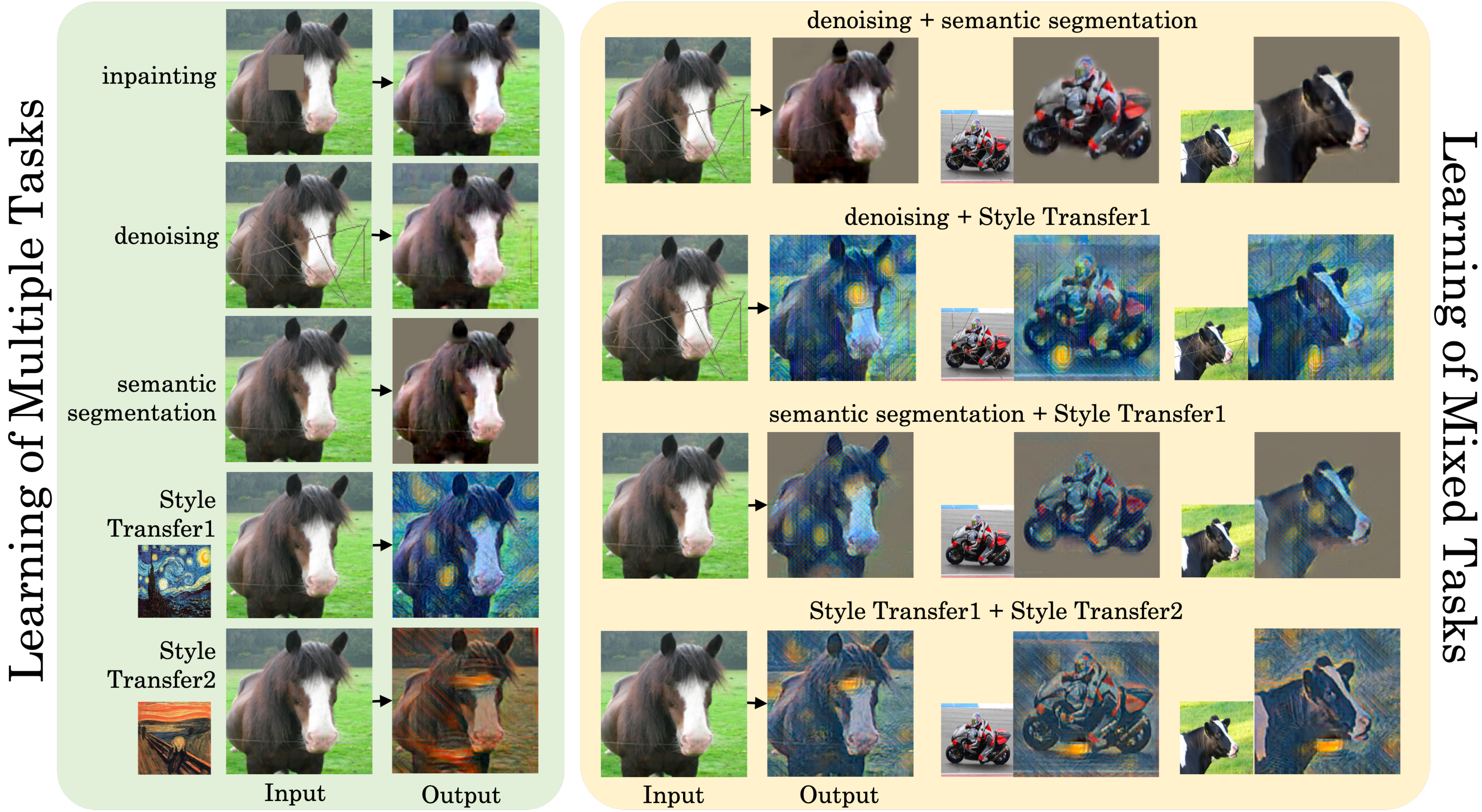
Training of Multiple and Mixed Tasks With A Single Network Using Feature Modulation

