









ICDM 2010




QMAS: Querying, Mining And Summarization of Multi-modal Databases

Robson Cordeiro[#], Fan Guo⁺, Donna Haverkamp^{*},
 James Horne^{*}, Ellen Hughes^{*}, Gunhee Kim⁺,
 Agma Traina[#], Caetano Traina Jr. [#] and Christos Faloutsos⁺

#

+

*





Microsoft
Research



2

Agenda

- Introduction
- QMAS
 - Feature Extraction
 - Low-labor Labeling
- Experiments
- Conclusions

3

Introduction

Problem 1: low-labor labeling

- **Given:**
 - Images

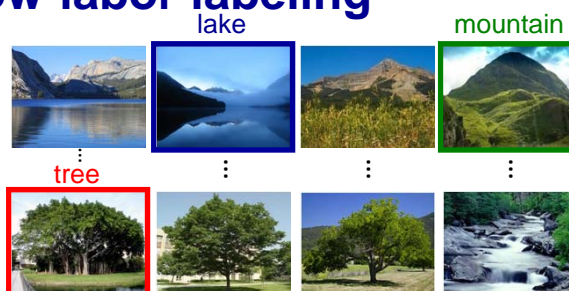


4

Introduction

Problem 1: low-labor labeling

- **Given:**
 - Images
 - A few labels



5

Introduction

Problem 1: low-labor labeling

- **Given:**

- Images
- A few labels



- **Find:**

- Labels for the remaining majority



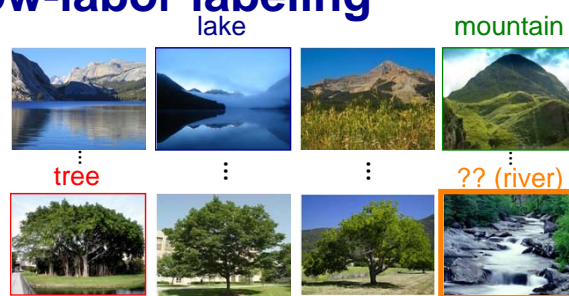
6

Introduction

Problem 1: low-labor labeling

- **Given:**

- Images
- A few labels



- **Find:**




- Labels for the remaining majority
- Images that may need non-predicted labels



7

Introduction

- **Our main Goals**
 - G1 – Speed**
 - G2 – Quality**
 - G3 – Low-labor labeling**







8

Introduction



- **Application on Satellite Imagery**

Annapolis, USA + 4 labels*





* geoeye.com

9

Agenda

- Introduction
- QMAS
 - Feature Extraction
 - Low-labor Labeling
- Experiments
- Conclusions

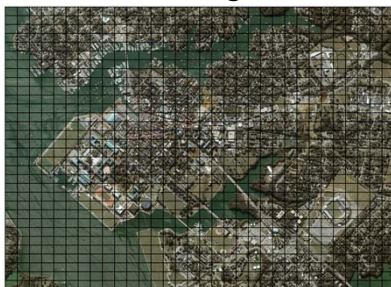


10

QMAS – Feature Extraction

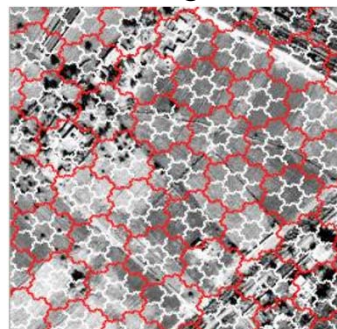
- **Step 1: Tiles**

Rectangles



or

Hexagons

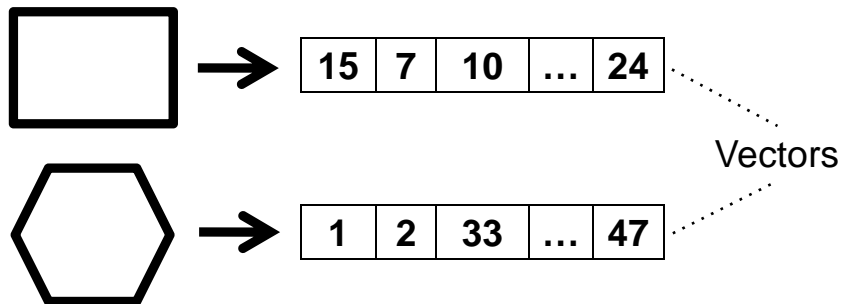


QMAS – Feature Extraction

11

- **Step 2: Features**

- Wavelets, Mean, Variance, Texture, etc.



