Visual Recognition using Deep Learning

Tips for Final Project Presentation

林彥宇 教授
Yen-Yu Lin, Professor

國立陽明交通大學 資訊工程學系
Computer Science, National Yang Ming Chiao Tung University

Presentation

- Your presentation/reports may include
  - Introduction
  - Related work
  - Proposed approach
  - Experimental results
  - Conclusions

- Presentation and reports need to include the link of your code

Presentation

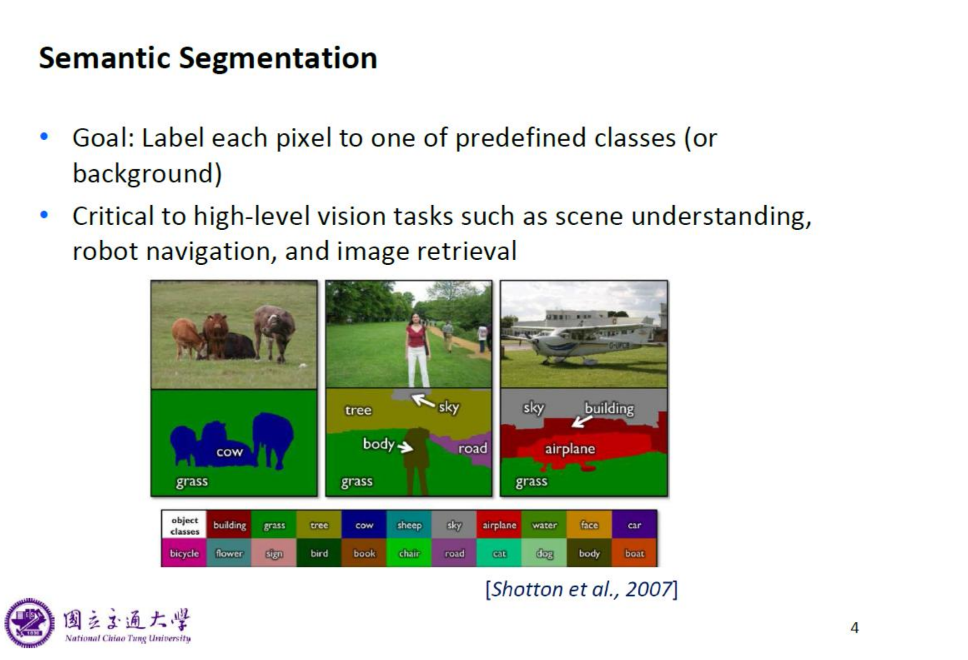
- Your presentation
    - Introduction
    - Related work
    - Proposed approach
    - Experimental results
    - Conclusions

Introduction

- Problem statement
- The importance of this problem
- The difficulties you address

Introduction

- Problem statement
- The importance of this problem
- The motivation or difficulty you address



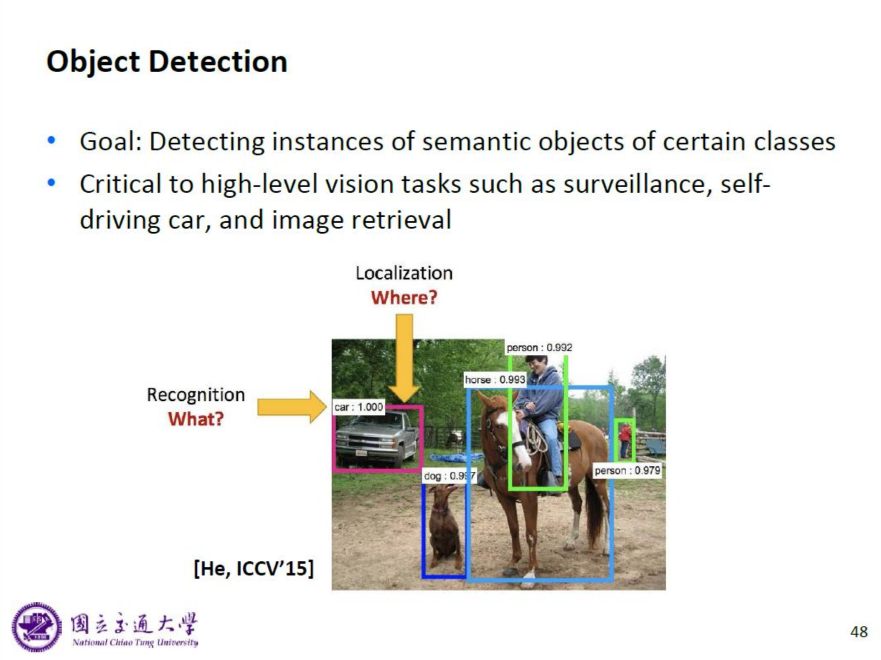
Semantic Segmentation

- Goal: Label each pixel to one of predefined classes (or background)
- Critical to high-level vision tasks such as scene understanding, robot navigation, and image retrieval

[*Shotton et al., 2007*]

國立交通大學 National Chiao Tung University

4



Object Detection

- Goal: Detecting instances of semantic objects of certain classes
- Critical to high-level vision tasks such as surveillance, self-driving car, and image retrieval

