**	41.5	**42.4**	71.9
				RKD (8)	random		1.6	0.7	3.8	1.6	0.7	0.4	2.4	0.7
				DR (9)	random		2.0	2.4	3.5	0.4	0.7	0.3	2.2	0.4

Table 8. *Asymmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52]. LAB: using labels in student model training. ASYM: Using asymmetric similarity (3) at training. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases. The results without a teacher in the top block correspond to symmetric testing (same as in Table 7) and are only added here for convenience.

STUDENT	d	TEACHER	LAB	LOSS	MINING	ASYM	MEDIUM				HARD			
							$\mathcal{R}$Oxf+$\mathcal{R}$1M		$\mathcal{R}$Par+$\mathcal{R}$1M		$\mathcal{R}$Oxf+$\mathcal{R}$1M		$\mathcal{R}$Par+$\mathcal{R}$1M	
							mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
VGG16	512		✓	Contr (4)	hard		42.6	68.1	45.4	94.1	19.0	29.4	19.1	64.9
ResNet101	2048		✓	Contr (4)	hard		45.2	71.1	52.3	95.3	19.9	34.9	24.7	73.3
MobileNetV2	512		✓	Contr (4)	hard		34.1	59.2	38.7	91.0	14.2	22.3	14.1	51.0
	2048		✓	Contr (4)	hard		37.4	66.2	42.0	91.0	17.6	28.1	17.2	57.7
EfficientNet-B3	512		✓	Contr (4)	hard		34.6	59.0	43.4	92.9	11.8	21.1	17.6	63.9
	2048		✓	Contr (4)	hard		36.6	63.1	45.4	94.4	17.4	23.7	19.2	65.4
MobileNetV2	512	VGG16	✓	Contr$^+$ (10)	hard	✓	35.1	59.3	**39.8**	90.1	**17.0**	23.1	**13.6**	52.0
			✓	Contr (4)	hard	✓	**36.7**	60.1	37.9	90.0	16.4	24.0	13.3	51.4
			✓	Triplet (5)	hard	✓	17.1	34.4	30.9	85.6	2.5	5.4	9.5	40.1
			✓	MS (6)	hard	✓	17.1	33.5	31.0	85.9	2.4	5.4	9.5	39.7
				Reg (7)	–	✓	32.6	56.5	37.0	89.7	13.5	21.9	11.8	47.0
				RKD (8)	random		29.2	51.8	34.1	85.7	13.0	17.4	9.5	39.7
				DR (9)	random		25.4	46.5	32.1	84.7	12.6	16.9	8.2	37.7
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	**45.1**	71.6	**47.5**	94.9	**22.0**	33.1	**18.8**	62.4
			✓	Contr (4)	hard	✓	42.1	65.7	45.9	93.7	20.8	30.9	18.4	62.3
			✓	Triplet (5)	hard	✓	28.3	50.4	42.9	90.9	5.5	12.4	15.7	51.7
			✓	MS (6)	hard	✓	24.8	46.9	39.5	88.4	6.4	11.9	14.2	51.0
				Reg (7)	–	✓	41.5	65.8	45.9	92.1	18.6	30.7	18.1	59.1
				RKD (8)	random		38.1	65.0	43.7	92.4	16.7	25.9	15.6	54.4
				DR (9)	random		23.6	45.7	29.6	83.6	11.1	13.8	7.9	35.7
EfficientNet-B3	512	VGG16	✓	Contr$^+$ (10)	hard	✓	35.7	58.7	**42.3**	91.9	13.8	23.2	15.2	55.3
			✓	Contr (4)	hard	✓	**36.8**	61.1	41.6	90.6	**16.4**	24.9	**15.6**	53.4
			✓	Triplet (5)	hard	✓	11.3	25.0	28.8	79.9	0.2	1.3	11.2	48.1
			✓	MS (6)	hard	✓	11.4	25.3	29.1	80.1	0.2	1.3	11.3	48.9
				Reg (7)	–	✓	35.5	61.0	40.3	90.9	15.8	23.6	14.0	53.1
				RKD (8)	random		26.1	46.9	39.5	90.3	6.7	12.6	13.6	50.6
				DR (9)	random		25.5	47.8	30.7	85.9	5.8	11.9	7.9	37.3
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	**47.3**	72.3	51.7	96.6	**23.2**	37.6	23.5	69.0
			✓	Contr (4)	hard	✓	46.4	72.1	**52.7**	96.7	22.0	34.0	**25.0**	70.3
			✓	Triplet (5)	hard	✓	16.2	31.6	35.3	84.0	0.5	2.3	14.9	53.3
			✓	MS (6)	hard	✓	16.3	32.1	35.9	84.0	0.5	2.4	15.1	53.9
				Reg (7)	–	✓	45.9	73.5	49.3	96.0	21.7	35.4	20.9	66.0
				RKD (8)	random		38.4	63.0	49.4	95.6	16.6	25.9	21.5	69.3
				DR (9)	random		29.1	49.7	35.8	88.4	9.8	14.6	10.3	43.3