Agenda

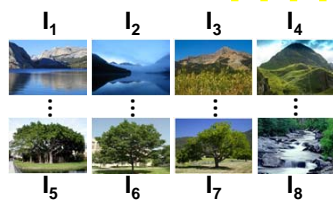
12

- Introduction
- QMAS
 - Feature Extraction
 - Low-labor Labeling
- Experiments
- Conclusions



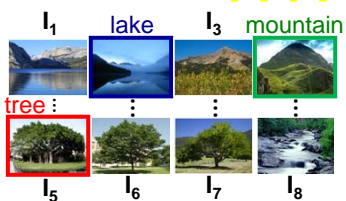
QMAS - Low-labor Labeling

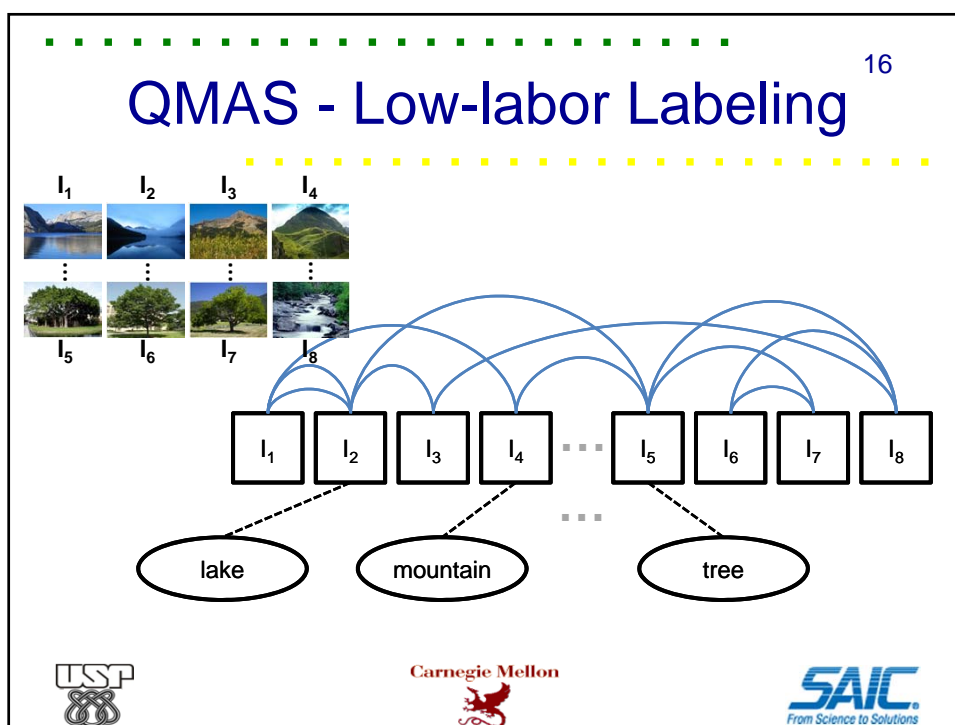
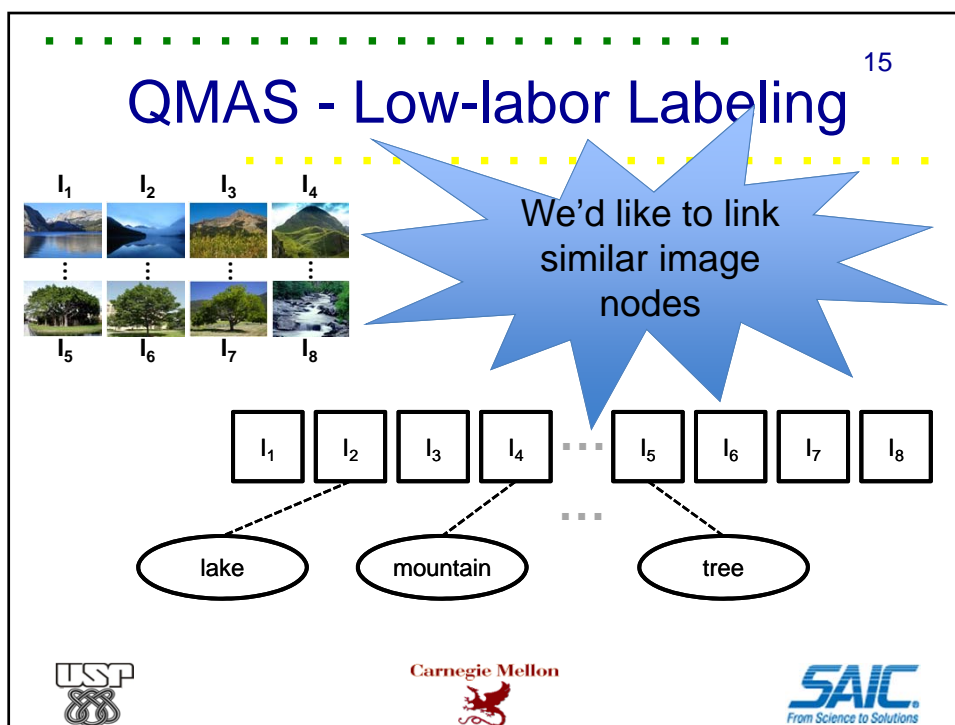
13