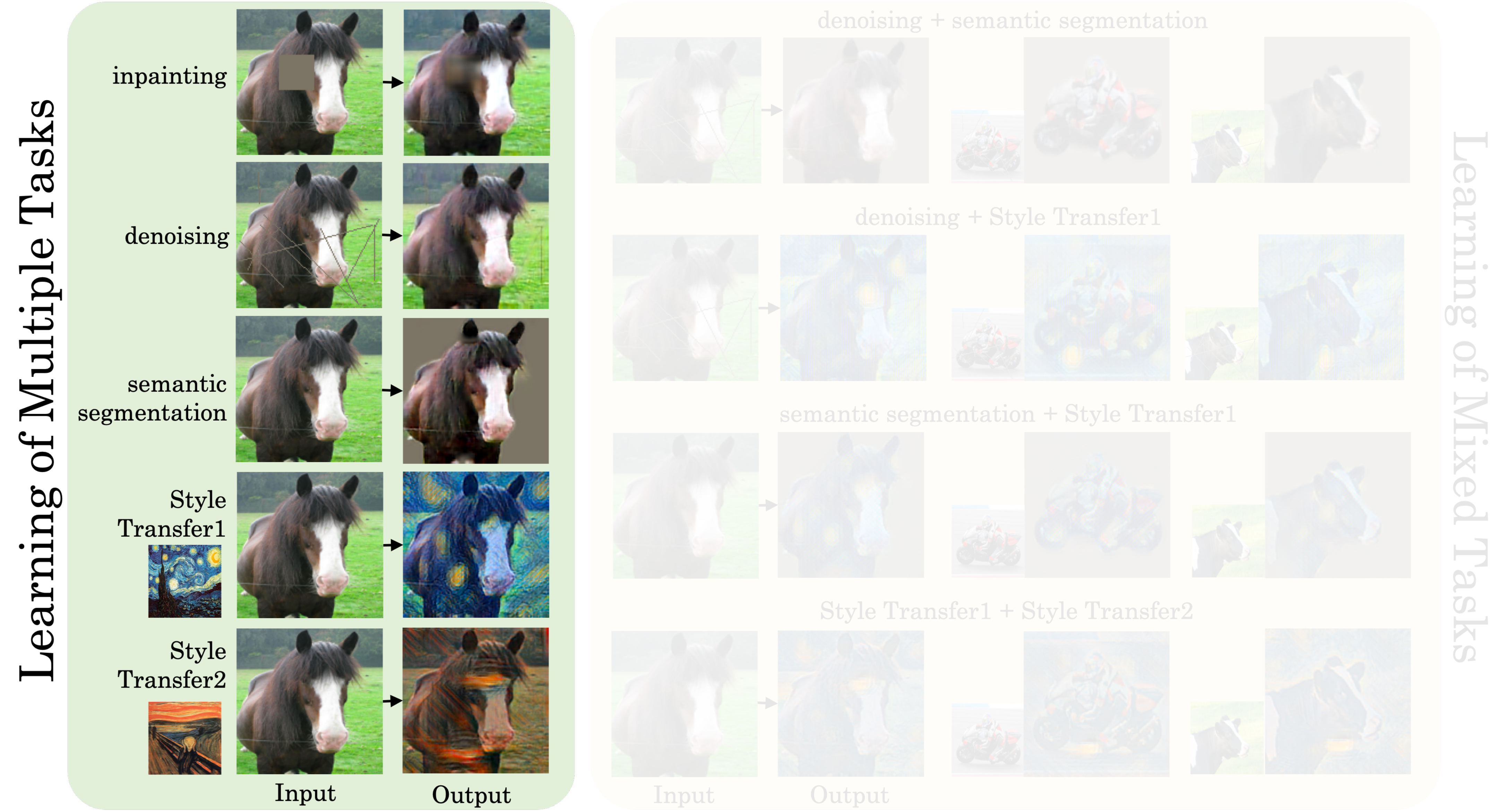
Mana Takeda[1], Gibran Benitez[1], and Keiji Yanai[1]
[1]The University of Electro-Communications, Tokyo

Abstract



Learning of Multiple Tasks
inpainting
denoising
semantic segmentation
Style Transfer1
Style Transfer2
Input
Output
Learning of Mixed Tasks
denoising + semantic segmentation
denoising + Style Transfer1
semantic segmentation + Style Transfer1
Style Transfer1 + Style Transfer2
Input
Output

Abstract



Learning of Multiple Tasks
inpainting
denoising
semantic segmentation
Style Transfer1
Style Transfer2
Input
Output
denoising + semantic segmentation
denoising + Style Transfer1
semantic segmentation + Style Transfer1
Style Transfer1 + Style Transfer2
Input
Output
Learning of Mixed Tasks

Abstract

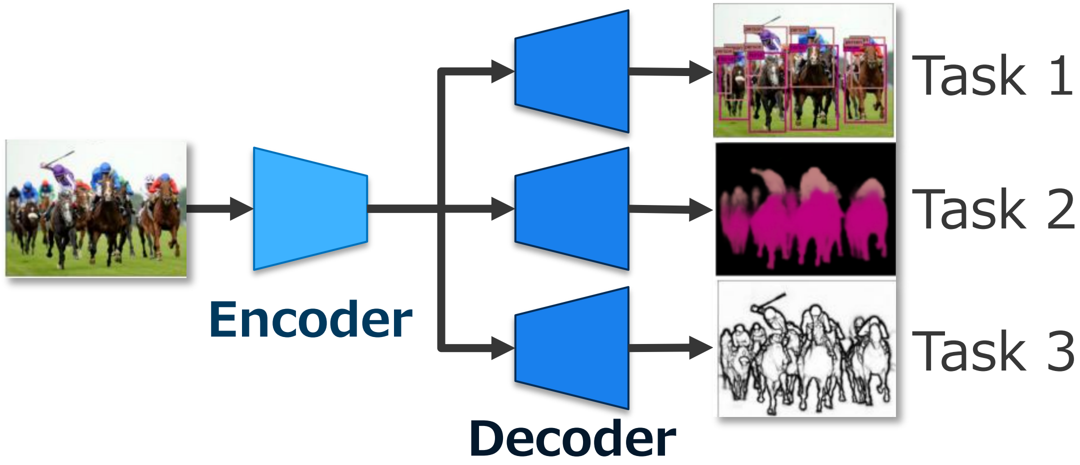
Learning of Multiple Tasks
inpainting
denoising
semantic segmentation
Style Transfer1
Style Transfer2
Input
Output


