Weakly-Supervised Semantic Segmentation

김 현 성

Vision & Display Systems Lab.

Dept. of Electronic Engineering, Sogang University

Outline

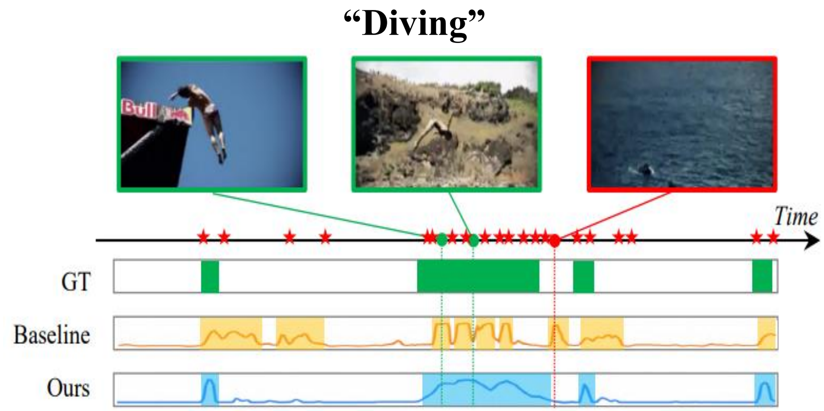
- Motivation
  - WS-SS & WS-TAL
- Background
  - How to do WS-SS?
  - CAM[3]
- Methods
  - CS-AE[1] (2021 ICCV)
  - RIB[2] (2021 NIPS)
- Reference

Motivation

- WS-SS(semantic segmentation) & WS-TAL(temporal action localization)

Only ***<u>“Class Label”</u>***

WS-SS



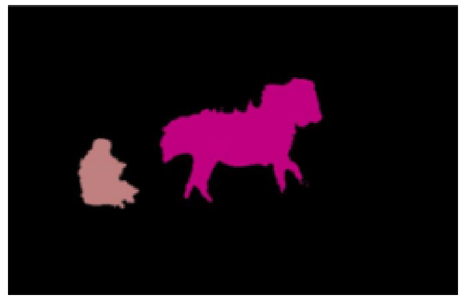
WS-TAL

“Classification” & “Boundary Detection”

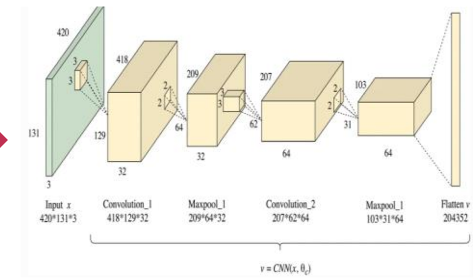
Background

- How to do WS-SS?

