中山大學
SUN YAT-SEN UNIVERSITY

Continual Object Detection via Prototypical Task Correlation Guided Gating Mechanism

通过原型任务相关性引导的门控机制实现连续目标检测

Continual Object Detection via Prototypical Task Correlation Guided Gating Mechanism

Binbin Yang[1*] Xinchi Deng[1*] Han Shi[2] Changlin Li[3] Gengwei Zhang[1]
Hang Xu[4] Shen Zhao[1] Liang Lin[1†] Xiaodan Liang[1]
[1]Sun Yat-sen University [2]The Hong Kong University of Science and Technology
[3]ReLER, AAII, UTS [4]Huawei Noah's Ark Lab
{yangbb3, dengxch5}@mail2.sysu.edu.cn, hshiac@cse.ust.hk, zhaosh35@mail.sysu.edu.cn,
{changlinli.ai, zgwdavid, chromexbjxh, xdliang328}@gmail.com, linliang@ieee.org

一、原型任务相关性引导的门控机制pRotOtypical taSk corrElaTion guided gaTing mechAnism (ROSETTA)

与以往工作中要为所有任务调参不同，文章提出了一个用于连续目标检测的简单且灵活的框架：原型任务相关性引导的门控机制ROSETTA。

其具体实现为，所有任务共享一个统一的框架。引入任务识别门来自动地为特定任务选择子模型。以这种方式，不同的知识可以通过存储他们相应的子模型的权重来被记忆。

二、原型任务相关性引导的门多样性控制器prototypical task correlation guided Gating Diversity Controller (GDC)

为了让原型任务相关性引导的门控机制ROSETTA自动地决定哪些经历是有用的，一个原型任务相关性引导的门多样性控制器GDC被引入，用来自适应地以类特定原型为基础调整门的多样性。

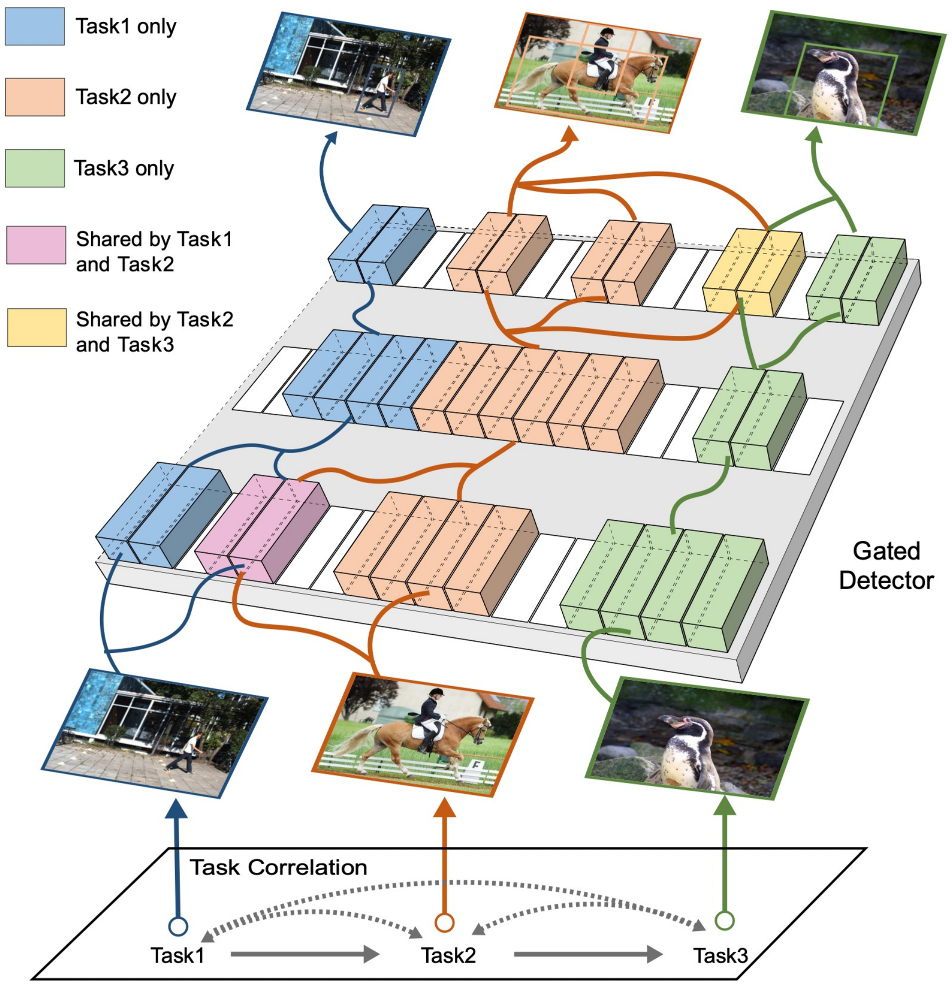
GDC计算类与类之间的相关性矩阵来描绘跨类任务的相关性，从而在出现显著的域差异时为新任务激活专有的门。

