How Transferable are Contrastive Representations?

Rogerio Feris

Principal Scientist and Manager

MIT-IBM Watson AI Lab, IBM Research

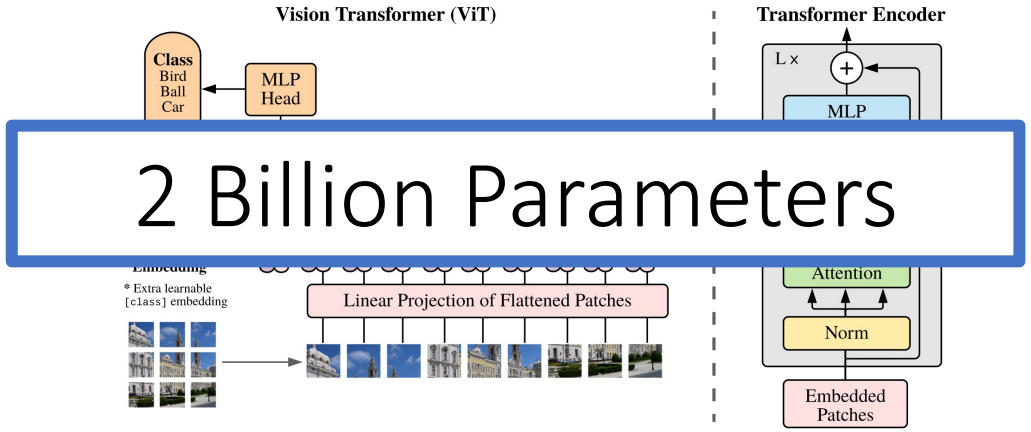
CVPR 2021 Workshop on Learning from Limited and Imperfect Data

June 20, 2021

Significant Progress in Visual Recognition Using Massive Labeled Datasets

3 Billion Images (weakly labelled)

+



2 Billion Parameters

90.45% Top-1 Accuracy in ImageNet

Xiaohua Zhai et al. “Scaling Vision Transformers”, Arxiv 2021

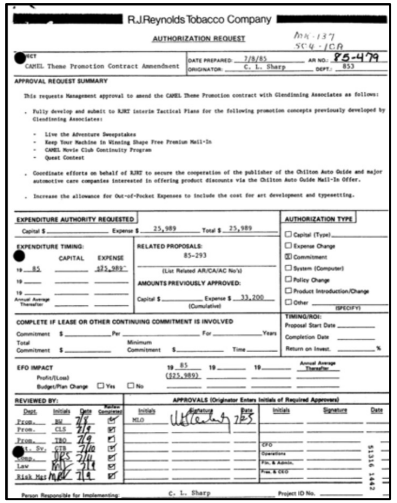
Doing Well on ImageNet is not Enough

Medical Diagnosis

Aerial Imagery

Scientific Imaging

Document Analysis



R.J.Reynolds Tobacco Company

AUTHORIZATION REQUEST

SUBJECT: CAMEL Theme Promotion Contract Amendment

DATE PREPARED: 7/8/85 AR NO.: 85-479

ORIGINATOR: C. L. Sharp DEPT.: 853

APPROVAL REQUEST SUMMARY

This requests Management approval to amend the CAMEL Theme Promotion contract with Glendinning Associates as follows:

- Fully develop and submit to RJRT interim Tactical Plans for the following promotion concepts previously developed by Glendinning Associates:
  - Live the Adventure Sweepstakes
  - Keep Your Machine In Winning Shape Free Premium Mail-In
  - CAMEL Movie Club Continuity Program
  - Quest Contest
- Coordinate efforts on behalf of RJRT to secure the cooperation of the publisher of the Chilton Auto Guide and major automotive care companies interested in offering product discounts via the Chilton Auto Guide Mail-In Offer.
- Increase the allowance for Out-of-Pocket Expenses to include the cost for art development and typesetting.

EXPENDITURE AUTHORITY REQUESTED

Capital $ _____ Expense $ 25,989 Total $ 25,989

EXPENDITURE TIMING: CAPITAL EXPENSE

19 85 _____ (25,989)

RELATED PROPOSALS: 85-293

(List Related AR/CAI/AC No's)

AMOUNTS PREVIOUSLY APPROVED:

Capital $ _____ Expense $ 33,200 (Cumulative)

AUTHORIZATION TYPE

☐ Capital (Type) _____
☐ Expense Change
☒ Commitment
☐ System (Computer)
☐ Policy Change
☐ Product Introduction/Change
☐ Other _____ (SPECIFY)

COMPLETE IF LEASE OR OTHER CONTINUING COMMITMENT IS INVOLVED

Commitment $ _____ Per _____ For _____ Years

Total Commitment $ _____ Minimum Commitment $ _____ Time _____

TIMING/ROI:

Proposal Start Date _____

Completion Date _____

Return on Invest. _____ %

EPO IMPACT

Profit/(Loss) 19 85 (325,989) 19 _____ 19 _____ Annual Average Thereafter

Budget/Plan Change ☐ Yes ☐ No

REVIEWED BY:

Dept.	Initials	Date	Notes Considered
Treas.	BV	7/9	BT
Prom.	CLS	7/8	BT
Treas.	TBO	7/9	BT
V. Pr.	DTS	7/10	BT
Comp.	BRC	7/11	BT
Law	BDJ	7/11	BT
Risk Mgt	MJBZ	7/9	BT

APPROVALS (Originator Enters Initials of Required Approvers)

	Initials	Signature	Date
M/D	[illegible]	[illegible]	[illegible]

Person Responsible for Implementing: C. L. Sharp Project ID No. _____

… and many other application domains!