Pseudo-labeling



Fully-supervised



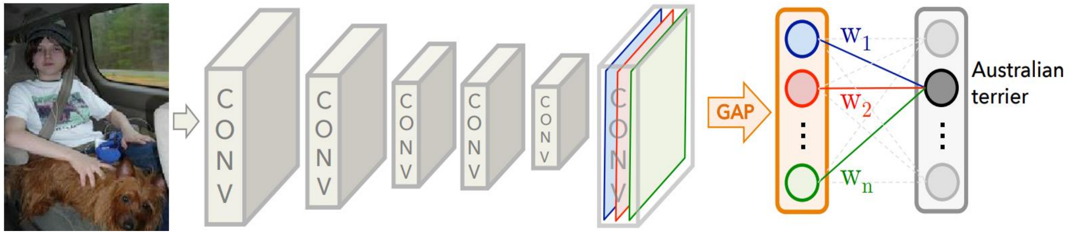
Input image

Pseudo mask

Deep neural network

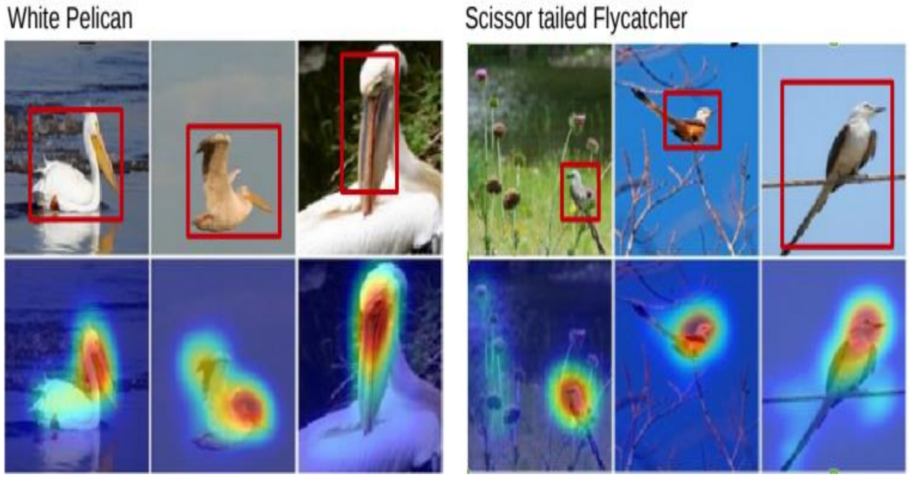
GT

Background

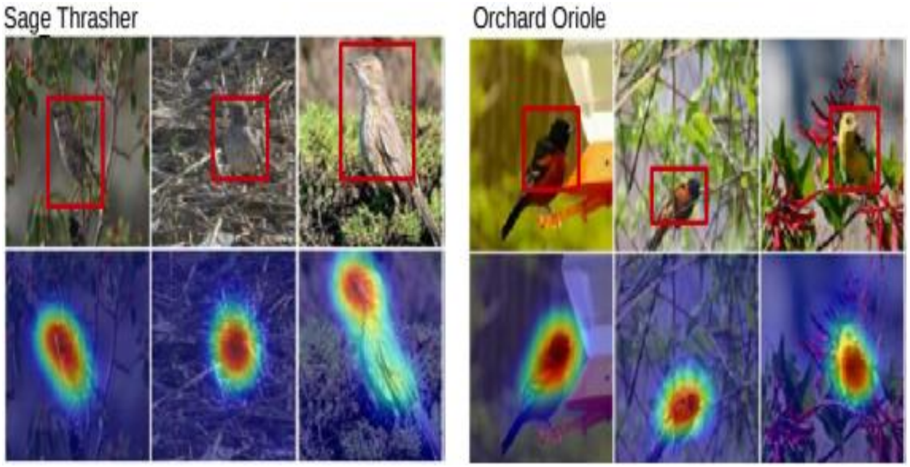


Structure for CAM

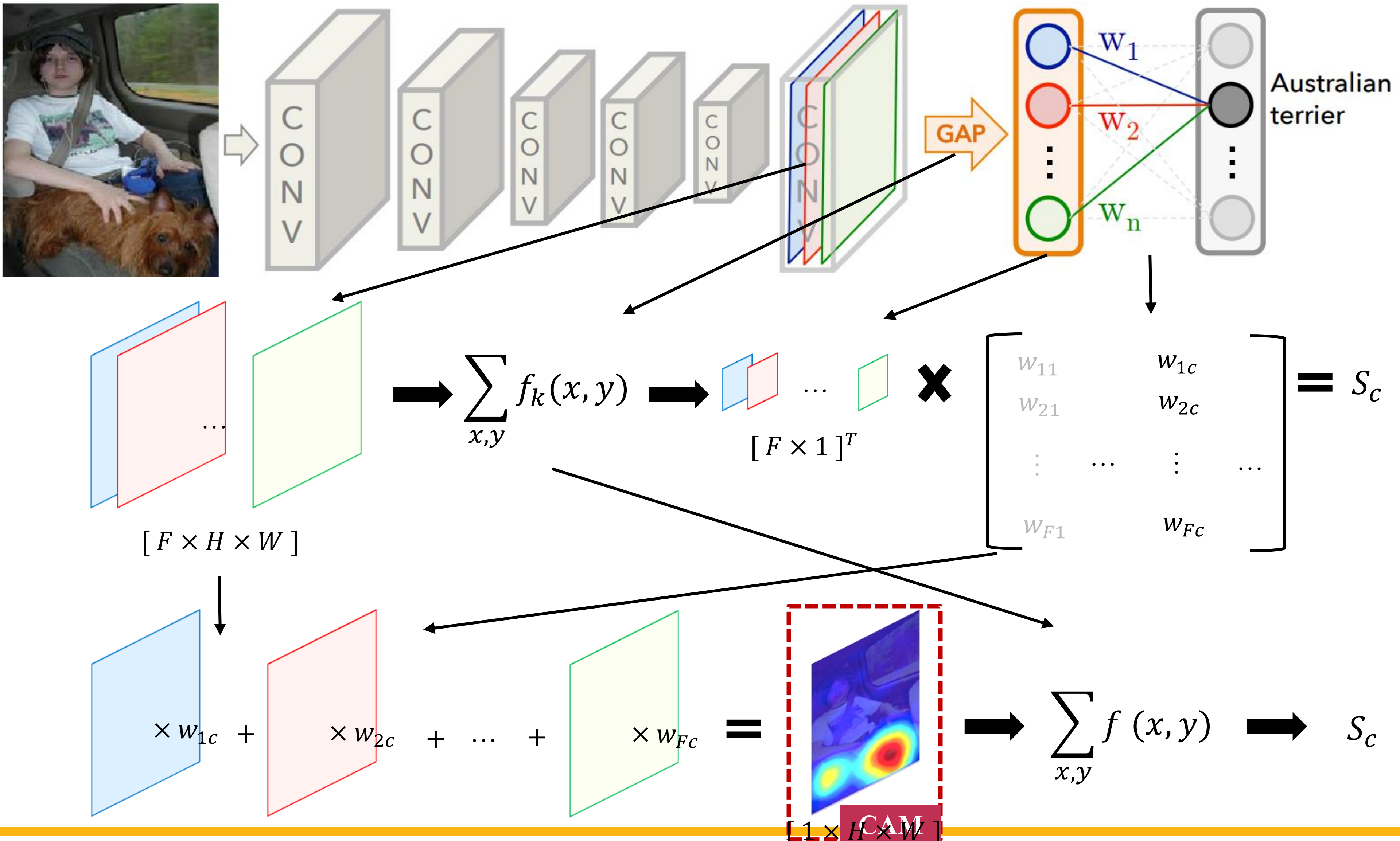
- CAM(class activation maps)
  - Discriminative image region about a specific category



Example



Example



CONV
CONV
CONV
CONV
CONV
CONV
GAP
w_1
w_2
w_n
Australian terrier
$\sum_{x,y} f_k(x,y)$
$[F \times 1]^T$
w_{11} w_{1c}
w_{21} w_{2c}
w_{F1} w_{Fc}
$= S_c$
$[F \times H \times W]$
$\times w_{1c} + \times w_{2c} + \cdots + \times w_{Fc} =$
CAM
$[1 \times H \times W]$
$\sum_{x,y} f(x,y)$
S_c

