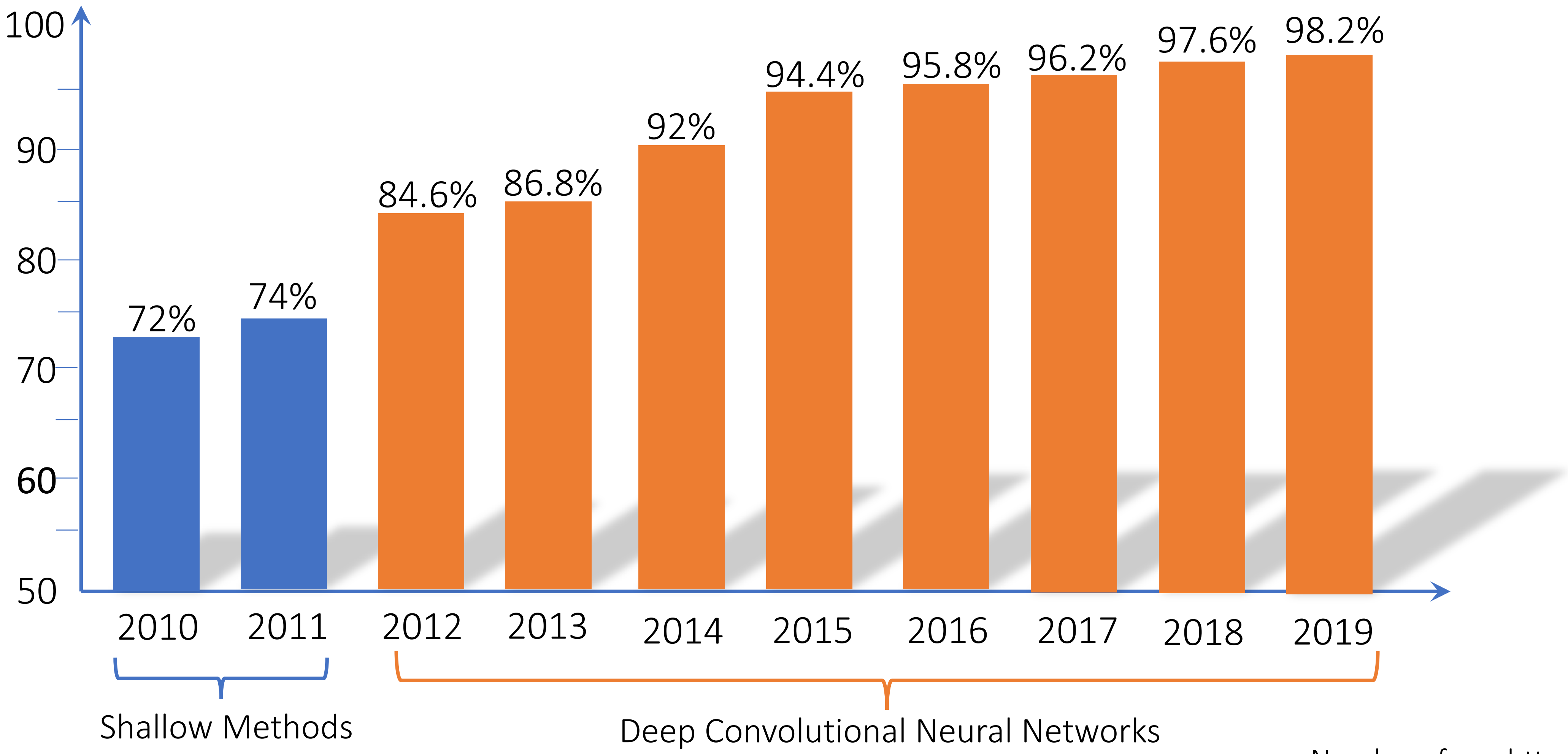
Visual Learning Beyond Natural Images

Rogerio Schmidt Feris

MIT-IBM Watson AI Lab

http://rogerioferis.com

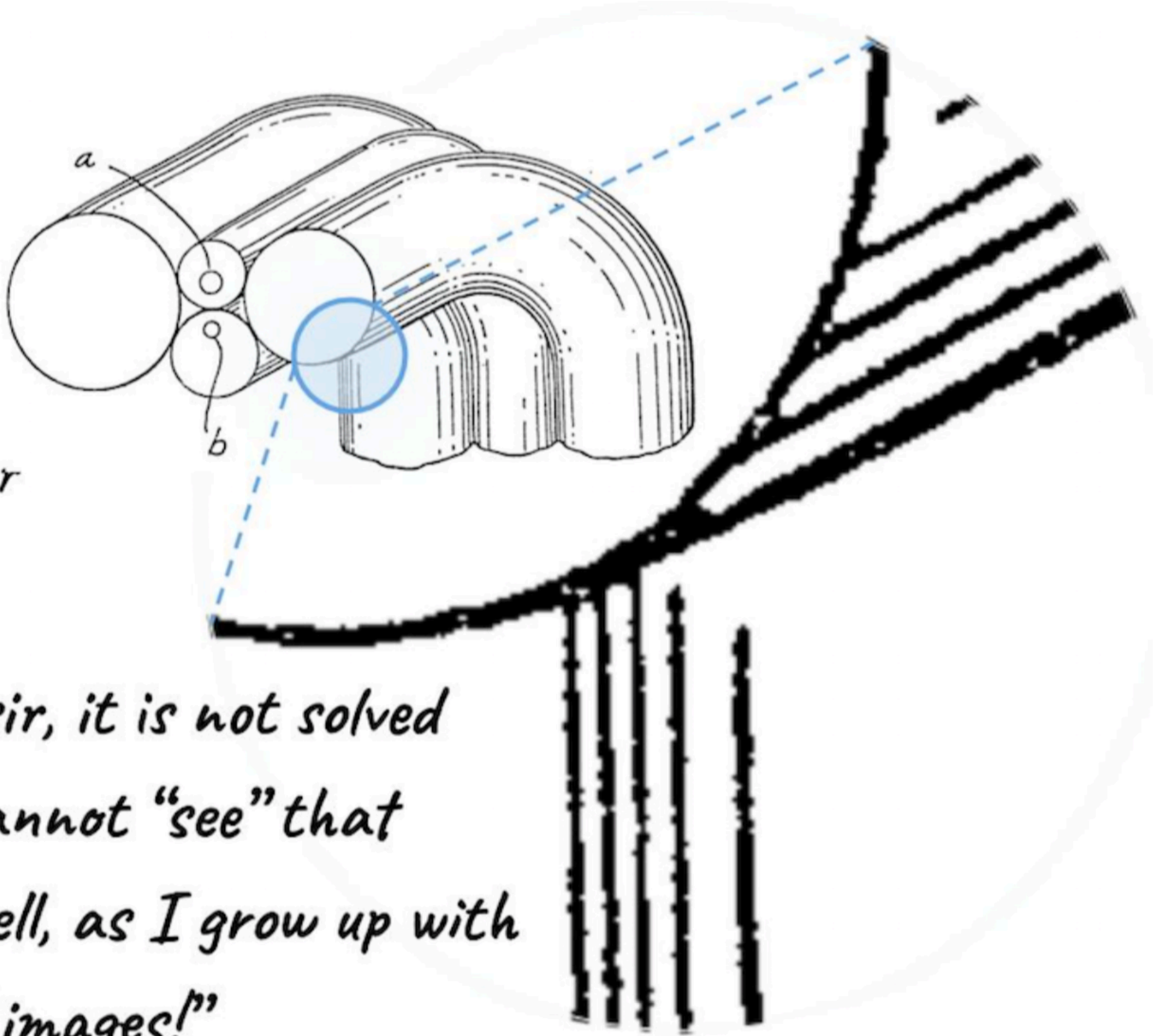
ImageNet Classification (top-5 accuracy)

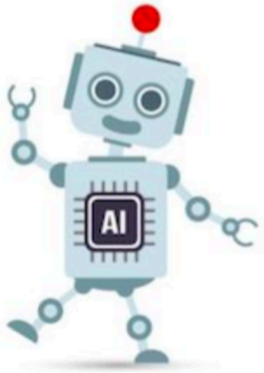


Numbers from http://paperswithcode.com

However…

"Bob, can you tell me whether this design has been patented or not?"



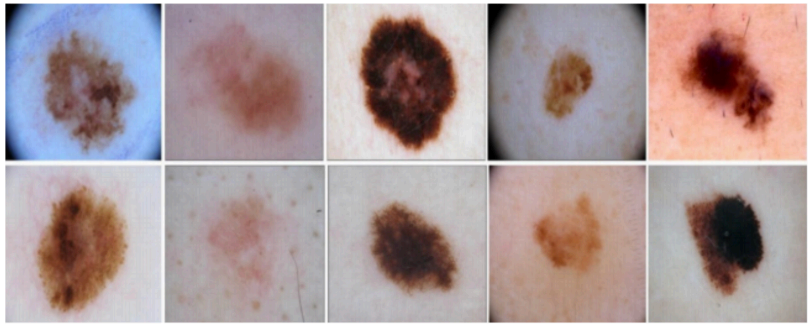


"sorry, sir, it is not solved yet; I cannot "see" that image well, as I grow up with natural images!"