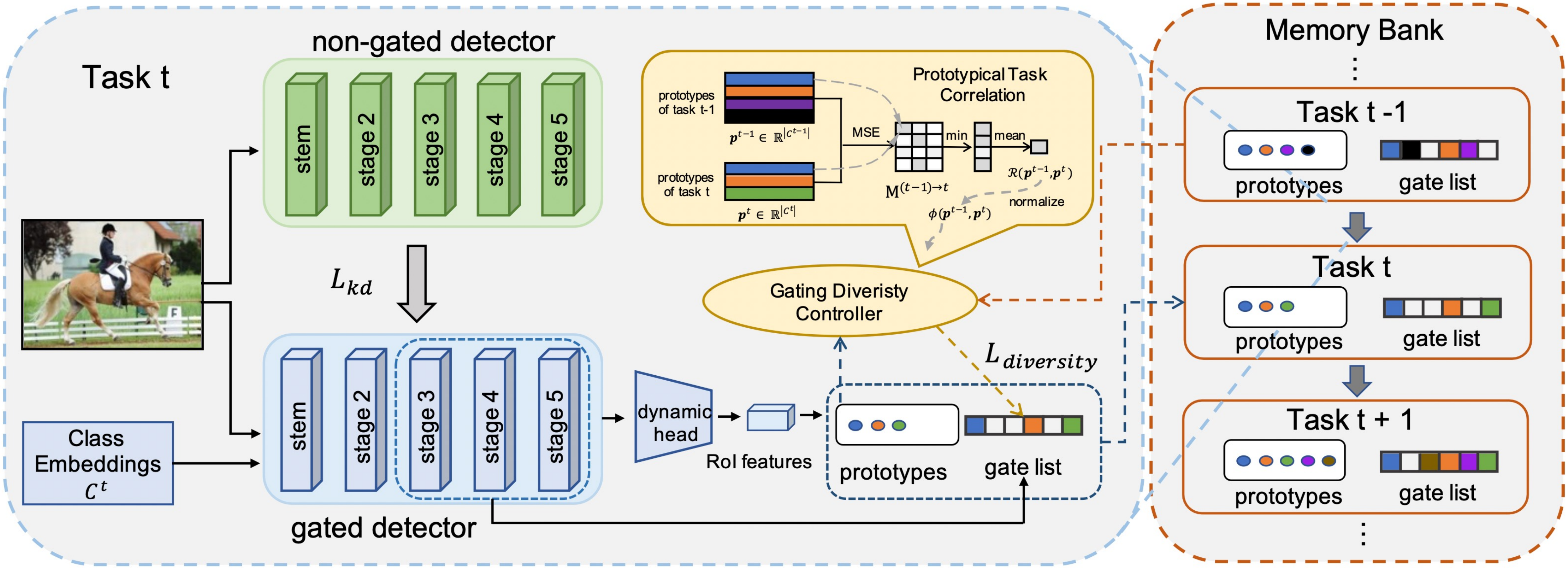
三、效果

在文章的实验条件下，模型在任务增量和类增量两种连续目标检测情况下都达到了SOTA。

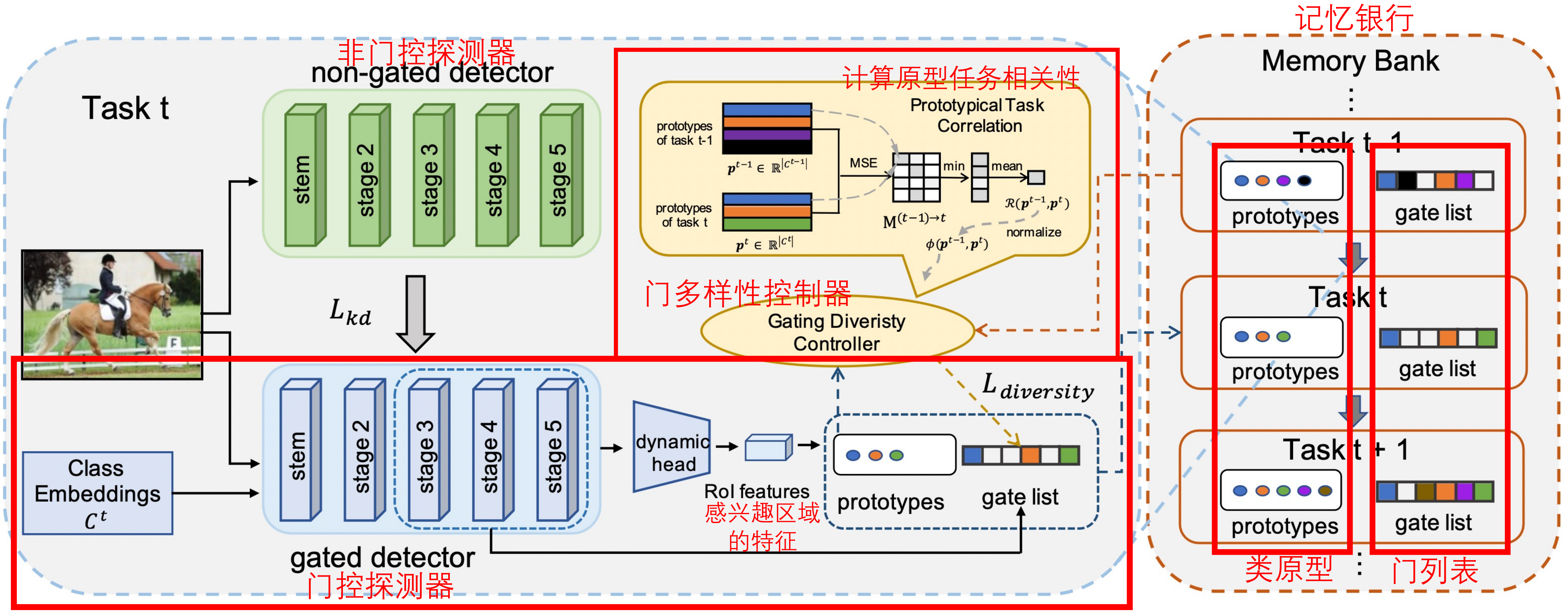
Abstract

Continual learning is a challenging real-world problem for constructing a mature AI system when data are provided in a streaming fashion. Despite recent progress in continual classification, the researches of continual object detection are impeded by the diverse sizes and numbers of objects in each image. Different from previous works that tune

