Cylindrical Convolutional Networks for Joint Object Detection and Viewpoint Estimation

Sunghun Joung, Seungryong Kim, Hanjae Kim, Minsu Kim,
Ig-Jae Kim, Junghyun Cho, and Kwnaghoon Sohn

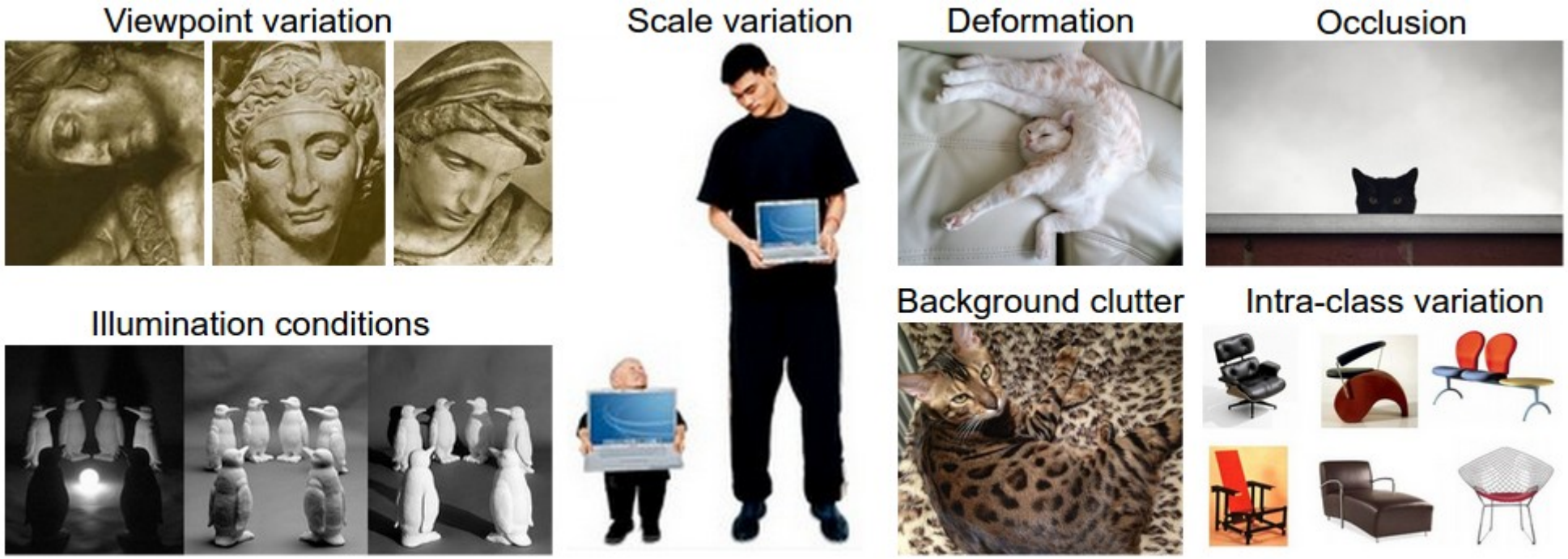
YONSEI UNIVERSITY EPFL

KOREA UNIVERSITY KIST 한국과학기술연구원
Korea Institute of Science and Technology

Introduction

Modeling Spatial Transformations

- "A long-standing problem in Visual Recognition"



→ Use visible features to model spatial transformation in a 2D space

[Image courtesy: Fei-Fei Li]