Figure from http://cvpr-dira.lipingyang.org

Visual Learning Beyond Natural Images

Healthcare



Sattelite

Document Analysis

[redacted] R.J.Reynolds Tobacco Company [redacted]

AUTHORIZATION REQUEST

MK-137
SC4-ICA

ECT
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

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

EXPENDITURE TIMING:

	CAPITAL	EXPENSE
19 85		$25,989
19		
19		
Annual Average Thereafter		

RELATED PROPOSALS:
85-293
(List Related AR/CA/AC No's)

AMOUNTS PREVIOUSLY APPROVED:
Capital $ ____ Expense $ 33,200
(Cumulative)

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. ____ %

EFO IMPACT

	19 85	19	19	Annual Average Thereafter
Profit/(Loss)	($25,989)			
Budget/Plan Change ☐ Yes ☐ No				

REVIEWED BY:

Dept.	Initials	Date	Review Completed
Prom.	BW	7/8	☑
Prom.	CLS	7/9	☑
Prom.	TBO	7/9	☑
t. Sv.	CTB	7/10	☑
Comp.	DRS	7/11	☑
Law	KMV	7/19	☑
Risk Mgt	MRV	7/9	☑

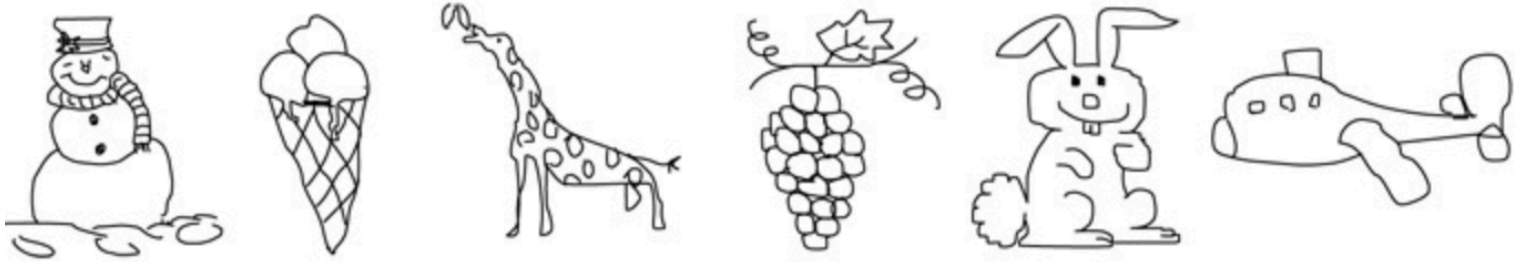
APPROVALS (Originator Enters Initials of Required Approvers)

Initials	Signature	Date	Initials	Signature	Date
MLO	[illegible]	7/23			
			CFO		
			Operations		
			Fin. & Admin.		
			Pres. & CEO		

51316 1442

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

Sketch Recognition



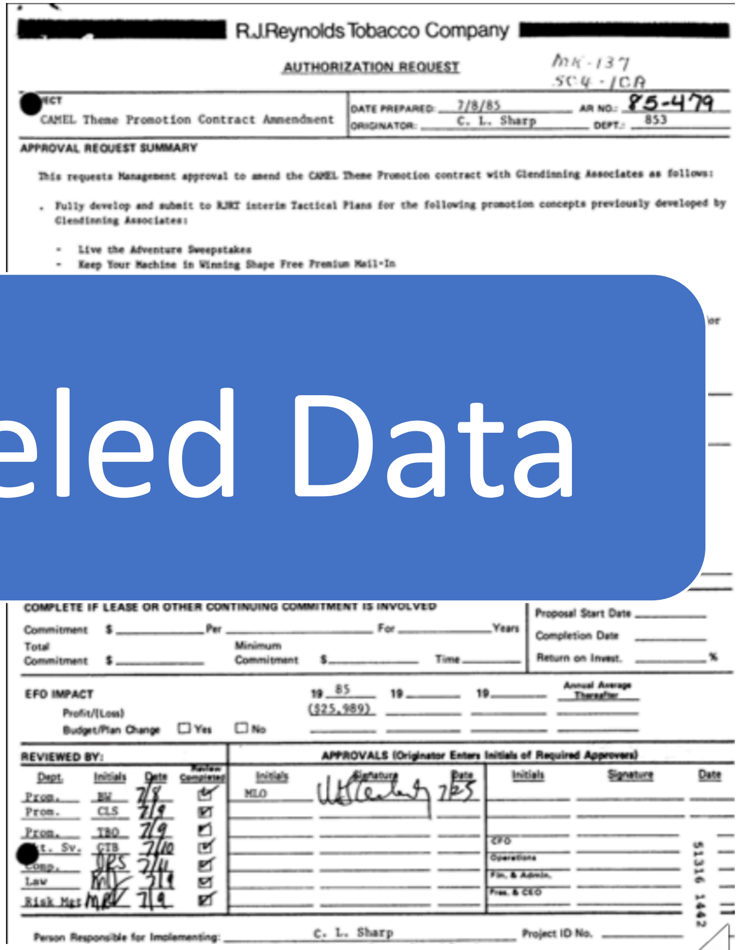
... and Much More !

Visual Learning Beyond Natural Images

Healthcare

Sattelite

Document Analysis



[redacted] R.J.Reynolds Tobacco Company [redacted]

AUTHORIZATION REQUEST

MK-137
5C4-ICA

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

COMPLETE IF LEASE OR OTHER CONTINUING COMMITMENT IS INVOLVED

Commitment $ Per For Years | Proposal Start Date
Total Commitment $ | Minimum Commitment $ Time | Completion Date | Return on Invest. %

EFO IMPACT 19 85 19 19 Annual Average Thereafter
Profit/(Loss) ($25,989)
Budget/Plan Change ☐ Yes ☐ No

REVIEWED BY: | APPROVALS (Originator Enters Initials of Required Approvers)

Dept.	Initials	Date	Review Completed	Initials	Signature	Date	Initials	Signature	Date
Prom.	BW	7/8	☑	MLO		7/25			
Prom.	CLS	7/9	☑						
Prom.	TBO	7/9	☑				CFO		
Mkt. Sv.	CTB	7/10	☑				Operations		
Comp.	DRS	7/11	☑				Fin. & Admin.		
Law		7/11	☑				Pres. & CEO		
Risk Mgt.		7/9	☑						

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

51316 1442

Problem: Limited Labeled Data

... and Much More !

A Broader Study of Cross-Domain Few-Shot Learning

Yunhui Guo, Noel C Codella, Leonid Karlinsky, James V Codella,
John R Smith, Kate Saenko, Tajana Rosing, Rogerio Feris

ECCV 2020

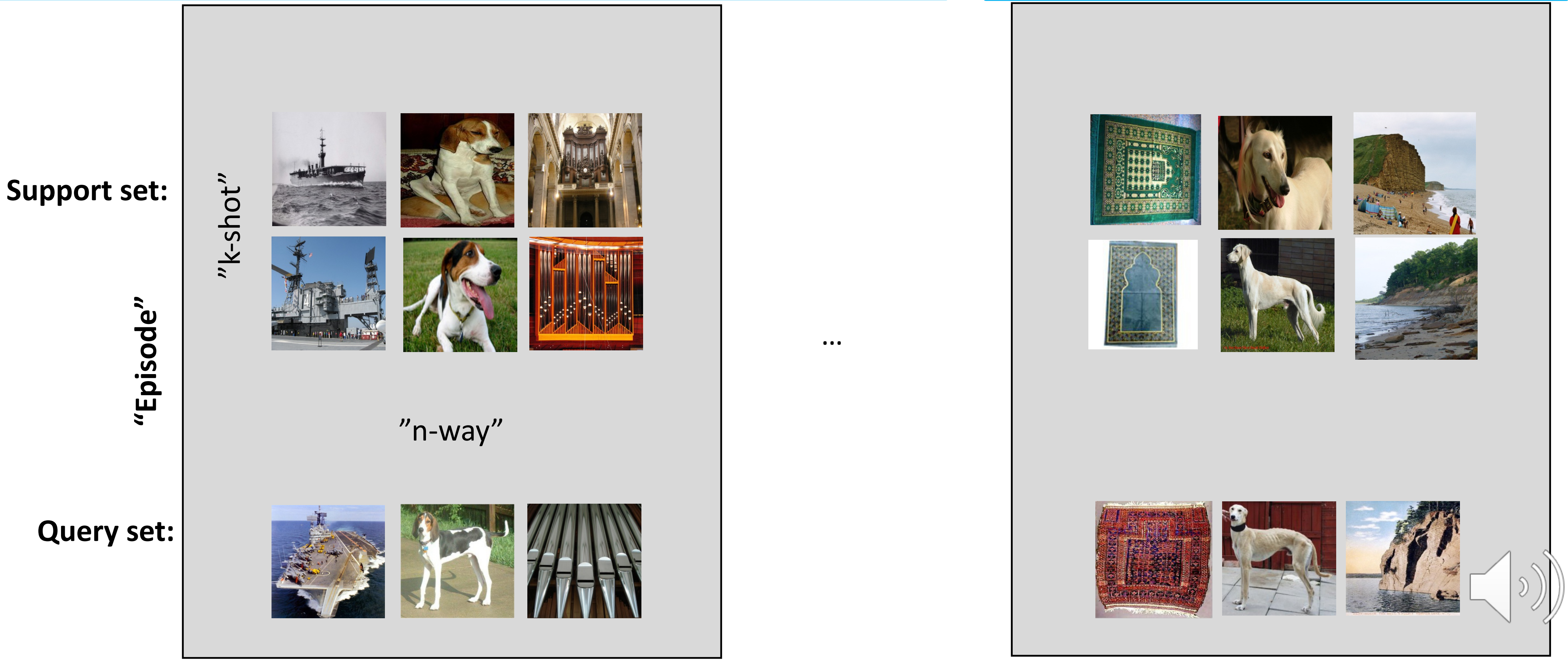
Few-Shot Learning Problem

Meta-Learning

Meta-learning: Learn how to learn with few examples in training tasks.
Performed via many trials of k-shot n-way tasks, while optimizing.

Meta-Testing

Evaluate method on novel tasks and
measure accuracy over many trials.