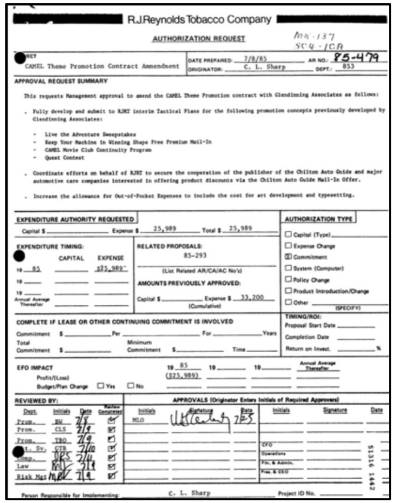
Doing Well on ImageNet is not Enough

Medical Diagnosis

Aerial Imagery

Scientific Imaging

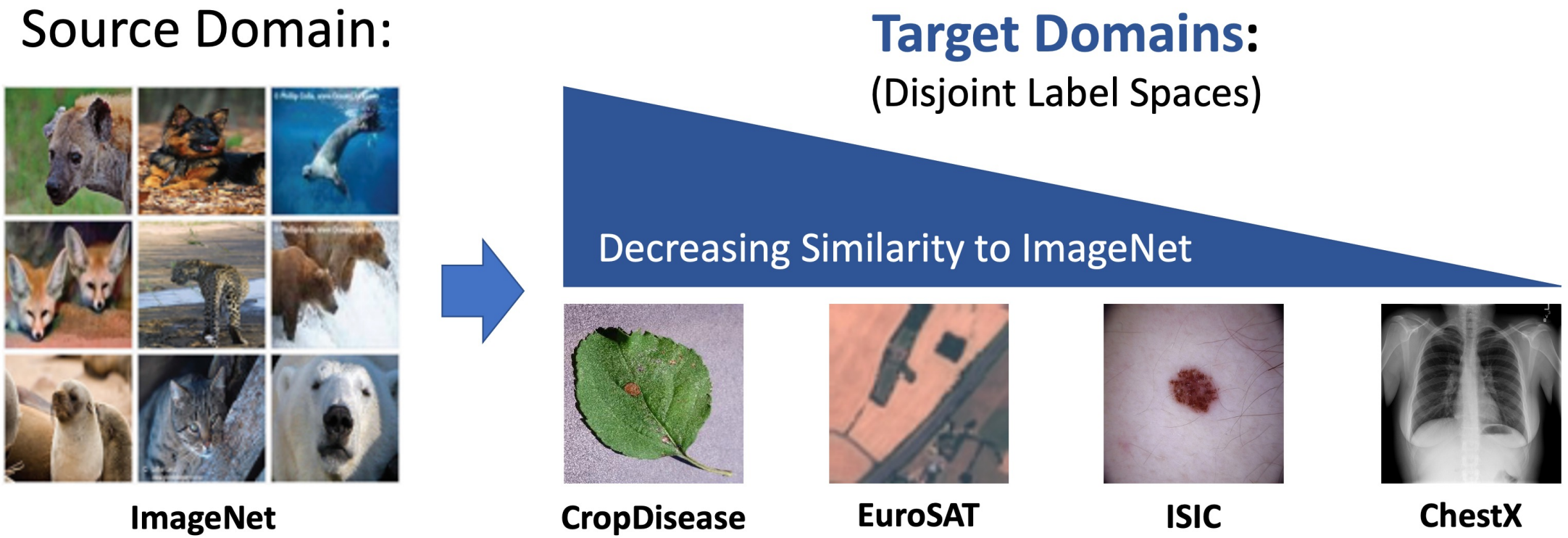
Document Analysis



Challenges:

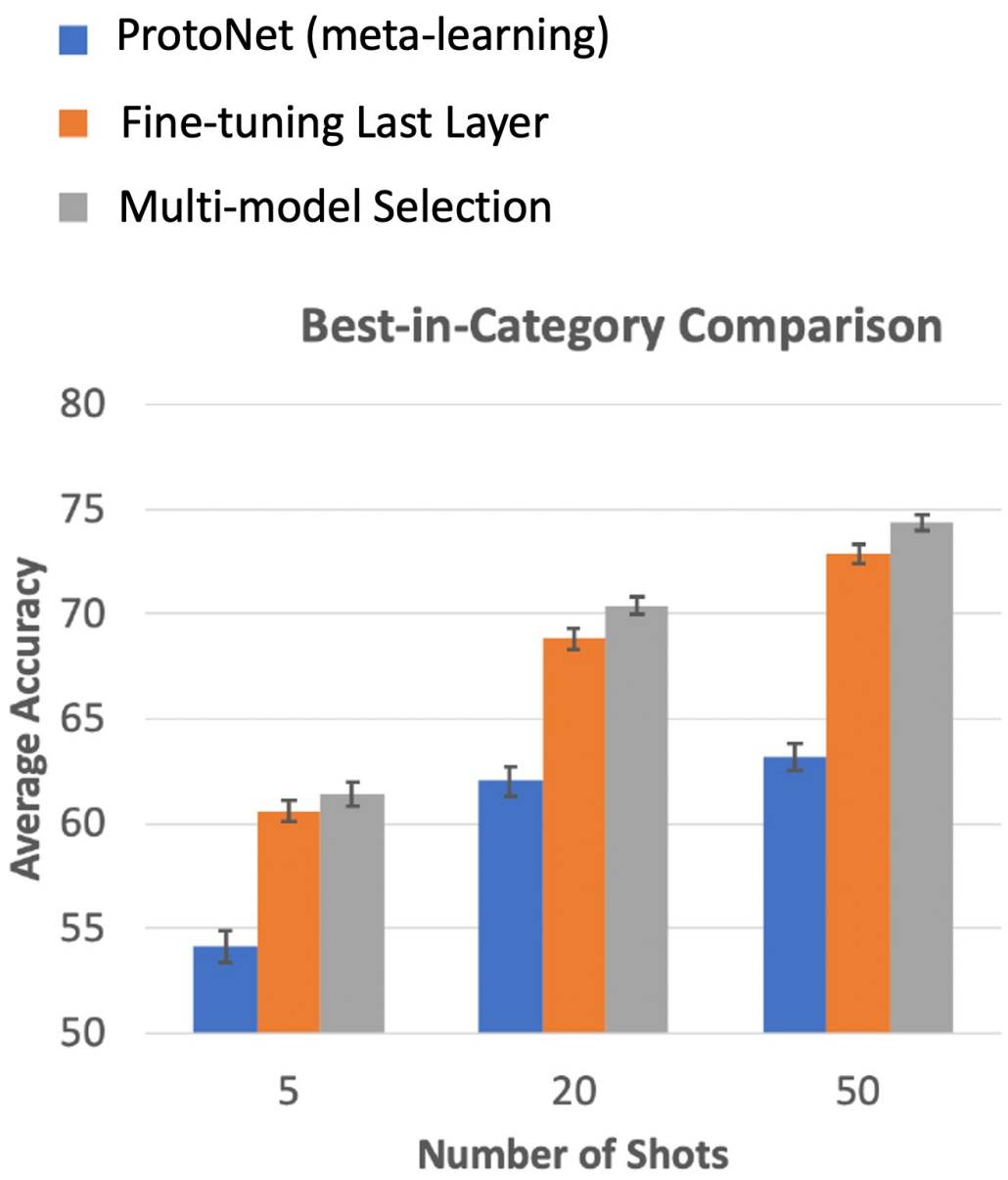
- Limited labeled data for many of these application domains
- Large domain mismatch from ImageNet

