Asymmetric metric learning for knowledge transfer

Mateusz Budnik and Yannis Avrithis

Inria Rennes-Bretagne Atlantique

CVPR 2021

Virtual, June 2021

asymmetric metric learning (AML)

- instance-level image retrieval
- asymmetric testing: database represented by large network, queries by lightweight network on device, no re-indexing
- asymmetric metric learning: use asymmetric representations at training in teacher-student setup
- applies to both symmetric and asymmetric testing
- combines of knowledge transfer with supervised metric learning

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

asymmetric metric learning (AML)

- instance-level image retrieval
- asymmetric testing: database represented by large network, queries by lightweight network on device, no re-indexing
- asymmetric metric learning: use asymmetric representations at training in teacher-student setup
- applies to both symmetric and asymmetric testing
- combines of knowledge transfer with supervised metric learning

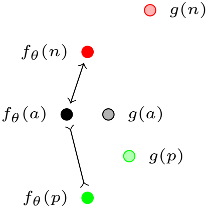
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

asymmetric metric learning (AML)

- instance-level image retrieval
- asymmetric testing: database represented by large network, queries by lightweight network on device, no re-indexing
- asymmetric metric learning: use asymmetric representations at training in teacher-student setup
- applies to both symmetric and asymmetric testing
- combines of knowledge transfer with supervised metric learning

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

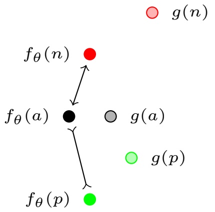
metric learning and knowledge transfer



symmetric

- labels used, teacher not used
- positive pairs of examples mutually attracted and negative pairs are repulsed in student space

metric learning and knowledge transfer



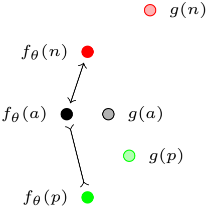
symmetric

- labels used, teacher not used (f_θ: student, g: teacher)
- contrastive $\ell_{\mathrm{C}}(a;\theta)$: independently, positive examples p close to anchor a, negative n farther from a by margin m in student space

$$\sum_{n \in N(a)} [s_\theta(a,n) - m]_+ - \sum_{p \in P(a)} s_\theta(a,p)$$

Hadsell, Chopra, Lecun. CVPR 2006. Dimensionality reduction by learning an invariant mapping.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



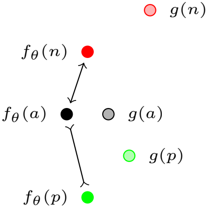
symmetric

- labels used, teacher not used (f_θ: student, g: teacher)
- triplet $\ell_{\mathrm{T}}(a;\theta)$: positive examples p closer to the anchor a than negative n by margin m in student space

$$\sum_{(p,n)\in L(a)} [s_\theta(a,n) - s_\theta(a,p) + m]_+$$

Wang, Song, Leung, Rosenberg, Wang, Philbin, Chen, Wu. CVPR 2014. Learning fine-grained image similarity with deep ranking.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



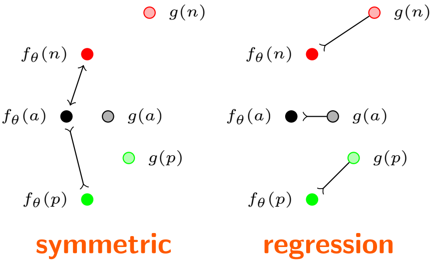
symmetric

- labels used, teacher not used (f_θ: student, g: teacher)
- multi-similarity $\ell_{\mathrm{MS}}(a;\theta)$: positives p (negatives n) farthest from (nearest to) anchor a receive the greatest relative weight

$$\frac{1}{\alpha}\log\left(1+\sum_{p\in P(a)} e^{-\alpha(s_\theta(a,p)-m)}\right)+\frac{1}{\beta}\log\left(1+\sum_{n\in N(a)} e^{\beta(s_\theta(a,n)-m)}\right)$$