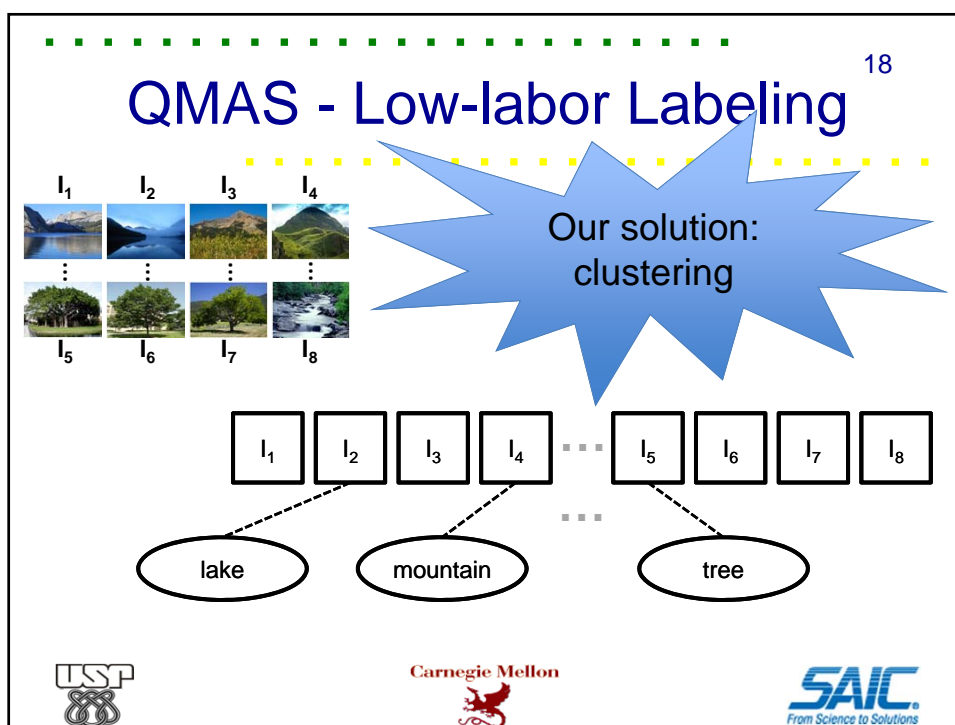
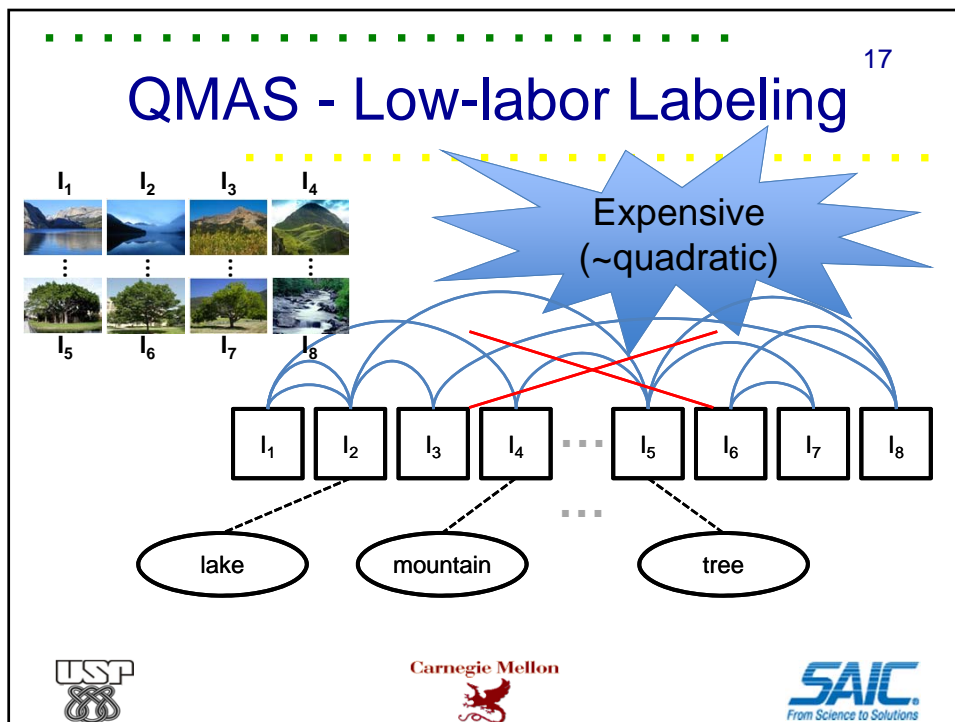