What happens when data is different from expected domain?

Meta-Learning (TRAIN)

"House Finch"

"Gazelle Hound"

"Triceratops"

k-way

n-shot

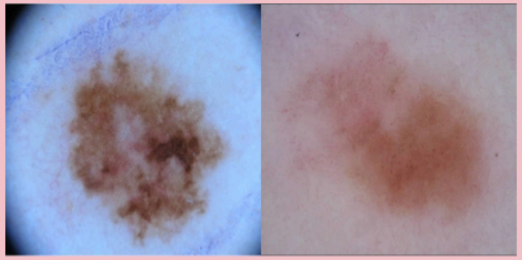
"Novel" Category (TEST – SAME DOMAIN)

"Golden Retriever"

Current research explores few-shot categories with exceptionally high similarity to training data!

"Novel" Category (TEST – DIFF. DOMAIN)

"Melanoma"



In practice, few-shot categories can wildly differ to training data, causing all existing techniques to break-down and perform worse than baseline simple methods.

tieredImageNet was first attempt to address

The **miniImageNet** dataset [33] is a standard benchmark for few-shot image classification benchmark, consisting of 100 randomly chosen classes from ILSVRC-2012 [24]. These classes are randomly split into 64, 16 and 20 classes for meta-training, meta-validation, and meta-testing respectively. Each class contains 600 images of size 84×84. Since the class splits were not released in the original publication [33], we use the commonly-used split proposed in [22].

The **tieredImageNet** benchmark [23] is a larger subset of ILSVRC-2012 [24], composed of 608 classes grouped into 34 high-level categories. These are divided into 20 categories for meta-training, 6 categories for meta-validation, and 8 categories for meta-testing. This corresponds to 351, 97 and 160 classes for meta-training, meta-validation, and meta-testing respectively. This dataset aims to minimize the semantic similarity between the splits. All images are of size 84 × 84.

But these are all still categories within the domain of natural images!

Significant progress on miniImagenet / tieredImageNet
(numbers below already outdated)

		miniImageNet 5-way		tieredImageNet 5-way	
model	**backbone**	**1-shot**	**5-shot**	**1-shot**	**5-shot**
Meta-Learning LSTM* [22]	64-64-64-64	43.44 ± 0.77	60.60 ± 0.71	-	-
Matching Networks* [33]	64-64-64-64	43.56 ± 0.84	55.31 ± 0.73	-	-
MAML [8]	32-32-32-32	48.70 ± 1.84	63.11 ± 0.92	51.67 ± 1.81	70.30 ± 1.75
Prototypical Networks*† [28]	64-64-64-64	49.42 ± 0.78	68.20 ± 0.66	53.31 ± 0.89	72.69 ± 0.74
Relation Networks* [29]	64-96-128-256	50.44 ± 0.82	65.32 ± 0.70	54.48 ± 0.93	71.32 ± 0.78
R2D2 [3]	96-192-384-512	51.2 ± 0.6	68.8 ± 0.1	-	-
Transductive Prop Nets [14]	64-64-64-64	55.51 ± 0.86	69.86 ± 0.65	59.91 ± 0.94	73.30 ± 0.75
SNAIL [18]	ResNet-12	55.71 ± 0.99	68.88 ± 0.92	-	-
Dynamic Few-shot [10]	64-64-128-128	56.20 ± 0.86	73.00 ± 0.64	-	-
AdaResNet [19]	ResNet-12	56.88 ± 0.62	71.94 ± 0.57	-	-
TADAM [20]	ResNet-12	58.50 ± 0.30	76.70 ± 0.30	-	-
Activation to Parameter† [21]	WRN-28-10	59.60 ± 0.41	73.74 ± 0.19	-	-
LEO† [25]	WRN-28-10	61.76 ± 0.08	77.59 ± 0.12	$\mathbf{66.33 \pm 0.05}$	$\mathbf{81.44 \pm 0.09}$
MetaOptNet-RR (ours)	ResNet-12	61.41 ± 0.61	77.88 ± 0.46	$\mathbf{65.36 \pm 0.71}$	$\mathbf{81.34 \pm 0.52}$
MetaOptNet-SVM (ours)	ResNet-12	62.64 ± 0.61	78.63 ± 0.46	$\mathbf{65.99 \pm 0.72}$	$\mathbf{81.56 \pm 0}$[illegible]
MetaOptNet-SVM-trainval (ours)†	ResNet-12	$\mathbf{64.09 \pm 0.62}$	$\mathbf{80.00 \pm 0.45}$	$\mathbf{65.81 \pm 0.74}$	$\mathbf{81.75 \pm 0.53}$

Let's get more realistic by gradually leaving domain: Proposed Cross-Domain Evaluation Benchmark

Source Domain:

ImageNet:
Perspective
Natural Images
Color

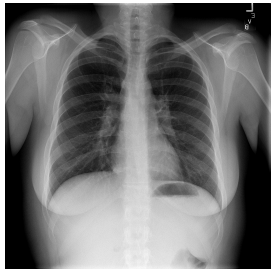
Target Domains:
(Disjoint Label Spaces)

Decreasing Similarity to ImageNet

CropDisease:
Perspective
Natural Images
Color

EuroSAT:
No Perspective
Natural Images
Color

ISIC:
No Perspective
Medical Images
Color



ChestX:
No Perspective
Medical Images
Grayscale

https://www.learning-with-limited-labels.com/challenge

Meta-learning doesn't generalize well. Fine-tuning is better. Ensembles are even better.

- We evaluate performance of current meta-learning, cross-domain few-shot learning, fine-tuning, and ensemble methods.
- Meta-learning approaches perform worst, even methods tailored for cross-domain.
- Fine-tuning is better.
- Ensemble approaches are best.

https://arxiv.org/pdf/1912.07200.pdf

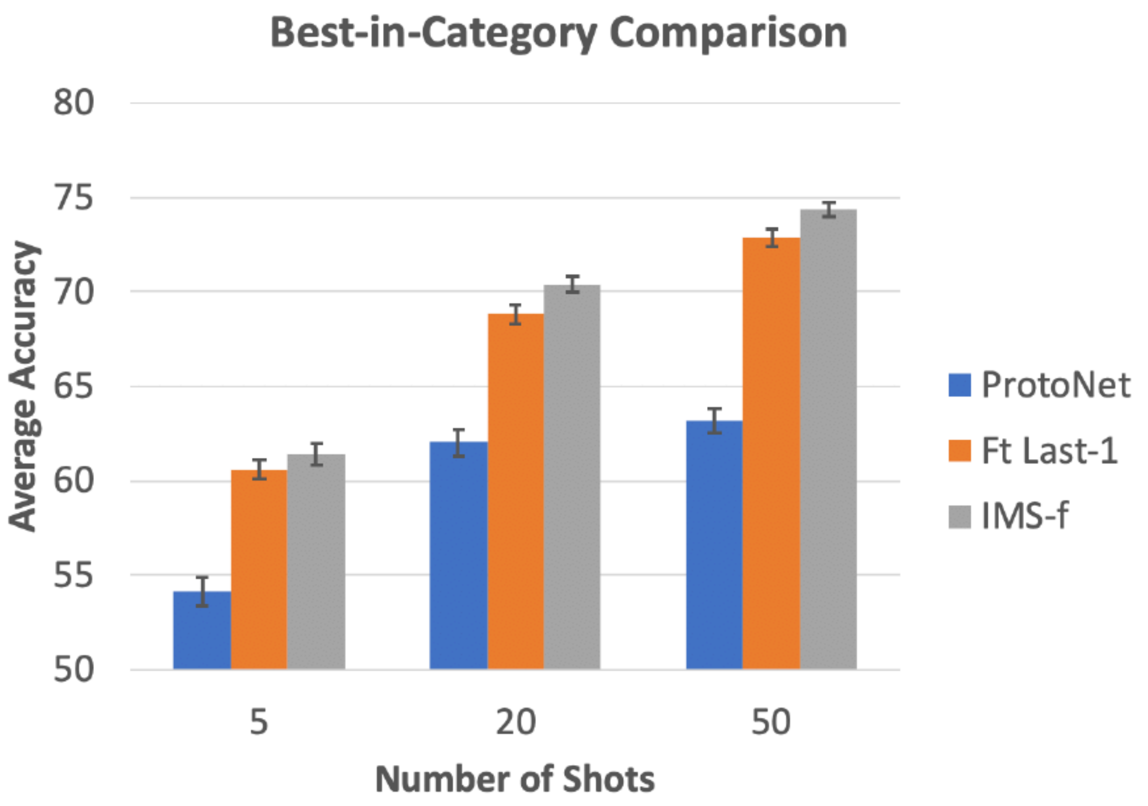


Fig. 4: Best meta-learning, single model, and multi-model transfer learning.