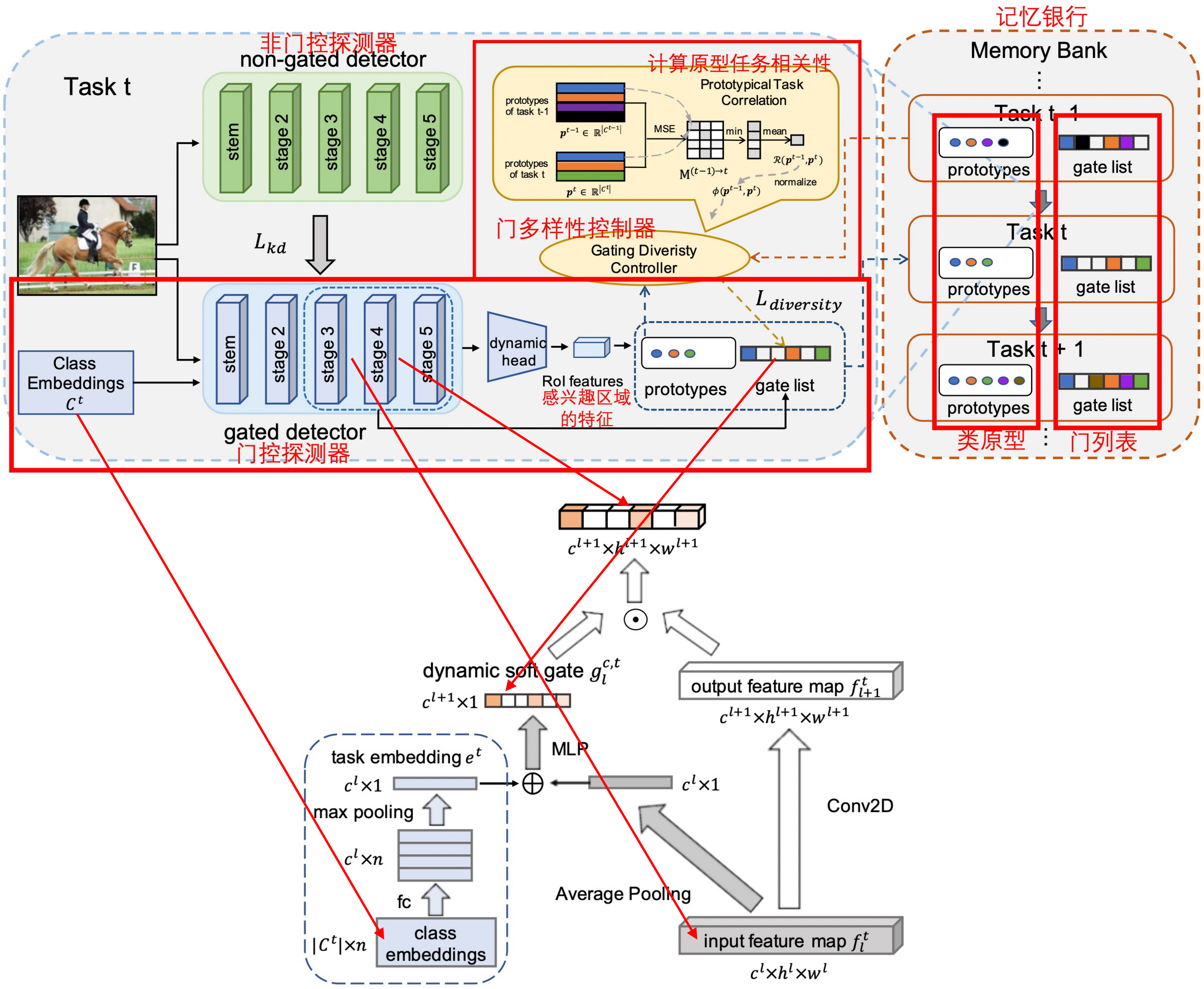




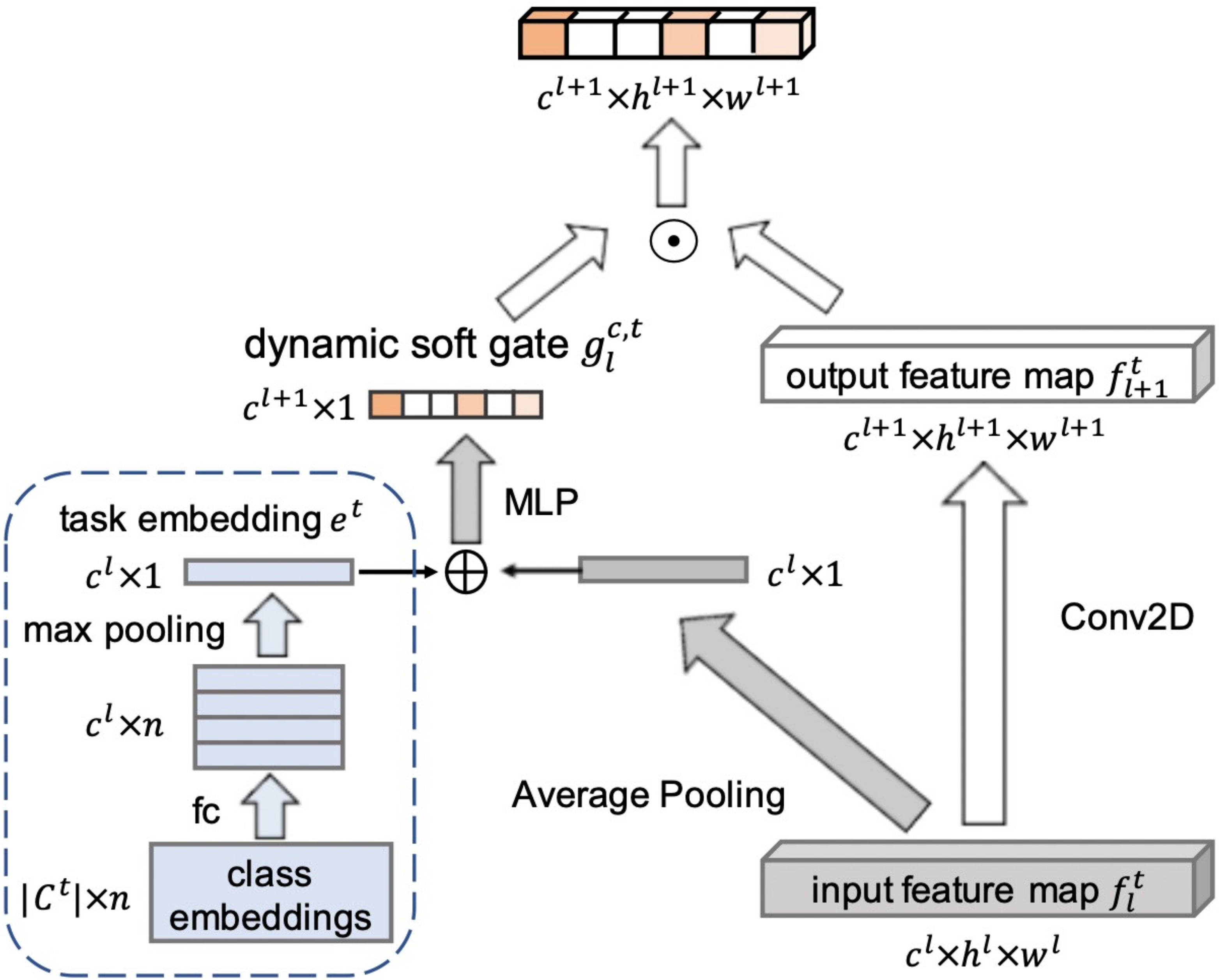
Task t
non-gated detector
stem
stage 2
stage 3
stage 4
stage 5
L_{kd}
Class Embeddings C^t
gated detector
dynamic head
RoI features
prototypes
gate list
Prototypical Task Correlation
prototypes of task t-1
$p^{t-1} \in \mathbb{R}^{|C^{t-1}|}$
prototypes of task t
$p^t \in \mathbb{R}^{|C^t|}$
MSE
min
mean
$M^{(t-1)\to t}$
$\mathcal{R}(p^{t-1}, p^t)$
normalize
$\phi(p^{t-1}, p^t)$
Gating Diveristy Controller
$L_{diversity}$
Memory Bank
Task t -1
Task t
Task t + 1