QMAS - Low-labor Labeling

14

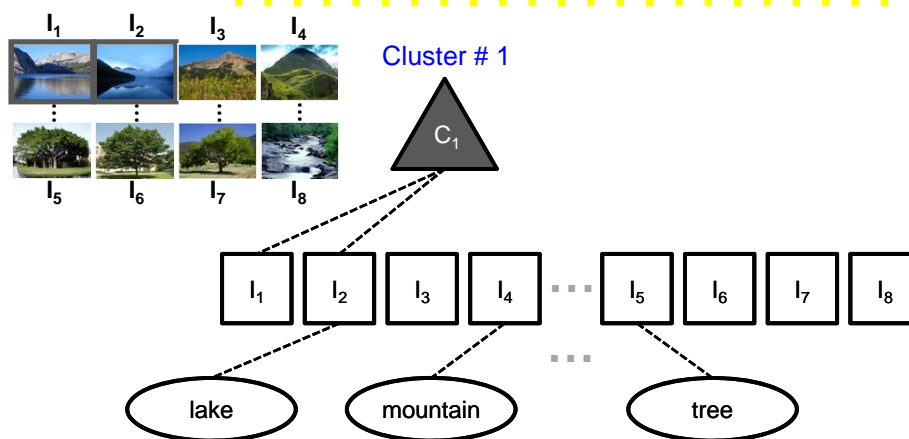






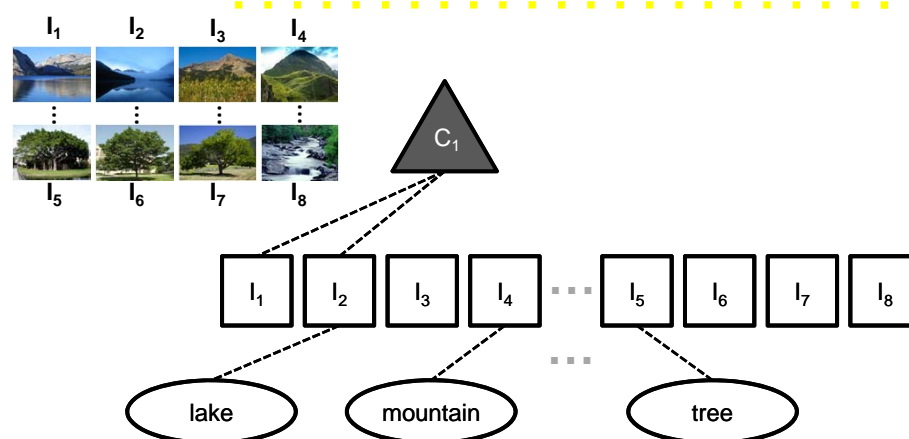
QMAS - Low-labor Labeling

19