denoising + semantic segmentation
denoising + Style Transfer1
semantic segmentation + Style Transfer1
Style Transfer1 + Style Transfer2
Input
Output
Learning of Mixed Tasks

Introduction

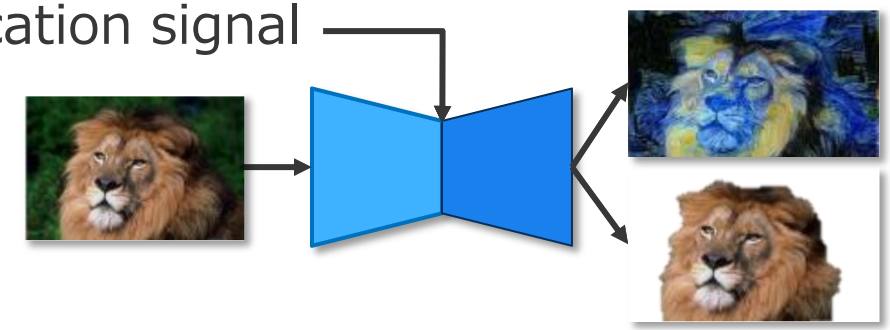
- Multi Task Learning
  - Run multiple tasks on a single network.
  - In general, MTL models require task-specific parts, in addition to the parts shared by all the tasks.
  - As the number of tasks increases, the network becomes larger.

- We propose a single network with negligibly small task-specific parts.