Task t
非门控探测器
non-gated detector
stem
stage 2
stage 3
stage 4
stage 5
L_{kd}
Class Embeddings C^t
gated detector
门控探测器
dynamic head
RoI features
感兴趣区域的特征
prototypes
gate list
$L_{diversity}$
门多样性控制器
Gating Diveristy Controller
计算原型任务相关性
Prototypical Task Correlation
prototypes of task t-1
$p^{t-1} \in \mathbb{R}^{|C^{t-1}|}$
prototypes of task t
$p^t \in \mathbb{R}^{|C^t|}$
MSE
min
mean
$M^{(t-1)\rightarrow t}$
$\mathcal{R}(p^{t-1}, p^t)$
normalize
$\phi(p^{t-1}, p^t)$
记忆银行
Memory Bank
Task t - 1
Task t
Task t + 1
类原型
门列表



记忆银行
非门控探测器
non-gated detector
Memory Bank
Task t
stem
stage 2
stage 3
stage 4
stage 5
计算原型任务相关性
Prototypical Task Correlation
prototypes of task t-1
prototypes of task t
MSE
min
mean
normalize
L_{kd}
门多样性控制器
Gating Diveristy Controller
Task t - 1
prototypes
gate list
Task t
Task t + 1
$L_{diversity}$
Class Embeddings C^t
dynamic head
RoI features
感兴趣区域的特征
gated detector
门控探测器
类原型
门列表
$c^{l+1}\times h^{l+1}\times w^{l+1}$
dynamic soft gate $g_l^{c,t}$
$c^{l+1}\times 1$
output feature map f_{l+1}^t
MLP
task embedding e^t
$c^l\times 1$
max pooling
$c^l\times n$
fc
$|C^t|\times n$
class embeddings
Conv2D
Average Pooling
input feature map f_l^t
$c^l\times h^l\times w^l$



$c^{l+1} \times h^{l+1} \times w^{l+1}$
dynamic soft gate $g_l^{c,t}$
$c^{l+1} \times 1$
output feature map f_{l+1}^t
$c^{l+1} \times h^{l+1} \times w^{l+1}$
MLP
task embedding e^t
$c^l \times 1$
$c^l \times 1$
max pooling
Conv2D
$c^l \times n$
fc
Average Pooling
$|C^t| \times n$
class embeddings
input feature map f_l^t
$c^l \times h^l \times w^l$

$$\boldsymbol{f}_{l+1}^{t} = G_l^t(\boldsymbol{f}_l^t, C^t) \odot (F_l^t(\boldsymbol{f}_l^t)), \tag{1}$$

f_l_t是特征图，输入到任务t的第l层
F_l_t是卷积操作
G_l是门模块

⊙是通道范围相乘操作
f_l+1_t是门控卷积的输出特征图

F_l_t(f_l_t)是卷积操作的输出
G_l_t(f_l_t,C_t)是第l层的通道门，g_l_1_t则是第一通道，g_l_2_t则是第二通道

e_t是任务嵌入
C_t是类嵌入

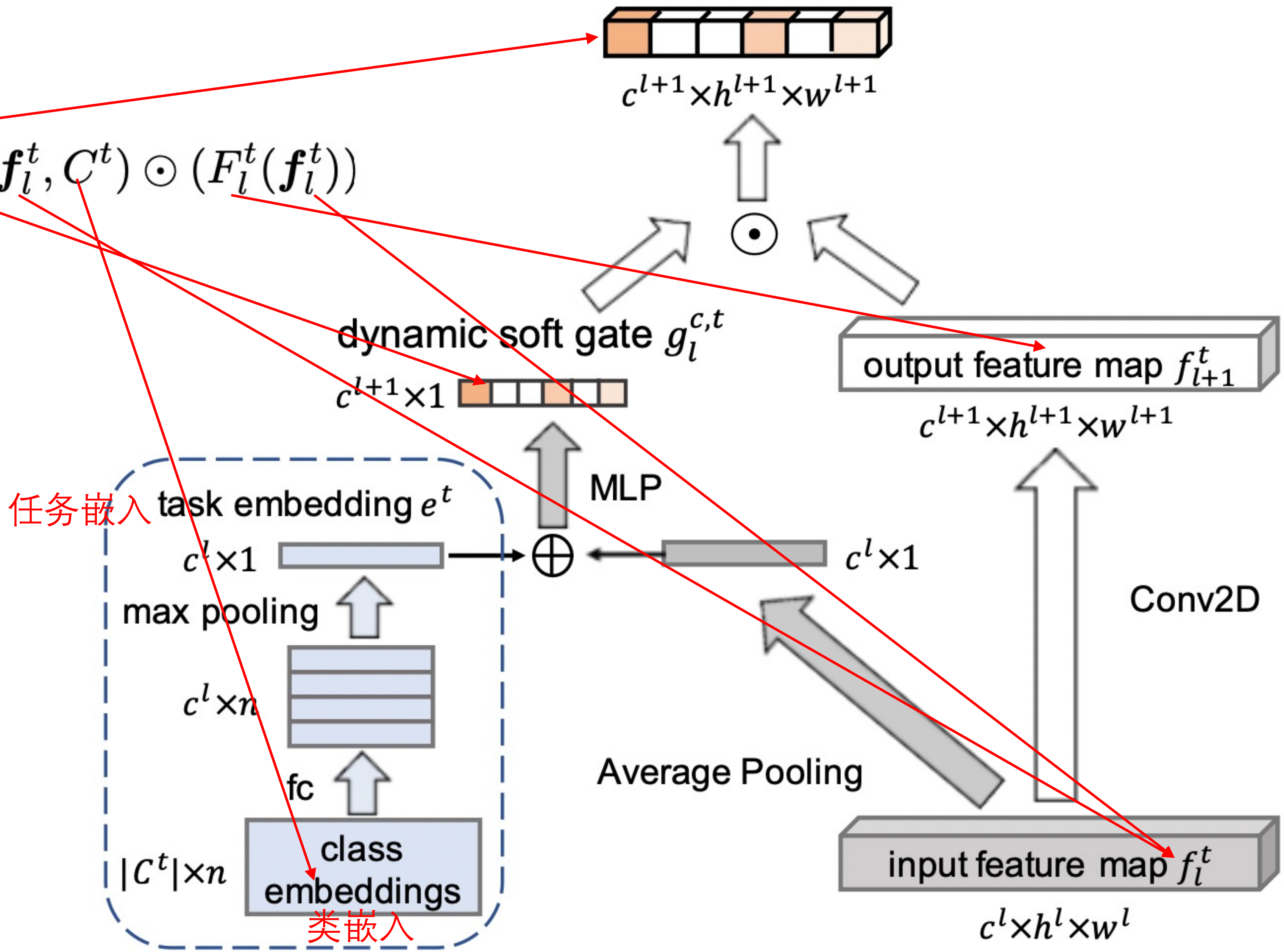
$$\gamma_l^{c,t} = \mathbb{E}_{(x,y)\sim D_{val}^t}[g_l^{c,t}], \tag{2}$$