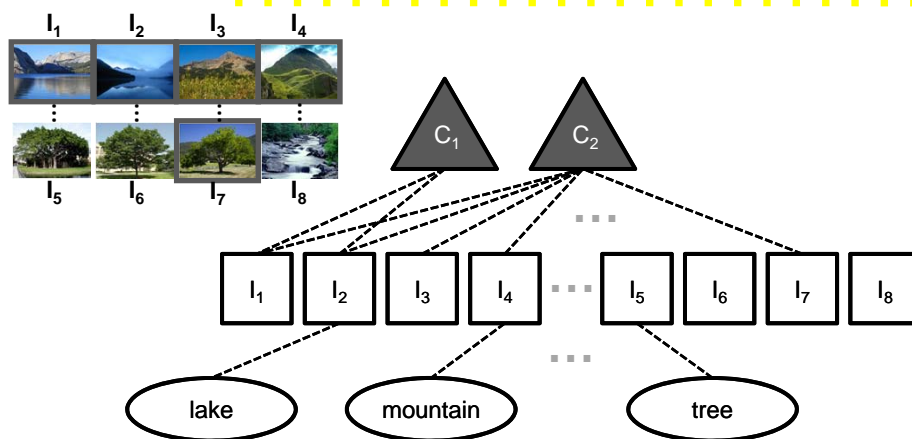


QMAS - Low-labor Labeling

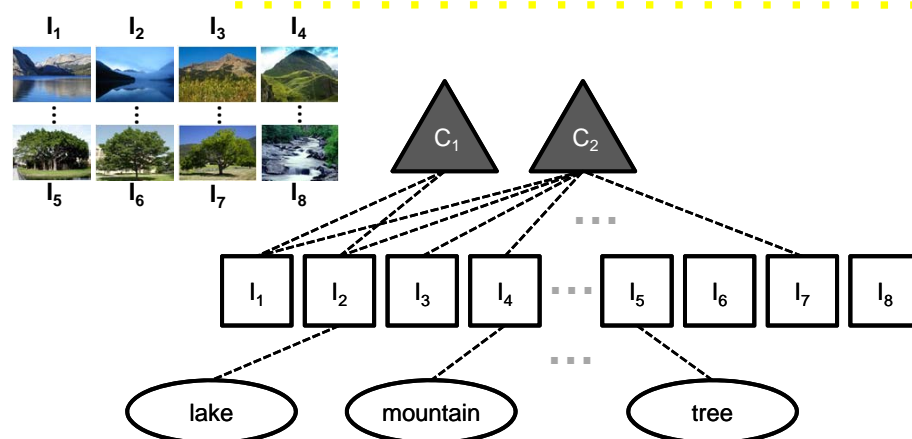
20



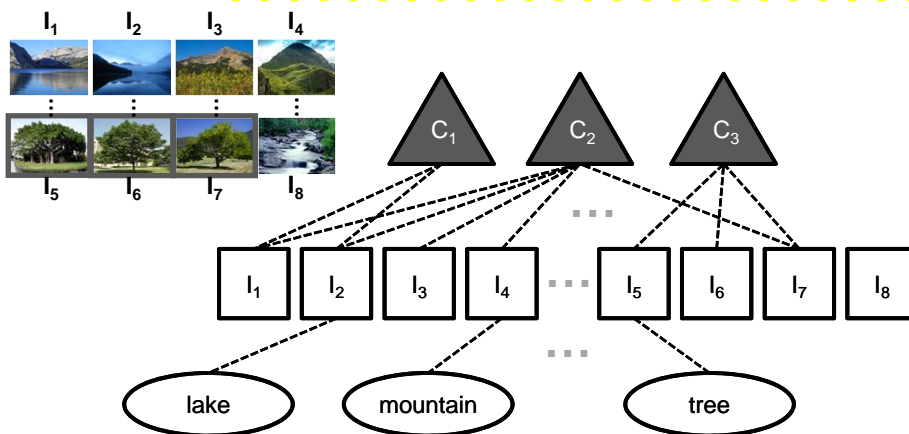
