URBAN ROBOTICS LAB

What I've learned from NeurIPS 2019

Lab Seminar (2020.01.09)

Yunsoo Kim

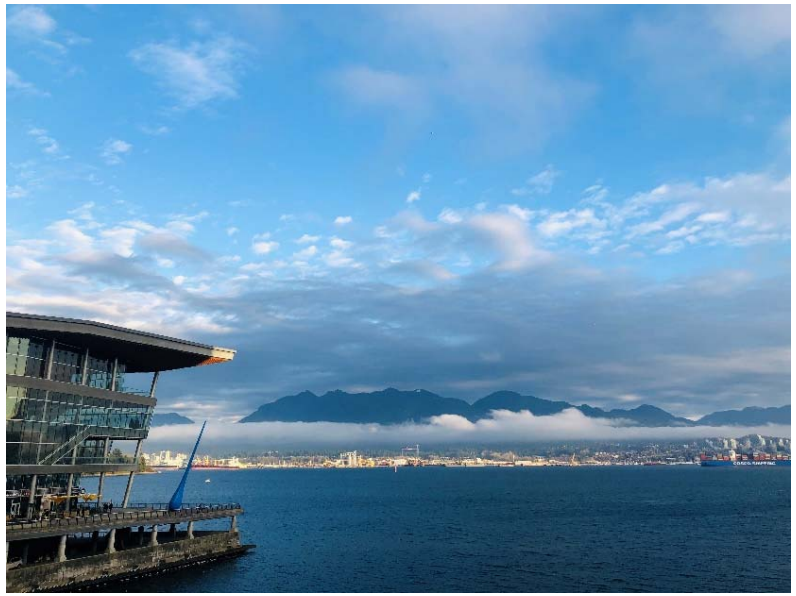
KAIST

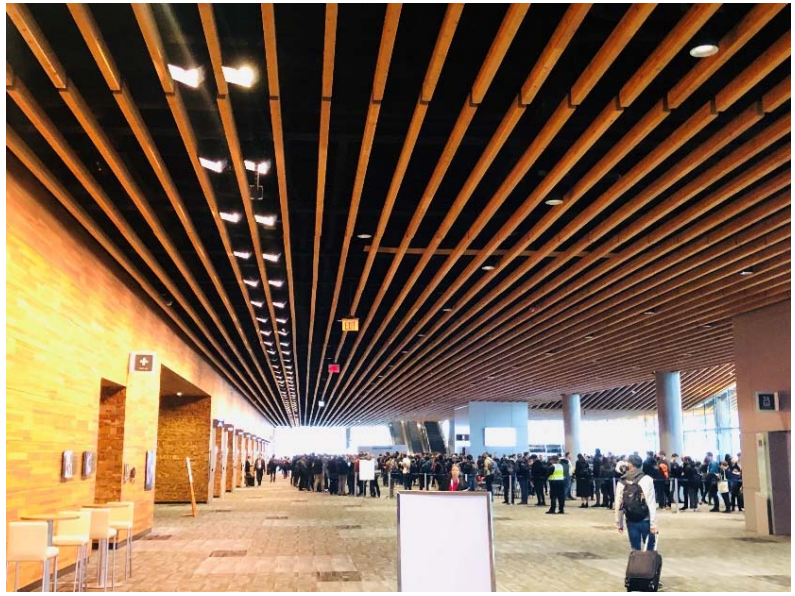
URBAN ROBOTICS LAB

NeurIPS (Neural Information Processing Systems)