$$\hat{g}_l^{c,t} = \mathbb{I}[g_l^{c,t} \geq \gamma_l^{c,t}], \tag{3}$$

$$\boldsymbol{f}_{l+1}^{t} = \hat{\boldsymbol{g}}_l^t \odot (F_l^t(\boldsymbol{f}_l^t)), \tag{4}$$

D_val_t是第t个任务的验证集
c是第l层的第c个通道
g_l_c是对应通道的门的值

$$\boldsymbol{f}_{l+1}^{t} = G_l^t(\boldsymbol{f}_l^t, C^t) \odot (F_l^t(\boldsymbol{f}_l^t))$$



$$q_l^t = \frac{\sum_{i=1}^{m_l^t} g_l^{c,t} \mathbb{I}[g_l^{c,t} \geq \eta]}{\sum_{i=1}^{m_l^t} g_l^{c,t}}, \tag{8}$$

$$\mathcal{L}_{diversity}^{l,t} = q_l^t \log q_l^t + (1 - q_l^t) \log(1 - q_l^t), \tag{7}$$

g_l_c_t是任务t时第l层第c通道的门值
q_l_t是任务t时第l层中新激活门占比的估计
m_l_t是为任务t保留的，在l层中先前未激活的通道总数
II是一个指示函数
η是一个阈值超参数
L_l_t是任务t在l层中的门多样性损失

$$\mathbf{M}_{i,j}^{m \to n} = \mathrm{MSE}(\boldsymbol{p}_i^m,\ \boldsymbol{p}_j^n),\ i \in C^m \wedge j \in C^n, \tag{9}$$

$$\mathcal{R}(\boldsymbol{p}_j^n, \boldsymbol{p}^m) = \begin{cases} \min\limits_{i \in C^m} & \mathbf{M}_{i,j}^{m \to n}, \quad m < n \\ \min\limits_{i \in C^m, i \neq j} & \mathbf{M}_{i,j}^{m \to n}, \quad m = n. \end{cases} \tag{10}$$

$$\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m) = \frac{1}{C^n} \sum_{j \in C^n} \mathcal{R}(\boldsymbol{p}_j^n, \boldsymbol{p}^m), m \leq n. \tag{11}$$

$$\phi(\boldsymbol{p}^n, \boldsymbol{p}^m) = \max\{\frac{\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m) - \mathcal{R}(\boldsymbol{p}^m, \boldsymbol{p}^m)}{\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m)}, 0\}. \tag{12}$$

$$\mathcal{L}_{diversity} = \frac{1}{L} \frac{1}{(t-1)} \sum_{l=1}^{L} \sum_{i=1}^{t-1} \phi(\boldsymbol{p}^t, \boldsymbol{p}^i) \mathcal{L}_{diversity}^{l,t}. \tag{13}$$

p_i_t是任务t的类i的原型
C_t是任务t的各类别的集合
式(9)计算任务m的类i与任务n的类j之间的相关性

p_m是任务m各类原型的集合
式(10)计算任务n的类j与任务p的相关性

式(11)计算任务n与任务m的相关性

式(12)计算0~1范围内的权重

式13计算任务t的跨任务相关性指引的门多样性损失

$$q_l^t = \frac{\sum_{i=1}^{m_l^t} g_l^{c,t}\mathbb{I}[g_l^{c,t} \geq \eta]}{\sum_{i=1}^{m_l^t} g_l^{c,t}}, \tag{8}$$

$$\mathcal{L}_{diversity}^{l,t} = q_l^t \log q_l^t + (1 - q_l^t)\log(1 - q_l^t), \tag{7}$$

g_l_c_t是任务t时第l层第c通道的门值
q_l_t是任务t时第l层中新激活门占比的估计
m_l_t是为任务t保留的，在l层中先前未激活的通道总数
II是一个指示函数
η是一个阈值超参数
L_l_t是任务t在l层中的门多样性损失