Cross-Domain Few-Shot Learning Benchmark



Guo et al. “A broader study of cross-domain few-shot learning” ECCV 2020

Guo et al. “A broader study of cross-domain few-shot learning” ECCV 2020

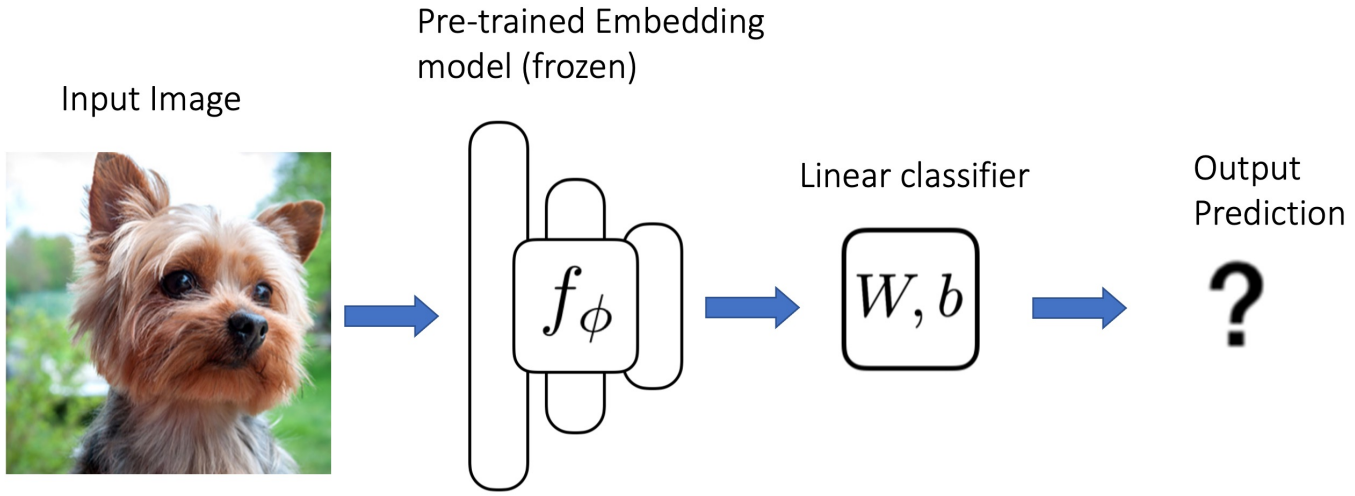


Key Takeaway

Simple fine-tuning outperforms all SOTA meta-learning methods by a large margin in this benchmark

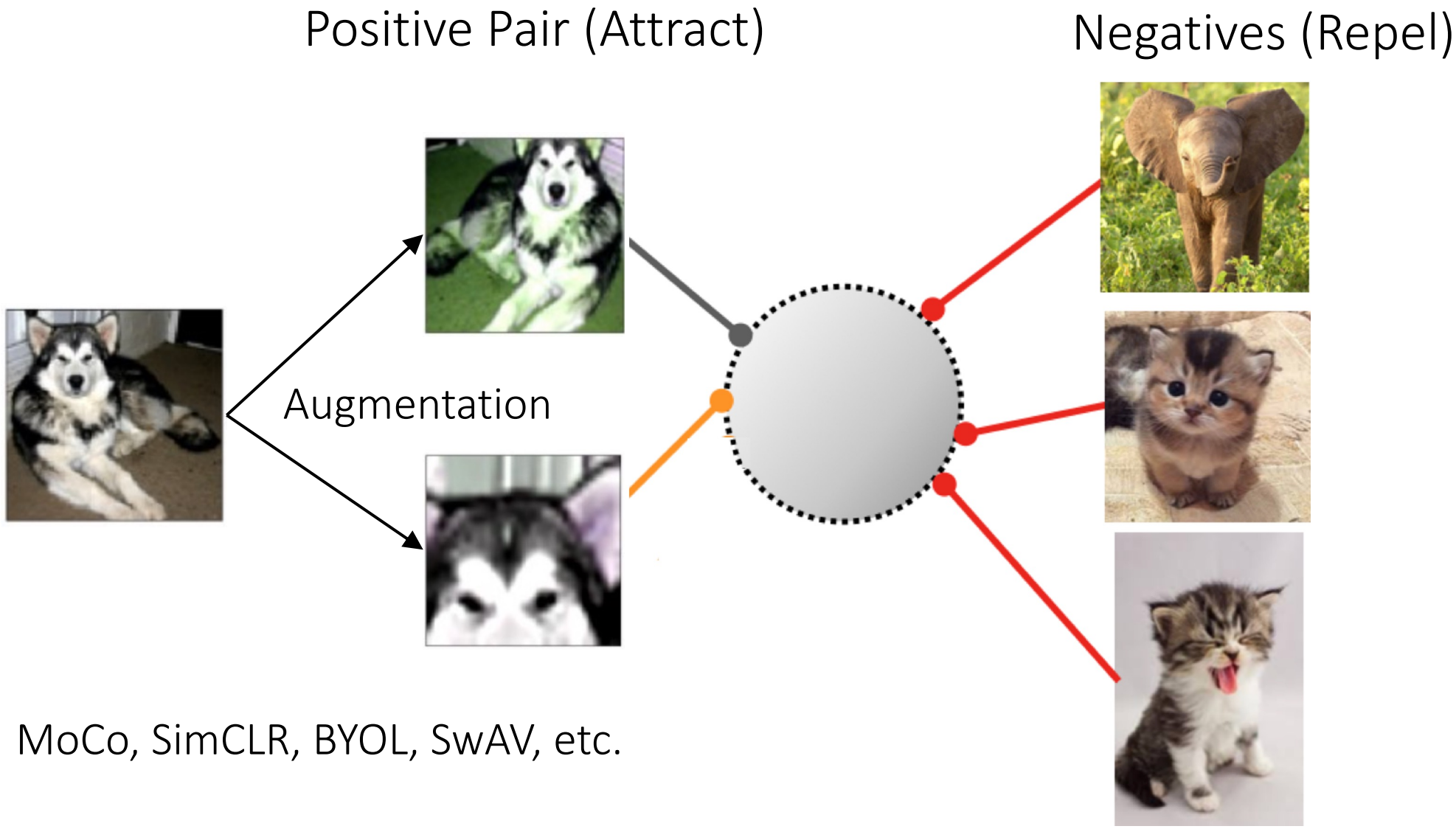
Rethinking Few-Shot Image Classification: a Good Embedding Is All You Need? [Tian et al. ECCV 2020]

