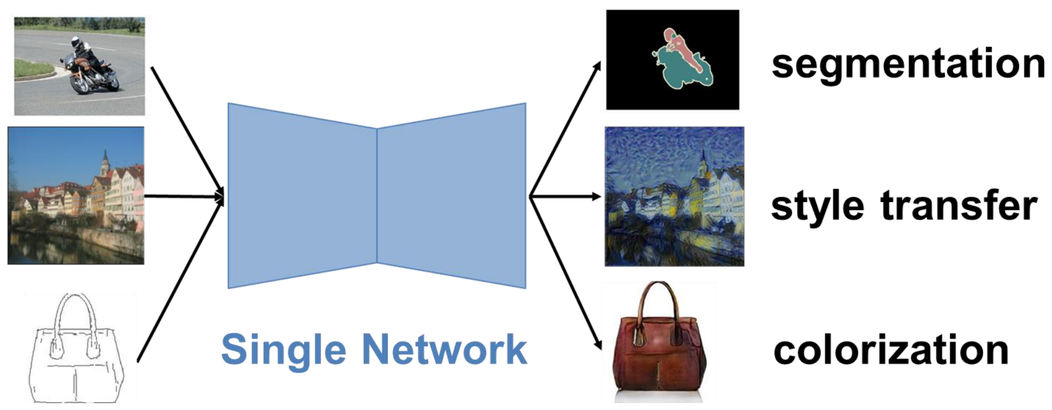
ACPR ID-103

Continual Learning of An Image Transformation Network Using Task-dependent Weight Selection Masks

Asato Matsumoto **Keiji Yanai** The University of Electro-Communications, Tokyo

Objective

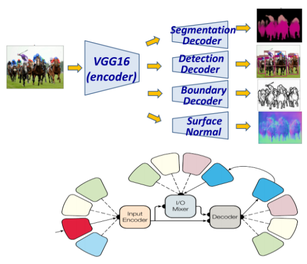
Continual Learning of An Image Transformation Network for heterogeneous tasks



Related Work

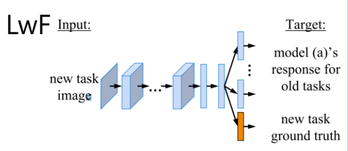
Simultaneous training of multiple tasks

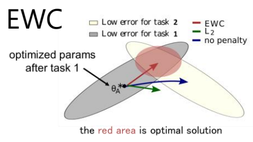
- Single encoder & task-specific decoders
  e.g. UberNet [Ikkinos CVPR2017]
- Multiple inputs & multiple outputs
  e.g. One Model To Learn Them All [Kaiser et al. arXiv 2017]

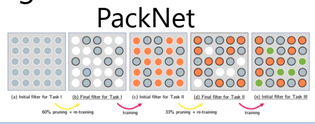


Continual learning (approaches for overcoming "catastrophic forgetting")

- **Rehearsal** [Hetherington et al. 1989]
  train new samples with old samples
- **Distillation** Learning without Forgetting [Li and Hoiem 2016]
  reproduce training labels of old tasks with trained model and use them for new trainig
- **Regularization** Elastic Weight Consolidation [Kirkpatrick et al. 2016]
  train weights for new tasks according to un-importantness of weights
- **Pruning** PackNet [Mallya et al. CVPR2018]
  fix trained weight for previous tasks and pruning un-important weights
- **Weight Selection** Piggyback [Mallya et al. ECCV2018]
  select task-specific weights from the fixed backbone network
  In this work, we adapt "Piggyback" as a basic approach.





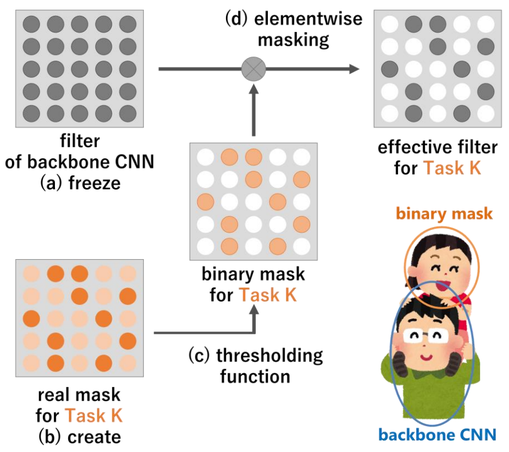


Method : Piggyback [Mallya et al. ECCV 2018]

Use task-dependent weight selection masks.

No catastrophic forgetting happen with
small additional binary masks and task-specific final layers.