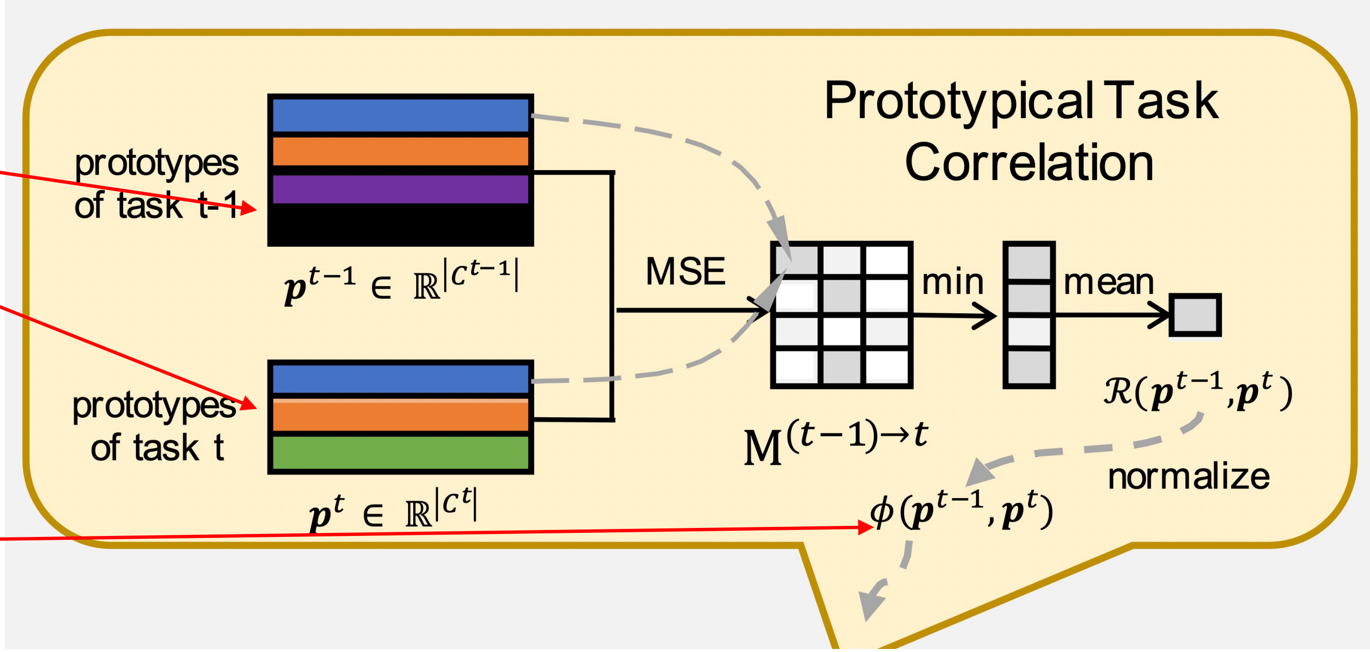
p_i_t是任务t的类i的原型

$$\mathbf{M}_{i,j}^{m\to n} = \mathrm{MSE}(\boldsymbol{p}_i^m,\ \boldsymbol{p}_j^n),\ i \in C^m \wedge j \in C^n, \tag{9}$$

$$\mathcal{R}(\boldsymbol{p}_j^n, \boldsymbol{p}^m) = \begin{cases} \min\limits_{i\in C^m} \quad \mathbf{M}_{i,j}^{m\to n}, & m < n \\ \min\limits_{i\in C^m, i\neq j} \mathbf{M}_{i,j}^{m\to n}, & m = n. \end{cases} \tag{10}$$

$$\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m) = \frac{1}{C^n}\sum_{j\in C^n} \mathcal{R}(\boldsymbol{p}_j^n, \boldsymbol{p}^m), m \leq n. \tag{11}$$

$$\phi(\boldsymbol{p}^n, \boldsymbol{p}^m) = \max\{\frac{\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m) - \mathcal{R}(\boldsymbol{p}^m, \boldsymbol{p}^m)}{\mathcal{R}(\boldsymbol{p}^n, \boldsymbol{p}^m)}, 0\}. \tag{12}$$



$$\mathcal{L}_{diversity} = \frac{1}{L}\frac{1}{(t-1)}\sum_{l=1}^{L}\sum_{i=1}^{t-1}\phi(\boldsymbol{p}^t, \boldsymbol{p}^i)\mathcal{L}_{diversity}^{l,t}. \tag{13}$$

式13计算任务t的跨任务相关性指引的门多样性损失

COCO-VOC				
Methods	COCO → VOC		VOC → COCO	
Joint Training (Faster R-CNN)	48.8	81.6	81.6	48.8
Fine-tuning (Faster R-CNN)	23.2	79.5	74.7	47.1
LwF Detection† [45]	26.6	73.0	-	
Feature Distillation† [43]	26.9	72.4	-	
Attention Distillation† [52]	28.5	73.0	-	
EWC† [24]	32.2	73.4	-	
MAS† [2]	32.7	73.4	-	
AFD† [34]	36.8	75.2	-	
EWC [24]	27.2	75.0	67.0	44.4
MAS [2]	28.1	74.8	69.0	43.9
AFD [34]	27.8	77.1	75.4	45.1
ROSETTA-Faster R-CNN(Ours)	**48.6**	**80.5**	**77.5**	**46.5**
Joint Training (Sparse R-CNN)	52.9	83.2	83.2	52.9
Fine-tuning (Sparse R-CNN)	35.7	81.2	74.9	49.9
ROSETTA-Sparse R-CNN(Ours)	49.5	82.3	79.5	48.3

KITTI-Kitchen				
Methods	KITTI → Kitchen		Kitchen → KITTI	
Joint Training (Faster R-CNN)	55.0	83.7	83.7	55.0
Fine-tuning (Faster R-CNN)	7.7	82.2	13.5	54.2
LwF Detection† [45]	39.4	69.9	59.9	54.7
Feature Distillation† [43]	35.0	69.4	62.7	54.4
Attention Distillation† [52]	39.8	71.0	64.2	52.8
EWC† [24]	48.3	65.5	68.4	52.8
MAS† [2]	42.8	71.7	67.7	**55.6**
AFD† [34]	48.1	72.4	68.6	53.4
EWC [24]	15.8	65.8	10.7	53.4
MAS [2]	8.9	70.3	11.5	54.0
AFD [34]	36.6	72.0	20.5	50.5
ROSETTA-Faster R-CNN(Ours)	**53.9**	**78.1**	**78.4**	54.7
Joint Training (Sparse R-CNN)	55.5	81.8	81.8	55.5
Fine-tuning (Sparse R-CNN)	20.2	79.8	18.6	53.6
ROSETTA-Sparse R-CNN(Ours)	53.3	78.3	78.2	54.5

10 + 10 setting	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
All 20	68.5	77.2	74.2	55.6	59.7	76.5	83.1	81.5	52.1	79.8	55.1	80.9	80.1	76.8	80.5	47.1	73.1	61.2	76.9	70.3	70.51
First 10	79.3	79.7	70.2	56.4	62.4	79.6	88.6	76.6	50.1	68.9	0	0	0	0	0	0	0	0	0	0	35.59
New 10	7.9	0.3	5.1	3.4	0	0	0.2	2.3	0.1	3.3	65	69.3	81.3	76.4	83.1	47.2	67.1	68.4	76.5	69.2	36.31
ILOD [45]	69.9	70.4	69.4	54.3	48	68.7	78.9	68.4	45.5	58.1	59.7	72.7	73.5	73.2	66.3	29.5	63.4	61.6	69.3	62.2	63.15
ILOD + Faster R-CNN	70.5	75.6	68.9	59.1	56.6	67.6	78.6	75.4	50.3	70.8	43.2	68.1	66.2	65.1	66.5	24.3	61.3	46.6	58.1	49.9	61.14
Faster ILOD [39]	72.8	75.7	71.2	60.5	61.7	70.4	83.3	76.6	53.1	72.3	36.7	70.9	66.8	67.6	66.1	24.7	63.1	48.1	57.1	43.6	62.16
ORE [23]	63.5	70.9	58.9	42.9	34.1	76.2	80.7	76.3	34.1	66.1	56.1	70.4	80.2	72.3	81.8	42.7	71.6	68.1	77	67.7	64.58
ROSETTA-Faster R-CNN	74.2	76.2	64.9	54.4	57.4	76.1	84.4	68.8	52.4	67.0	62.9	63.3	79.8	72.8	78.1	40.1	62.3	61.2	72.4	66.8	**66.80**