- A simple baseline outperforms complex SOTA meta-learning methods



- Similar conclusion by other works: [W. Chen et al, 2019], [G. Dhillon et al, 2020], [Y. Guo et al, 2020], [Y. Chen et al, 2020]

Learning a Good Embedding: Contrastive Representations

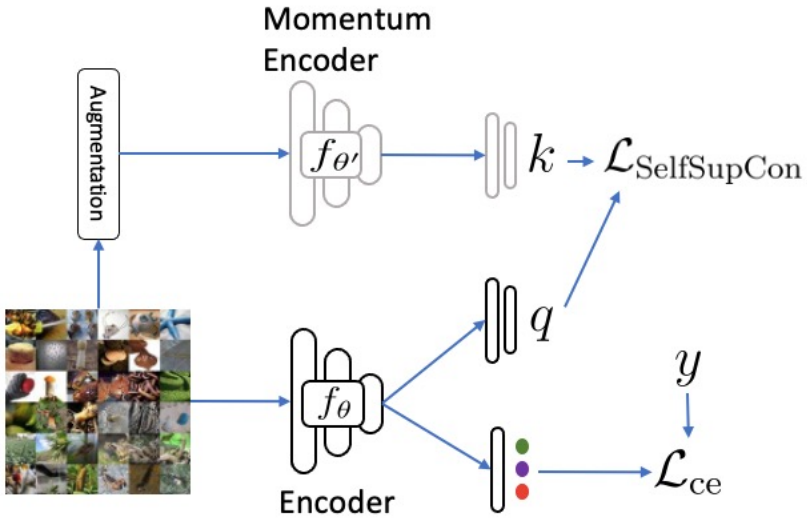


MoCo, SimCLR, BYOL, SwAV, etc.

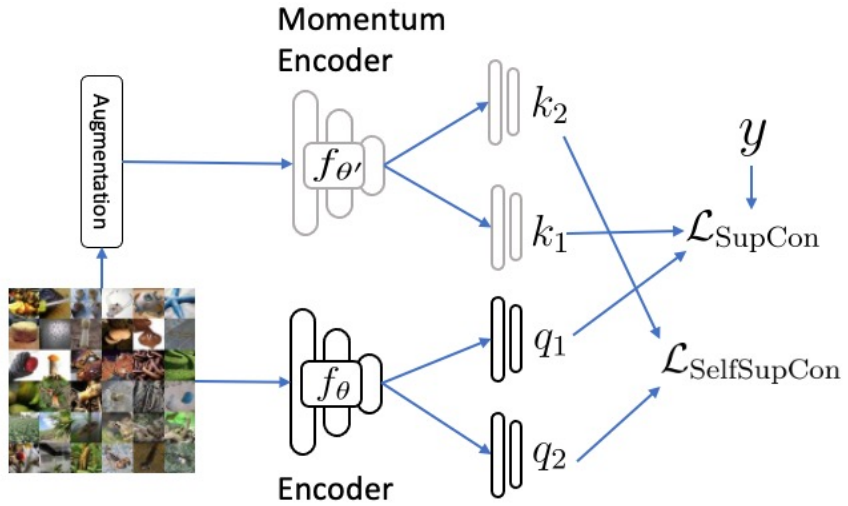
How well contrastive representations (learned on ImageNet) transfer to various other domains?

Experimental Setup – Loss Functions

- Supervised Cross-Entropy (CE)
- Self-Supervised Contrastive (SelfSupCon) based on MoCo V2
- Supervised Contrastive (SupCon)
- Supervised + Self-Supervised Contrastive (CE+SelfSupCon)
- Supervised Contrastive + Self-Supervised Contrastive (SupCon+ SelfSupCon)



CE+SelfSupCon



SupCon+ SelfSupCon

[Islam et al, 2021]

Experimental Setup – Datasets and Protocols

- Base dataset: ImageNet 1K
- Downstream tasks:
  - 12 datasets from various domains, including Natural, Satellite, Symbolic, Medical, Illustrative, and Texture data
- Model: ResNet-50
- Evaluation protocol:
  - Linear evaluation over fixed network
  - Fine-tune whole network