Vancouver
CANADA

8th, Dec
~ 14th, Dec





NeurIPS 2019

> 12000

Number of Registrations

NeurIPS 2019

6743

Papers Submitted

NeurIPS 2019

1428

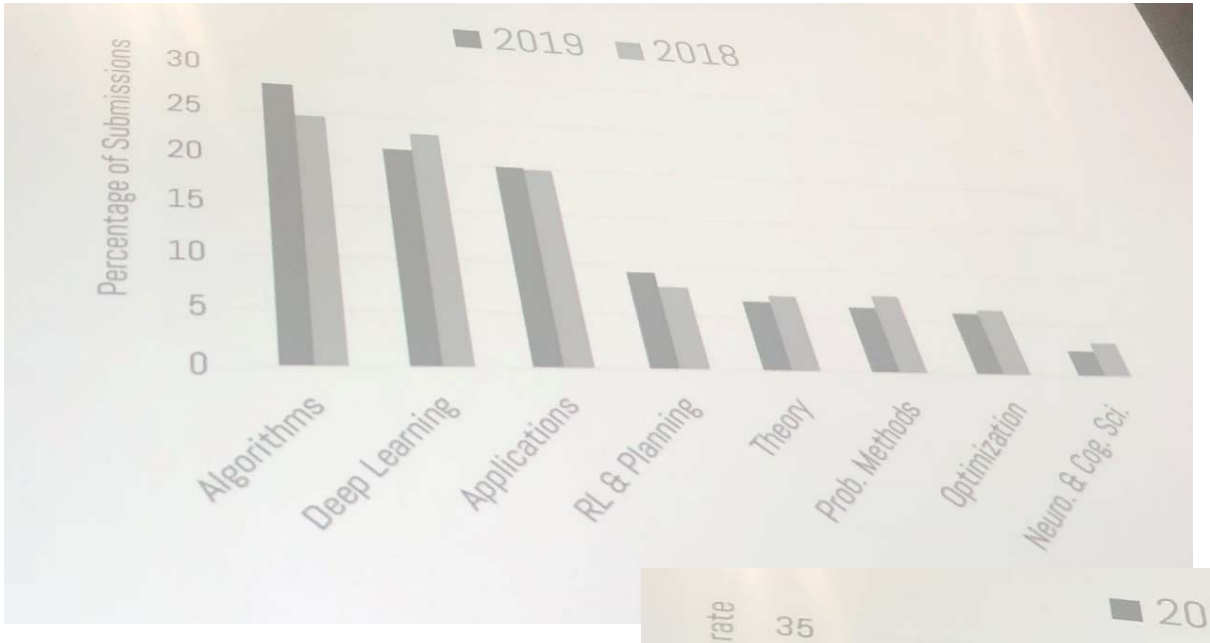
Papers Accepted