Introduction

Viewpoint Variation

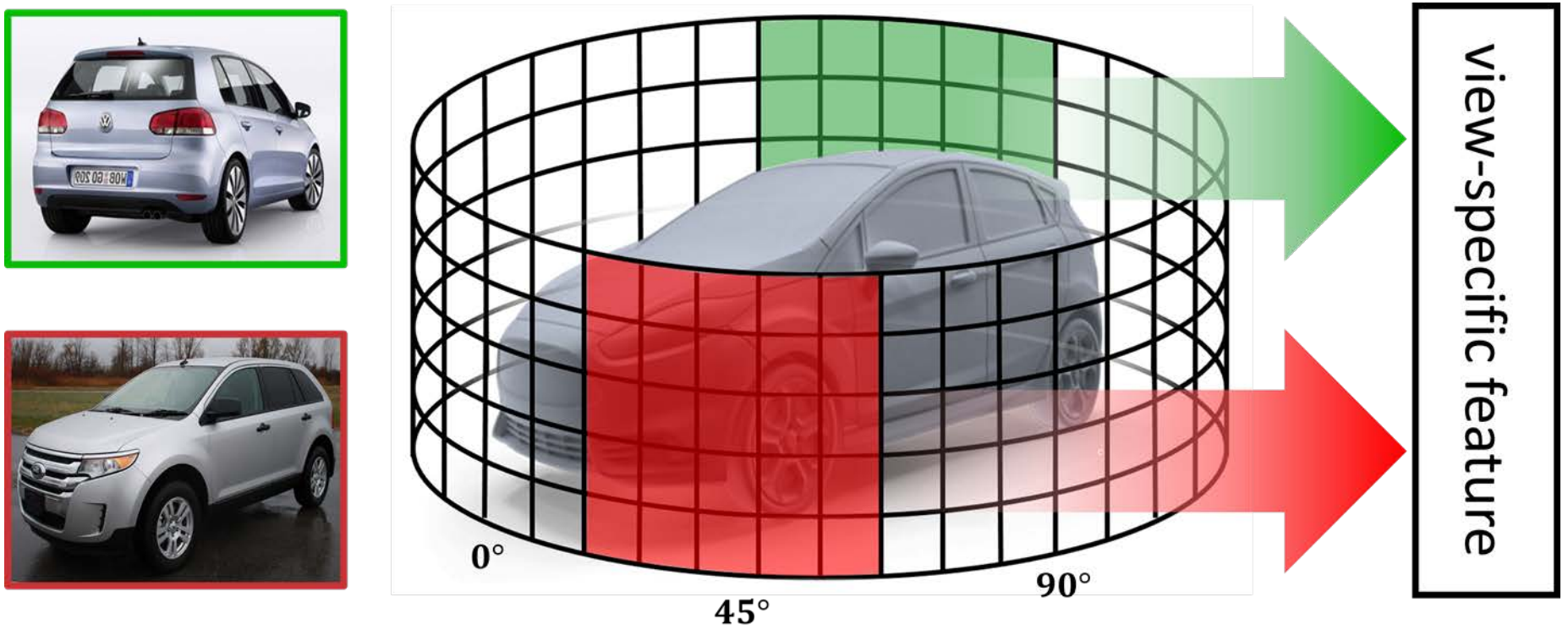
- Objects in a 2D space are a projection of 3D ones

→ How to extract **view-specific feature** conditioned on the object viewpoint?