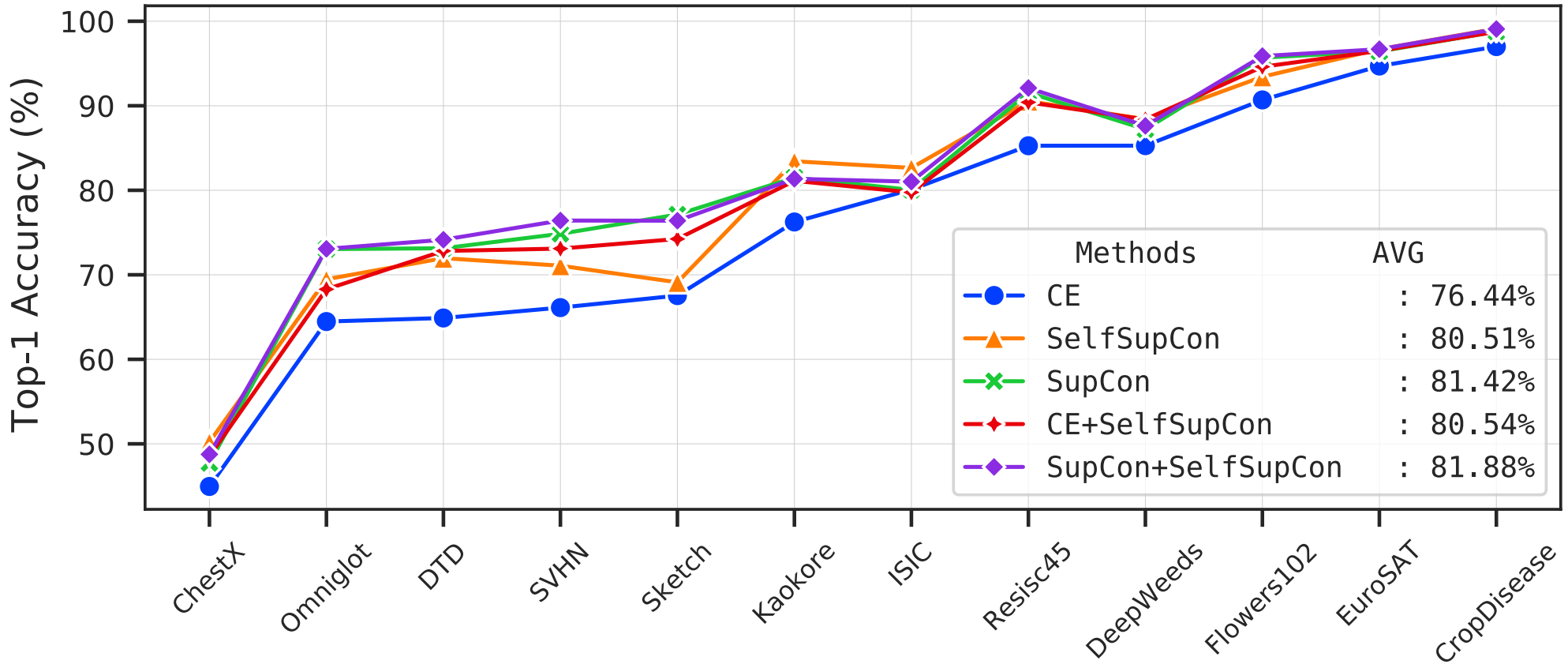
[Islam et al, 2021]

Category	Dataset	Train Size	Test Size	Classes
Natural	CropDisease [30]	43456	10849	38
	Flowers [32]	6149	1020	102
	DeepWeeds [33]	12252	5257	9
Satellite	EuroSAT [19]	18900	8100	10
	Resisc45 [8]	22005	9495	45
Symbolic	Omniglot [28]	9226	3954	1623
	SVHN [31]	73257	26032	10
Medical	ISIC [10]	7007	3008	7
	ChestX [42]	18090	7758	7
Illustrative	Kaokore [38]	6568	821	4
	Sketch [41]	35000	15889	1000
Texture	DTD [9]	3760	1880	47

[Islam et al, 2021]

Experimental Results – Linear Evaluation

- Contrastive models consistently perform better in transfer learning than cross-entropy
- Combining self-supervised contrastive loss with cross-entropy or supervised contrastive loss improves transfer learning performance



[Islam et al, 2021]

Experimental Results – Fine-Tuning

- Contrastive pretrained methods are only slightly more effective than cross-entropy when fine-tuning all network parameters, especially with less data

	CropDisease	DeepWeeds	Flowers102	EuroSAT	Resisc45	ISIC	ChestX	Omniglot	SVHN	Kaokore	Sketch	DTD	Mean
Full dataset													
CE	99.94	97.22	98.04	98.80	95.98	89.86	56.17	90.29	96.95	90.26	79.51	72.07	88.76
SelfSupCon	99.91	97.32	97.45	98.88	96.24	90.13	56.93	91.27	97.19	89.40	77.46	72.61	88.73
SupCon	99.93	96.75	98.24	98.69	95.87	88.93	55.68	90.64	96.99	88.19	80.67	74.31	88.74
CE+SelfSupCon	99.89	97.17	98.24	98.93	95.87	89.23	55.70	90.49	97.00	89.16	81.00	73.62	**88.86**
SupCon+SelfSupCon	99.90	96.98	97.65	98.84	96.10	89.16	56.73	91.07	96.99	89.52	80.55	75.53	**89.08**
1000 training samples													
CE	93.01	86.63	80.10	95.22	79.47	79.36	41.25	41.81	77.04	81.00	16.89	60.48	69.35
SelfSupCon	93.54	88.55	81.47	95.78	80.84	79.09	42.39	45.80	83.55	81.36	10.24	61.91	70.38
SupCon	93.40	86.19	82.06	95.10	81.90	79.06	41.72	41.88	80.57	79.78	16.47	64.10	70.19
CE+SelfSupCon	92.87	87.64	82.84	95.19	81.63	79.69	41.49	41.91	80.56	79.90	16.17	65.11	**70.42**
SupCon+SelfSupCon	93.29	87.14	82.16	95.58	82.88	79.02	41.69	41.55	81.70	80.51	15.29	65.05	**70.49**

[Islam et al, 2021]

Experimental Results - Few-shot classification

- Experimental Setup:
  - Mini-ImageNet as base dataset, ResNet-18 model
  - Linear evaluation (logistic regression on top of frozen model), average score over 600 episodes