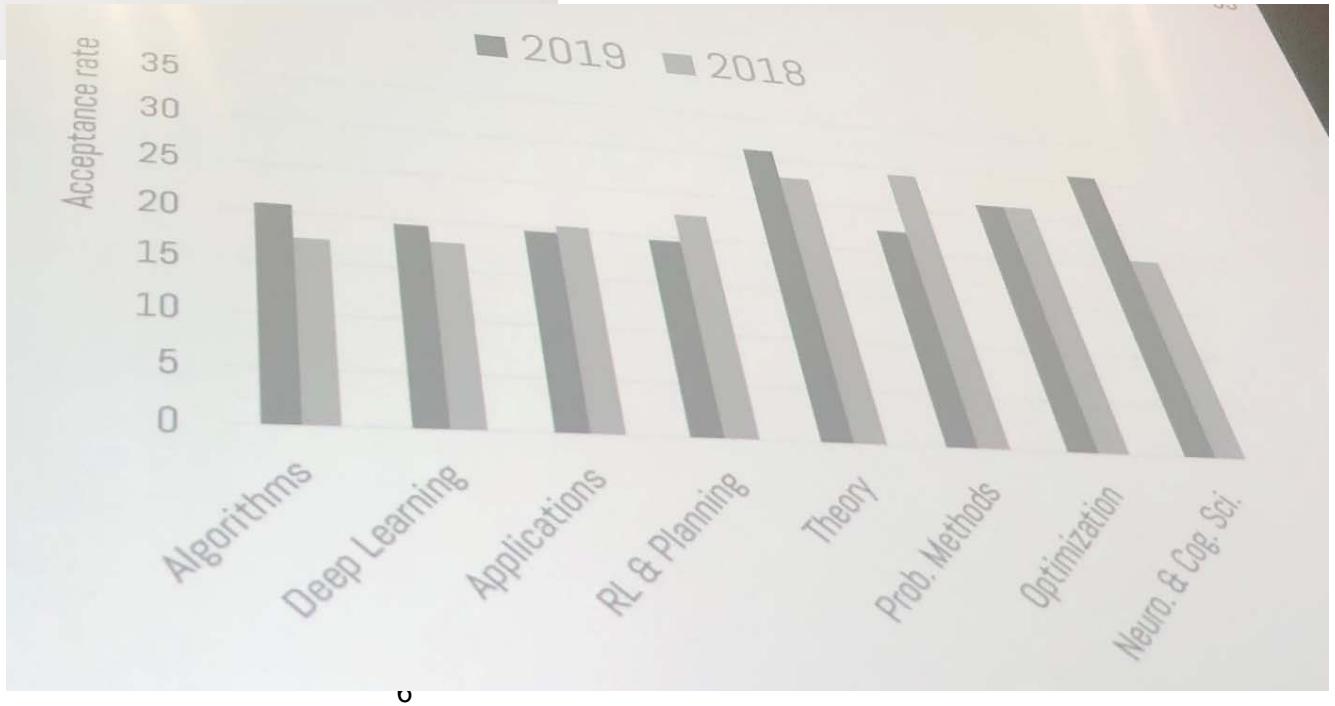
NeurIPS 2019

21%

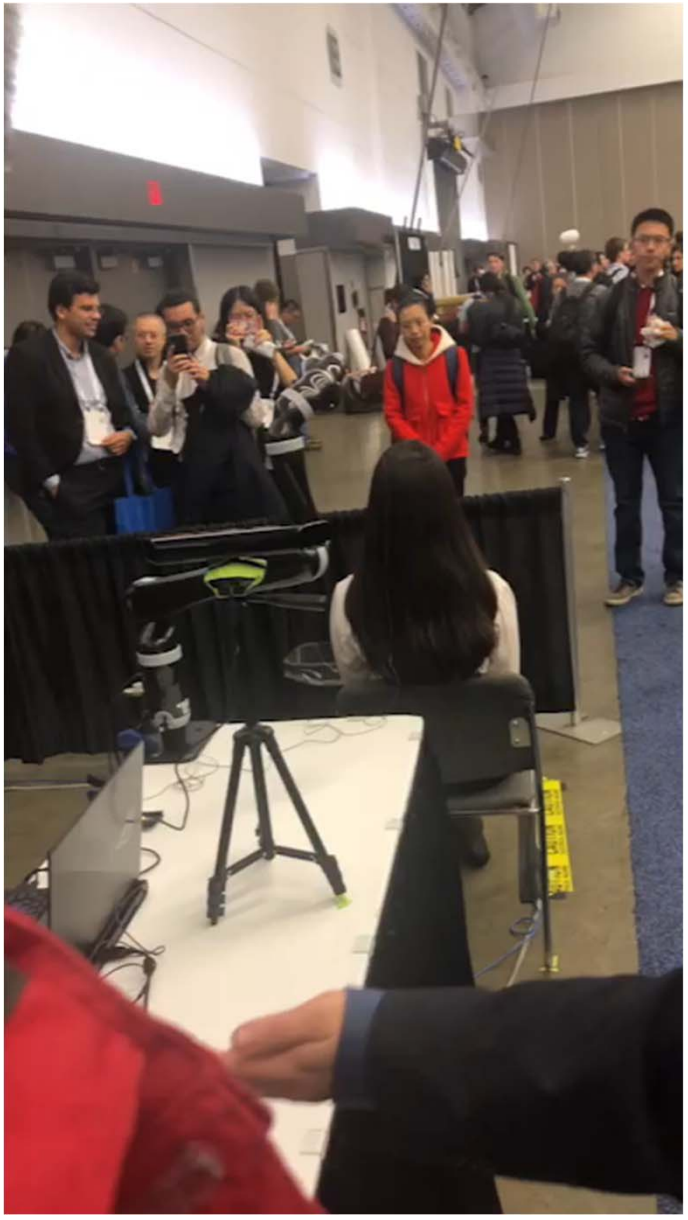
Acceptance rate

NeurIPS 2019

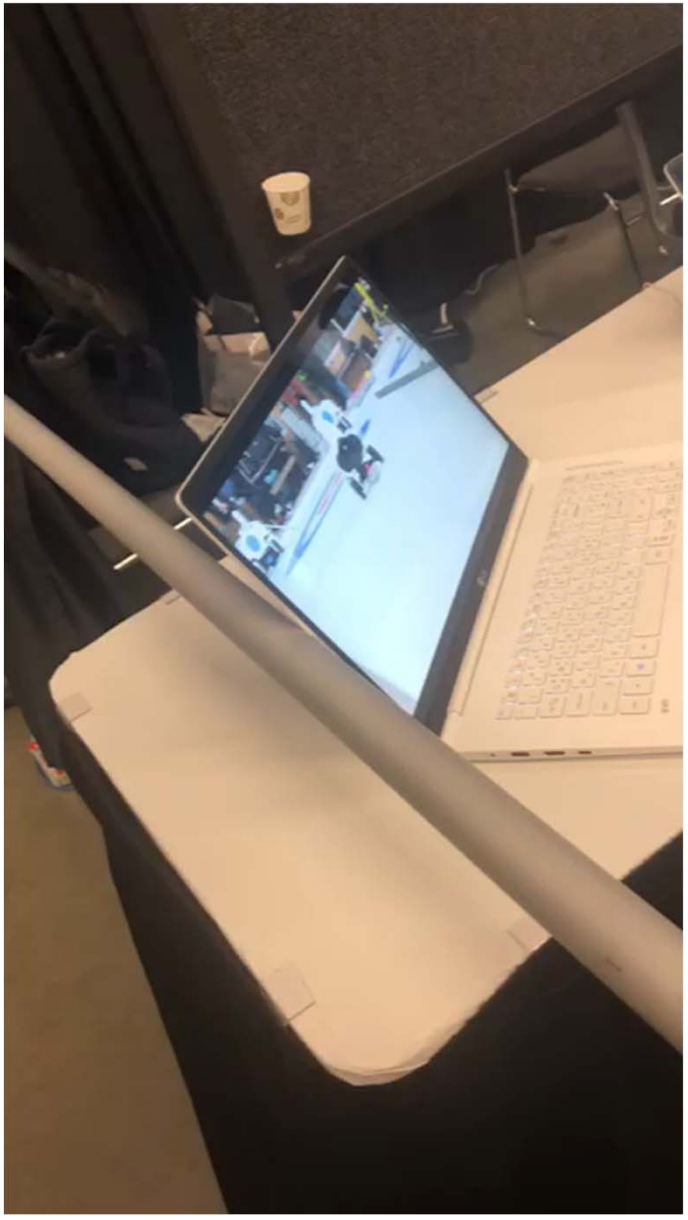




Demos – Hair-Brushing Machine



Demos – Curling Robot (KOREA univ.)