General multi-task learning model

1. Learning multiple heterogeneous image transfer tasks in a single network.

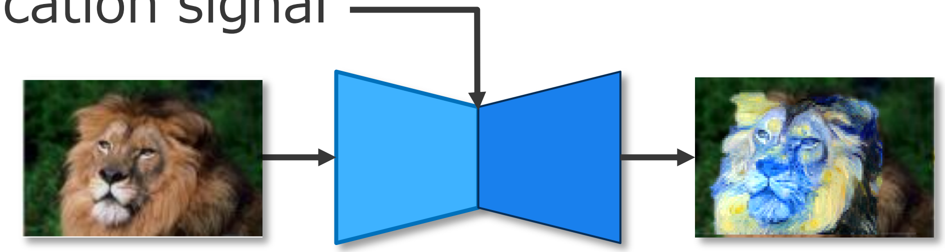


2. Learning mixed-task in the proposed network using synthesized mixed-task training samples.

Objective 2

1. Learning multiple heterogeneous image transfer tasks in a single network.

2. Learning mixed-task in the proposed network using synthesized mixed-task training samples.



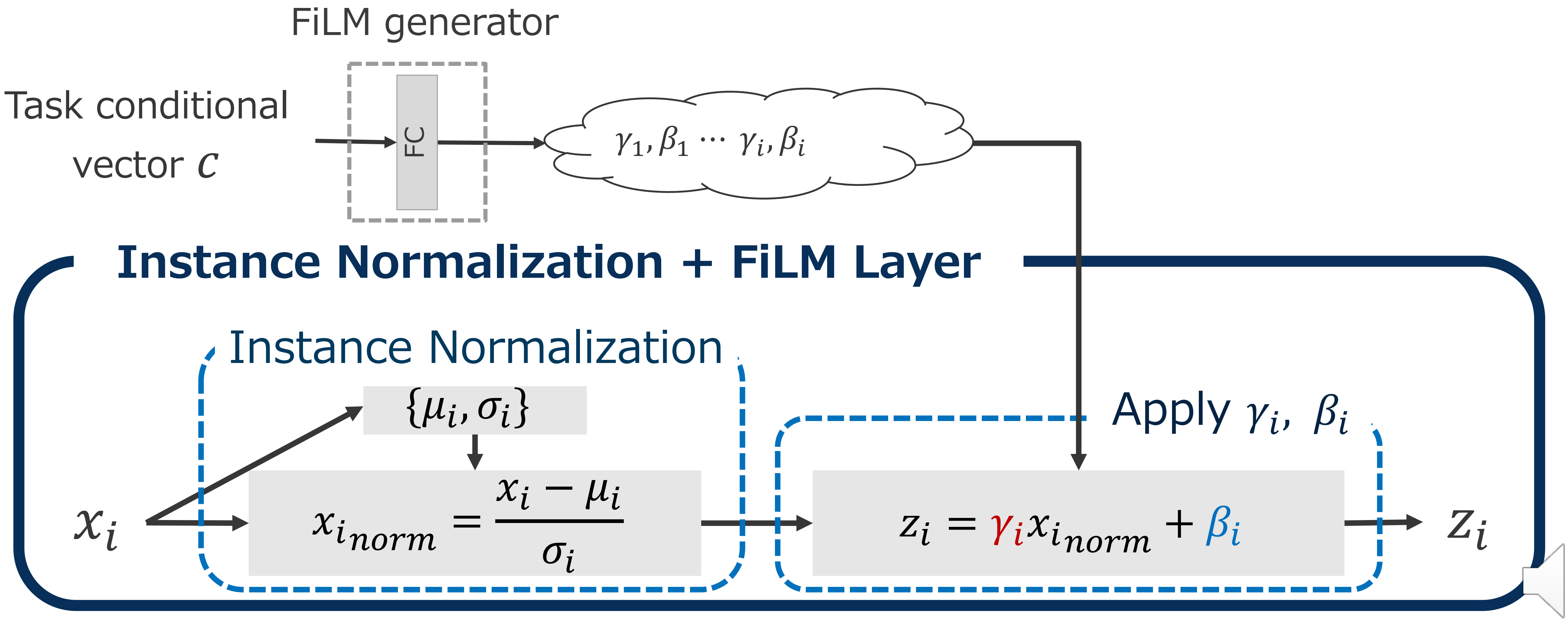
Overview of the method – Task conditional vector

- Task conditional vector $\boldsymbol{c} = [c_1,\ c_2,\ \cdots,\ c_n]$
  - c : Task Strength $(0 \leq c \leq 1)$
  - n : Number of tasks

Ex.） If the number of tasks n is 3, $\mathbf{c} = [c_1,\ c_2,\ c_3]$

- Learn only Task 1 : $[1,\ 0,\ 0]$
- Mixed learning of Task 1 and Task 3 : $[1,\ 0,\ 1]$

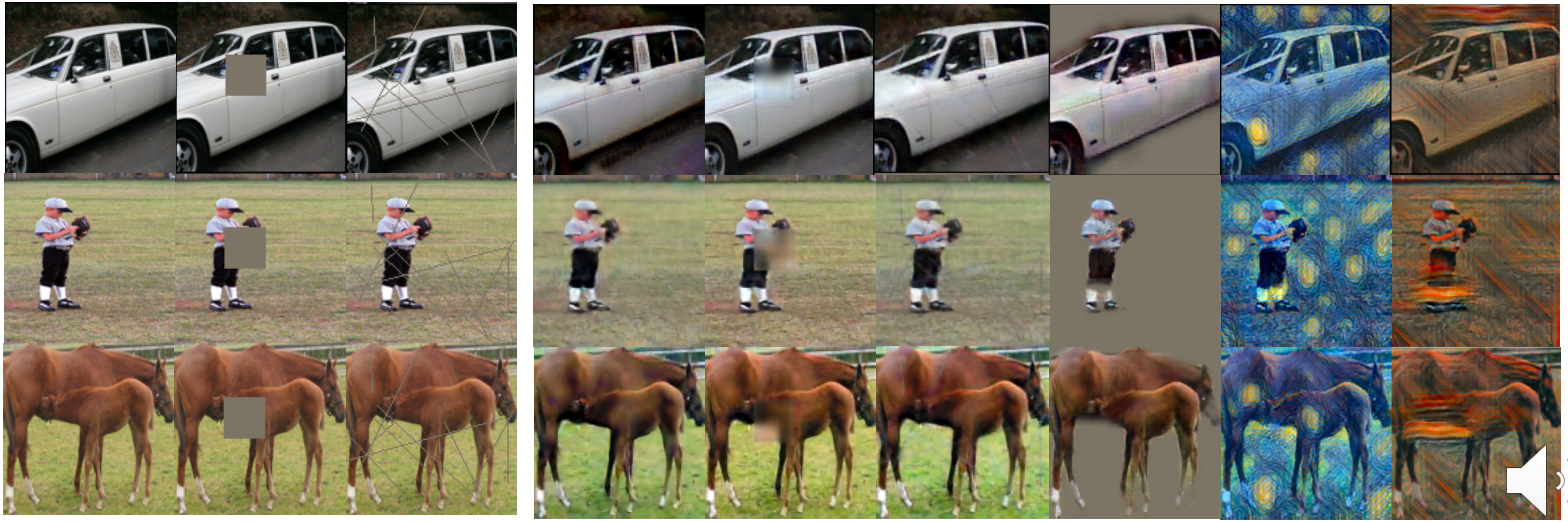
- After normalizing the input features by Instance Normalization, affine transformation with FiLM parameters is applied.