	Mini-IN*	CropDisease	DeepWeeds	Flowers102	EuroSAT	Resisc45	ISIC	ChestX	Omniglot	SVHN	Kaokore	Sketch	DTD	Mean
5-shot														
CE	72.47	86.58	48.33	81.31	78.51	72.86	44.28	26.16	94.26	27.98	37.63	64.64	57.64	60.01
SelfSupCon	67.71	83.29	49.90	84.29	81.65	72.36	45.20	26.91	93.61	27.10	42.75	66.32	61.35	61.23
SupCon	75.20	83.44	47.74	82.93	80.94	74.48	42.97	26.23	96.78	33.22	45.10	74.24	65.42	62.79
CE+SelfSupCon	76.13	84.68	50.02	86.88	82.63	75.11	44.66	27.93	96.19	31.36	45.32	72.38	67.21	**63.70**
SupCon+SelfSupCon	72.81	84.26	50.35	86.72	82.57	74.94	45.82	28.19	96.72	30.67	45.26	71.09	65.08	**63.47**
20-shot														
CE	80.81	92.51	58.43	89.00	84.77	82.08	52.88	29.86	97.76	35.69	46.87	76.43	67.85	67.84
SelfSupCon	76.95	90.42	58.99	90.77	87.72	82.20	53.57	32.01	97.70	34.70	52.71	77.94	71.00	69.14
SupCon	83.32	90.73	55.90	90.45	87.71	83.40	51.11	31.06	98.87	44.98	54.94	85.23	74.11	70.71
CE+SelfSupCon	84.24	91.27	58.63	92.70	89.13	84.54	52.80	33.42	98.63	41.60	54.72	83.32	75.78	**71.38**
SupCon+SelfSupCon	82.07	91.54	59.24	92.51	89.04	84.18	54.08	33.92	98.71	40.84	55.25	82.84	74.58	**71.39**

*: test on the novel classes.

[Islam et al, 2021]

Experimental Results - Few-shot classification

- Experimental Setup:

  Mini-ImageNet as base dataset, ResNet-18 model

  Linear evaluation (logistic regression on top of frozen model), average score over 600 episodes

	Mini-IN*	CropDisease	DeepWeeds	Flowers102	EuroSAT	Resisc45	ISIC	ChestX	Omniglot	SVHN	Kaokore	Sketch	DTD	Mean
5-shot														
CE	72.47	86.58	48.33	81.31	78.51	72.86	44.28	26.16	94.26	27.98	37.63	64.64	57.64	60.01
SelfSupCon	67.71	83.29	49.90	84.29	81.65	72.36	45.20	26.91	93.61	27.10	42.75	66.32	61.35	61.23
SupCon	75.20	83.44	47.74	82.93	80.94	74.48	42.97	26.23	96.78	33.22	45.10	74.24	65.42	62.79
CE+SelfSupCon	76.13	84.68	50.02	86.88	82.63	75.11	44.66	27.93	96.19	31.36	45.32	72.38	67.21	**63.70**
SupCon+SelfSupCon	72.81	84.26	50.35	86.72	82.57	74.94	45.82	28.19	96.72	30.67	45.26	71.09	65.08	**63.47**
20-shot														
CE	80.81	92.51	58.43	89.00	84.77	82.08	52.88	29.86	97.76	35.69	46.87	76.43	67.85	67.84
SelfSupCon	76.95	90.42	58.99	90.77	87.72	82.20	53.57	32.01	97.70	34.70	52.71	77.94	71.00	69.14
SupCon	83.32	90.73	55.90	90.45	87.71	83.40	51.11	31.06	98.87	44.98	54.94	85.23	74.11	70.71
CE+SelfSupCon	84.24	91.27	58.63	92.70	89.13	84.54	52.80	33.42	98.63	41.60	54.72	83.32	75.78	**71.38**
SupCon+SelfSupCon	82.07	91.54	59.24	92.51	89.04	84.18	54.08	33.92	98.71	40.84	55.25	82.84	74.58	**71.39**

*: test on the novel classes.

Contrastive representations perform better than cross-entropy across domains

[Islam et al, 2021]

Experimental Results - Few-shot classification

This conclusion does not hold in mini-ImageNet (in-domain), where self-supervised contrastive representations have worse performance

	Mini-IN*	CropDisease	DeepWeeds	Flowers102	EuroSAT	Resisc45	ISIC	ChestX	Omniglot	SVHN	Kaokore	Sketch	DTD	Mean
5-shot														
CE	72.47	86.58	48.33	81.31	78.51	72.86	44.28	26.16	94.26	27.98	37.63	64.64	57.64	60.01
SelfSupCon	67.71	83.29	49.90	84.29	81.65	72.36	45.20	26.91	93.61	27.10	42.75	66.32	61.35	61.23
SupCon	75.20	83.44	47.74	82.93	80.94	74.48	42.97	26.23	96.78	33.22	45.10	74.24	65.42	62.79
CE+SelfSupCon	76.13	84.68	50.02	86.88	82.63	75.11	44.66	27.93	96.19	31.36	45.32	72.38	67.21	**63.70**
SupCon+SelfSupCon	72.81	84.26	50.35	86.72	82.57	74.94	45.82	28.19	96.72	30.67	45.26	71.09	65.08	**63.47**
20-shot														
CE	80.81	92.51	58.43	89.00	84.77	82.08	52.88	29.86	97.76	35.69	46.87	76.43	67.85	67.84
SelfSupCon	76.95	90.42	58.99	90.77	87.72	82.20	53.57	32.01	97.70	34.70	52.71	77.94	71.00	69.14
SupCon	83.32	90.73	55.90	90.45	87.71	83.40	51.11	31.06	98.87	44.98	54.94	85.23	74.11	70.71
CE+SelfSupCon	84.24	91.27	58.63	92.70	89.13	84.54	52.80	33.42	98.63	41.60	54.72	83.32	75.78	**71.38**
SupCon+SelfSupCon	82.07	91.54	59.24	92.51	89.04	84.18	54.08	33.92	98.71	40.84	55.25	82.84	74.58	**71.39**

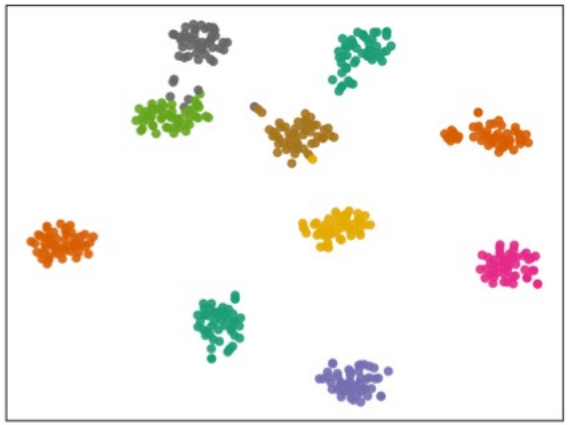
*: test on the novel classes.