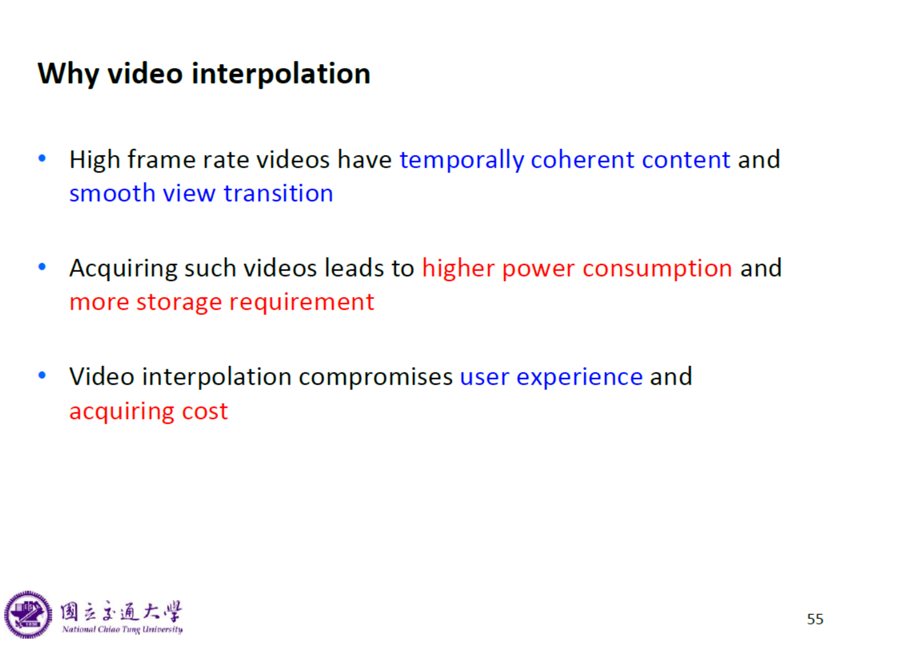
[He, ICCV'15]

國立交通大學 National Chiao Tung University

48

Introduction

- Problem statement
- The importance of this problem
- The motivation or difficulty you address

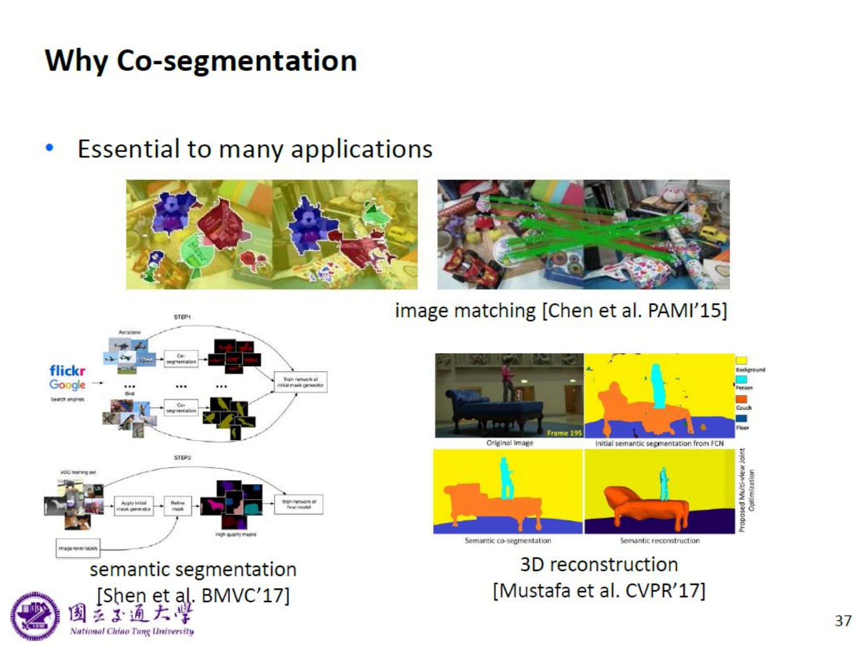


Why video interpolation

- High frame rate videos have temporally coherent content and smooth view transition
- Acquiring such videos leads to higher power consumption and more storage requirement
- Video interpolation compromises user experience and acquiring cost

國立交通大學
National Chiao Tung University

55



Why Co-segmentation

- Essential to many applications

image matching [Chen et al. PAMI'15]

semantic segmentation
[Shen et al. BMVC'17]

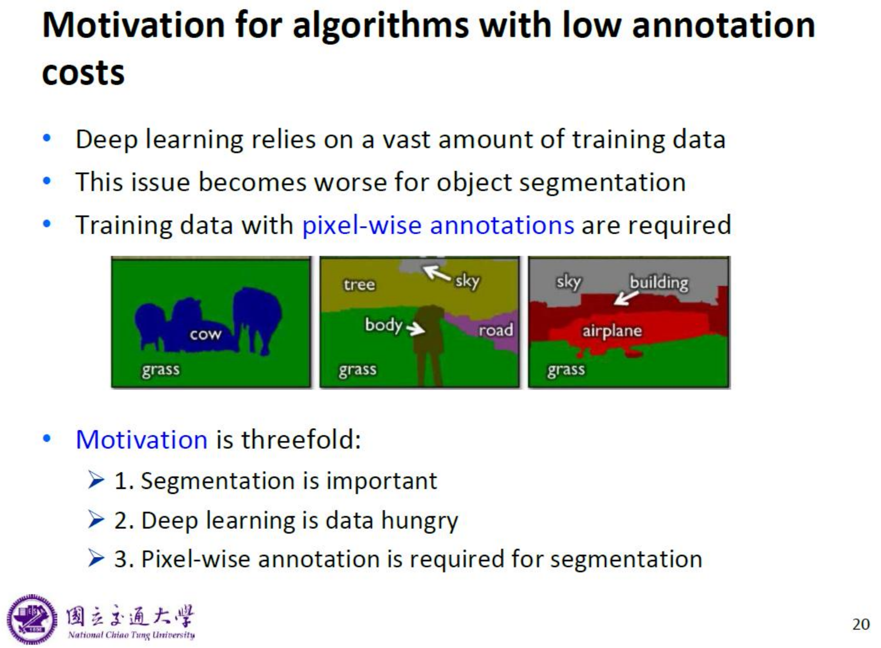
3D reconstruction
[Mustafa et al. CVPR'17]