Details of results on each dataset

Best Meta-learning approach   Prior state-of-art

Methods	ChestX			ISIC		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
MatchingNet	22.40% ± 0.7%	23.61% ± 0.86%	22.12% ± 0.88%	36.74% ± 0.53%	45.72% ± 0.53%	54.58% ± 0.65%
MatchingNet+FWT	21.26% ± 0.31%	23.23% ± 0.37%	23.01% ± 0.34%	30.40% ± 0.48%	32.01% ± 0.48%	33.17% ± 0.43%
MAML	23.48% ± 0.96%	27.53% ± 0.43%	-	40.13% ± 0.58%	52.36% ± 0.57%	-
ProtoNet	24.05% ± 1.01%	28.21% ± 1.15%	29.32% ± 1.12%	39.57% ± 0.57%	49.50% ± 0.55%	51.99% ± 0.52%
ProtoNet+FWT	23.77% ± 0.42%	26.87% ± 0.43%	30.12% ± 0.46%	38.87% ± 0.52%	43.78% ± 0.47%	49.84% ± 0.51%
RelationNet	22.96% ± 0.88%	26.63% ± 0.92%	28.45% ± 1.20%	39.41% ± 0.58%	41.77% ± 0.49%	49.32% ± 0.51%
RelationNet+FWT	22.74% ± 0.40%	26.75% ± 0.41%	27.56% ± 0.40%	35.54% ± 0.55%	43.31% ± 0.51%	46.38% ± 0.53%
MetaOpt	22.53% ± 0.91%	25.53% ± 1.02%	29.35% ± 0.99%	36.28% ± 0.50%	49.42% ± 0.60%	54.80% ± 0.54%

Methods	EuroSAT			CropDiseases		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
MatchingNet	64.45% ± 0.63%	77.10% ± 0.57%	54.44% ± 0.67%	66.39% ± 0.78%	76.38% ± 0.67%	58.53% ± 0.73%
MatchingNet+FWT	56.04% ± 0.65%	63.38% ± 0.69%	62.75% ± 0.76%	62.74% ± 0.90%	74.90% ± 0.71%	75.68% ± 0.78%
MAML	71.70% ± 0.72%	81.95% ± 0.55%	-	78.05% ± 0.68%	89.75% ± 0.42%	-
ProtoNet	73.29% ± 0.71%	82.27% ± 0.57%	80.48% ± 0.57%	79.72% ± 0.67%	88.15% ± 0.51%	90.81% ± 0.43%
ProtoNet+FWT	67.34% ± 0.76%	75.74% ± 0.70%	78.64% ± 0.57%	72.72% ± 0.70%	85.82% ± 0.51%	87.17% ± 0.50%
RelationNet	61.31% ± 0.72%	74.43% ± 0.66%	74.91% ± 0.58%	68.99% ± 0.75%	80.45% ± 0.64%	85.08% ± 0.53%
RelationNet+FWT	61.16% ± 0.70%	69.40% ± 0.64%	73.84% ± 0.60%	64.91% ± 0.79%	78.43% ± 0.59%	81.14% ± 0.56%
MetaOpt	64.44% ± 0.73%	79.19% ± 0.62%	83.62% ± 0.58%	68.41% ± 0.73%	82.89% ± 0.54%	91.76% ± 0.38%

Table 1: The results of meta-learning methods on the proposed benchmark.

Details of results on each dataset

Best transfer-learning approach

Methods	**ChestX**			**ISIC**		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
Random	21.80% ± 1.03%	25.69% ± 0.95%	26.19% ± 0.94%	37.91% ± 1.39%	47.24% ± 1.50%	50.85% ± 1.37%
Fixed	25.35% ± 0.96%	30.83% ± 1.05%	36.04% ± 0.46%	43.56% ± 0.60%	52.78% ± 0.58%	57.34% ± 0.56%
Ft All	25.97% ± 0.41%	31.32% ± 0.45%	35.49% ± 0.45%	48.11% ± 0.64%	59.31% ± 0.48%	66.48% ± 0.56%
Ft Last-1	25.96% ± 0.46%	31.63% ± 0.49%	37.03% ± 0.50%	47.20% ± 0.45%	59.95% ± 0.45%	65.04% ± 0.47%
Ft Last-2	26.79% ± 0.59%	30.95% ± 0.61%	36.24% ± 0.62%	47.64% ± 0.44%	59.87% ± 0.35%	66.07% ± 0.45%
Ft Last-3	25.17% ± 0.56%	30.92% ± 0.89%	37.27% ± 0.64%	48.05% ± 0.55%	60.20% ± 0.33%	66.21% ± 0.52%
Transductive Ft	26.09% ± 0.96%	31.01% ± 0.59%	36.79% ± 0.53%	49.68% ± 0.36%	61.09% ± 0.44%	67.20% ± 0.59%

Methods	**EuroSAT**			**CropDiseases**		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
Random	58.00% ± 2.01%	68.93% ± 1.47%	71.65% ± 1.47%	69.68% ± 1.72%	83.41% ± 1.25%	86.56% ± 1.42%
Fixed	75.69% ± 0.66%	84.13% ± 0.52%	86.62% ± 0.47%	87.48% ± 0.58%	94.45% ± 0.36%	96.62% ± 0.25%
Ft All	79.08% ± 0.61%	87.64% ± 0.47%	90.89% ± 0.36%	89.25% ± 0.51%	95.51% ± 0.31%	97.68% ± 0.21%
Ft Last-1	80.45% ± 0.54%	87.92% ± 0.44%	91.41% ± 0.46%	88.72% ± 0.53%	95.76% ± 0.65%	97.87% ± 0.48%
Ft Last-2	79.57% ± 0.51%	87.67% ± 0.46%	90.93% ± 0.45%	88.07% ± 0.56%	95.68% ± 0.76%	97.64% ± 0.59%
Ft Last-3	78.04% ± 0.77%	87.52% ± 0.53%	90.83% ± 0.42%	89.11% ± 0.47%	95.31% ± 0.7%	97.45% ± 0.46%
Transductive Ft	81.76% ± 0.48%	87.97% ± 0.42%	92.00% ± 0.56%	90.64% ± 0.54%	95.91% ± 0.72%	97.48% ± 0.56%

Table 2: The results of different variants of single model fine-tuning on the proposed benchmark.

Intelligent selective ensembles outperform naïve ensembles.

Best ensemble approach

Methods	**ChestX**			**ISIC**		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
All embeddings	26.74% ± 0.42%	32.77% ± 0.47%	38.07% ± 0.50%	46.86% ± 0.60%	58.57% ± 0.59%	66.04% ± 0.56%
IMS-f	25.50% ± 0.45%	31.49% ± 0.47%	36.40% ± 0.50%	45.84% ± 0.62%	61.50% ± 0.58%	68.64% ± 0.53%

Methods	**EuroSAT**			**CropDiseases**		
	5-way 5-shot	5-way 20-shot	5-way 50-shot	5-way 5-shot	5-way 20-shot	5-way 50-shot
All embeddings	81.29% ± 0.62%	89.90% ± 0.41%	92.76% ± 0.34%	90.82% ± 0.48%	96.64% ± 0.25%	98.14% ± 0.18%
IMS-f	83.56% ± 0.59%	91.22% ± 0.38%	93.85% ± 0.30%	90.66% ± 0.48%	97.18% ± 0.24%	98.43% ± 0.16%

Table 4: The results of using all embeddings, and the *Incremental Multi-model Selection* (IMS-f) based on fine-tuned pre-trained models on the proposed benchmark.

Similar Conclusions Hold for Document Analysis

Experiments with the RVL-CDIP dataset (document classification)

	5-way 5-shot	**5-way 20-shot**	**5-way 1-shot**	**5-way 50-shot**
MatchinetNet	42.18% +- 0.72%	43.41% +- 0.64%	29.45% +- 0.59%	27.10% +- 0.48%
ProtoNet	49.92% +- 0.81%	55.74% +- 0.74%	35.27% +- 0.72%	57.20% +- 0.75%
MetaOpt	41.11% +- 0.72%	54.60% +- 0.75%	33.86% +- 0.71%	62.97% +- 0.72%
RelationNet	44.09% +- 0.71%	52.39% +- 0.70%	35.66% +- 0.79%	55.57% +- 0.66%
MAML	34.48% +- 0.69%	36.49% +- 0.67%	36.23% +- 0.80%	-
Fixed	51.31% +- 0.78%	61.82% +- 0.72%	35.03% +- 0.74%	66.30% +- 0.69%
Fine-tune	55.93% +- 0.79%	64.78% +- 0.75%	37.01% +- 0.76%	69.24% +- 0.70%
Ft last 1	55.11% +- 0.83%	63.58% +- 0.75%	36.44% +- 0.78%	67.36% +- 0.71%
Ft last 2	55.65% +- 0.77%	63.44% +- 0.74%	36.24% +- 0.80%	67.31% +- 0.70%
Ft last 3	55.55% +- 0.83%	63.80% +- 0.76%	36.40% +- 0.73%	67.80% +- 0.66%
Mean centroid	55.25% +- 0.79%	62.64% +- 0.76%	40.62% +- 0.81%	65.40% +- 0.75%
Cosine Classifier	55.42% +- 0.77%	63.02% +- 0.73%	41.17% +- 0.85%	65.98% +- 0.70%
IMS	53.46% +- 0.85%	63.18% +- 0.77%	38.92% +- 0.79%	69.07% +- 0.75%

We have shown that fine-tuning methods outperform meta-learning methods for cross domain few-shot learning

How to choose which layers to fine-tune for a given dataset?

Where to fine-tune in a deep network?

- Fine-tune just the last layer?
- Fine-tune the last K layers?
- Fine-tune all network parameters?
- Fine-tune a non-contiguous set of layers?
- How to make these choices for high capacity models with 10s, or 100s, or 1000s of layers?

Where to fine-tune in a deep network?

- Fine-tune just the last layer?
- Fine-tune the last K layers?
- Fine-tune all network parameters?
- Fine-tune a non-contiguous set of layers?
- How to make these choices for high capacity models with 10s, or 100s, or 1000s of layers?