(1) For the first, train and fix the weight of backbone network
(2) For the second task or more, train mask weights and obtain the task-specific mask
(3) At evaluation time, use effective filter by element-wise masking

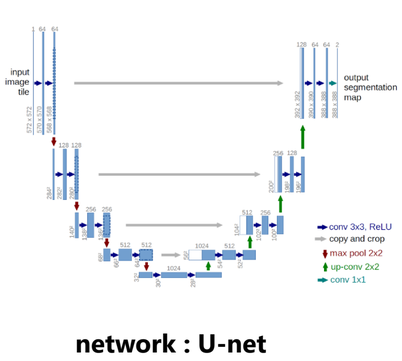


Dataset	Classifier Only	PackNet [7] ↓	PackNet [7] ↑	Piggyback (ours)	Individual Networks
ImageNet	28.42 (9.61)	29.33 (9.99)		28.42 (9.61)	28.42 (9.61)
CUBS	36.49	22.30	29.69	20.99	21.30
Stanford Cars	54.66	15.81	21.66	11.87	12.49
Flowers	20.01	10.33	10.25	7.19	7.35
WikiArt	49.53	32.80	31.48	29.91	29.84
Sketch	58.53	28.62	24.88	22.70	23.54
# Models (Size)	1 (537 MB)	1 (587 MB)		1 (621 MB)	6 (3,222 MB)

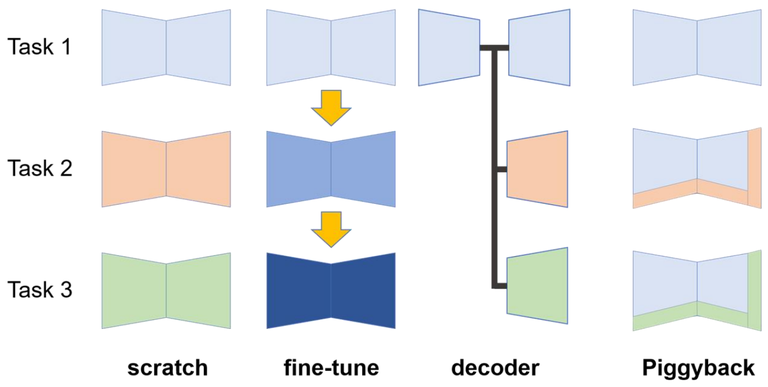
Table 2: Errors obtained by starting from an ImageNet-trained VGG-16 network and then using various methods to learn new fine-grained classification

Baselines

Scratch independent model
Fine-tune fine-tune a single identical network
Decoder shared encoder and task-dependent decoder
Piggyback adapt Piggyback to encoder-decoder network



network : U-net



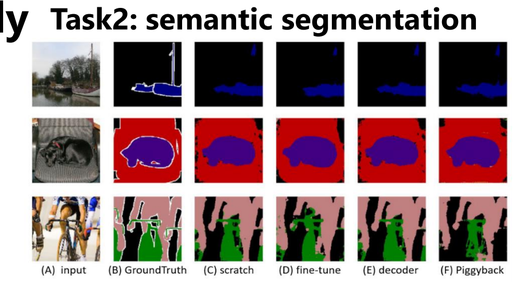
Experiment

Apply Piggyback to image-to-image tasks

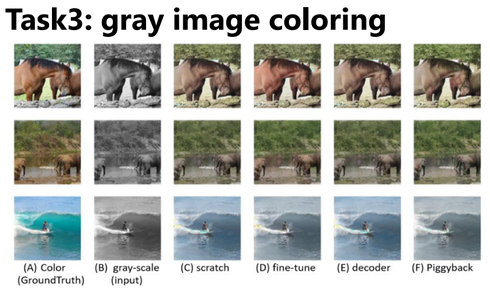
- **6 tasks to be learned continuously**

task number	task category	dataset	evaluation
Task 1	semantic segmentation	MS COCO	mIoU(%)
Task 2	semantic segmentation	Pascal VOC 2012	mIoU(%)
Task 3	gray image coloring	MS COCO	SSIM
Task 4	style transfer (Gogh)	MS COCO	SSIM
Task 5	style transfer (Munk)	MS COCO	loss
Task 6	edge image coloring	edges2handbags	MSE

Task2: semantic segmentation



Task3: gray image coloring

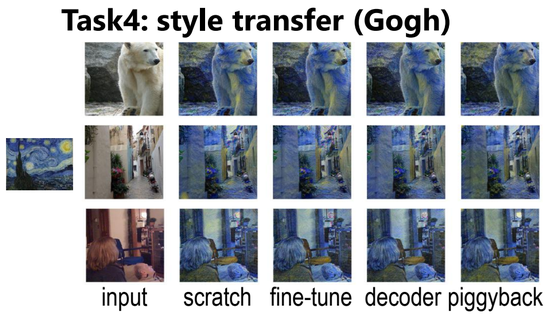


- **Result**

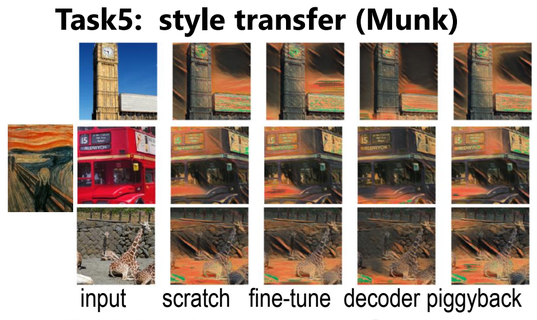
"Piggyback" is effective for Encoder-Decoder net