國立交通大學
National Chiao Tung University

37

Interduction

- Problem statement
- The importance of this problem
- The motivation or difficulty you address



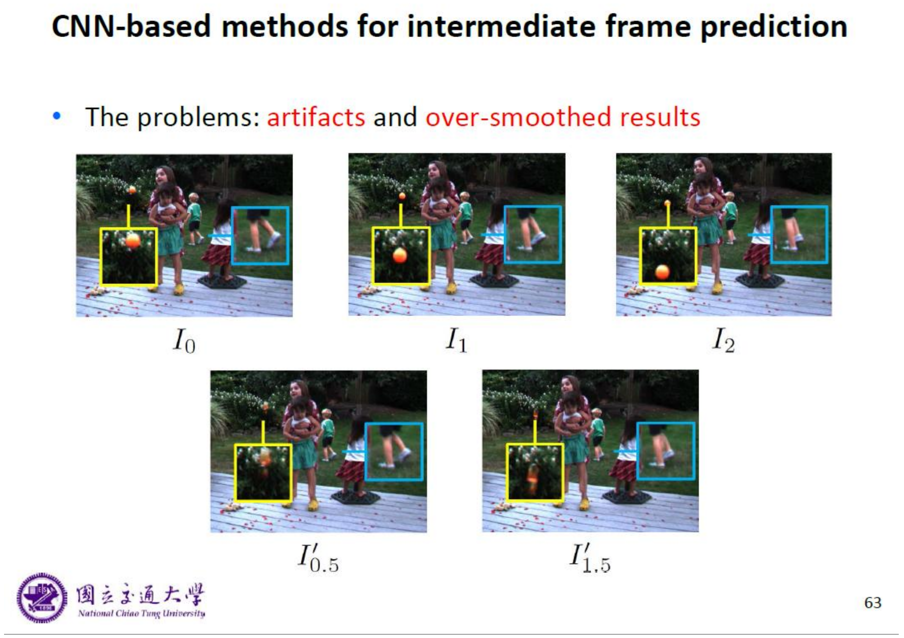
Motivation for algorithms with low annotation costs

- Deep learning relies on a vast amount of training data
- This issue becomes worse for object segmentation
- Training data with pixel-wise annotations are required

- Motivation is threefold:
  - 1. Segmentation is important
  - 2. Deep learning is data hungry
  - 3. Pixel-wise annotation is required for segmentation

國立交通大學 National Chiao Tung University

20



CNN-based methods for intermediate frame prediction

- The problems: artifacts and over-smoothed results

國立交通大學 National Chiao Tung University

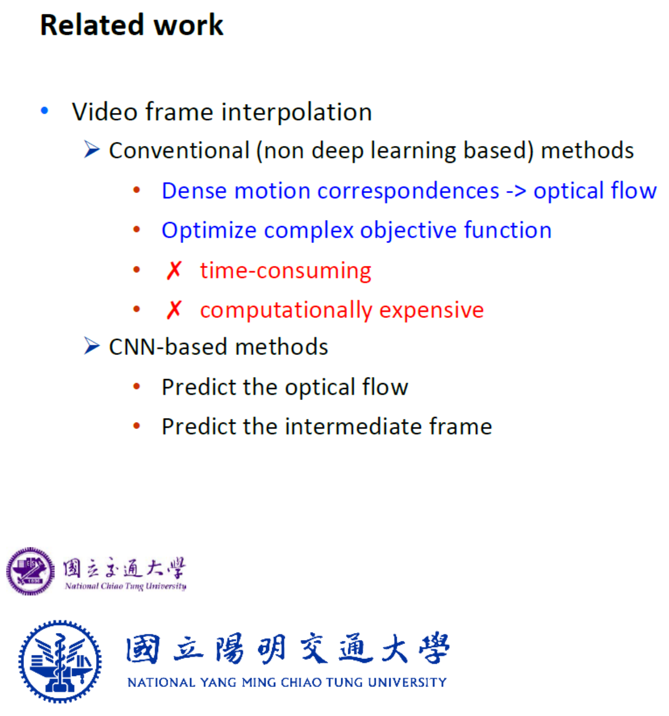
63

Presentation

- Your presentation
  - Introduction
  - Related work
  - Proposed approach
  - Experimental results
  - Conclusions