Demos

2 Concepts what I interestingly heard

- Saliency Map & TCAV

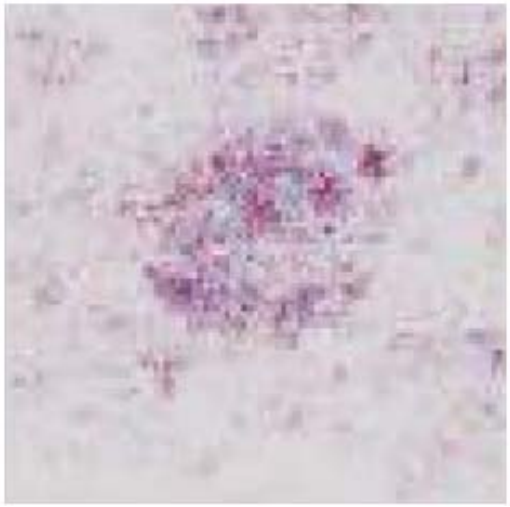
- Federated Learning

Saliency Map

- Evidence of prediction

A trained
machine learning model
(e.g., neural network)

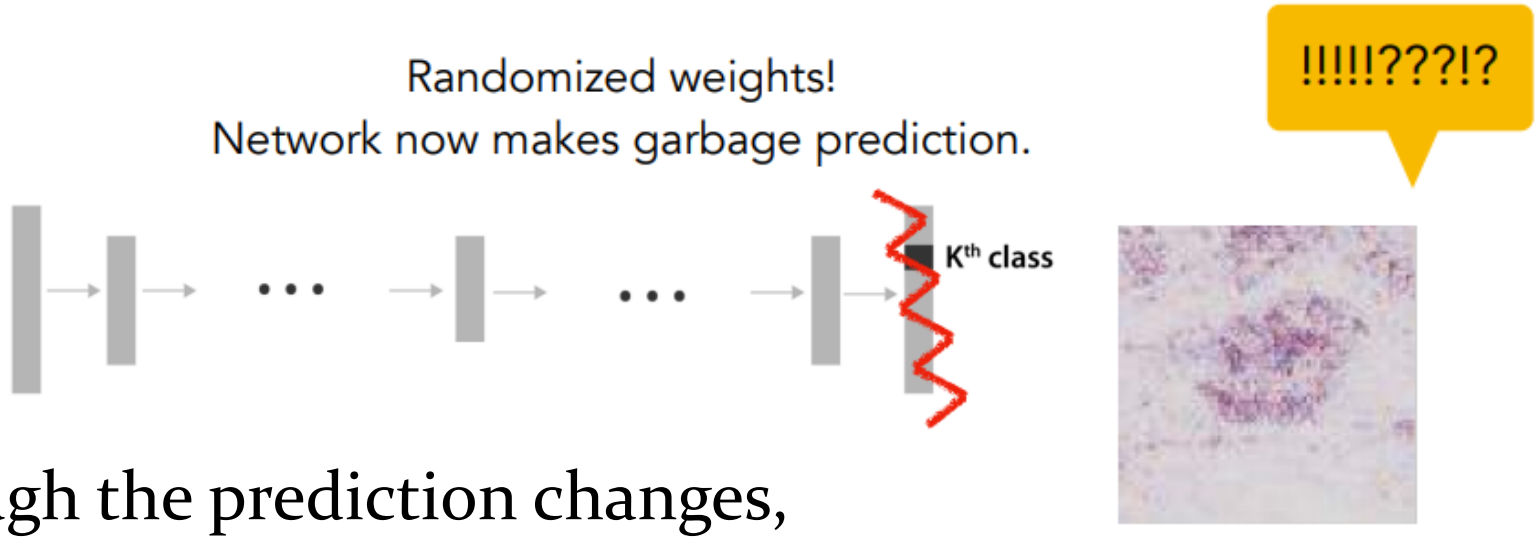
$p(z)$



Saliency map

Saliency Map

What if the prediction changes? (what if the prediction is random?)



Even though the prediction changes,
explanations are not that changed

Saliency Map

Saliency map is not the reason for the predicted label,
but show which part the machine sees.

→ Can we quantitatively measure importance of user-chosen concepts?