Experiment 1: Learning of Multiple Different Tasks

- Multiple different image translation tasks can be learned with a single FiLM-based model.

Input (Task 0,3,4,5)	Input (Task 1)	Input (Task 2)	Task 0 (reconstruction)	Task 1 (inpainting)	Task 2 (denoising)	Task 3 (Semantic Segmentation)	Task 4 (Style Transfer 1)	Task 5 (Style Transfer 2)

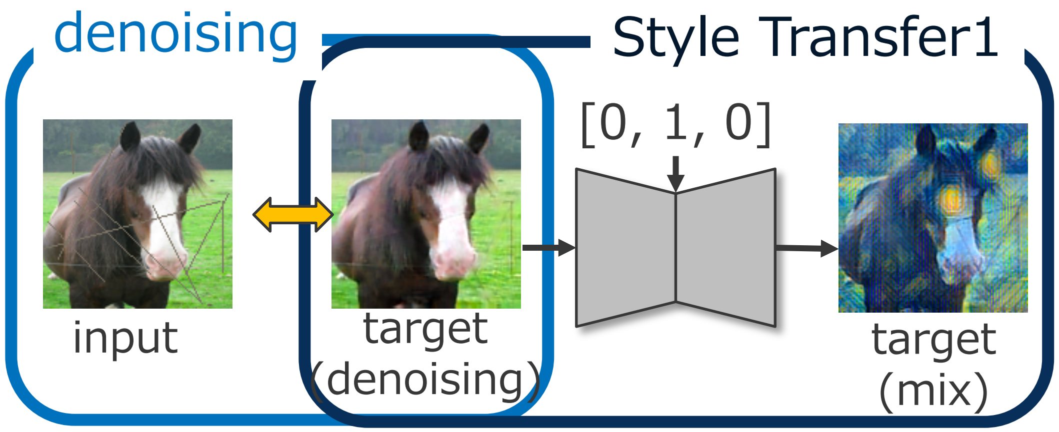


Experiment 2: Learning of Mixed Tasks - Setting up the experiment

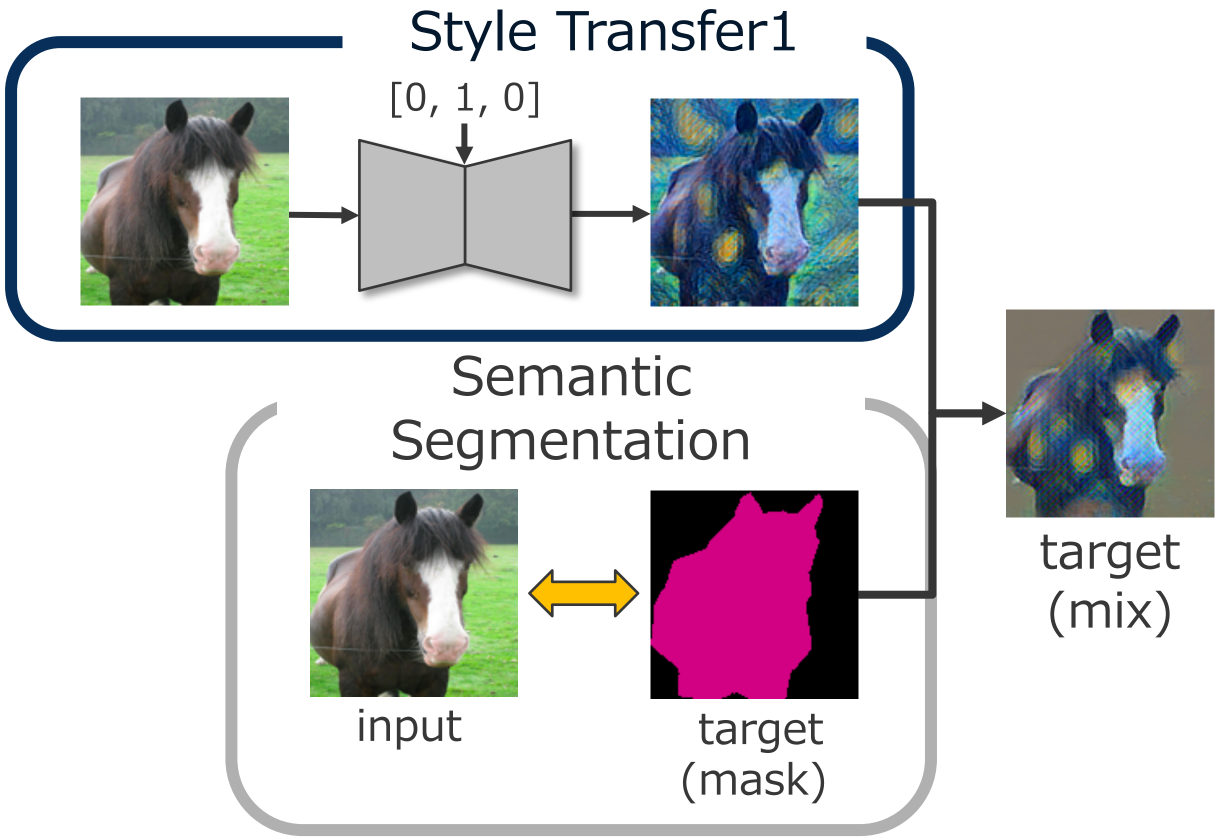
- For verification of mixed-task learning, four tasks were used.

