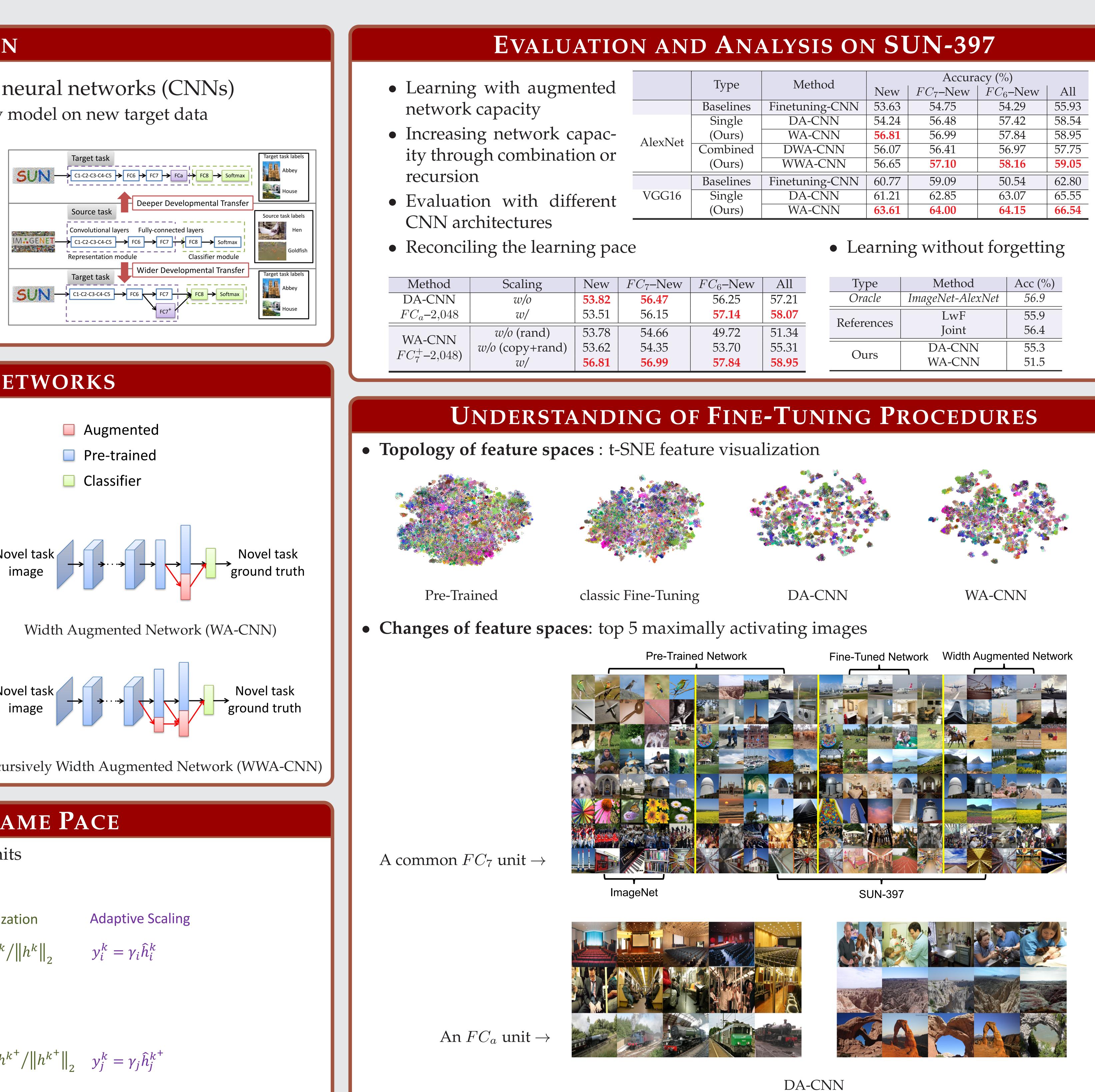
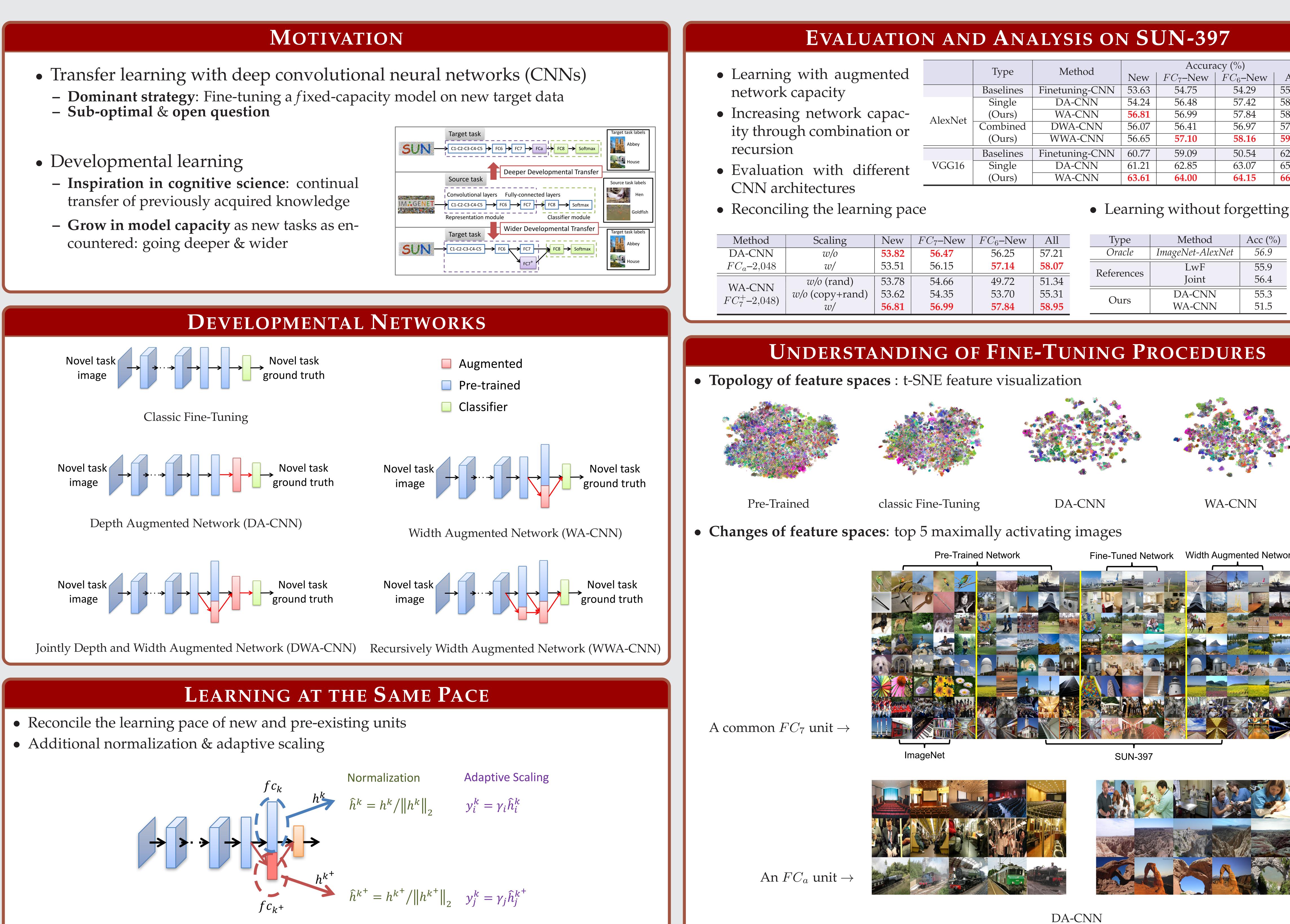
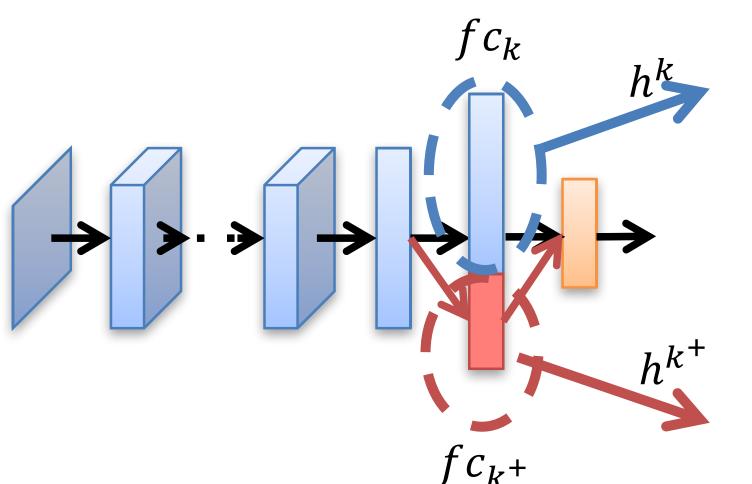