Background

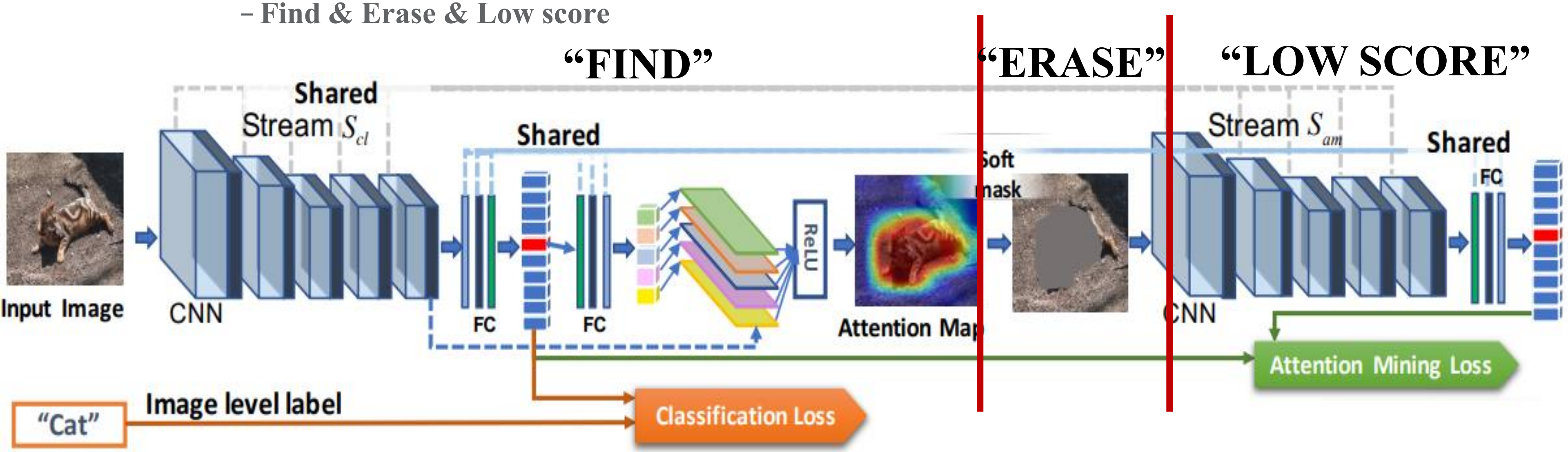
- CAM(class activation maps)

only “**DISCRIMINATIVE**” region…

Input RGB **GT** **CAM**

Methods

- CS-AE(class-specific adversarial erasing) [1]
  - What is AE?
    - Find & Erase & Low score

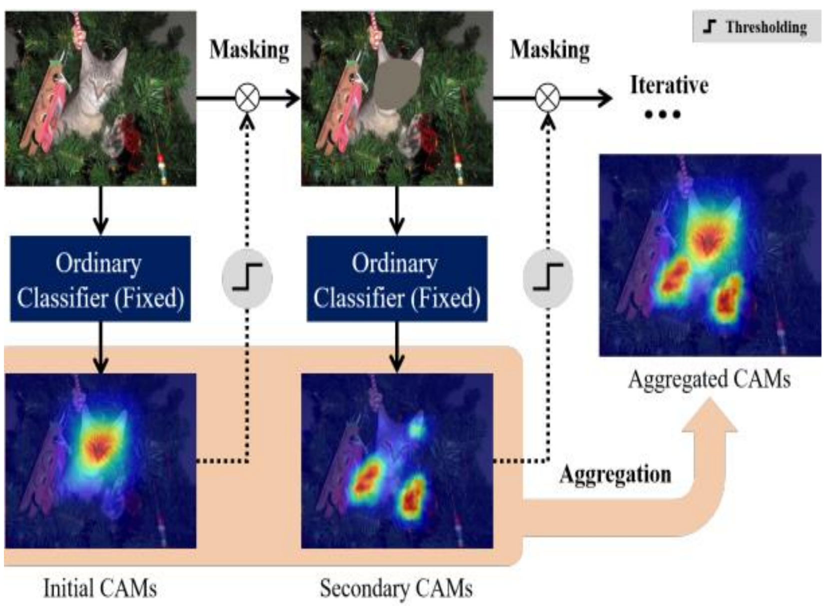


GAIN[4] : Tell me where to look?

  - Problem?
    - Network shared
    - Class-agnostic

Methods

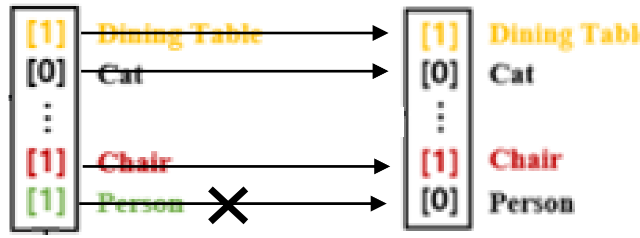
- CS-AE(class-specific adversarial erasing) [1]
  - Motivation
    - Ordinary classifiers can find not only discriminative **but also the non-discriminative region**.
  - Experiment