Number	Task name	Conditional vector
Task 0	reconstruction	[0, 0, 0]
Task 1	denoising	[1, 0, 0]
Task 2	semantic segmentation	[0, 1, 0]
Task 3	Style Transfer 1	[0, 0, 1]
Mix 1	denoising + Style Transfer 1	[1, 0, 1]
Mix 2	denoising + semantic segmentation	[1, 1, 0]
Mix 3	semantic segmentation + Style Transfer 1	[0, 1, 1]

Experiment 2: Learning of Mixed Tasks - How to create a target



(a) sequential mixing



(b) mixing by region masking

Experiment 2: Learning of Mixed Tasks - Results

- The results are almost the same as ground truth (GT), which means mixed- task learning succeeded.

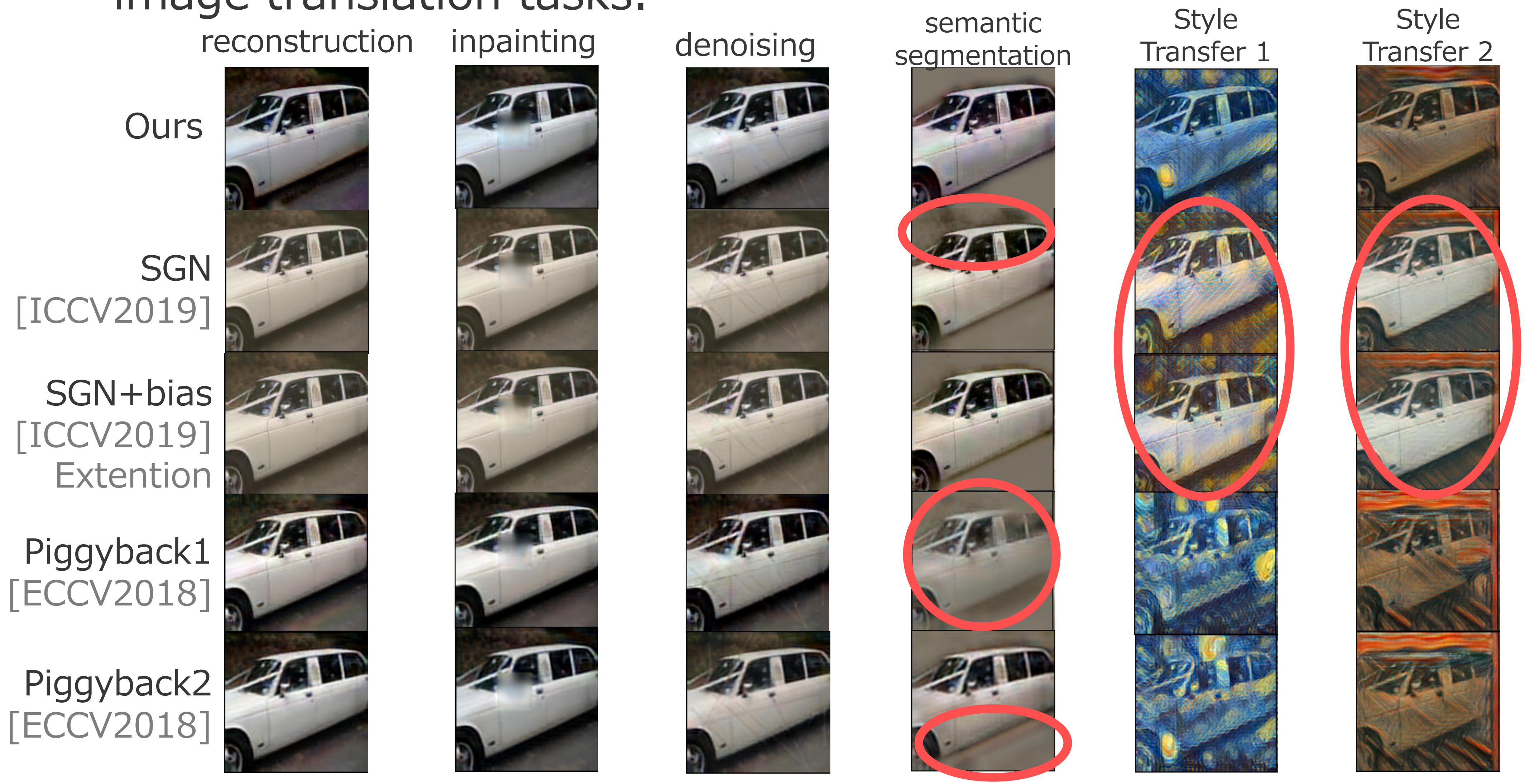
MIX 1
A : denoising
B : Style Transfer 1