15 + 5 setting	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
First 15	74.2	79.1	71.3	60.3	60	80.2	88.1	80.2	48.8	74.6	61	76	85.3	78.2	83.4	0	0	0	0	0	55.03
New 5	3.7	0.5	6.3	4.6	0.9	0	8.8	3.9	0	0.4	0	0	16.4	0.7	0	41	55.7	49.2	59.1	67.8	15.95
ILOD [45]	70.5	79.2	68.8	59.1	53.2	75.4	79.4	78.8	46.6	59.4	59	75.8	71.8	78.6	69.6	33.7	61.5	63.1	71.7	62.2	65.87
ILOD + Faster R-CNN	63.5	76.3	70.7	53.1	55.8	67.1	81.5	80.3	49.6	73.8	62.1	77.1	79.7	74.2	73.9	37.1	59.1	61.7	68.6	61.3	66.35
Faster ILOD [39]	66.5	78.1	71.8	54.6	61.4	68.4	82.6	82.7	52.1	74.3	63.1	78.6	80.5	78.4	80.4	36.7	61.7	59.3	67.9	59.1	67.94
ORE [23]	75.4	81	67.1	51.9	55.7	77.2	85.6	81.7	46.1	76.2	55.4	76.7	86.2	78.5	82.1	32.8	63.6	54.7	77.7	64.6	68.51
ROSETTA-Faster R-CNN	76.5	77.5	65.1	56.0	60.0	78.3	85.5	78.7	49.5	68.2	67.4	71.2	83.9	75.7	82.0	43.0	60.6	64.1	72.8	67.4	**69.17**

19 + 1 setting	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	train	tv	mAP
First 19	77.8	81.7	69.3	51.6	55.3	74.5	86.3	80.2	49.3	82	63.6	76.8	80.9	77.5	82.4	42.9	73.9	70.4	70.4	0	67.34
Last 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	64	3.2
ILOD [45]	69.4	79.3	69.5	57.4	45.4	78.4	79.1	80.5	45.7	76.3	64.8	77.2	80.8	77.5	70.1	42.3	67.5	64.4	76.7	62.7	68.25
ILOD + Faster R-CNN	60.9	74.6	70.8	56	51.3	70.7	81.7	81.5	49.45	78.3	58.3	79.5	79.1	74.8	75.7	42.8	74.7	61.2	67.2	65.1	67.72
Faster ILOD [39]	64.2	74.7	73.2	55.5	53.7	70.8	82.9	82.6	51.6	79.7	58.7	78.8	81.8	75.3	77.4	43.1	73.8	61.7	69.8	61.1	68.56
ORE [23]	67.3	76.8	60	48.4	58.8	81.1	86.5	75.8	41.5	79.6	54.6	72.8	85.9	81.7	82.4	44.8	75.8	68.2	75.7	60.1	68.89
ROSETTA-Faster R-CNN	75.3	77.9	65.3	56.2	55.3	79.6	84.6	72.9	49.2	73.7	68.3	71.0	78.9	77.7	80.7	44.0	69.6	68.5	76.1	68.3	**69.64**

Table 2. Comparisons with existing class-incremental object detectors with Faster R-CNN backbone on three different settings: "10 + 10", "15 + 5", "19 + 1". For example, "First 15" means training on the first 15 classes of Pascal VOC 2007 with Faster R-CNN backbone and "New 5" refers to fine-tuning on the new 5 classes. The best results are denoted in boldface.

	COCO	→ VOC	→ KITTI	→ Kitchen	Average
Fine-tuning	4.3	16.3	45.9	87.0	38.4
ROSETTA(Ours)	49.5	82.3	62.5	87.3	**70.4**
	KITTI	→ COCO	→ VOC	→ Kitchen	Average
Fine-tuning	24.9	5.9	24.6	85.2	35.2
ROSETTA(Ours)	53.3	48.9	80.3	84.5	**66.8**

Table 3. Results of sequential training on 4 tasks.

$\mathcal{L}_{sparsity}$	$\mathcal{L}_{kd}$	$\mathcal{L}_{diversity}$	KITTI	→ Kitchen	Gates			
					only `task1`	overlap	only `task2`	not used
✗	✗	✗	51.7	70.2	6.8%	48.3%	9.4%	35.5%
✓	✗	✗	51.0	70.5	6.1%	26.4%	1.2%	66.3%
✓	✓	✗	53.3	73.3	9.6%	32.9%	8.1%	49.4%
✓	✓	✓	53.3	78.3	14.3%	28.2%	20.3%	37.2%

Table 4. Ablation studies on KITTI→Kitchen.