- countered: going deeper & wider







$$= h^k / \left\| h^k \right\|_2 \qquad y_i^k$$

$$y_i^k = \gamma_i \hat{h}_i^k$$

$$\hat{h}^{k^+} = h^{k^+} / \|h^{k^+}\|_2 \quad y_j^k = \gamma_j \hat{h}_j^{k^-}$$

GROWING A BRAIN: FINE-TUNING BY INCREASING MODEL CAPACITY

Deva Ramanan Martial Hebert Yu-Xiong Wang Email: {yuxiongw,dramanan,hebert}@cs.cmu.edu

GENERALIZATION TO OTHER TASKS AND DATASETS

- Networks: Width augmented networks (WA-CNN)
- Target tasks: Scene classification & fine-grained recognition & action recognition

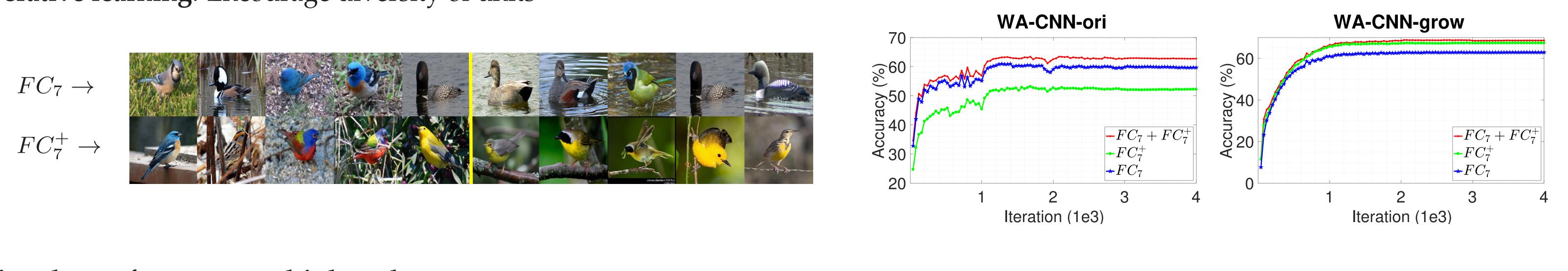
	MIT-67		102 Flowers		CUB200-2	011	Stanford-40	
Туре	Approach	Acc(%)	Approach	Acc(%)	Approach	Acc(%)	Approach	Acc(%)
	Ours	61.2	Ours	75.3	Ours	62.9	Ours	57.7
ImageNet CNNs	Caffe	59.5	CNN-SVM	74.7	CNN-SVM	53.3	Deep Standard	58.9
\sim	—		CNNaug-SVM	86.8	CNNaug-SVM	61.8		
	Caffe-DAG	64.6	LSVM	87.1	LSVM	61.4	Deep Optimized	66.4
Task Customized			MsML	89.5	DeCaf+DPD	65.0		
CNNs	Places	68.2	MPP	91.3	MsML	67.9		
			Deep Optimized	91.3				
Data Augmented CNNs	Combined-AlexNet	58.8	Combined-AlexNet	83.3			Combined-AlexNet	56.4
Multi-Task CNNs	Joint	63.9			Joint	56.6		
IVIUIU-IASK CININS	LwF	64.5			LwF	57.5		
Ours	WA-CNN	66.3	WA-CNN	92.8	WA-CNN	69.0	WA-CNN	67.5

A SINGLE UNIVERSAL HIGHER CAPACITY MODEL?

• Should standard models have used higher capacity even for the source task?

Dataset	CNN	WA-CNN-scratch	WA-CNN-grow (Ours)	Dataset	CNN-FT	WA-CNN-ori	WA-CNN-grow (Ours)
			0 1	MIT-67	61.2	62.3	66.3
ImageNet	56.9	57.6	57.8	CUB200-2011	62.9	63.2	69.0

• **Cooperative learning**: Encourage diversity of units



Continual transfer across multiple tasks					
	WA-CNN (Ours)		Baselines		
Scenarios	ImageNet→MIT67	ImageNet	Places	ImageNet-VGG	
Acc(%)	66.3	79.3	68.2	74.7	

- Newly added units should have a pace of learning that is comparable to the pre-existing units.
- process when learning from continuously evolving data streams and tasks.

-39)'/	
Accura	acy (%)	
–New	FC_6 –New	All
4.75	54.29	55.93
6.48	57.42	58.54
6.99	57.84	58.95
6.41	56.97	57.75
7.10	58.16	59.05
9.09	50.54	62.80
2.85	63.07	65.55
1 00	6/ 15	66 51

/lethod	Acc (%)
Net-AlexNet	56.9
LwF	55.9
Joint	56.4
A-CNN	55.3
A-CNN	51.5



CONCLUSIONS

• Increasing model capacity significantly helps existing units better adapt and specialize to the target task.

• A developmental view of CNN optimization, where model capacity is progressively grown throughout a lifelong learning