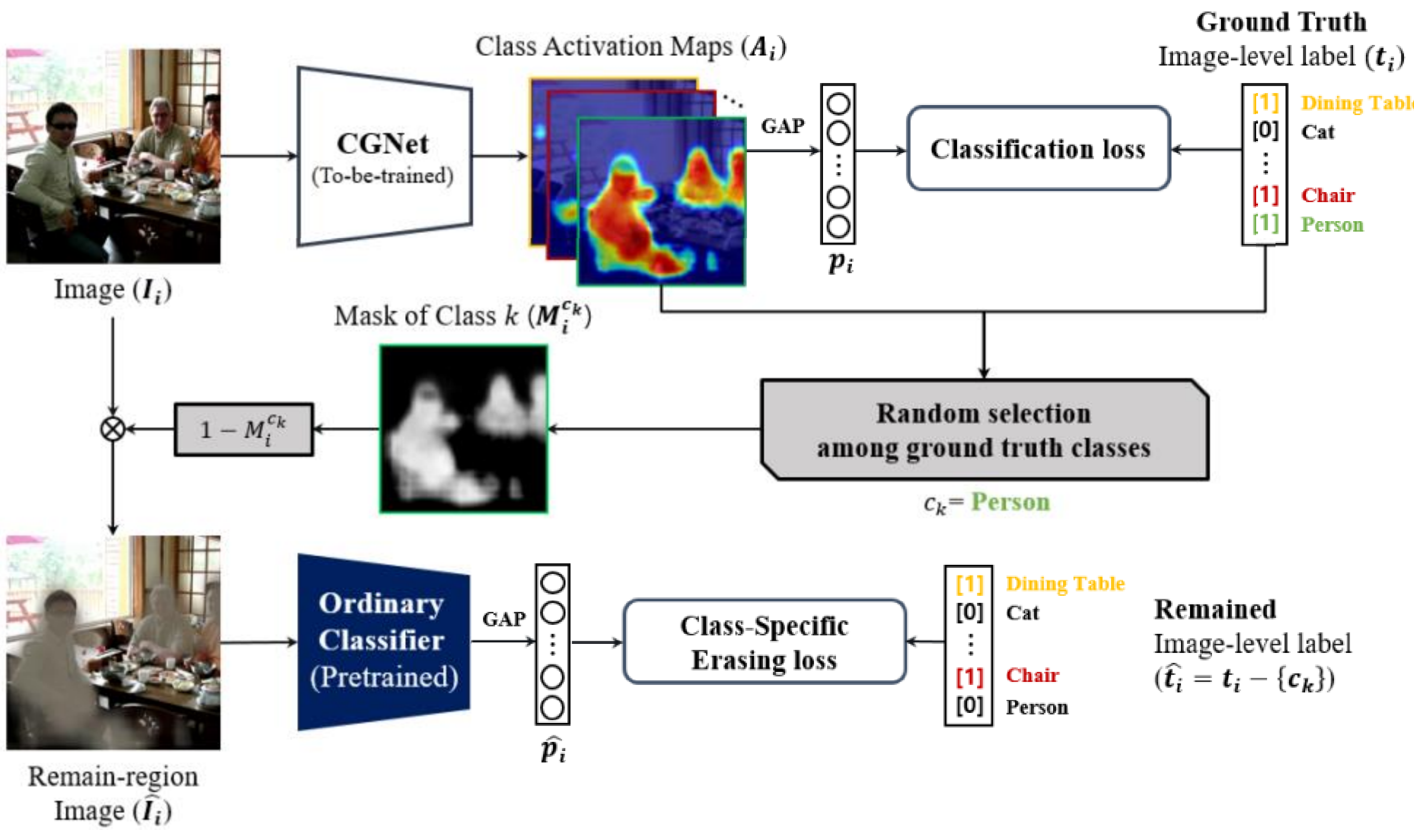
No activation on non-discriminative region