Related work

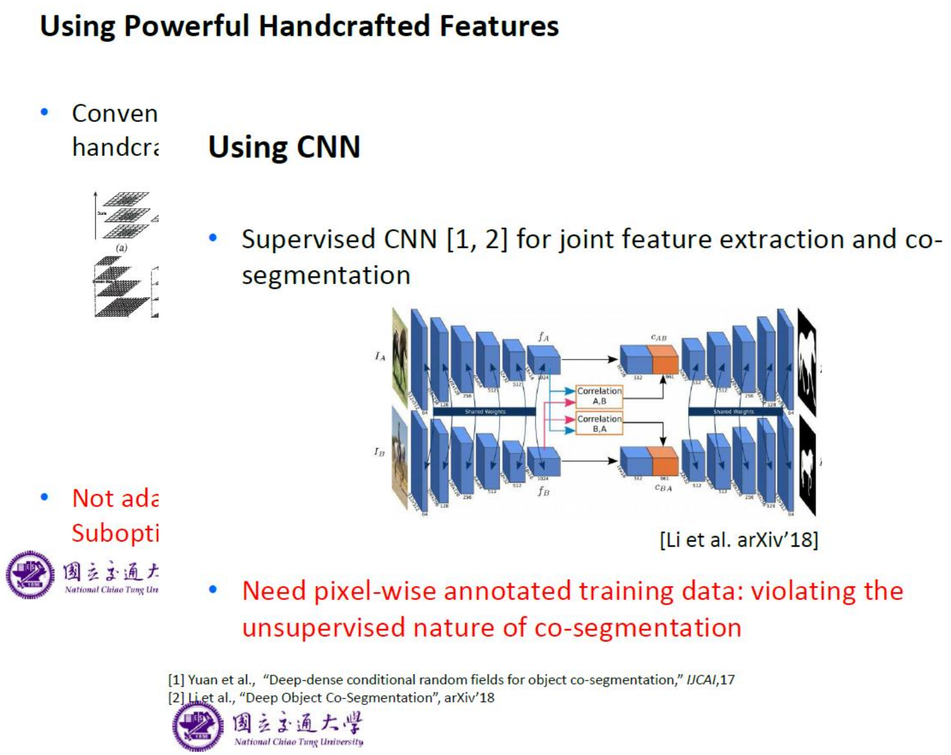
- Divide the related work/methods into groups
- For each group,
  - ➤ Give a high-level description about methods of this group
  - ➤ Summarize the pros and cons for each group



Related work

- Video frame interpolation
  - ➤ Conventional (non deep learning based) methods
    - Dense motion correspondences -> optical flow
    - Optimize complex objective function
    - ✗ time-consuming
    - ✗ computationally expensive
  - ➤ CNN-based methods
    - Predict the optical flow
    - Predict the intermediate frame

國立交通大學 National Chiao Tung University



Using Powerful Handcrafted Features

- Conven handcra
- Not ada Subopti

(a)

國立交通大 National Chiao Tung Un

Using CNN

- Supervised CNN [1, 2] for joint feature extraction and co-segmentation

[Li et al. arXiv'18]

- Need pixel-wise annotated training data: violating the unsupervised nature of co-segmentation

[1] Yuan et al., "Deep-dense conditional random fields for object co-segmentation," *IJCAI*,17
[2] Li et al., "Deep Object Co-Segmentation", arXiv'18

國立交通大學 National Chiao Tung University

Presentation

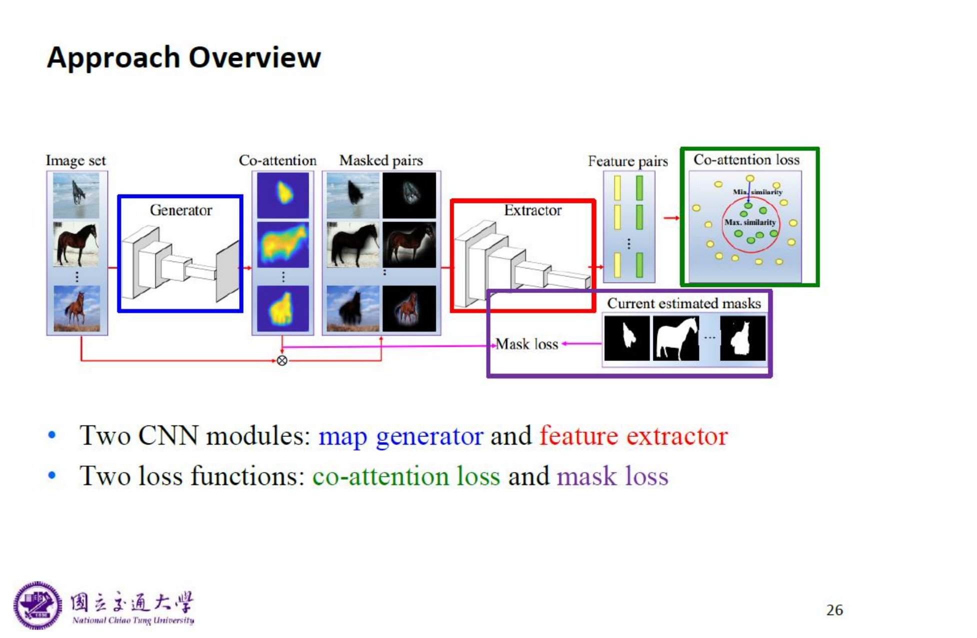
- Your presentation
  - Introduction
  - Related work
  - Proposed approach
  - Experimental results
  - Conclusions