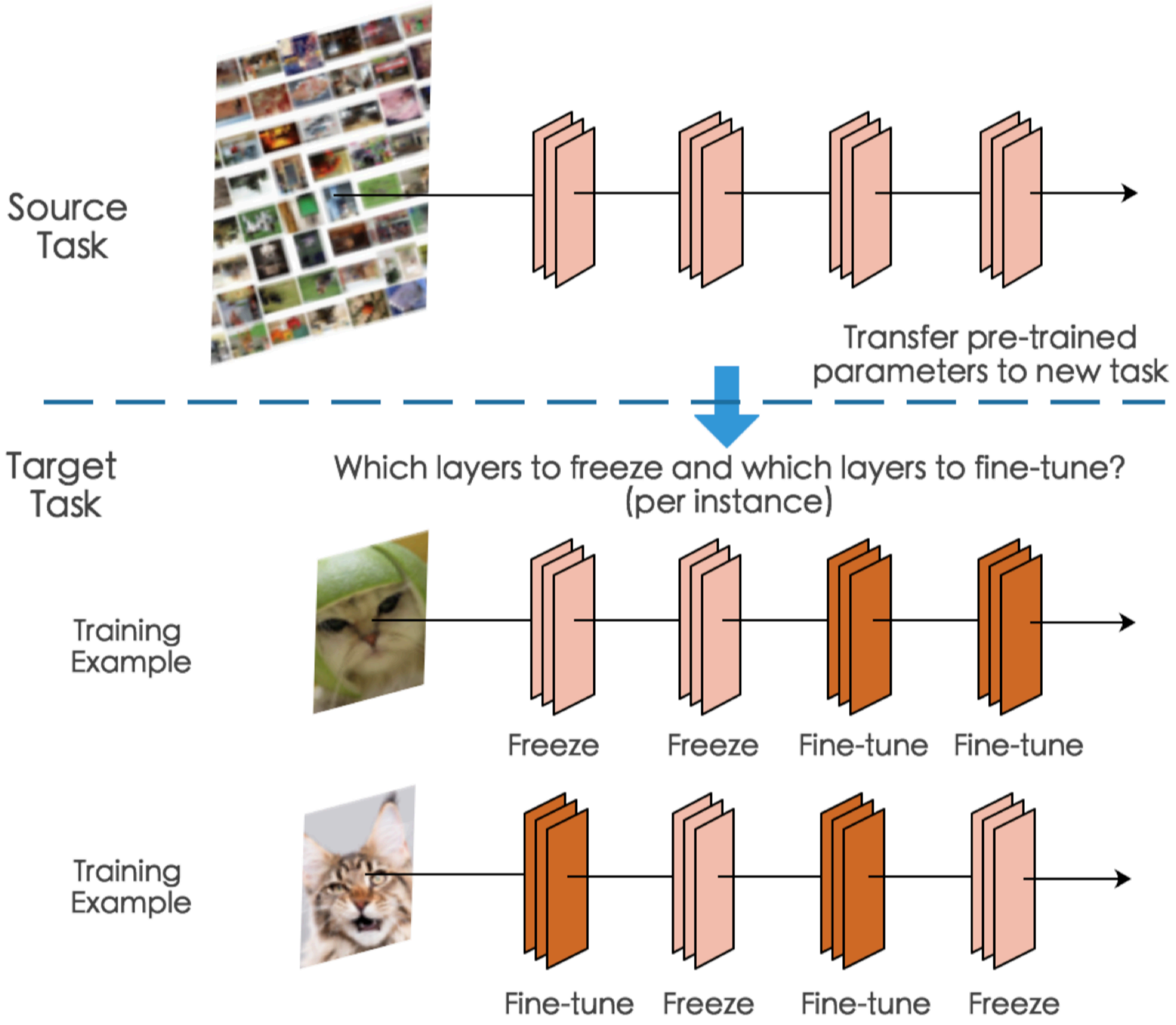
It depends on the dataset, pre-trained model, ...

Fine-tuning is an art !

SpotTune: Transfer Learning through Adaptive Fine-Tuning

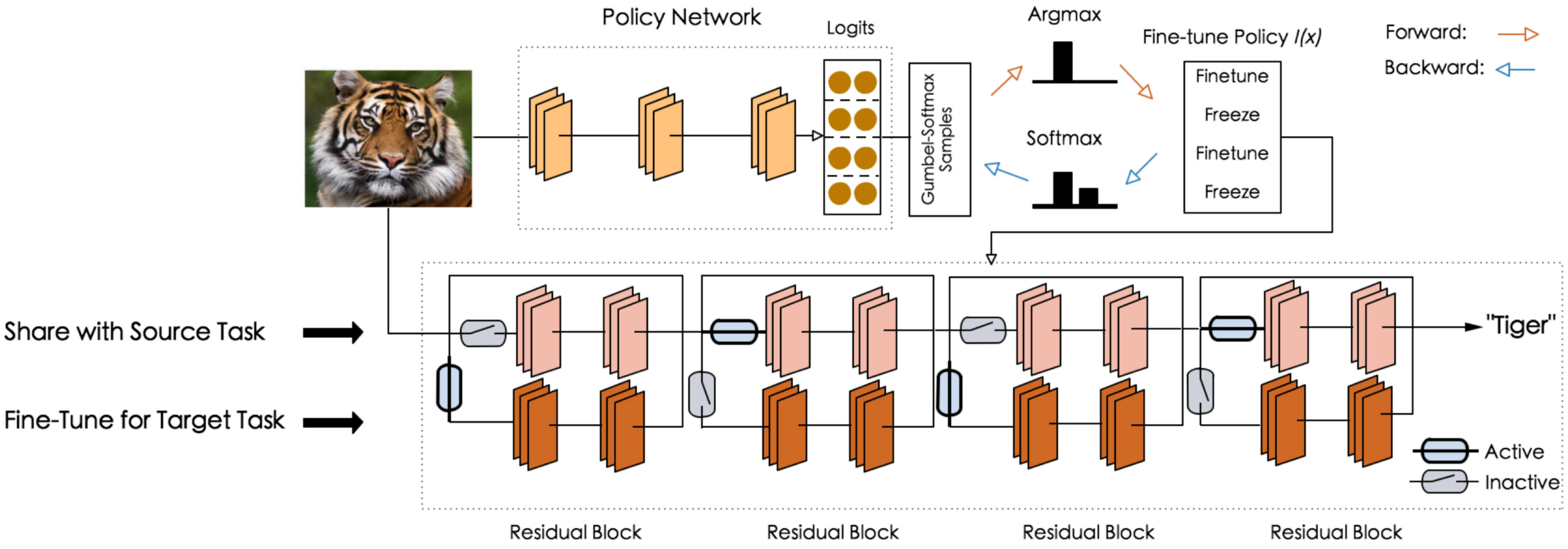
Yunhui Guo, Honghui Shi, Abhishek Kumar, Kristen Grauman,
Tajana Rosing, Rogerio Feris

CVPR 2019



[Guo et al, CVPR 2019]

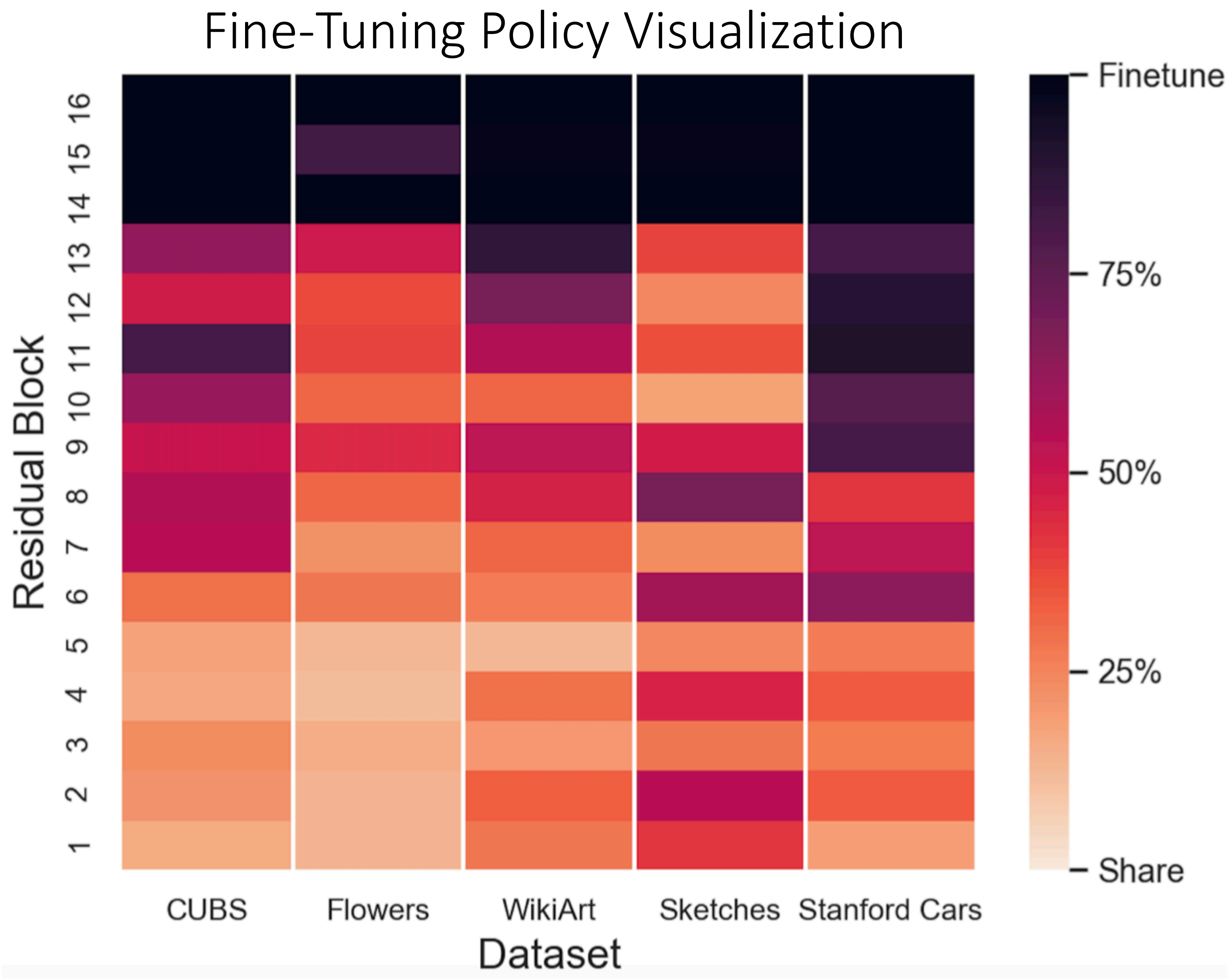
SpotTune: Transfer Learning through Adaptive Fine-Tuning



* General approach to any architecture (ResNet, VGG, ...)