Imbalanced activation

Methods

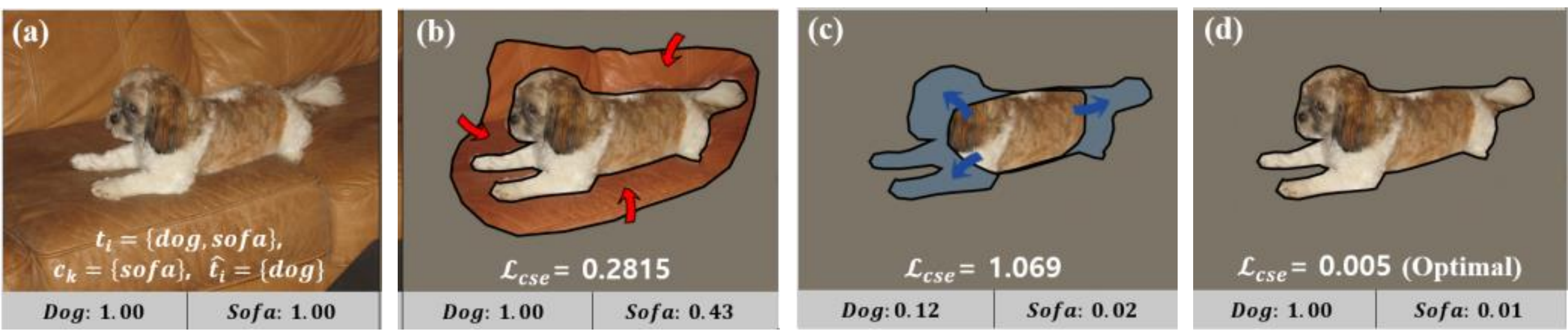


- CS-AE(class-specific adversarial erasing) [1]
  - Proposed method



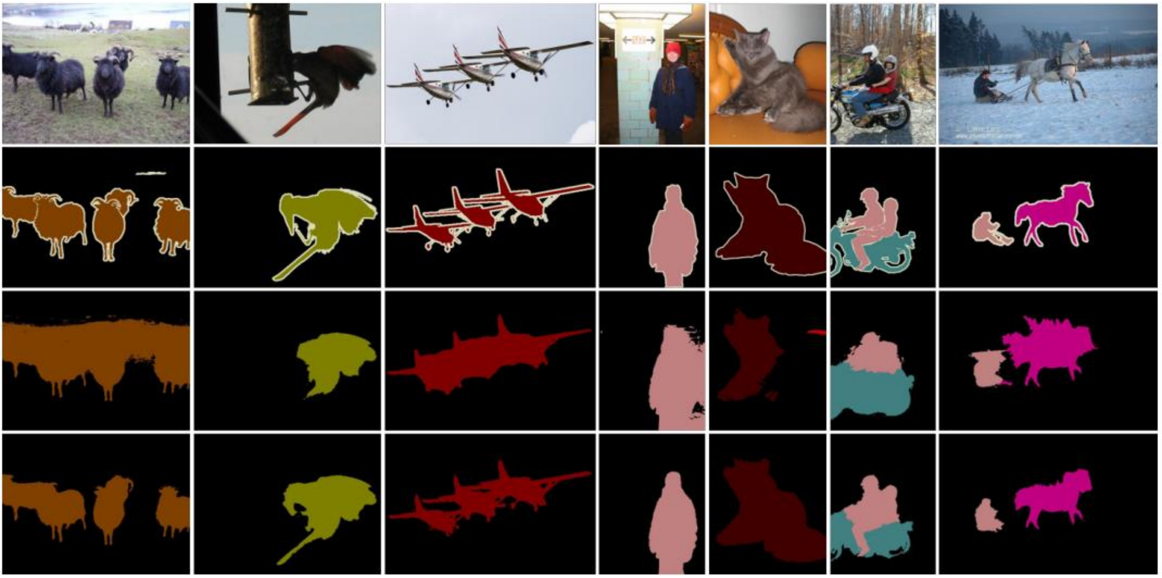
Methods

- CS-AE(class-specific adversarial erasing) [1]
  - Effect
    - Detect class boundaries precisely
    - Experiment
      - Goal : Find the mask of the SOFA!



Methods

- CS-AE(class-specific adversarial erasing) [1]
  - Results



Qualitative results

Methods	*Backbone*	*Sup.*	*Pub.*	*Val*	*Test*
AdvErasing [34]	VGG16	$\mathcal{I}$	CVPR17	55.0	55.7
GAIN [22]	VGG16	$\mathcal{I}$	CVPR18	55.3	56.8
AffinityNet [2]	ResNet38	$\mathcal{I}$	CVPR18	61.7	63.7
ICD [10]	ResNet101	$\mathcal{I}$	CVPR20	64.1	64.3
IRNet [1]	ResNet50	$\mathcal{I}$	CVPR19	63.5	64.8
SSDD [28]	ResNet38	$\mathcal{I}$	ICCV19	64.9	65.5
SEAM [33]	ResNet38	$\mathcal{I}$	CVPR20	64.5	65.7
Sub-category [3]	ResNet101	$\mathcal{I}$	CVPR20	66.1	65.9
RRM [37]	ResNet101	$\mathcal{I}$	AAAI20	66.3	66.5
BES [5]	ResNet101	$\mathcal{I}$	ECCV20	65.7	66.6
Ours	ResNet38	$\mathcal{I}$	-	**68.4**	<u>68.2</u>

Quantitative results

  - Conclusion
    - Capability of the ordinary classifier
    - The effect of class-specific adversarial erasing

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Goal?
    - To find **non-discriminative** region also
  - How?
    - Let's reduce **Information Bottleneck**

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - Mutual information, $I(X;Y)$
      - Indicates how much two variables are dependent

$$\boldsymbol{I(X;Y)} = \underbrace{H(X) + H(Y)}_{\text{regard as constant}} - \boldsymbol{H(X,Y)}$$

X \ Y	0	1
0	1/2	0
1	0	1/2

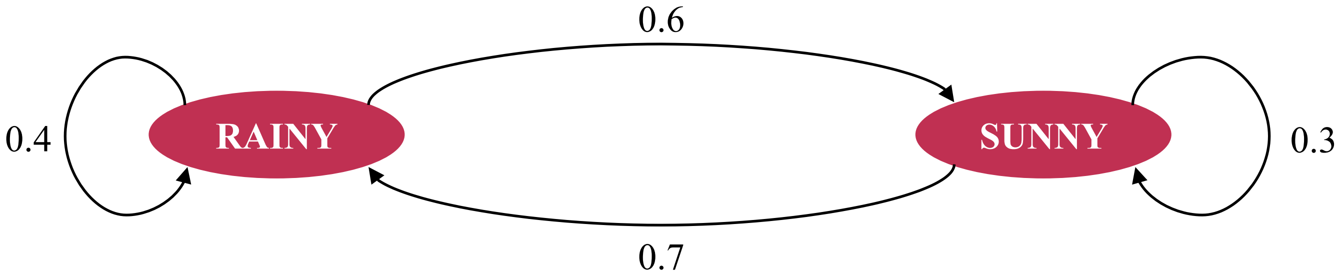
$$\boldsymbol{H(X,Y)} = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$$

X \ Y	0	1
0	1/4	1/4
1	1/4	1/4

$$\boldsymbol{H(X,Y)} = \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 = 2$$

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - Markov chain
      - Discrete-time Stochastic Process which has Markov Chain property

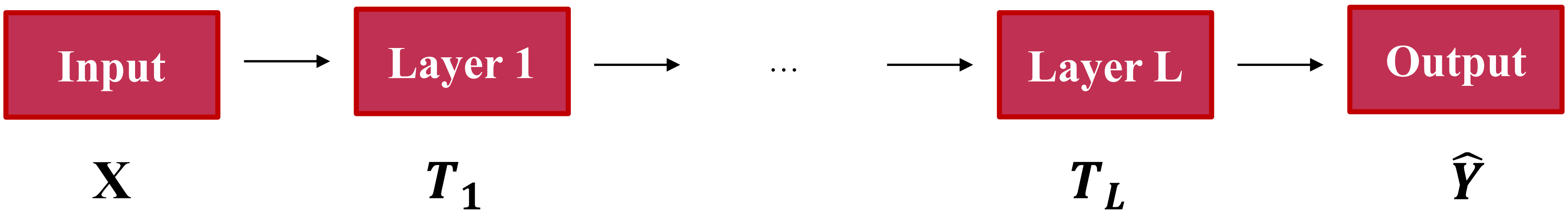


    - Data Processing Inequality
      - for random variables X, Y, Z that form **Markov Chain** $X \rightarrow Y \rightarrow Z$

$$\boldsymbol{I(X;\ Y) \geq I(X;Z)}$$

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - Markov chain in **Deep Neural Network**