Table 9. *Symmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52] with $\mathcal{R}$1M distractors. LAB: using labels in student model training. ASYM: Using asymmetric similarity (3) at training. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases.

STUDENT	d	TEACHER	LAB	LOSS	MINING	ASYM	MEDIUM				HARD			
							$\mathcal{R}$Oxf+$\mathcal{R}$1M		$\mathcal{R}$Par+$\mathcal{R}$1M		$\mathcal{R}$Oxf+$\mathcal{R}$1M		$\mathcal{R}$Par+$\mathcal{R}$1M	
							mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
VGG16	512		✓	Contr (4)	hard		42.6	68.1	45.4	94.1	19.0	29.4	19.1	64.9
ResNet101	2048		✓	Contr (4)	hard		45.2	71.1	52.3	95.3	19.9	34.9	24.7	73.3
MobileNetV2	512		✓	Contr (4)	hard		34.1	59.2	38.7	91.0	14.2	22.3	14.1	51.0
	2048		✓	Contr (4)	hard		37.4	66.2	42.0	91.0	17.6	28.1	17.2	57.7
EfficientNet-B3	512		✓	Contr (4)	hard		34.6	59.0	43.4	92.9	11.8	21.1	17.6	63.9
	2048		✓	Contr (4)	hard		36.6	63.1	45.4	94.4	17.4	23.7	19.2	65.4
MobileNetV2	512	VGG16	✓	Contr$^+$ (10)	hard	✓	14.4	27.0	22.5	61.9	7.2	11.0	7.4	31.6
			✓	Contr (4)	hard	✓	8.2	24.4	16.5	49.4	1.4	5.7	4.7	22.9
			✓	Triplet (5)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			✓	MS (6)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				Reg (7)	–	✓	**22.6**	40.1	**29.0**	77.9	**10.4**	16.7	**10.7**	43.0
				RKD (8)	random		0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
				DR (9)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	18.0	38.5	28.8	80.7	6.3	12.6	8.8	39.1
			✓	Contr (4)	hard	✓	5.5	23.7	13.3	41.7	1.0	1.9	3.4	17.1
			✓	Triplet (5)	hard	✓	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
			✓	MS (6)	hard	✓	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
				Reg (7)	–	✓	**26.5**	46.0	**34.6**	86.0	**7.8**	16.1	**12.7**	49.0
				RKD (8)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				DR (9)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EfficientNet-B3	512	VGG16	✓	Contr$^+$ (10)	hard	✓	18.6	34.6	25.8	68.0	7.8	13.9	8.9	35.0
			✓	Contr (4)	hard	✓	5.9	17.1	13.8	42.1	0.6	2.4	3.1	17.0
			✓	Triplet (5)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			✓	MS (6)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				Reg (7)	–	✓	**24.7**	45.4	**32.8**	83.4	**12.2**	18.4	**12.5**	50.9
				RKD (8)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				DR (9)	random		0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
	2048	ResNet101	✓	Contr$^+$ (10)	hard	✓	16.8	37.7	27.2	75.4	2.6	10.1	8.8	35.6
			✓	Contr (4)	hard	✓	5.4	16.6	10.5	38.7	0.6	3.0	2.2	15.0
			✓	Triplet (5)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			✓	MS (6)	hard	✓	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				Reg (7)	–	✓	**29.7**	51.8	**39.0**	88.0	**10.4**	21.2	**16.0**	53.4
				RKD (8)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
				DR (9)	random		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 10. *Asymmetric testing* on $\mathcal{R}$Oxford5k and $\mathcal{R}$Paris6k [52] with $\mathcal{R}$1M distractors. LAB: using labels in student model training. ASYM: Using asymmetric similarity (3) at training. Best mAP highlighted per teacher-student pair. GeM pooling and learned whitening [54] used in all cases. The results without a teacher in the top block correspond to symmetric testing (same as in Table 9) and are only added here for convenience.