QMAS - Low-labor Labeling ²¹



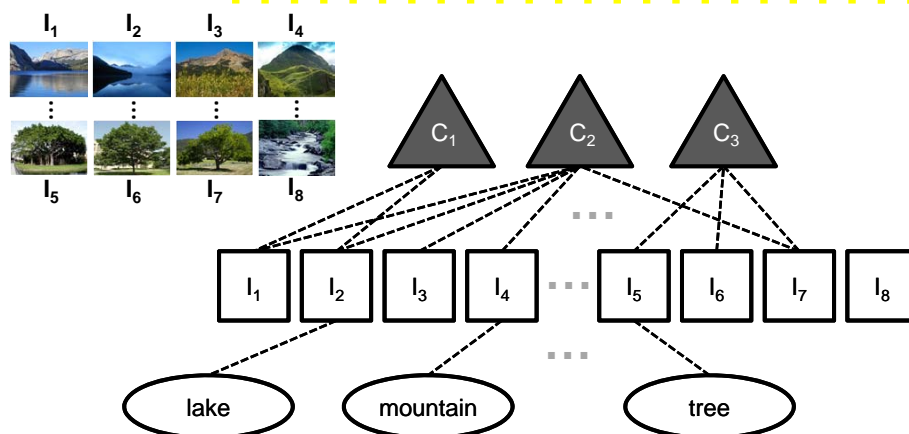
QMAS - Low-labor Labeling ²²



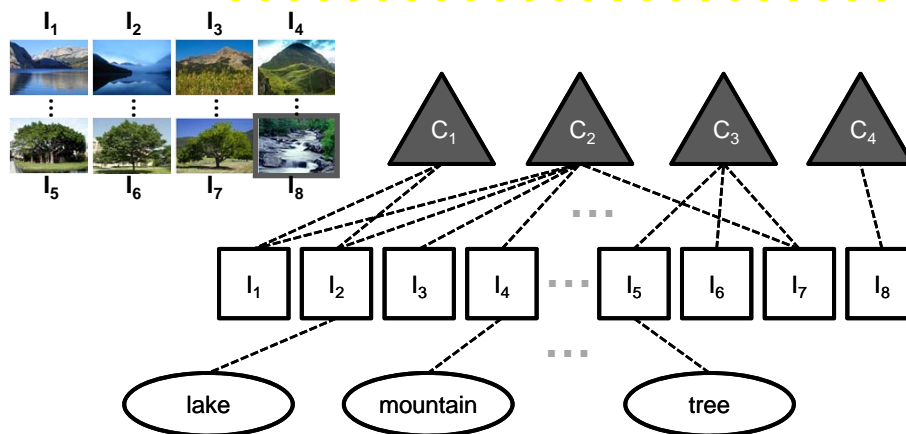
QMAS - Low-labor Labeling ²³