Proposed approach

- Overview: Network figure
- Details of your approach

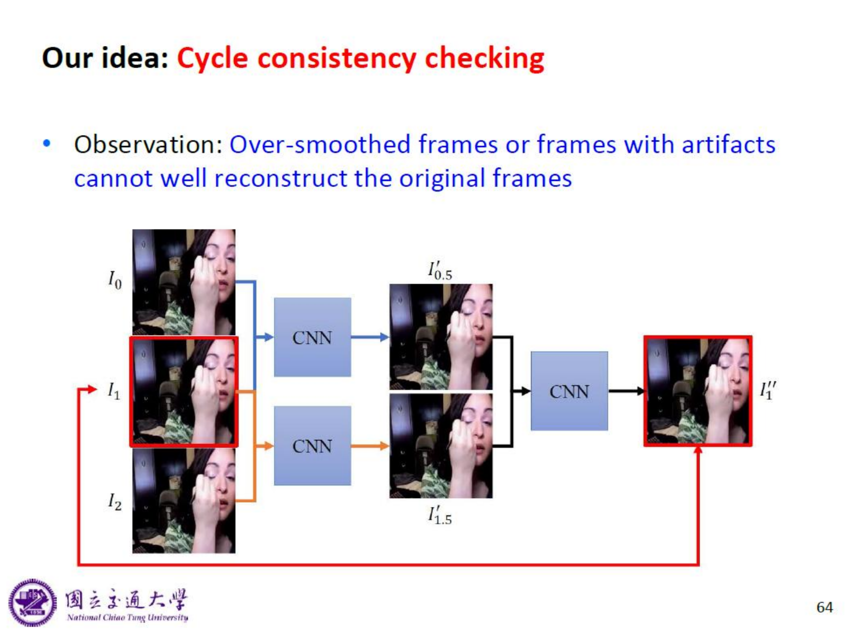
Proposed approach

- Overview: Network figure
- Details of your approach



Approach Overview

- Two CNN modules: map generator and feature extractor
- Two loss functions: co-attention loss and mask loss

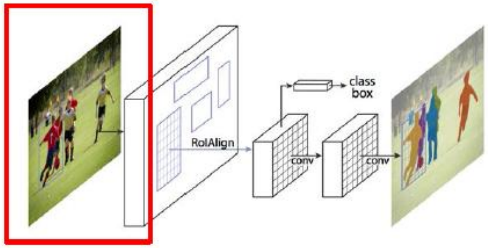


Our idea: Cycle consistency checking

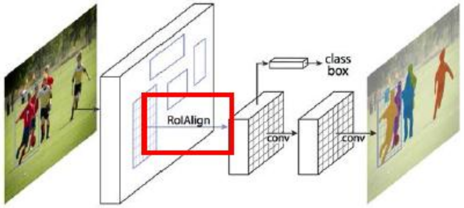
- Observation: Over-smoothed frames or frames with artifacts cannot well reconstruct the original frames

Proposed approach

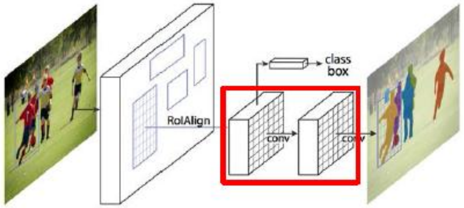
- Overview: Network figure with loss function
- Details of your approach