	scratch	fine-tune	decoder	Piggyback
Task 1 (mIoU(%))	21.47			
Task 2 (mIoU(%))	58.59	64.87	61.63	61.45
Task 3 (SSIM)	0.9138	0.9148	0.9121	0.9058
Task 4 (SSIM)	0.3678	0.3555	0.3595	0.3501
Task 5 (total loss)	447480	490490	544348	521476
Task 6 (MSE)	211.96	207.76	237.53	232.02
Task 1 after Task 2	-	0.70 (catastrophic forgetting)	21.47 (NOT catastrophic forgetting)	21.47
Task 2 after Task 3	-	1.87 (catastrophic forgetting)	61.63 (NOT catastrophic forgetting)	61.45
Task 3 after Task 4	-	0.5321 (catastrophic forgetting)	0.9121 (NOT catastrophic forgetting)	0.9058
Model Size (MB)	338.4 (56.4*6)	338.4 (56.4*6)	158.9 (56.4+20.5*5)	65.4 (56.4+1.8*5)

Task4: style transfer (Gogh)



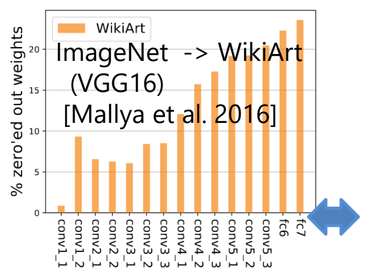
Task5: style transfer (Munk)

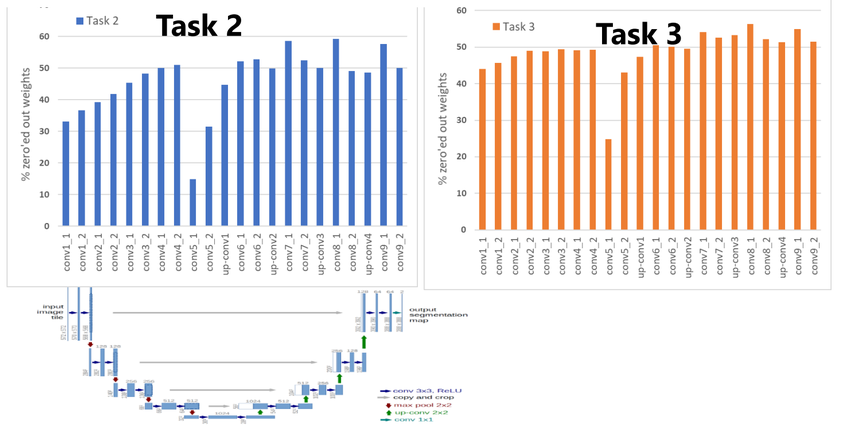


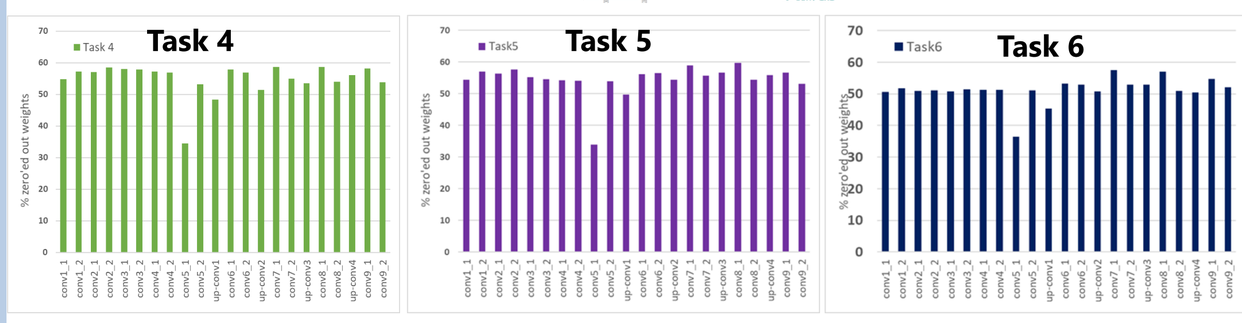
Task6: edge image coloring

- **The ratio of Zero'ed out weights**

the zero'ed out weights of image translation is higher than classification







- **The similarity matrix between binary masks (a ratio of same value)**

	Task 1 segmentation	Task 2 segmentation	Task 3 gray coloring	Task 4 style transfer	Task 5 style transfer	Task 6 edge coloring
Task 2	0.5075	-	-	-	-	-
Task 3	0.5042	0.5054	-	-	-	-
Task 4	0.4326	0.5034	0.5020	-	-	-
Task 5	0.4529	0.5029	0.5025	0.5210	-	-
Task 6	0.4847	0.5063	0.5026	0.5093	0.5077	-

Future work

- Reducing the size of binary mask
- Analyzing the trained mask