Problem Formulation

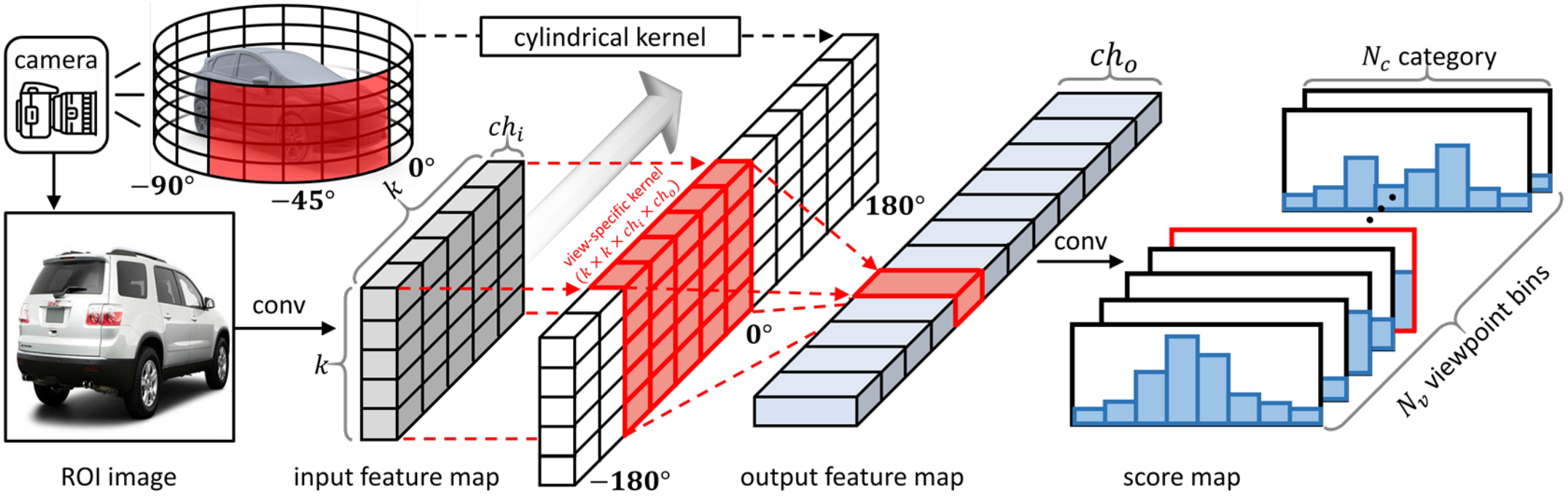
Cylindrical Convolutional Networks (CCNs)

- Exploit **cylindrical convolutional kernel**, where each subset is **view-specific kernel** to capture structural information at each viewpoint



Proposed Method

CCNs for Joint Object Detection and Viewpoint Estimation



- Input: Input feature maps of single-view image
- Output: Multiple view-specific feature to predict object category scores at each viewpoint

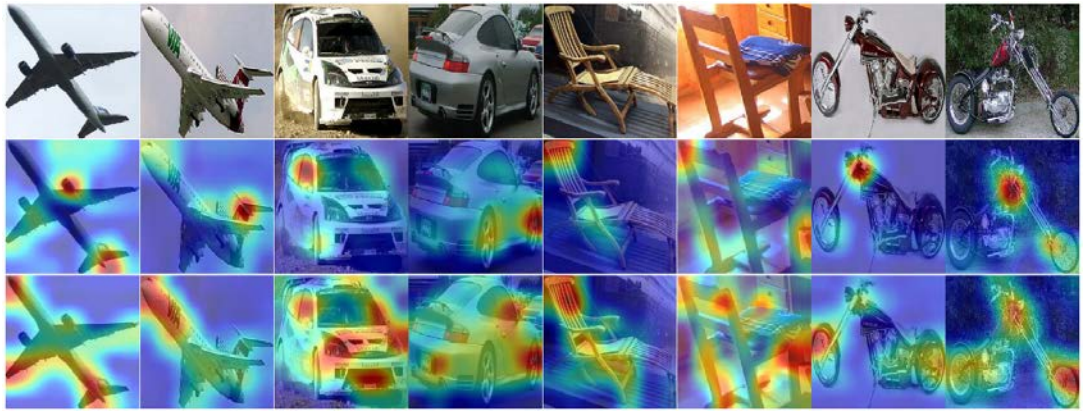
Experiments

Ablation Study

- Analysis of CCNs components

Method		Cateogory		Viewpoint	
N_v	CCNs	top-1	top-3	$\text{Acc}_{\pi/6}$	Mederr
24		0.91	0.97	0.56	23.5
18	✓	0.95	0.99	0.63	17.3
24	✓	**0.95**	**0.99**	**0.66**	**15.5**
30	✓	0.94	0.98	0.63	17.7

- Network visualization



Results

- Pascal 3D+ dataset

- KITTI dataset

Thank you!

Sunghun Joung
Ph.D. Candidate
School of Electrical and Electronic Engineering
Yonsei University
E-mail: sunghunjoung@yonsei.ac.kr