"***Markov chain***"

$$I(X;\ T_1) \geq I(X;\ T_2) \geq\ \dots\ \geq I(X; Y)$$

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - *Information Bottleneck* trade-off

$\uparrow: to\ increase\ accuracy$

$$T^* = argmin_T\ I(X;\ T)\ -\ \beta I(T;\ Y)$$

$\downarrow: to\ avoid\ over\ fitting$

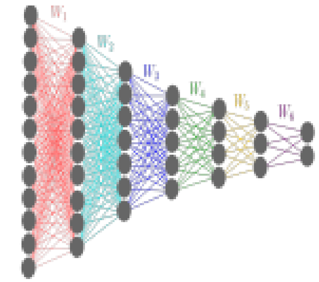
Optimal T /
Minimum sufficient features /
Discriminative features

Information bottleneck trade-off

Obtained by

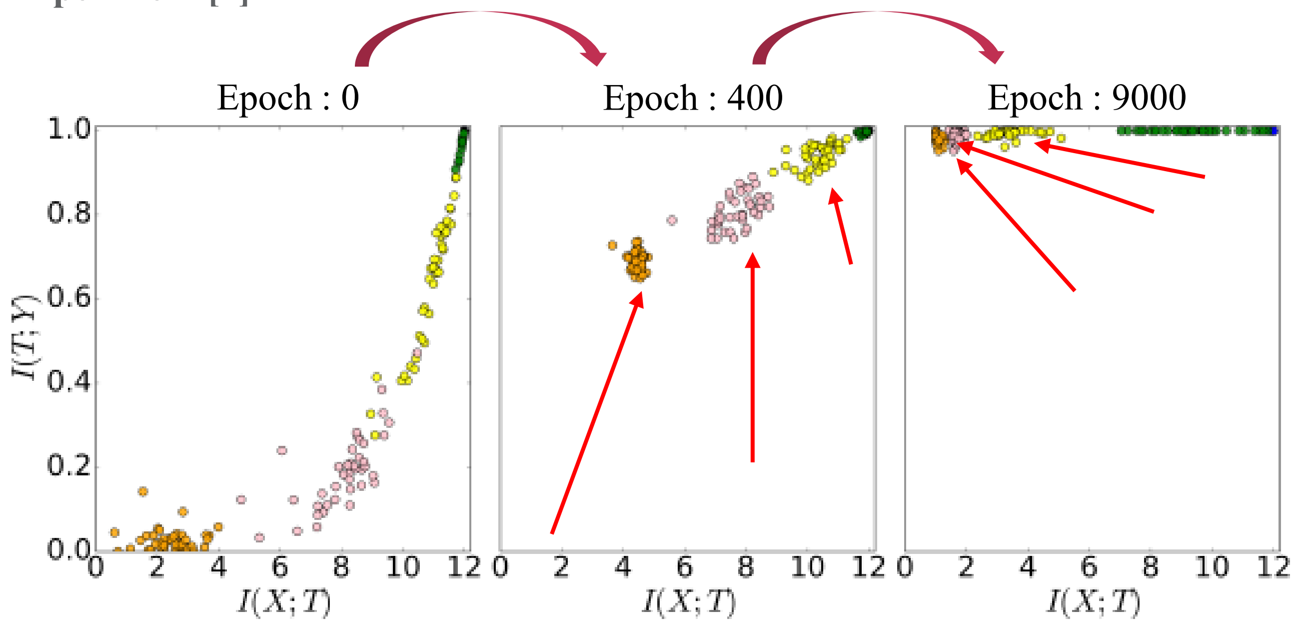
Methods

- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - Experiment [4]