[Islam et al, 2021]

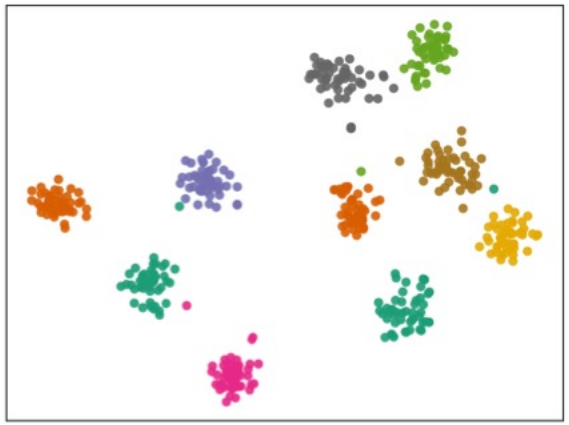
Why contrastive representations transfer better across domains?

Analysis – ImageNet t-SNE visualization

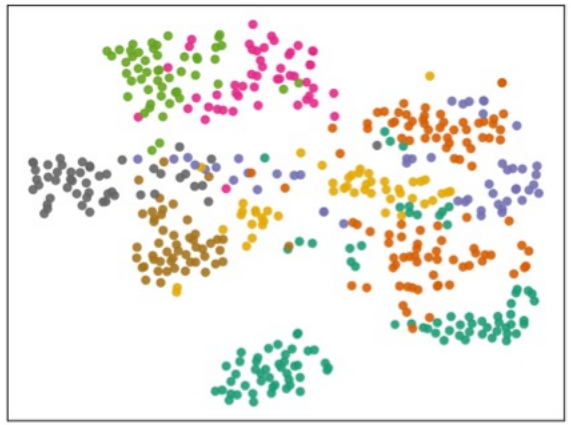
- Contrastive models have higher intra-class class variation than cross-entropy models, which may facilitate transfer across domains (see Zhao et al, ICLR 2021)



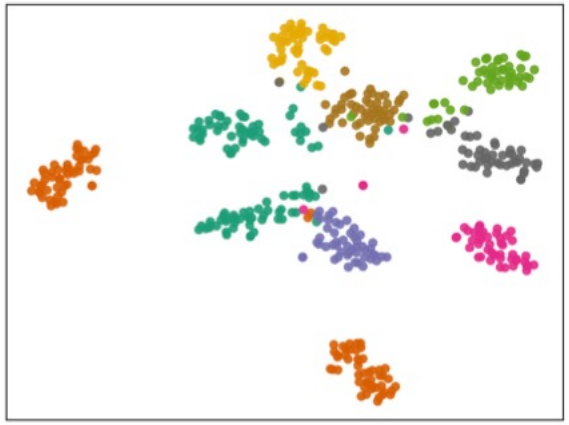
CE



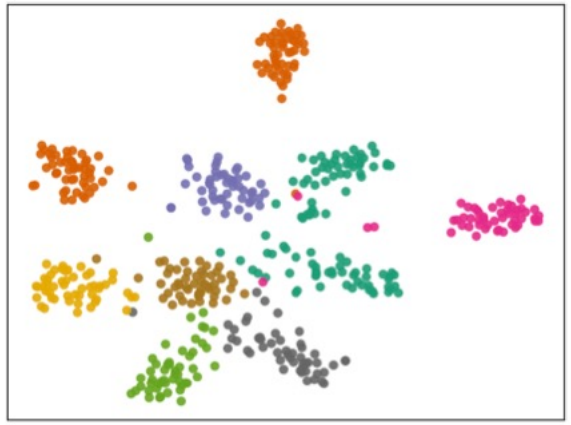
CE+SelfSupCon



SelfSupCon



SupCon

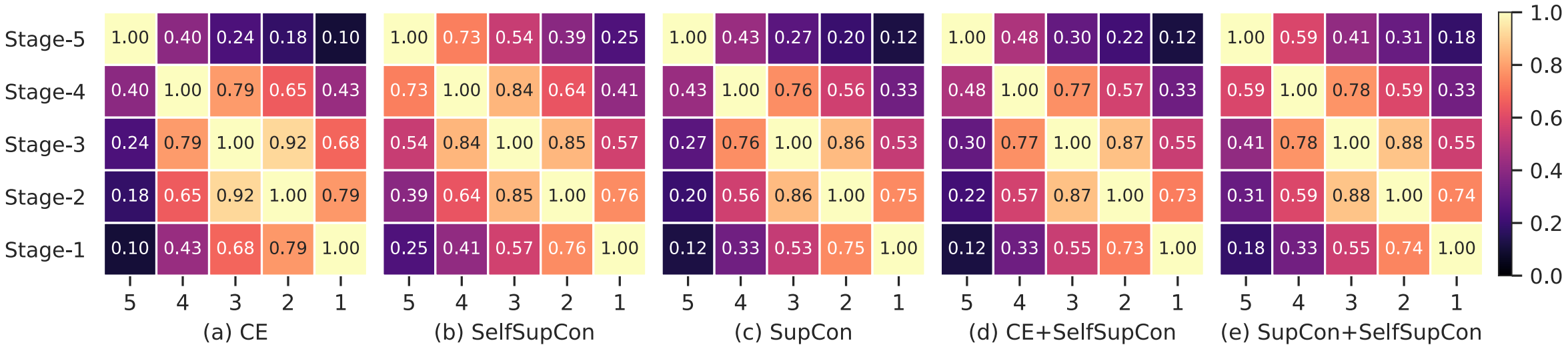


SupCon+SelfSupCon

[Islam et al, 2021]