MIX 2
A : denoising
B : semantic segmentation

MIX 3
A : semantic segmentation
B : Style Transfer1

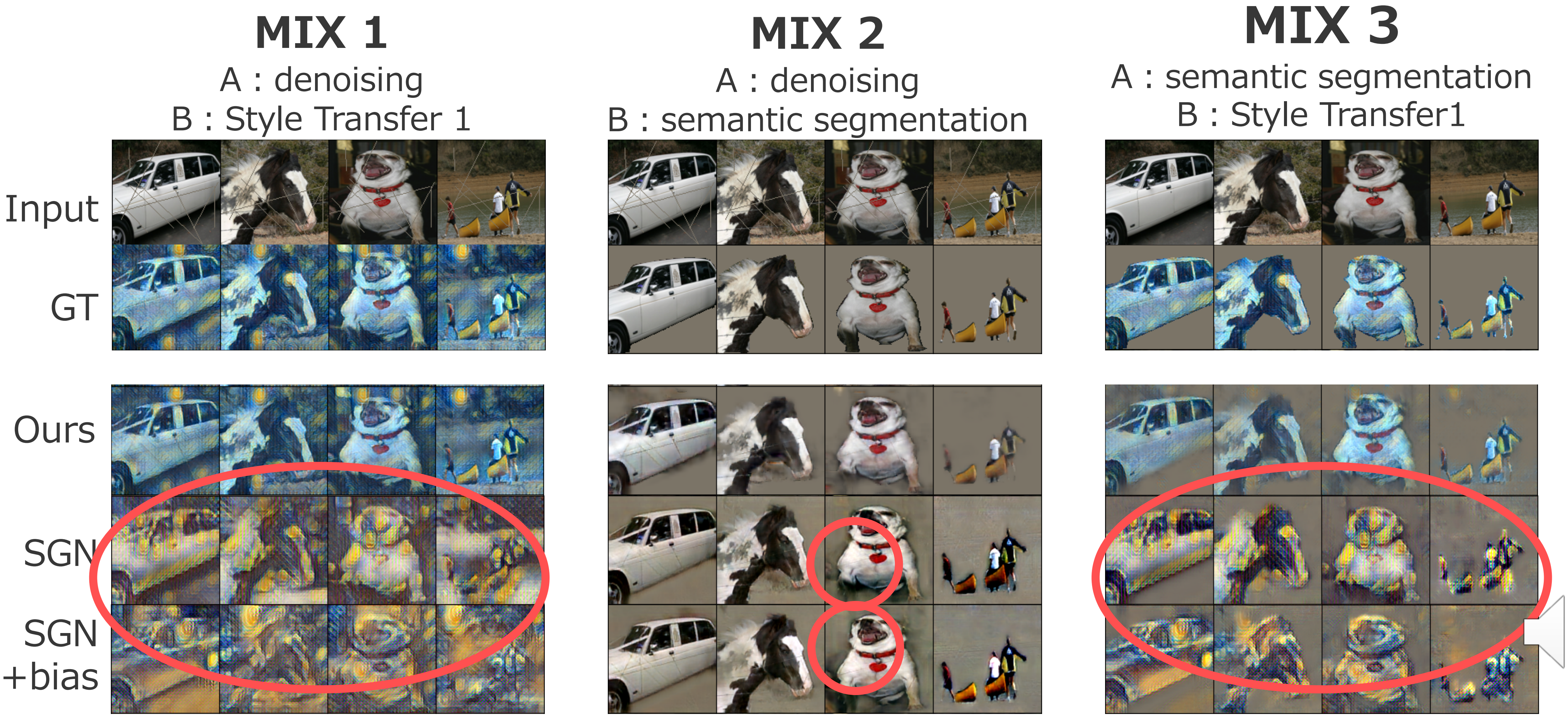
Experiment 3: Comparison to the Baselines

- This shows their outputs when learning of multiple heterogeneous image translation tasks.



Experiment 3: Comparison to the Baselines

- This compares the output of the proposed method and the baseline for mixed-task learning.