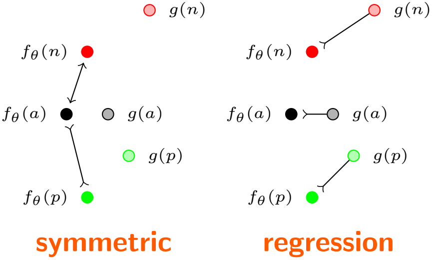
Wang, Han, Huang, Dong, Scott. CVPR 2019. Multi-similarity loss with general pair weighting for deep metric learning.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



- labels not used, teacher used
- examples in student space attracted to corresponding examples in teacher space

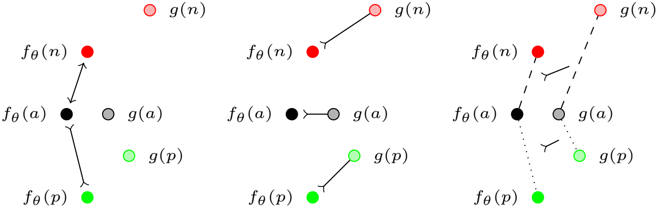
metric learning and knowledge transfer



- labels not used, teacher used (f_θ: student, g: teacher)
- regression $\ell_R(a;\theta)$: representations of same example a by two models f_θ, g close to each other, where g is fixed

$$-s_\theta^{\text{asym}}(a, a) = -\operatorname{sim}(f_\theta(a), g(a))$$

metric learning and knowledge transfer

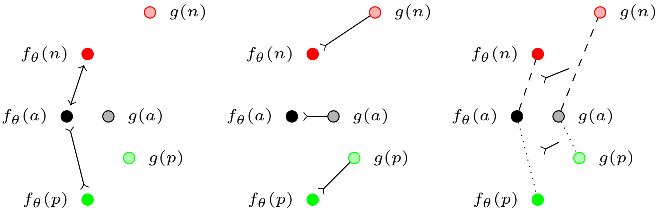


symmetric **regression** **relational**

- labels not used, teacher used
- pairwise / groupwise relations like distances, angles or ranks encouraged to be compatible in both spaces

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



symmetric **regression** **relational**

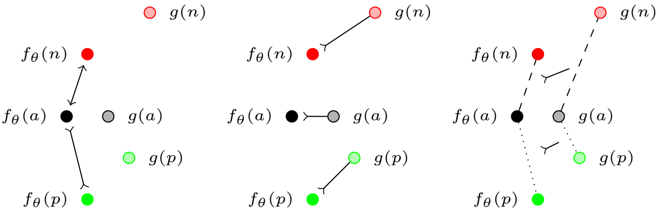
- labels not used, teacher used (f_θ: student, g: teacher)
- relational distillation $\ell_{\mathrm{RKD}}(a;\theta)$: measurements $\psi(\mathbf{a}, \mathbf{x}, \dots)$ of same examples $(a, x, \dots)$ by two models f_θ, g close to each other

$$\sum_{(x,\dots)\in U(a)^n} -\operatorname{sim}(\psi(f_\theta(a), f_\theta(x), \dots), \psi(g(a), g(x), \dots))$$

e.g. distance $\|\mathbf{a} - \mathbf{x}\|$, angle $\operatorname{sim}(\mathbf{a} - \mathbf{x}, \mathbf{a} - \mathbf{y})$; regression $\psi(\mathbf{a}) := \mathbf{a}$

Park, Kim, Lu, Cho. CVPR 2019. Relational knowledge distillation.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