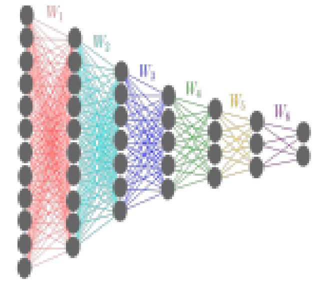
Network with tanh function

Fitting phase　　Compression phase

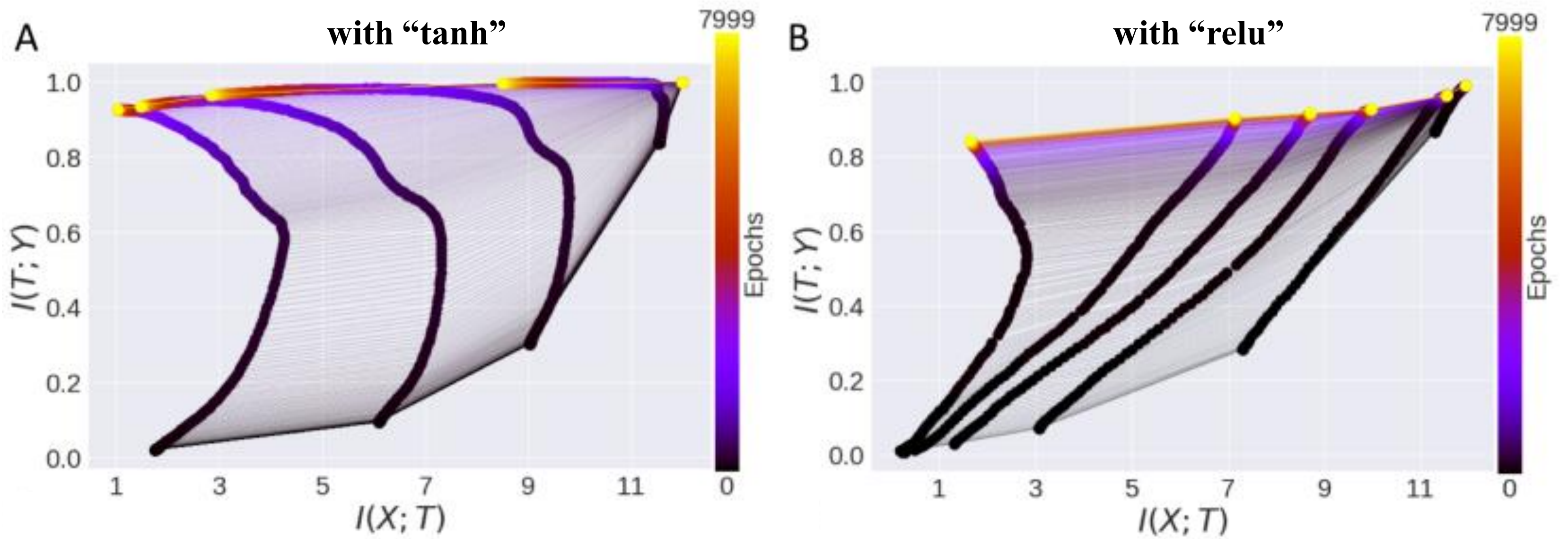


WHY..?

Methods



- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - – *Because of **double-sided saturation non-linearities..?*** [5]

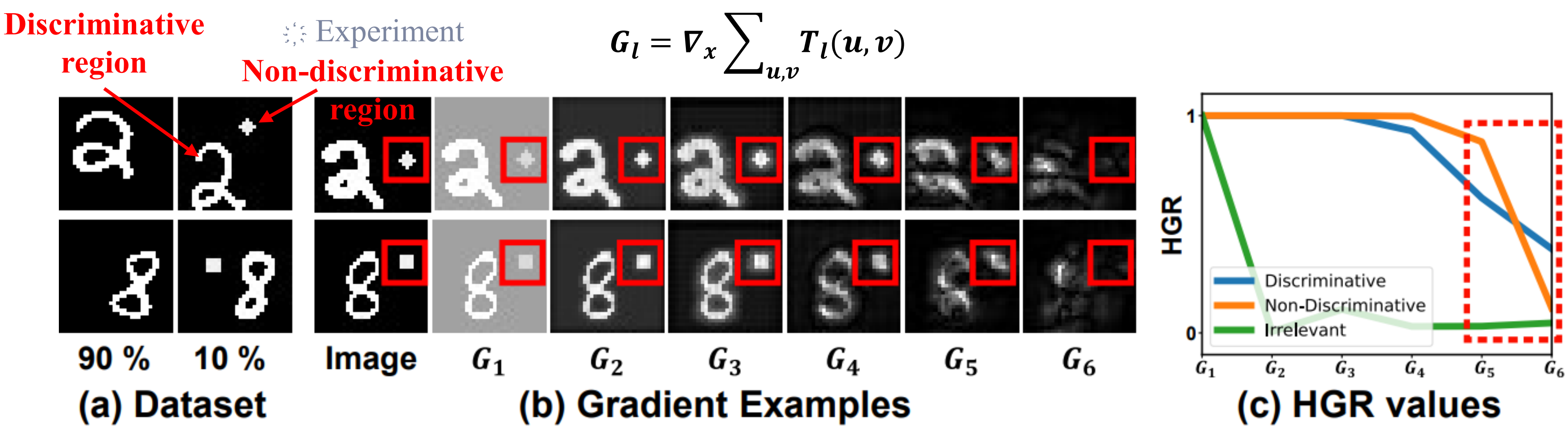


No compression phase is visible!

Methods

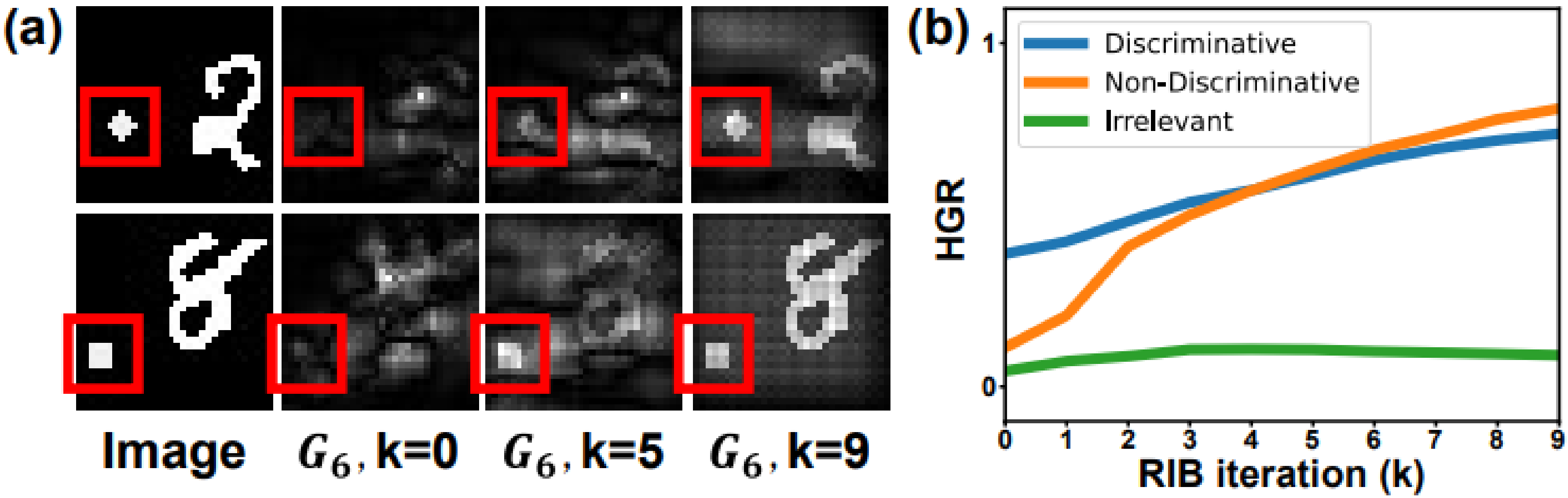
- RIB (Reducing Information Bottleneck) [2]
  - Information Bottleneck
    - *Conclusion*
      - *Information bottleneck is* ***increased*** *by* ***double-sided activation function***
      - *Information bottleneck is* ***more severe in the last layer than in the intermediate layer***
    - *Proof*
      - Experiment