→ TCAV (Testing with Concept Activation Vectors)

TCAV

A trained machine learning model (e.g., neural network)

$p(z)$

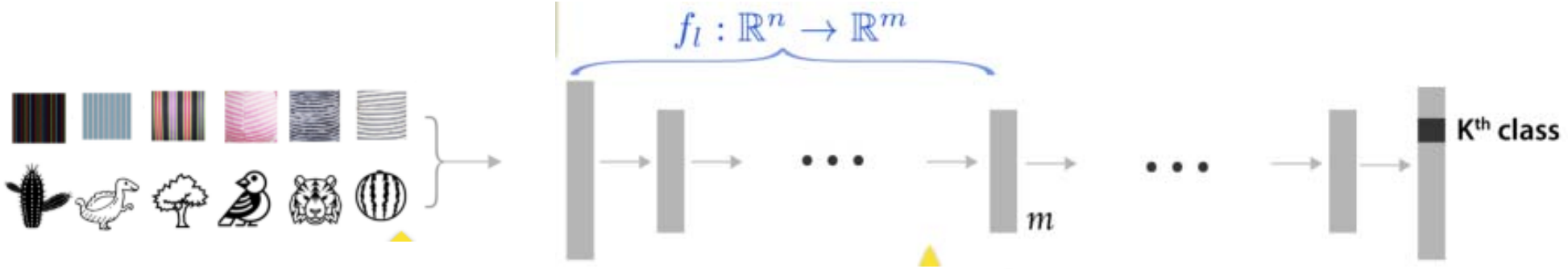
zebra-ness

Was striped concept important to this zebra image classifier?

TCAV score for **Zebra**

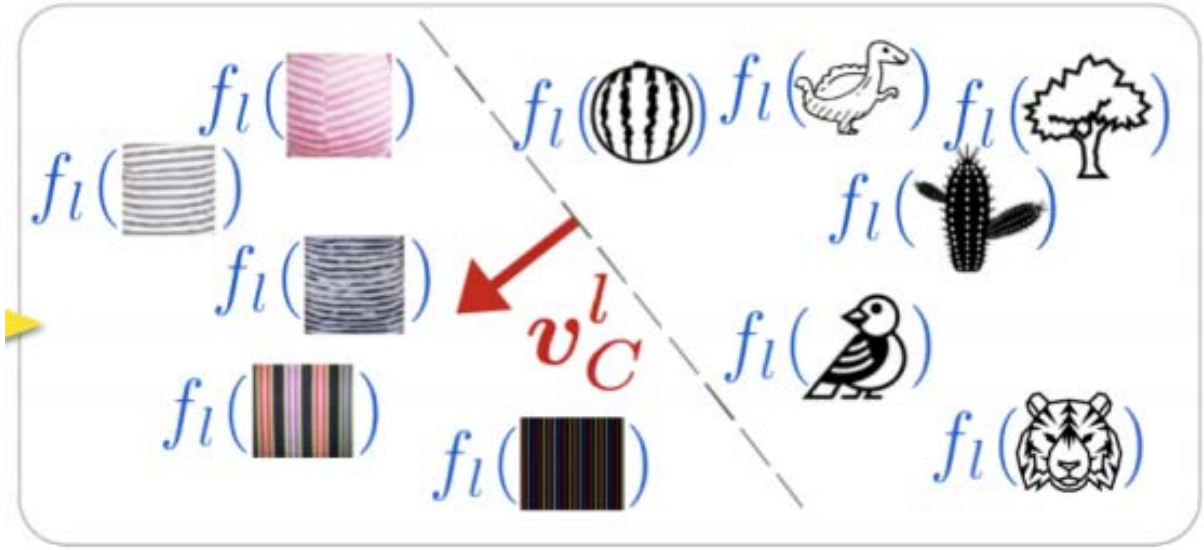
TCAV provides **quantitative importance** of a concept **if and only if** your network learned about it.

TCAV



Examples of concepts & random images

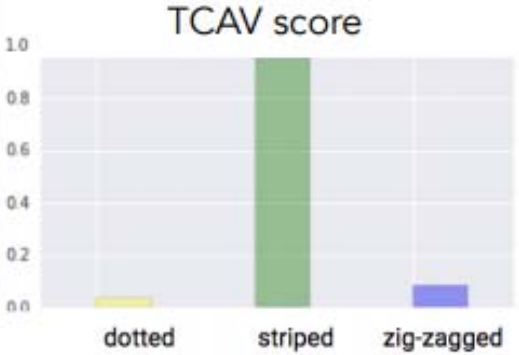
Gather the activations

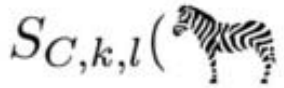


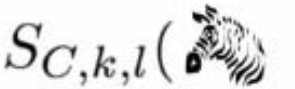
Train classifier : get vector from random to concept

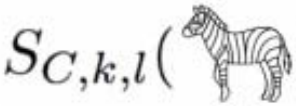
TCAV

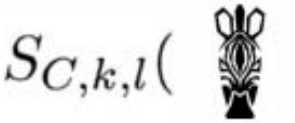
TCAV



$S_{C,k,l}($  $)$

$S_{C,k,l}($  $)$

$S_{C,k,l}($  $)$

$S_{C,k,l}($  $)$

zebra-ness → striped CAV →

$$\frac{\partial p(z)}{\partial \boldsymbol{v}_C^l} = S_{C,k,l}(\boldsymbol{x})$$

$$\text{TCAVQ}_{C,k,l} = \frac{|\{\boldsymbol{x} \in X_k : S_{C,k,l}(\boldsymbol{x}) > 0\}|}{|X_k|}$$