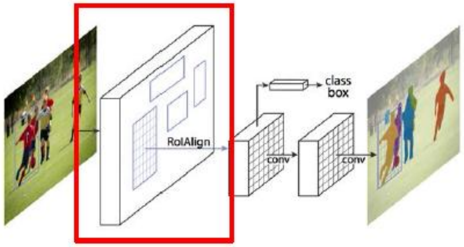
1. feature extractor



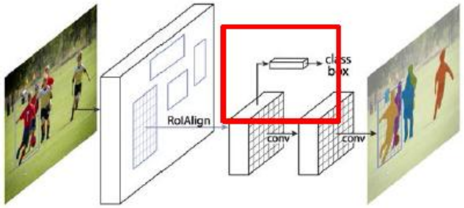
3. ROIAlign



5. segmentation branch



2. region proposal



4. detection branch

Presentation

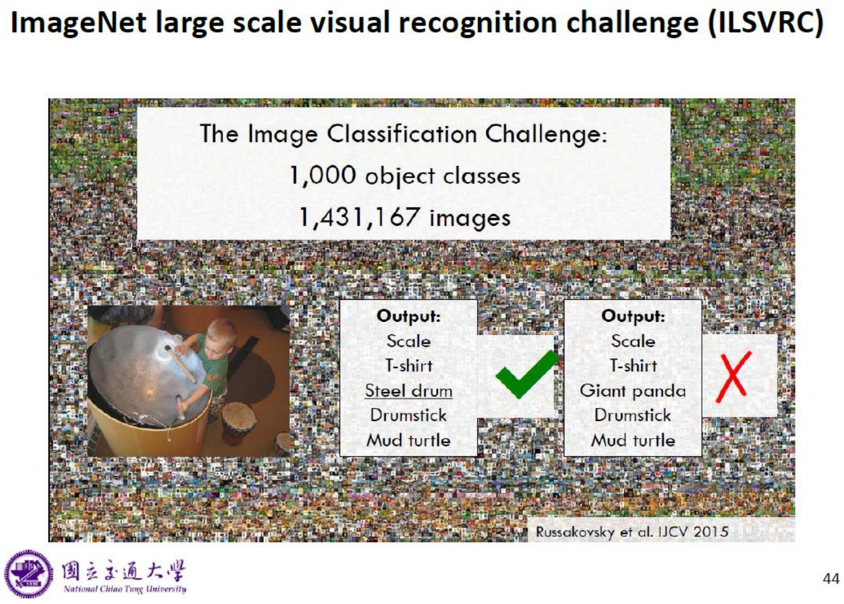
- Your presentation
  - Introduction
  - Related work
  - Proposed approach
  - Experimental results
  - Conclusions

Experiment results

- Dataset(s) and metric(s) for evaluation
- Comparison with state-of-the-arts
- Ablation studies

Experiment results

- Dataset(s) and metric(s) for evaluation
- Comparison with state-of-the-arts
- Ablation studies



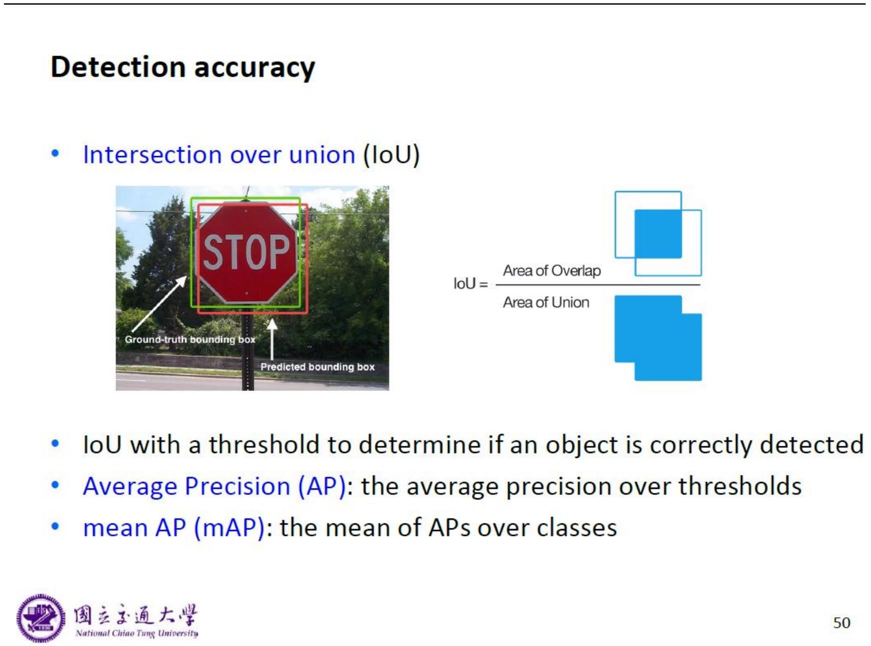
ImageNet large scale visual recognition challenge (ILSVRC)

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Russakovsky et al. IJCV 2015

國立交通大學 National Chiao Tung University

44



Detection accuracy

- Intersection over union (IoU)

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

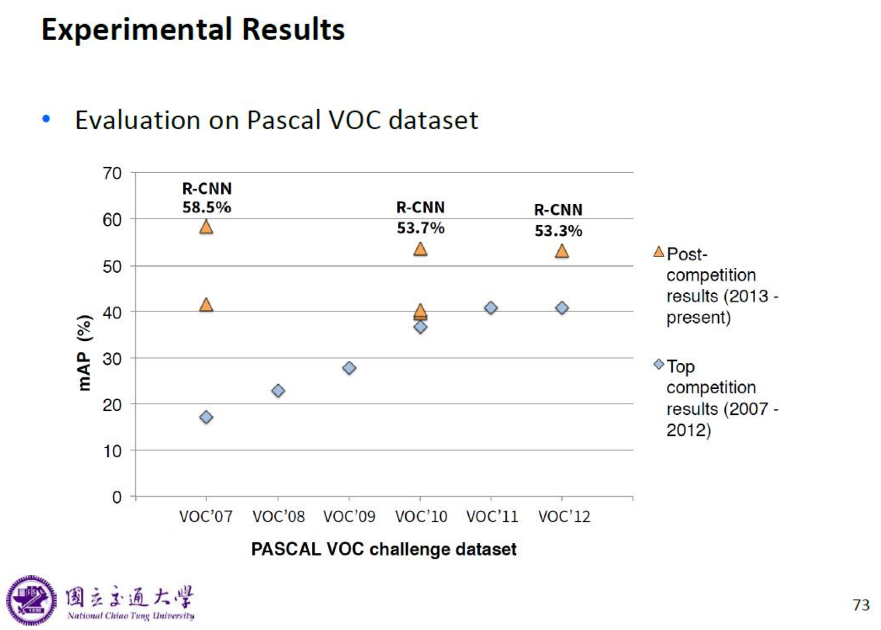
- IoU with a threshold to determine if an object is correctly detected
- Average Precision (AP): the average precision over thresholds
- mean AP (mAP): the mean of APs over classes