$$G_l = \nabla_x \sum_{u,v} T_l(u,v)$$



Methods

- RIB (Reducing Information Bottleneck) [2]
  - Solution
    - Let's not use **Sigmoid!**

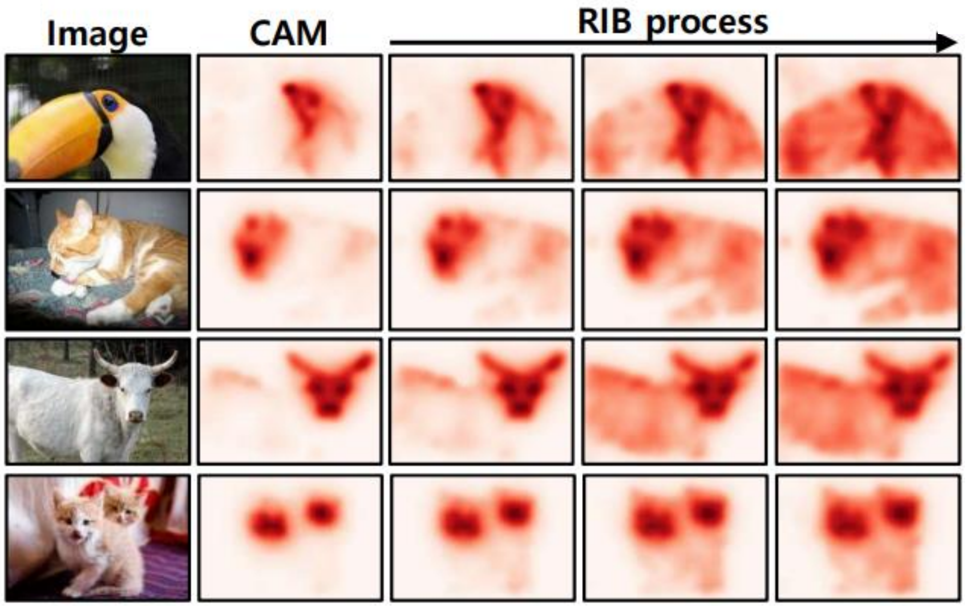


Methods	Backbone	Sup.	Pub.	Val	Test
AdvErasing [34]	VGG16	$\mathcal{I}$	CVPR17	55.0	55.7
GAIN [22]	VGG16	$\mathcal{I}$	CVPR18	55.3	56.8
AffinityNet [2]	ResNet38	$\mathcal{I}$	CVPR18	61.7	63.7
ICD [10]	ResNet101	$\mathcal{I}$	CVPR20	64.1	64.3
IRNet [1]	ResNet50	$\mathcal{I}$	CVPR19	63.5	64.8
SSDD [28]	ResNet38	$\mathcal{I}$	ICCV19	64.9	65.5
SEAM [33]	ResNet38	$\mathcal{I}$	CVPR20	64.5	65.7
Sub-category [3]	ResNet101	$\mathcal{I}$	CVPR20	66.1	65.9
RRM [37]	ResNet101	$\mathcal{I}$	AAAI20	66.3	66.5
BES [5]	ResNet101	$\mathcal{I}$	ECCV20	65.7	66.6
Ours	ResNet38	$\mathcal{I}$	-	**68.4**	<u>68.2</u>

Quantitative results of [1]

Methods

- RIB (Reducing Information Bottleneck) [2]
  - Result



Qualitative results

Method	*val*	*test*
Supervision: Bounding box labels		
Song *et al.* CVPR '19 [50]	70.2	-
BBAM CVPR '21 [34]	73.7	73.7
Supervision: Image class labels		
IRN CVPR '19 [2]	63.5	64.8
SEAM CVPR '20 [55]	64.5	65.7
BES ECCV '20 [10]	65.7	66.6
Chang *et al.* CVPR '20 [7]	66.1	65.9
RRM AAAI '20 [61]	66.3	66.5
CONTA NeurIPS '20 [62]	66.1	66.7
RIB (Ours)	**68.3**	**68.6**

Quantitative results

Reference

1. Kweon, Hyeokjun, et al. "Unlocking the potential of ordinary classifier: Class-specific adversarial erasing framework for weakly supervised semantic segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
2. Lee, Jungbeom, et al. "Reducing Information Bottleneck for Weakly Supervised Semantic Segmentation." Advances in Neural Information Processing Systems 34 (2021).
3. Zhou, Bolei, et al. "Learning deep features for discriminative localization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
4. Shwartz-Ziv, Ravid, and Naftali Tishby. "Opening the black box of deep neural networks via information." arXiv preprint arXiv:1703.00810 (2017).
5. Saxe, Andrew M., et al. "On the information bottleneck theory of deep learning." Journal of Statistical Mechanics: Theory and Experiment 2019.12 (2019): 124020.