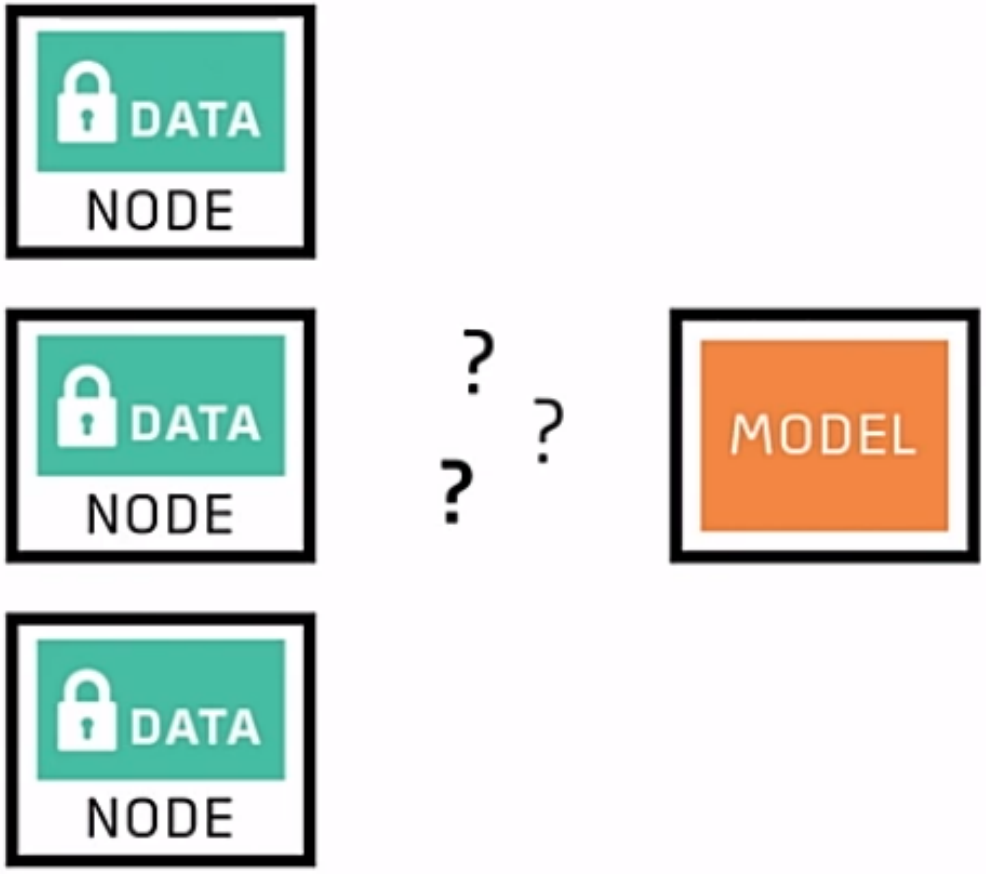
Directional derivative with CAV

Federated Learning (Distributed learning)

For better life, need to use deep learning without an invasion of privacy. (ex. Health care, etc..) → Need to get diverse data