[Guo et al, CVPR 2019]

SpotTune: Transfer Learning through Adaptive Fine-Tuning



SpotTune automatically identifies the right fine-tuning policy for each dataset, for each training example.

[Guo et al, CVPR 2019]

SpotTune: Transfer Learning through Adaptive Fine-Tuning

Model	CUBS	Stanford Cars	Flowers	WikiArt	Sketches
Feature Extractor	74.07%	70.81%	85.67%	61.60%	75.50%
Standard Fine-tuning	81.86%	89.74%	93.67%	75.60%	79.58%
Stochastic Fine-tuning	81.03%	88.94%	92.95%	73.06%	78.30%
Fine-tuning last-3	81.54%	88.21%	89.03%	72.68 %	77.72%
Fine-tuning last-2	80.34%	85.36%	91.81%	70.82%	78.37%
Fine-tuning last-1	78.68%	81.73%	89.99%	68.96%	77.20%
Random Policy	81.63 %	88.57%	93.44%	73.82%	78.30%
Fine-tuning ResNet-101	82.13%	90.32%	94.21%	**76.52%**	78.92%
L^2-SP	83.69%	91.08%	95.21%	75.38%	79.60%
Progressive Neural Nets	83.08 %	91.59%	95.55%	75.41%	79.71%
SpotTune (running fine-tuned blocks)	82.36%	92.04%	93.49%	67.27%	78.88%
SpotTune (Global-k)	83.48%	90.51%	**96.60%**	75.63%	80.02%
SpotTune	**84.03 %**	**92.40%**	96.34%	75.77%	**80.20%**

[Guo et al, CVPR 2019]

SpotTune: Transfer Learning through Adaptive Fine-Tuning

	#par	ImNet	Airc.	C100	DPed	DTD	GTSR	Flwr	OGlt	SVHN	UCF	Score
Scratch	10x	59.87	57.10	75.73	91.20	37.77	96.55	56.30	88.74	96.63	43.27	1625
Scratch+ [37]	11x	59.67	59.59	76.08	92.45	39.63	96.90	56.66	88.74	96.78	44.17	1826
Feature Extractor	1x	59.67	23.31	63.11	80.33	55.53	68.18	73.69	58.79	43.54	26.80	544
Fine-tuning [38]	10x	60.32	61.87	82.12	92.82	55.53	99.42	81.41	89.12	96.55	51.20	3096
BN Adapt. [5]	1x	59.87	43.05	78.62	92.07	51.60	95.82	74.14	84.83	94.10	43.51	1353
LwF [26]	10x	59.87	61.15	82.23	92.34	58.83	97.57	83.05	88.08	96.10	50.04	2515
Series Res. adapt. [37]	2x	60.32	61.87	81.22	93.88	57.13	99.27	81.67	89.62	96.57	50.12	3159
Parallel Res. adapt. [38]	2x	60.32	64.21	81.92	94.73	58.83	99.38	84.68	89.21	96.54	50.94	3412
Res. adapt. (large) [37]	12x	67.00	67.69	84.69	94.28	59.41	97.43	84.86	89.92	96.59	52.39	3131
Res. adapt. decay [37]	2x	59.67	61.87	81.20	93.88	57.13	97.57	81.67	89.62	96.13	50.12	2621
Res. adapt. finetune all [37]	2x	59.23	63.73	81.31	93.30	57.02	97.47	83.43	89.82	96.17	50.28	2643
DAN [39]	2x	57.74	64.12	80.07	91.30	56.54	98.46	86.05	89.67	96.77	49.48	2851
PiggyBack [31]	1.28x	57.69	65.29	79.87	96.99	57.45	97.27	79.09	87.63	97.24	47.48	2838
SpotTune	11x	60.32	63.91	80.48	96.49	57.13	99.52	85.22	88.84	96.72	52.34	**3612**

SpotTune sets the new state of the art on the Visual Decathlon Challenge

[Guo et al, CVPR 2019]

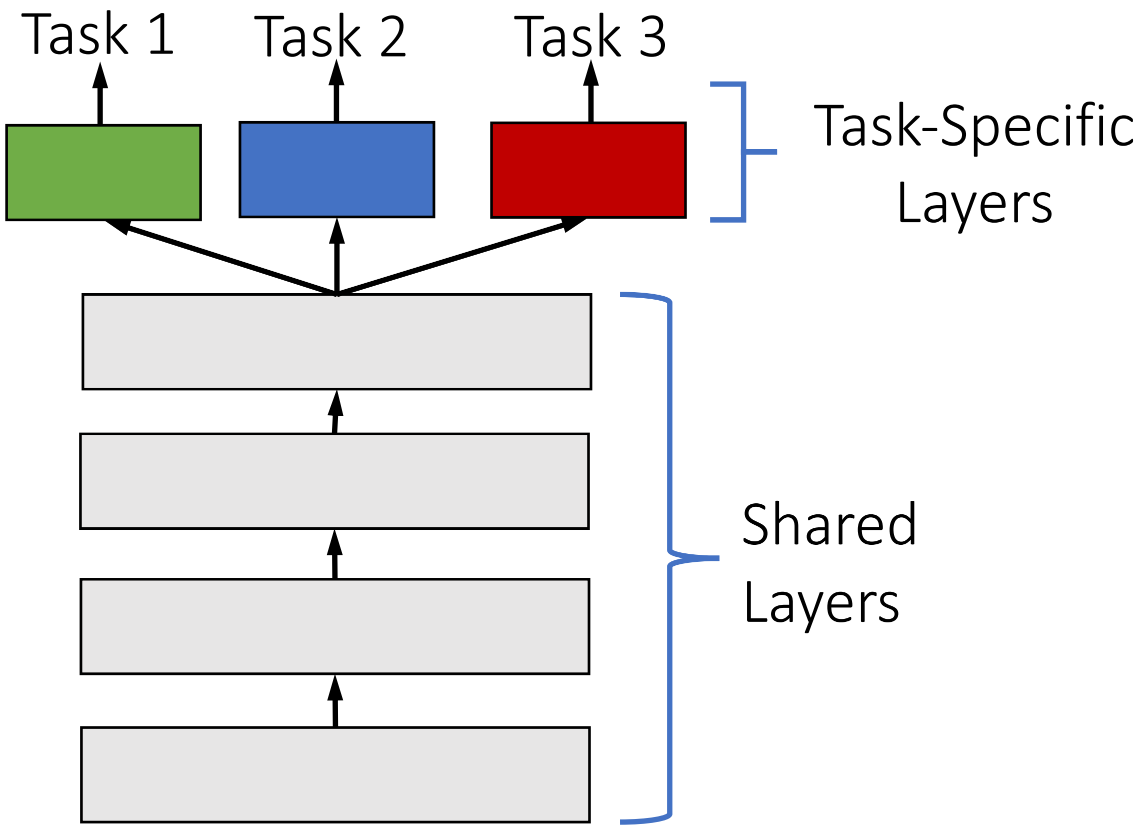
AdaShare: Learning What to Share for Efficient Multi-Task Learning

Ximeng Sun, Rameswar Panda, Rogerio Feris, Kate Saenko

NeurIPS 2020

Hard Parameter Sharing

- Hand-designed architectures composed of base layers that are shared across tasks and specialized branches that learn task-specific features.