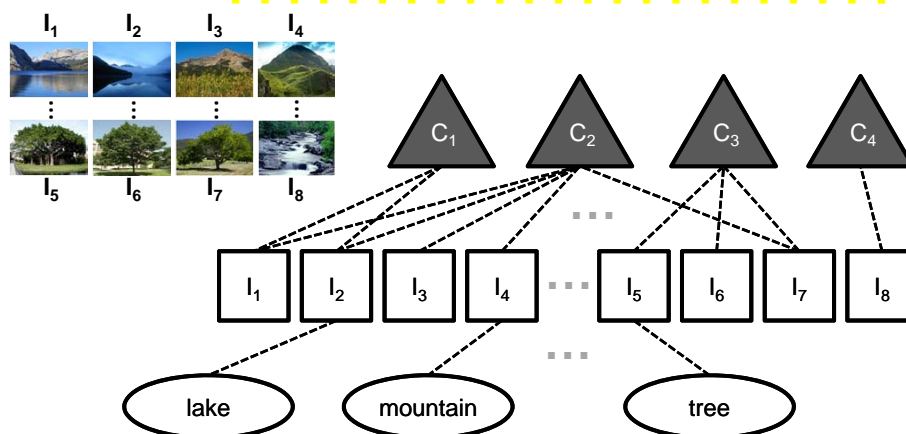
QMAS - Low-labor Labeling ²⁴



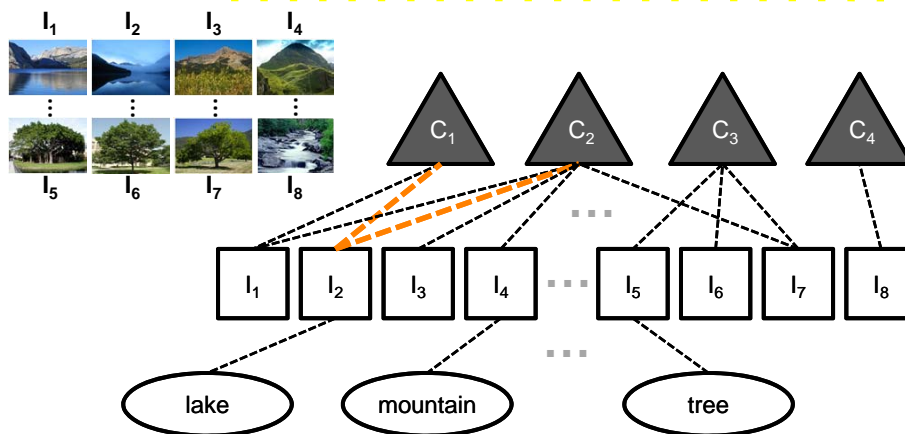
QMAS - Low-labor Labeling ²⁵