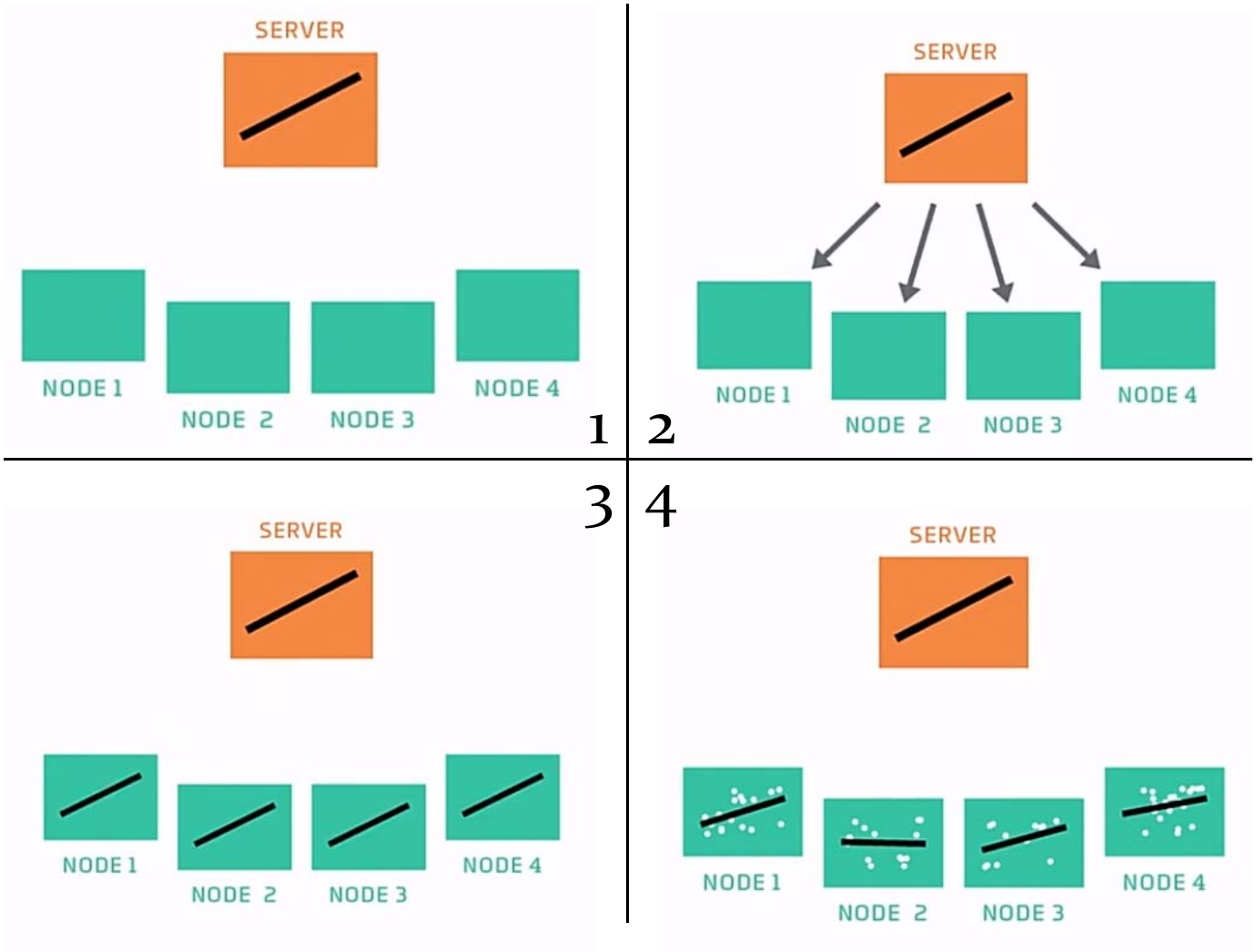
Federated Learning (Distributed learning)

- Improve the model based on the data from the edge devices



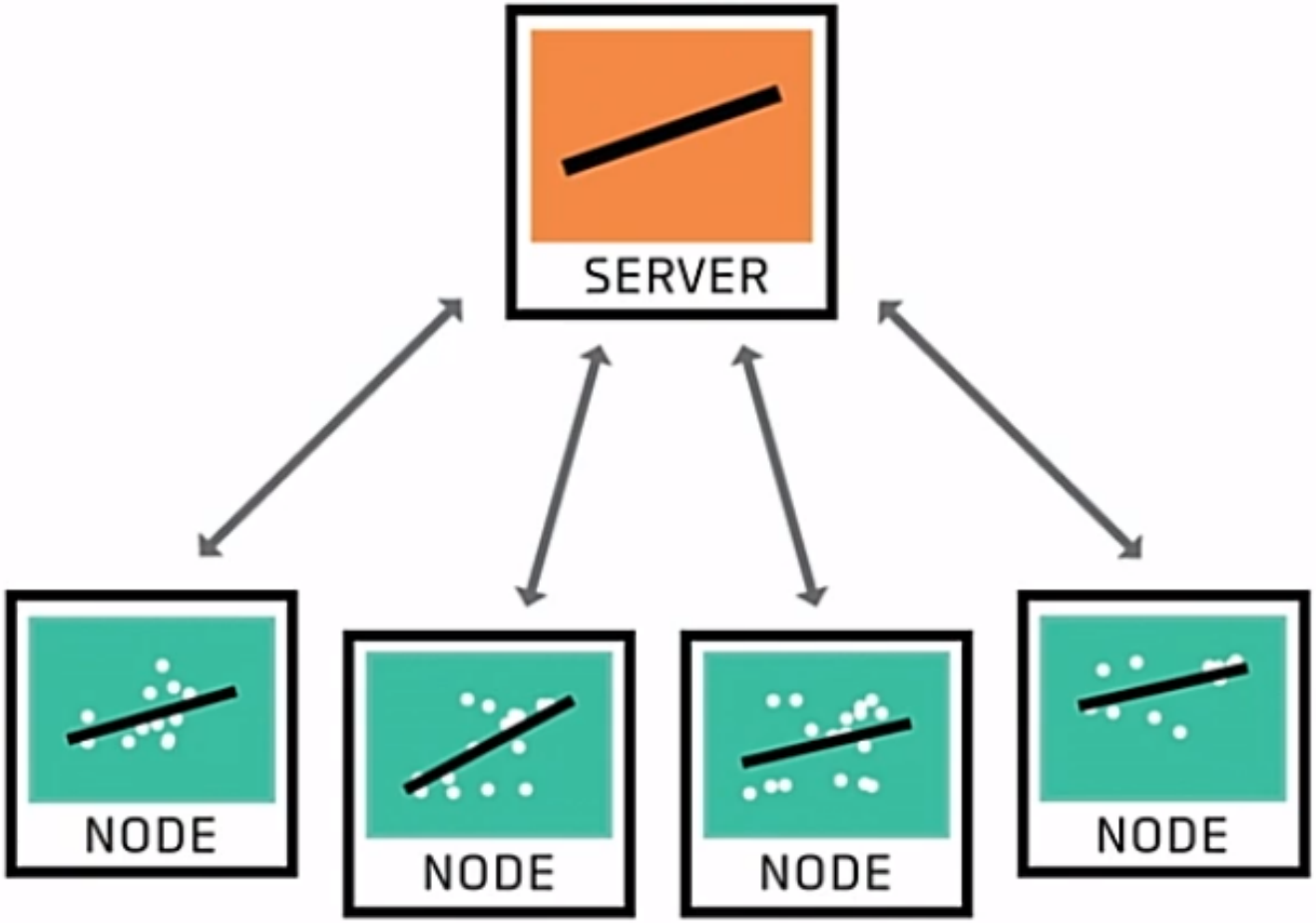
Federated Learning (Distributed learning)