Experiment 3: Comparison to the Baselines

	Ours	SGN	SGN+bias	Piggyback1	Piggyback2
reconstruction (PSNR↑)	**8.514**	3.786	3.777	7.964	7.908
inpainting (PSNR↑)	**10.306**	3.445	3.451	9.101	8.139
denoising (PSNR↑)	**10.762**	3.[illegible]	[illegible]	[illegible]	[illegible]2
semantic segmentation (IoU↑)	0.5907	0.[illegible]	[illegible]	[illegible]	[illegible]14
Style Transfer 1(FID↓)	**281.3**	299.0	307.1	333.4	331.4
Style Transfer 2 (FID↓)	**235.3**	263.6	250.1	297.6	323.3
denoising + Style Transfer 1 (FID↓)	**304.3**	349.1	343.4	-	-
denoising+ semantic segmentation (PSNR↑)	**6.716**	2.442	2.470	-	-
semantic segmentation + Style Transfer 1 (FID↓, IoU↑)	**313.0** **0.5457**	3[illegible] 0[illegible]	[illegible]	[illegible]	-
Model size (num. of prams)	1,698,435	1,765,363	1,902,243	**1,679,235**	**1,679,23[illegible]**

Ours achieved the best evaluation scores in almost all the tasks

small model size

Conclusions

- we performed learning of multiple different image translation tasks and their mixed tasks with the single FiLM-based network.
- Our method realized mixed task learning in addition to learning multiple individual tasks.
- In future work, we plan to add more tasks such as various kinds of image do- main translation tasks and mix them.

Motivation to use FiLM

- The effectiveness of the method has been demonstrated in various image transformation tasks.

  → It has high versatility and can be used to learn various image transformation tasks.

- FiLM allows the combination of multiple styles in Style Transfer.

  →Possibility to combine several different image transformation tasks.

style1 style2

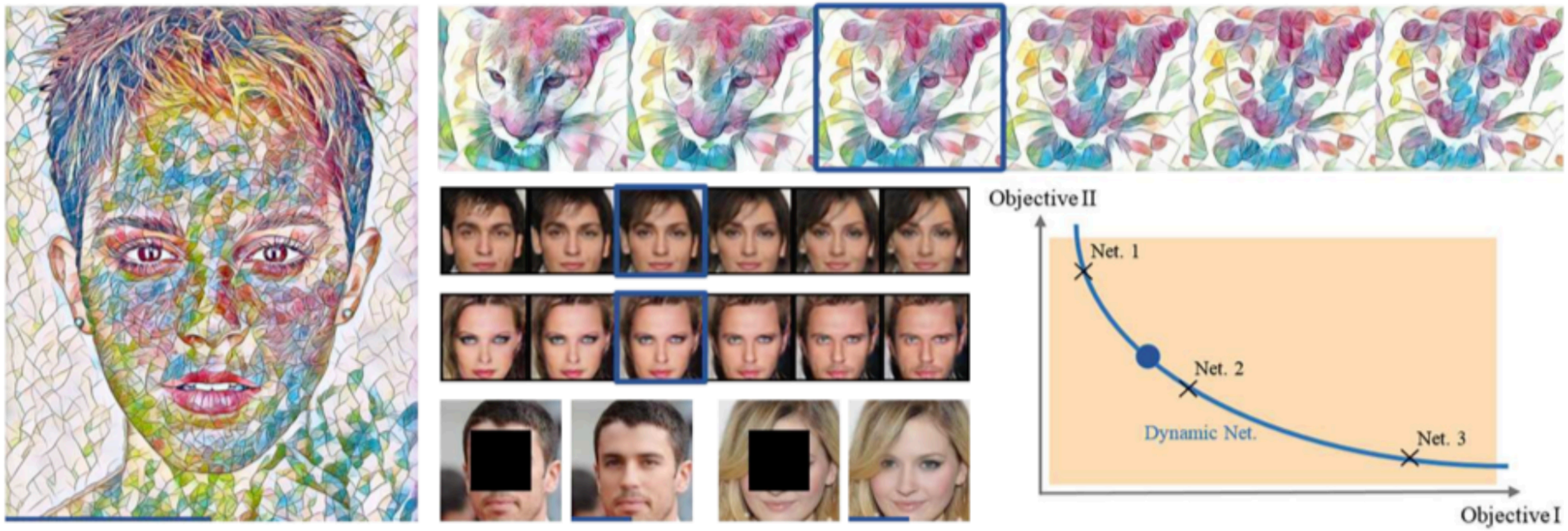
style3 style4

[3] Dumoulin et al.: A learned representation for artistic style, ICLR 2017.

Why the value of a task conditional vector can manipulate the strength of a transformation

- A similar example of manipulating the strength of a transformation with conditional signals is Dynamic-Net.

- We consider the objective space of various image transformation tasks to be a linearly coupled space of pre- and post-task execution.



Alon Shoshan, Roey Mechrez, and Lihi Zelnik-Manor. Dynamic-net: Tuning the objective without re-training. In Proc.of IEEE International Conference on Computer Vision, 2019.

Mixed-task learning in the presence of many tasks

- When there are many tasks, it is hard to learn all combinations of them.

- Therefore, our goal is to be able to generate task combinations during inference by simply learning each task individually.