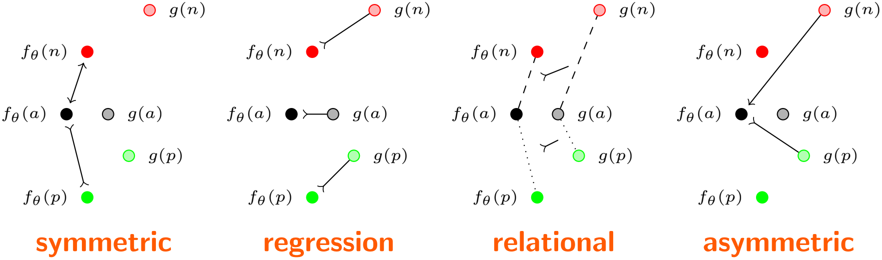
symmetric **regression** **relational**

- labels not used, teacher used (f_θ: student, g: teacher)
- DarkRank $\ell_{\mathrm{DR}}(a;\theta)$: examples $y \in V(a,x)$ mapped farther from anchor a than x in teacher space do the same in student space:

$$-\sum_{x \in U(a)} \left(s_\theta^{\mathrm{sym}}(a,x) - \log \sum_{y \in V(a,x)} e^{s_\theta^{\mathrm{sym}}(a,y)} \right)$$

Chen, Wang, Zhang. AAAI 2018. DarkRank: Accelerating deep metric learning via cross sample similarities transfer.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

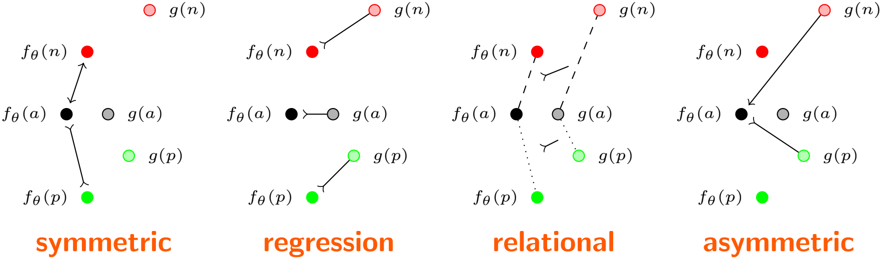
metric learning and knowledge transfer



- both labels and teacher used
- anchors in student space attracted to positives and repulsed from negatives in teacher space

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

metric learning and knowledge transfer



- both labels and teacher used (f_θ: student, g: teacher)
- Asymmetric Metric Learning (AML): simply use

$$s_\theta^{\mathrm{asym}}(a, x) := \mathrm{sim}(f_\theta(a), g(x))$$

with any supervised metric learning loss like ℓ_{C}, ℓ_{T}, ℓ_{MS}

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

best loss functions

- regression (Reg)

$$\ell_{\mathrm{R}}(a;\theta) := -s_\theta^{\mathrm{asym}}(a,a) = -\operatorname{sim}(f_\theta(a), g(a))$$

- asymmetric contrastive (Contr)

$$\ell_{\mathrm{C}}(a;\theta) := \sum_{n\in N(a)} [s_\theta(a,n) - m]_+ - \sum_{p\in P(a)} s_\theta(a,p)$$

- asymmetric contrastive + regression (Contr$^+$)

$$\ell_{\mathrm{C}^+}(a;\theta) := \sum_{n\in N(a)} [s_\theta(a,n) - m]_+ - \sum_{p\in P(a)} s_\theta(a,p) \quad - s_\theta(a,a)$$

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

best loss functions

- regression (Reg)

$$\ell_{\mathrm{R}}(a;\theta) := \boxed{-s_\theta^{\mathrm{asym}}(a,a)} = -\operatorname{sim}(f_\theta(a), g(a))$$

- asymmetric contrastive (Contr)

$$\ell_{\mathrm{C}}(a;\theta) := \boxed{\sum_{n \in N(a)} [s_\theta(a,n) - m]_+ - \sum_{p \in P(a)} s_\theta(a,p)}$$

- asymmetric contrastive + regression (Contr$^+$)

$$\ell_{\mathrm{C}^+}(a;\theta) := \boxed{\sum_{n \in N(a)} [s_\theta(a,n) - m]_+ - \sum_{p \in P(a)} s_\theta(a,p)} \boxed{-\ s_\theta(a,a)}$$