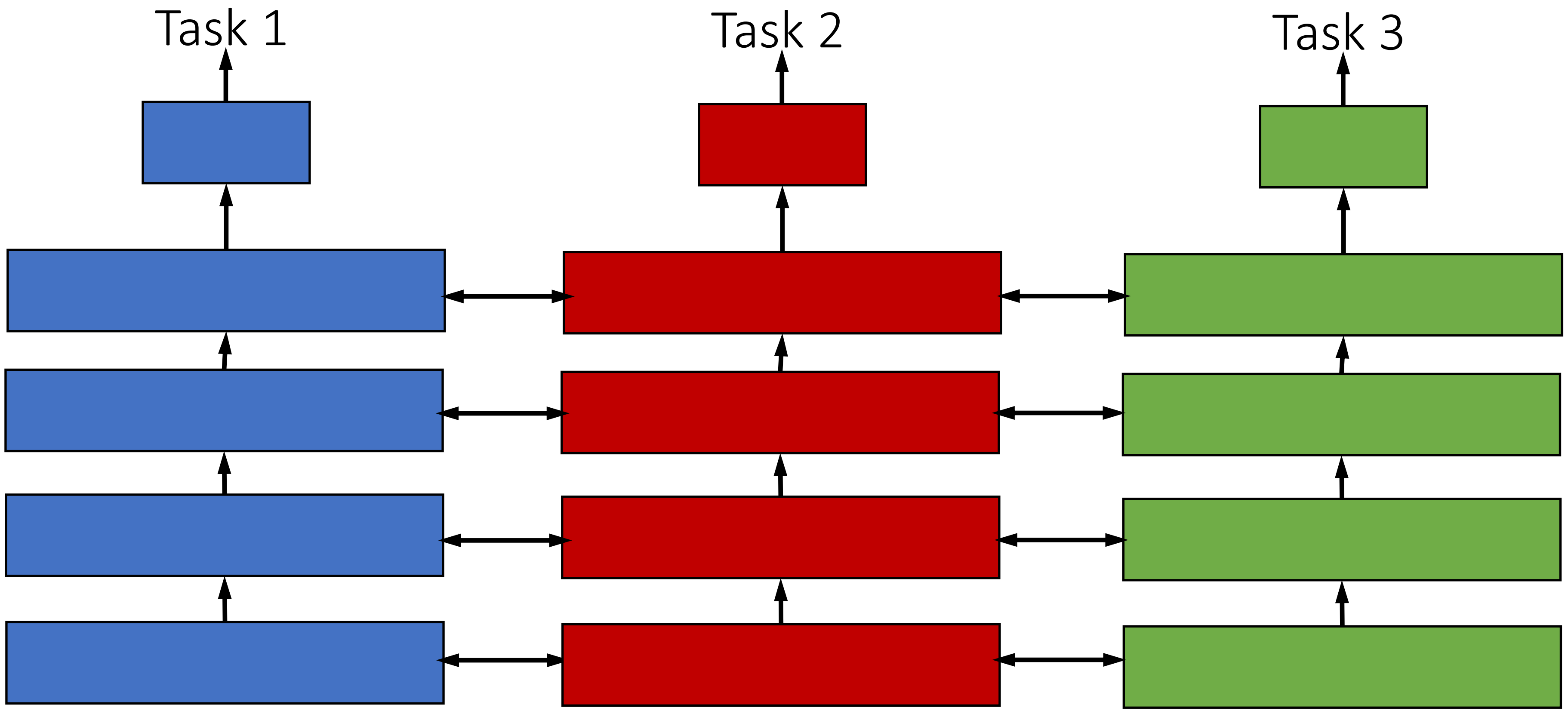


- Performance depends on "where to branch" in the network [Misra et al, 2016]
- The space of possible branching architectures is combinatorially large !

Soft Parameter Sharing

- Network column for each task and a mechanism for feature sharing between columns.

Number of parameters grow linearly with the number of tasks !

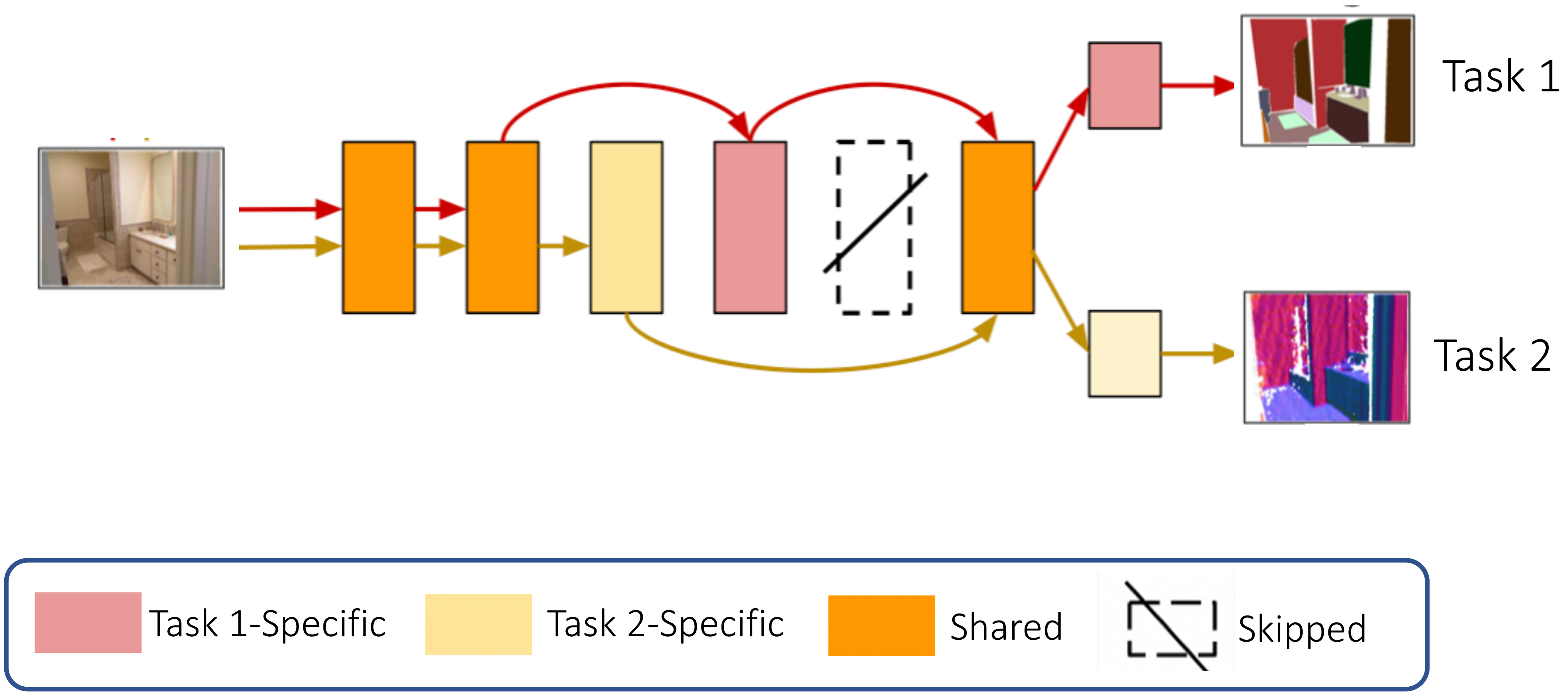


Problem

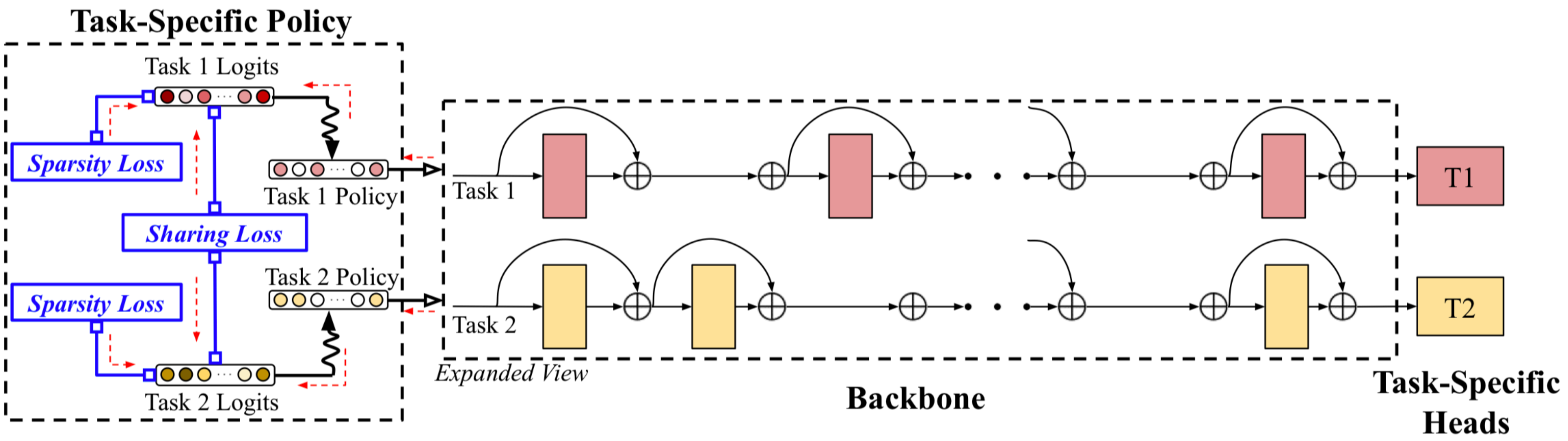
Can we determine which layers in the network should be shared across which tasks and which layers should be task-specific to achieve the best accuracy/memory footprint trade-off for scalable and efficient multi-task learning?

Proposed Approach: AdaShare

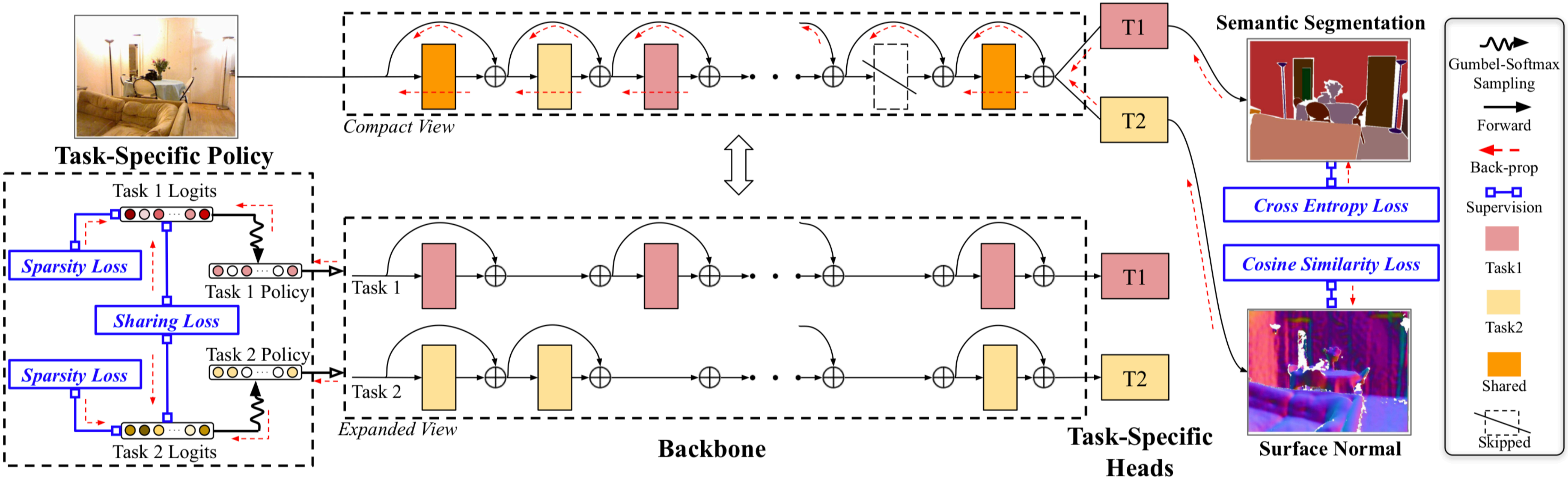
- Single network that supports separate execution paths for different tasks