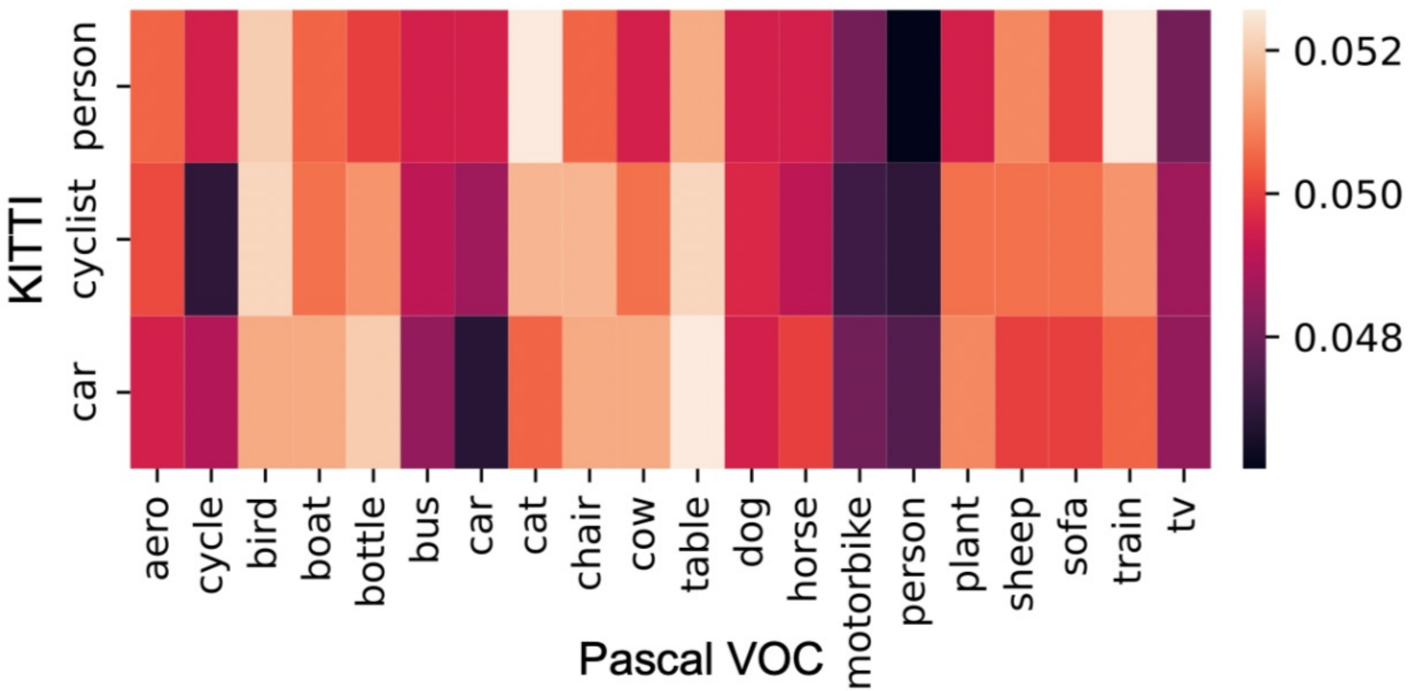


Figure 4. Visualization of the normalized prototypical correlation matrix of VOC→KITTI, *i.e.* $\mathbf{M}^{\text{VOC}\rightarrow\text{KITTI}}$. Darker color indicates higher similarity between the classes of these two tasks.

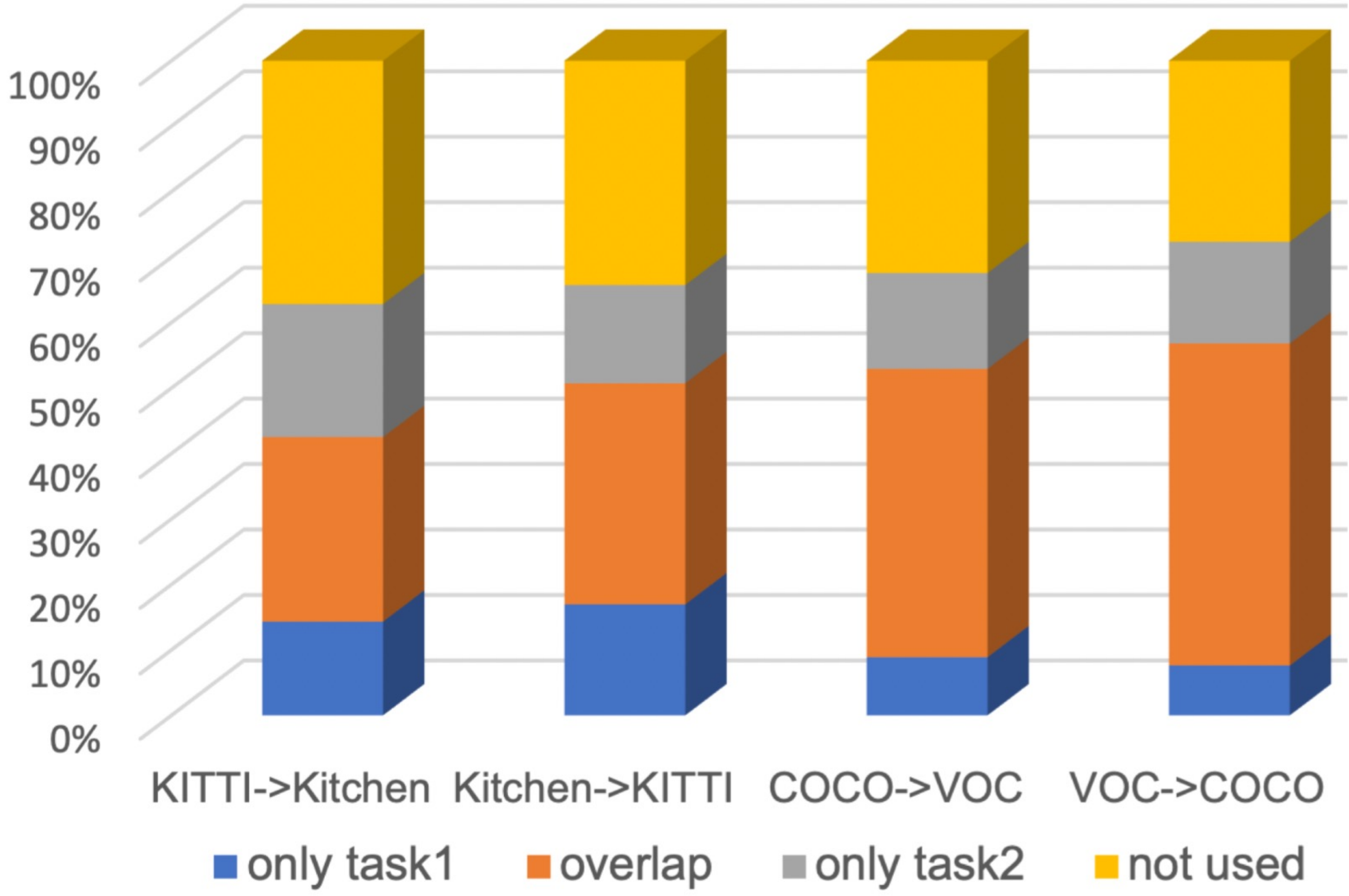


Figure 5. Analysis of gates for task-incremental detection.