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

test set: revisited Oxford and Paris

- 11 + 11 landmarks, 70 + 70 queries, 5k + 6k images, easy/hard
- 1M distractor images
- performance measured by mAP: positive ranked first

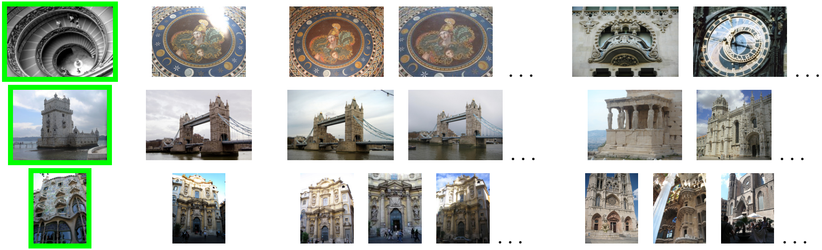
Radenovic, Iscen, Tolias, Avrithis, Chum. CVPR 2018. Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

training set: SfM120k (positives)

- camera position (closest to query)
- number of inliers (co-observed 3D points with query)
- according to SIFT descriptors

Radenovic, Tolias, Chum. ECCV 2016. CNN Image Retrieval Learns From BoW: Unsupervised Fine-Tuning with Hard Examples.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

training set: SfM120k (negatives)



- k-nearest neighbors from non-matching clusters
- at most one image per cluster
- according to learned descriptors

Radenovic, Tolias, Chum. ECCV 2016. CNN Image Retrieval Learns From BoW: Unsupervised Fine-Tuning with Hard Examples.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

network models

Network	Teacher	d	GFLOPS	Param(M)
ResNet101		2048	42.85	42.50
EfficientNet-B3		1536	5.36	10.70
	ResNet101	2048	6.26	13.84

- teacher: ResNet101 (RN101)
- student: EfficientNet-B3 (EN-B3), dimensions d adapted to teacher
- $7\times$ less FLOPS
- $3\times$ less parameters

Tan and Le. ICML 2019. EfficientNet: Rethinking model scaling for convolutional neural networks.
Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr$^+$	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr$^+$: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr$^+$	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr$^+$: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr^{+}	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr^{+}: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr$^+$	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr$^+$: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr^{+}	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr^{+}: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr^+	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr^+: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr$^+$	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr$^+$: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

symmetric testing

Stu	d	Tea	Lab	Mining	Asym	Loss	Medium $\mathcal{R}$Oxf	Medium $\mathcal{R}$Par	Hard $\mathcal{R}$Oxf	Hard $\mathcal{R}$Par
RN101	2048		✓	hard		Contr	65.4	76.7	40.1	55.2
EN-B3	512		✓	hard		Contr	53.8	70.9	26.2	46.0
	2048		✓	hard		Contr	59.6	75.1	33.3	51.9
EN-B3	2048	RN101	✓	hard	✓	Contr$^+$	**66.8**	77.1	**42.5**	**55.5**
			✓	hard	✓	Contr	66.3	**77.4**	41.3	**55.5**
			✓	hard	✓	Triplet	39.5	69.4	11.6	45.8
			✓	hard	✓	MS	39.9	69.7	11.7	46.2
				–	✓	Reg	64.9	74.4	40.5	52.4
				random		RKD	56.3	73.0	30.5	50.4
				random		DR	40.3	69.9	11.8	46.4

- Contr, Contr$^+$: student beats teacher
- Reg: second best, slightly below teacher
- everything else fails (worse than student alone)

Budnik and Avrithis. CVPR 2021. Asymmetric Metric Learning for Knowledge Transfer.

Mateusz

Yannis

paper

https://arxiv.org/abs/2006.16331

code

https://github.com/budnikm/asymmetric_metric_learning

more

https://avrithis.net