國立交通大學 National Chiao Tung University

50

Experiment results

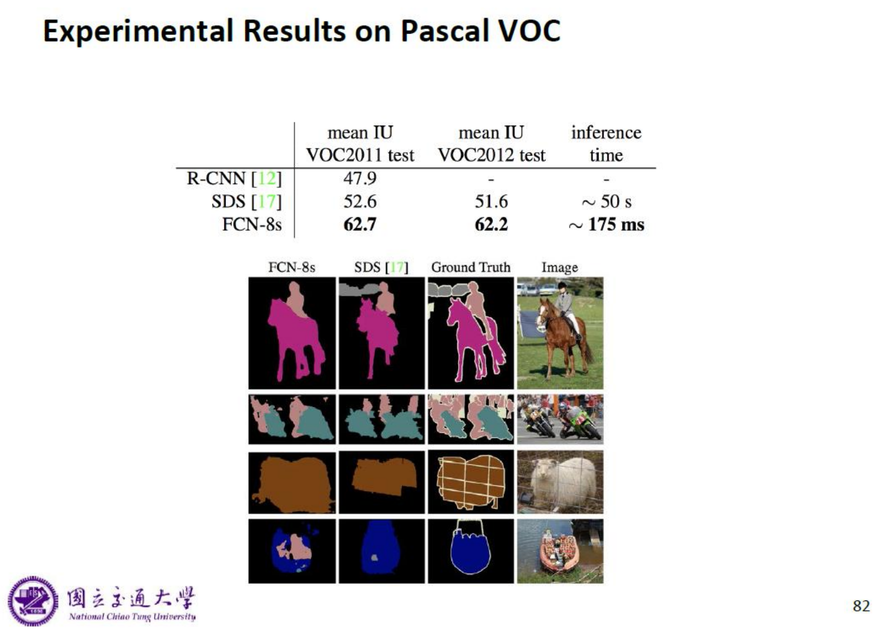
- Dataset(s) and metric(s) for evaluation
- Comparison with state-of-the-arts
- Ablation studies



Experimental Results

- Evaluation on Pascal VOC dataset

國立交通大學
National Chiao Tung University



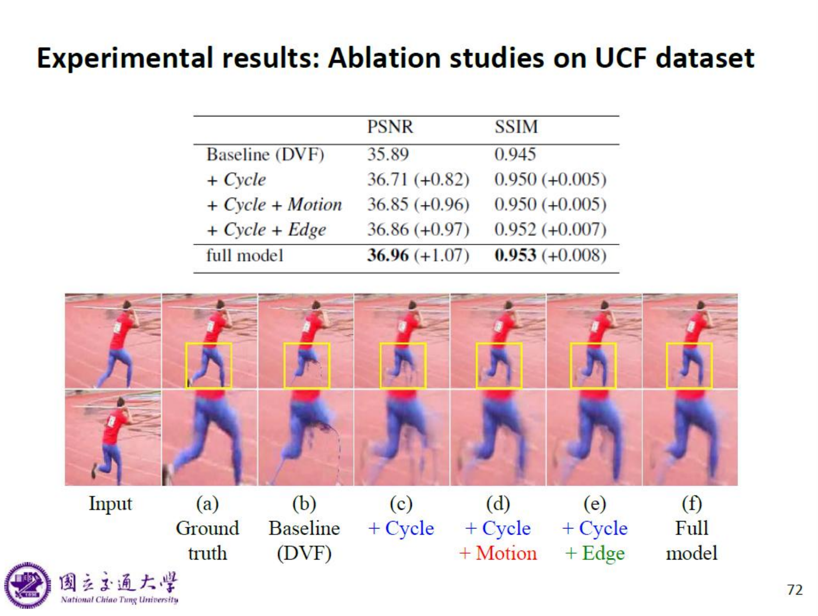
Experimental Results on Pascal VOC

	mean IU VOC2011 test	mean IU VOC2012 test	inference time
R-CNN [12]	47.9	-	-
SDS [17]	52.6	51.6	~ 50 s
FCN-8s	**62.7**	**62.2**	**~ 175 ms**

國立交通大學
National Chiao Tung University

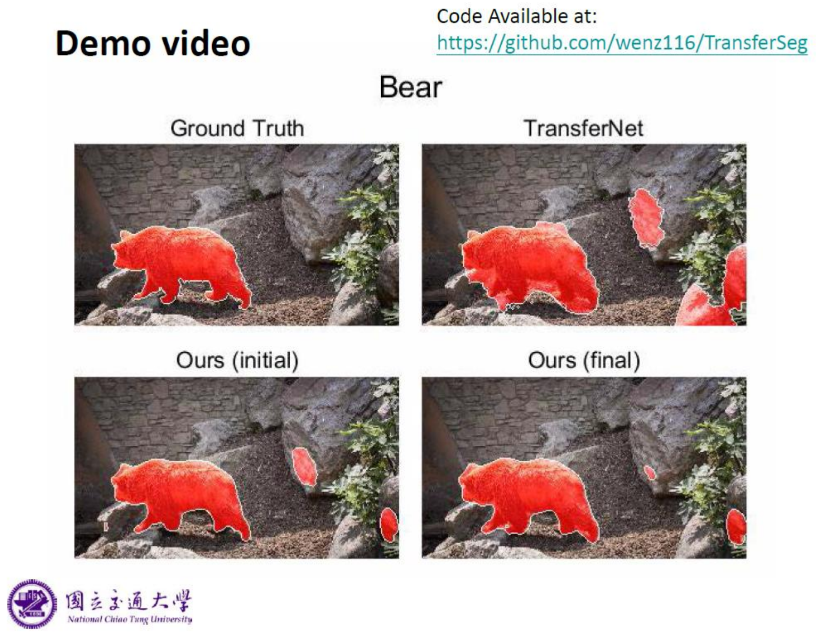
Experiment results

- Dataset(s) and metric(s) for evaluation
- Comparison with state-of-the-arts
- Ablation studies