AdaShare: Learning what to Share in Multi-Task Learning



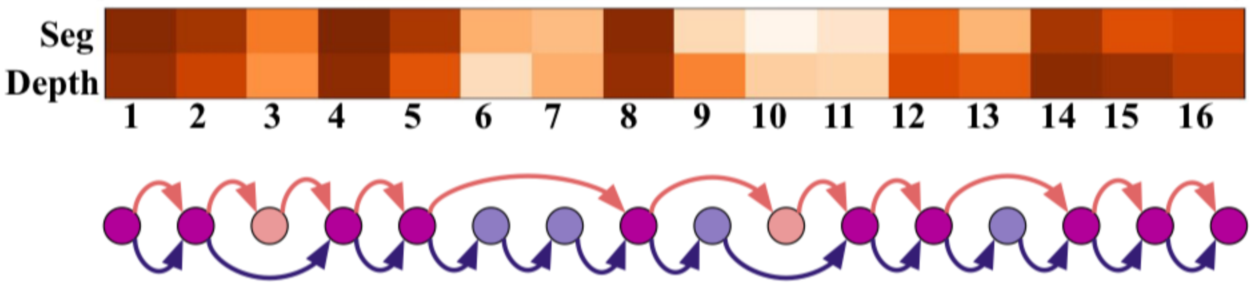
AdaShare: Learning what to Share in Multi-Task Learning



AdaShare: Experimental Results

- **CityScapes [2 tasks].** *AdaShare* achieves the best performance on 5 out of 7 metrics using less than 1/2 parameters of most baselines.

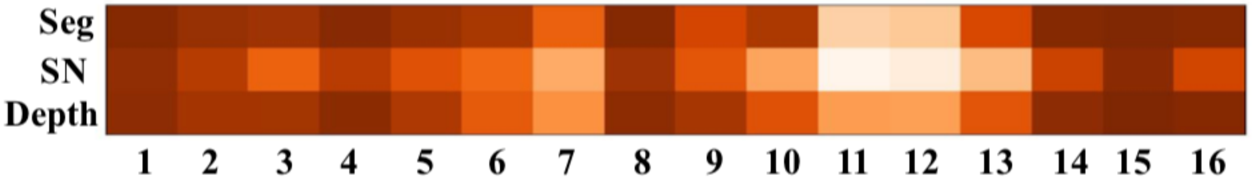
Model	# Params ↓	Semantic Seg.		Depth Prediction				
		mIoU ↑	Pixel Acc ↑	Error↓		δ, within ↑		
				Abs	Rel	1.25	1.25^2	1.25^3
Single-Task	2	40.2	74.7	0.017	0.33	70.3	86.3	93.3
Multi-Task	**1**	37.7	73.8	0.018	0.34	72.4	88.3	94.2
Cross-Stitch	2	40.3	74.3	**0.015**	**0.30**	74.2	89.3	**94.9**
Sluice	2	39.8	74.2	0.016	0.31	73.0	88.8	94.6
NDDR-CNN	2.07	**41.5**	74.2	0.017	0.31	74.0	89.3	94.8
MTAN	2.41	40.8	74.3	**0.015**	0.32	75.1	89.3	94.6
AdaShare	**1**	**41.5**	**74.9**	0.016	0.33	**75.5**	**89.8**	**94.9**

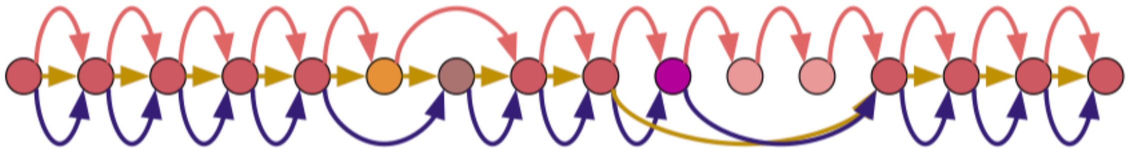


AdaShare: Experimental Results

- **NYU v2 [3 tasks].** AdaShare achieves the best performance on 10 out of 12 metrics using less than 1/3 parameters of most baselines.

Model	# Params ↓	Semantic Seg.		Surface Normal Prediction					Depth Prediction				
		mIoU ↑	Pixel Acc ↑	Error ↓		θ, within ↑			Error ↓		δ, within ↑		
				Mean	Median	11.25°	22.5°	30°	Abs	Rel	1.25	1.25^2	1.25^3
Single-Task	3	27.5	58.9	17.5	15.2	34.9	73.3	85.7	0.62	0.25	57.9	85.8	95.7
Multi-Task	**1**	24.1	57.2	**16.6**	13.4	42.5	73.2	84.6	0.58	0.23	62.4	88.2	96.5
Cross-Stitch	3	25.4	57.6	17.2	14.0	41.4	70.5	82.9	0.58	0.23	61.4	88.4	95.5
Sluice	3	23.8	56.9	17.2	14.4	38.9	71.8	83.9	0.58	0.24	61.9	88.1	96.3
NDDR-CNN	3.15	21.6	53.9	17.1	14.5	37.4	**73.7**	**85.6**	0.66	0.26	55.7	83.7	94.8
MTAN	3.11	26.0	57.2	**16.6**	13.0	43.7	73.3	84.4	0.57	0.25	62.7	87.7	95.9
AdaShare	**1**	**30.2**	**62.4**	**16.6**	**12.9**	**45.0**	71.7	83.0	**0.55**	**0.20**	**64.5**	**90.5**	**97.8**

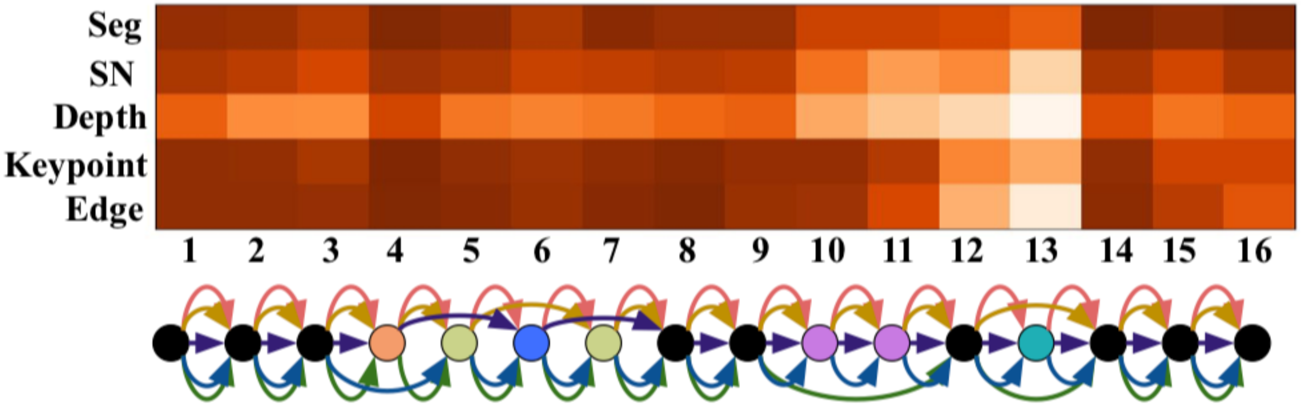




AdaShare: Experimental Results

- **Tiny-Taskonomy [5 Tasks].** AdaShare outperforms the baselines on 3 out of 5 tasks using less than 1/5 parameters of most baselines.

Models	# Params ↓	Seg ↓	SN ↑	Depth ↓	Keypoint ↓	Edge ↓
Single-Task	5	0.575	**0.707**	**0.022**	0.197	0.212
Multi-Task	**1**	0.587	0.702	0.024	0.194	0.201
Cross-Stitch	5	0.560	0.684	**0.022**	0.202	0.219
Sluice	5	0.610	0.702	0.023	**0.192**	0.198
NDDR-CNN	5.41	**0.539**	0.705	0.024	0.194	0.206
MTAN	4.51	0.637	0.702	0.023	0.193	0.203
AdaShare	**1**	0.566	**0.707**	0.025	**0.192**	**0.193**



Visual Learning Beyond Natural Images: Summary

- Naïve fine-tuning outperforms current meta-learning approaches for cross-domain (beyond natural images) few-shot learning
- The optimal set of layers to fine-tune is dependent on the dataset. SpotTune automatically decides which layers of a model should be shared with the pre-trained model and which layers should be fine-tuned
- Deciding what features should be shared is also crucial for joint multi-task learning. AdaShare selects specific computational paths for each task to maximize accuracy and efficiency.

Thank you !

- Y. Guo, N. Codella, L. Karlinsky, J. Smith, T. Rosing and R. S. Feris. ”A new Benchmark for Evaluation of Cross-domain Few-shot Learning." ECCV 2020
(* equal contribution) [code available]

- Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing and R. S. Feris. “SpotTune: Transfer Learning Through Adaptive Fine-Tuning” CVPR 2019 [code available]

- X. Sun, R. Panda and R. S. Feris. “AdaShare: Learning What to Share for Efficient Deep Multi-Task Learning”. NeurIPS 2020