Analysis – Centered Kernel Alignment

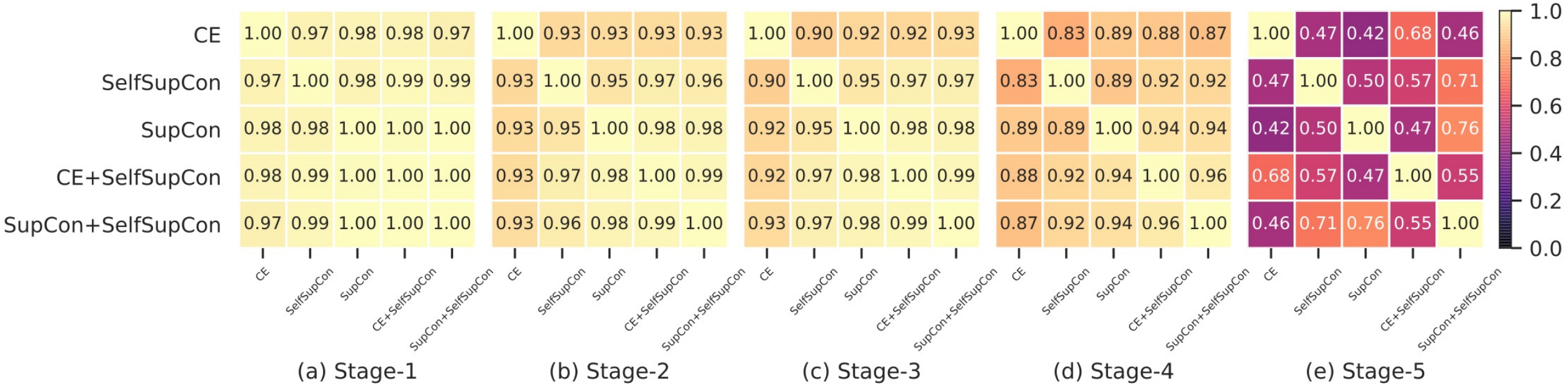
- Centered Kernel Alignment scores between different stages of the same model
- ★ Final layers of a contrastive model are more similar to initial layers, compared to cross-entropy which are more specialized



[Islam et al, 2021]

Analysis – Centered Kernel Alignment

- Centered Kernel Alignment scores between different models considering the same stage
- ☆ Different models learn similar representation in the initial layers but diverge drastically in the final layers



[Islam et al, 2021]

See more experiments and analysis in our paper:

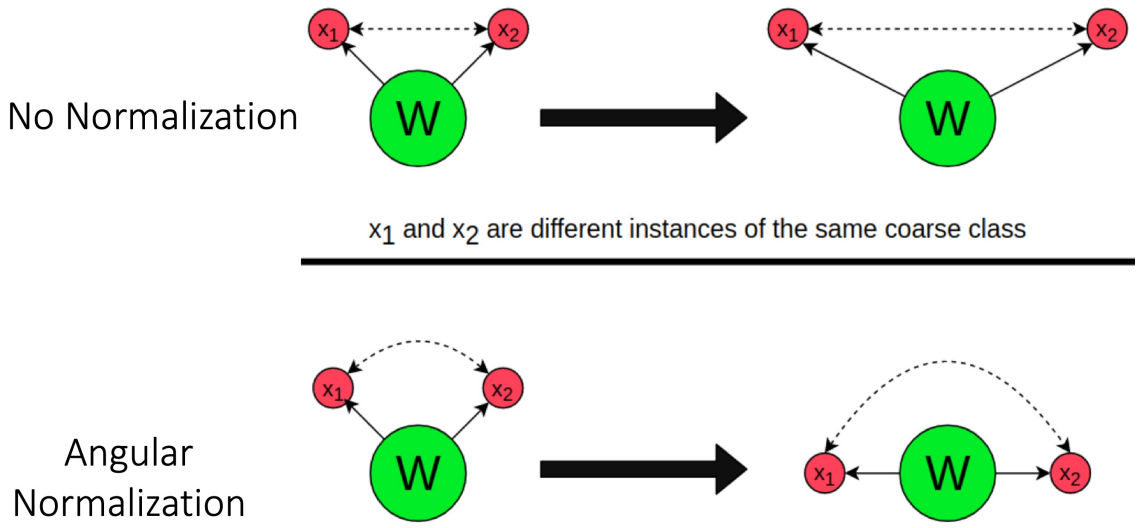
Ashraful Islam et al. “A Broad Study on the Transferability of Visual Representations with Contrastive Learning” Arxiv 2021

- Loss weighting effect on ImageNet and transfer accuracy
- Performance with respect to longer training
- Additional results on object detection and instance segmentation
- Robustness to image corruption
- Etc.

See also our related CVPR 2021 oral paper:

Guy et al. “Fine-grained Angular Contrastive Learning with Coarse Labels”. CVPR 2021

- Coarse-to-fine few-shot learning (labels only for coarse-grained categories)
- Supervised loss (coarse-grained categories) + self-supervised loss (coarse and fine-grained categories) using a novel angular normalization module



Summary

- Do contrastive representations transfer better? Yes (but not always)
  - Contrastive models consistently perform better than standard cross-entropy models in linear evaluation mode across a wide range of domains
  - When fine-tuning the network parameters, performance of contrastive representations is similar to cross-entropy models. In few-shot linear evaluation mode, performance may be worse if the same domain is considered
- Combining self-supervised contrastive loss with cross-entropy or supervised contrastive loss improves transfer learning performance
- Analysis of why contrastive representations transfer better across domains (intra-class variation, Center Kernel Alignment scores)

References

- Yunhui Guo, Noel C. Codella, Leonid Karlinsky, James V. Codella, John R. Smith, Kate Saenko, Tajana Rosing, Rogerio Feris, "A Broader Study of Cross-domain Few-shot Learning". ECCV 2020

- Ashraful Islam, Richard Chen, Rameswar Panda, Leonid Karlinsky, Richard Radke, Rogerio Feris. "A Broad Study on the Transferability of Visual Representations with Contrastive Learning" Arxiv 2021

- Guy Bukchin, Eli Schwartz, Kate Saenko, Ori Shahar, Rogerio Feris, Raja Giryes, Leonid Karlinsky"Fine-grained Angular Contrastive Learning with Coarse Labels". CVPR 2021