Experimental results: Ablation studies on UCF dataset

	PSNR	SSIM
Baseline (DVF)	35.89	0.945
+ *Cycle*	36.71 (+0.82)	0.950 (+0.005)
+ *Cycle* + *Motion*	36.85 (+0.96)	0.950 (+0.005)
+ *Cycle* + *Edge*	36.86 (+0.97)	0.952 (+0.007)
full model	**36.96** (+1.07)	**0.953** (+0.008)



Demo video

Code Available at:
https://github.com/wenz116/TransferSeg

Presentation

- Your presentation
  - ➢ Introduction
  - ➢ Related work
  - ➢ Proposed approach
  - ➢ Experimental results
  - ➢ Conclusions

Conclusions

- Summarize your work
- Summarize what you learned/found in the final project

Presentation & Reports & Code

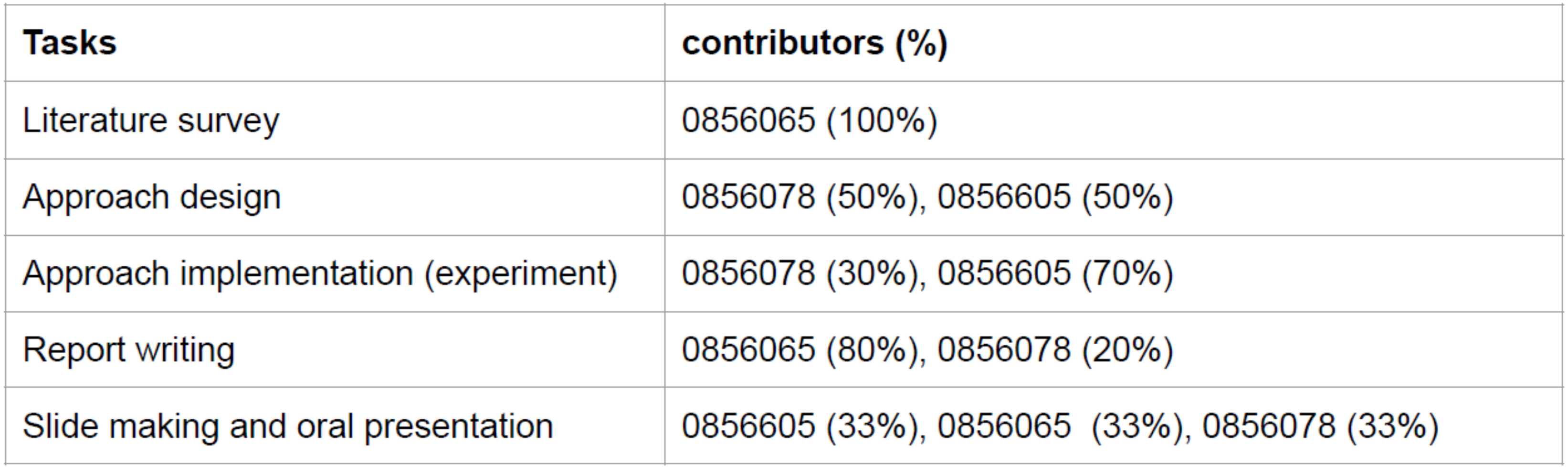
- Your presentation/reports should include
  - GitHub/ GitLab link of your code
  - Introduction
  - Related work
  - Proposed approach
  - Experiment results
  - Conclusions
- Meeting all aforementioned requirements gets 80% of the scores for this part

Presentation & Reports & Code

- Your presentation/reports should include
  - The link to your code
  - Introduction: How you advance this field/topic?
  - Related work: What are the advantages of your method over all existing methods?
  - Proposed approach: How to design your approach to achieve the advantages you claim? Is your method technically sound?
  - Experiment results: Does your approach achieve SOTA results? Do ablation studies support the claimed advantages?
  - Conclusions: Any new and insightful findings or conclusions?
- Meeting all aforementioned requirements gets 80% of the scores for this part

Team member contribution

- Specify the contribution made by each team member to each of the following five tasks in the report:



Tasks	contributors (%)
Literature survey	0856065 (100%)
Approach design	0856078 (50%), 0856605 (50%)
Approach implementation (experiment)	0856078 (30%), 0856605 (70%)
Report writing	0856065 (80%), 0856078 (20%)
Slide making and oral presentation	0856605 (33%), 0856065 (33%), 0856078 (33%)

Thank You for Your Attention!

Yen-Yu Lin (林彥宇)

Email: lin@cs.nycu.edu.tw

URL: https://www.cs.nycu.edu.tw/members/detail/lin