- Federated averaging



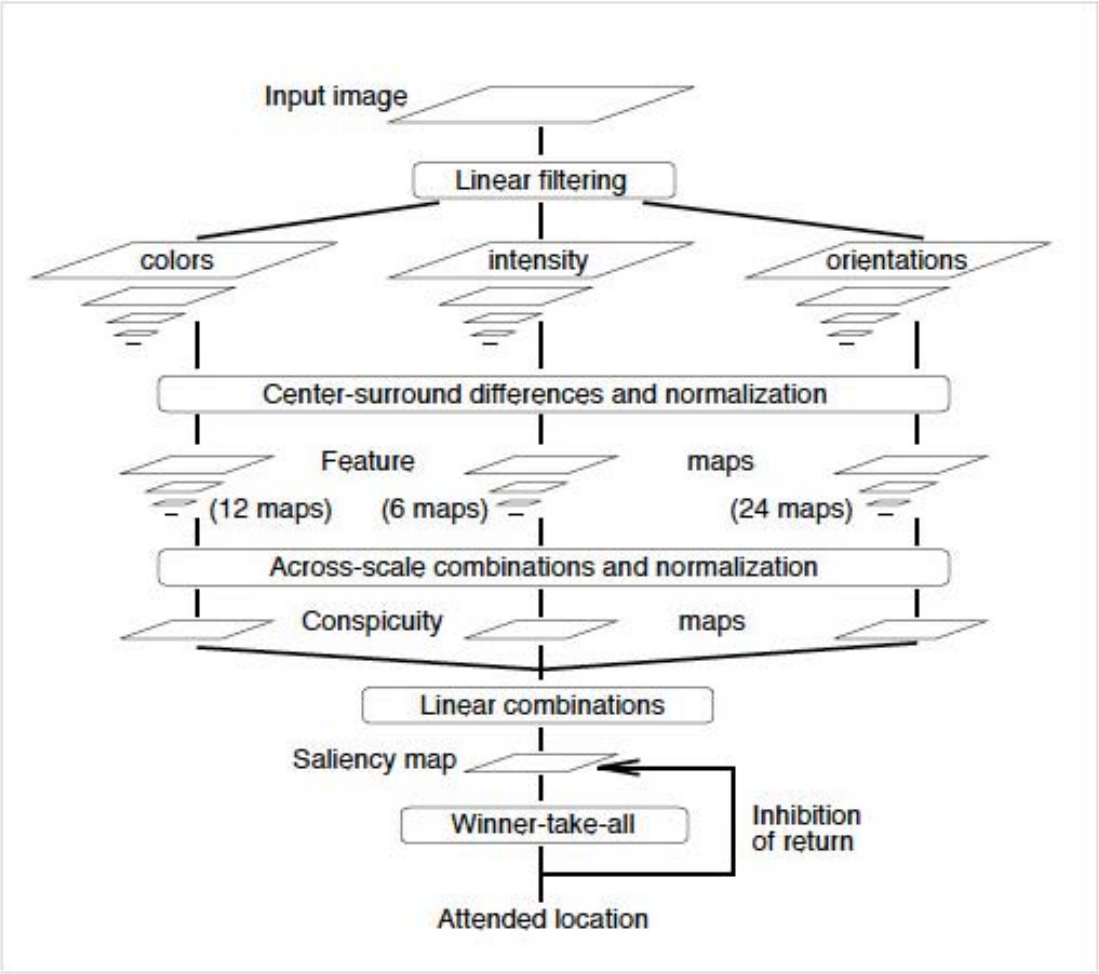
Federated Learning (Distributed learning)

- Federated averaging



Take an average of the models

Thank you


Input image
Linear filtering
colors
intensity
orientations
Center-surround differences and normalization
Feature maps
(12 maps)
(6 maps)
(24 maps)
Across-scale combinations and normalization
Conspicuity maps
Linear combinations
Saliency map
Winner-take-all
Inhibition of return
Attended location