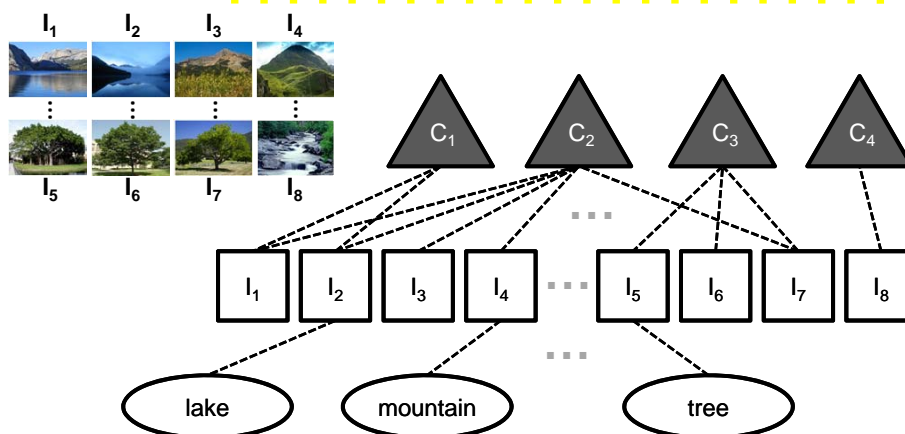
QMAS - Low-labor Labeling ²⁶

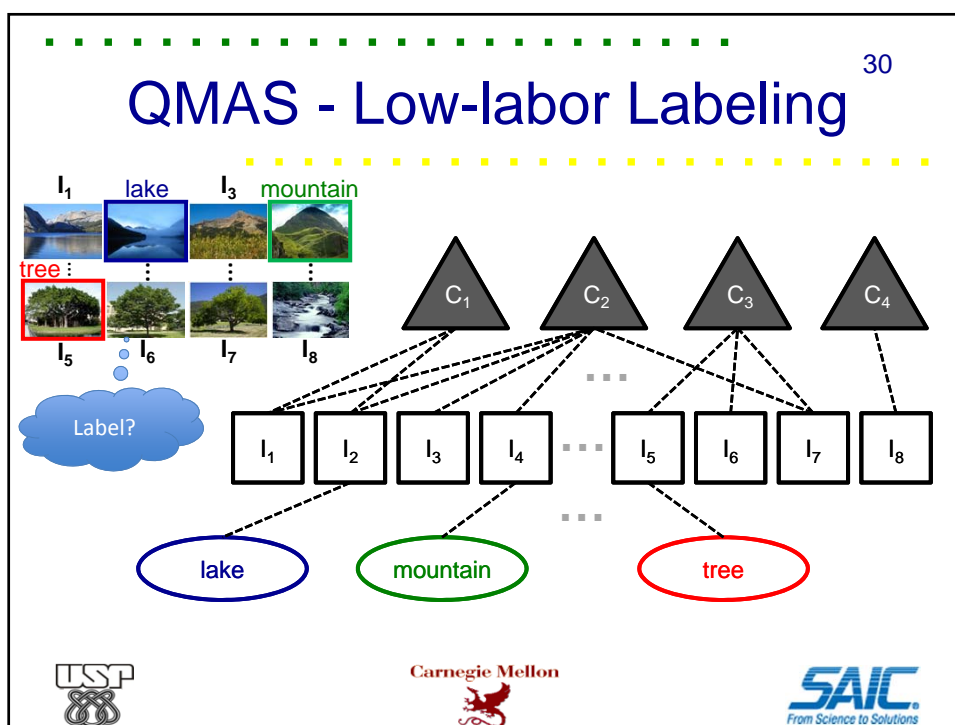
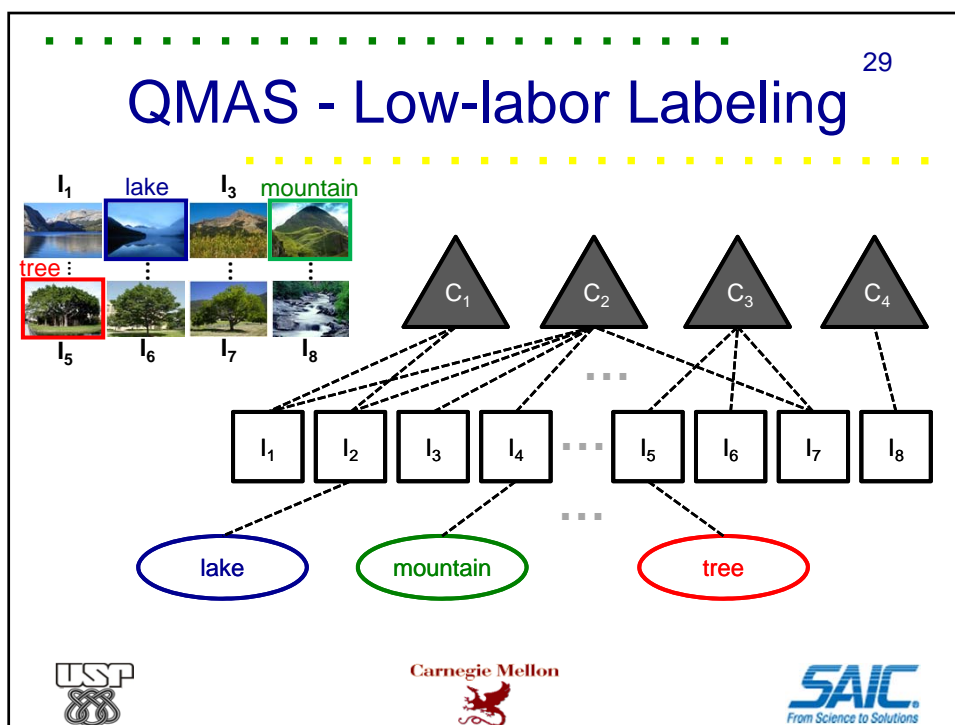


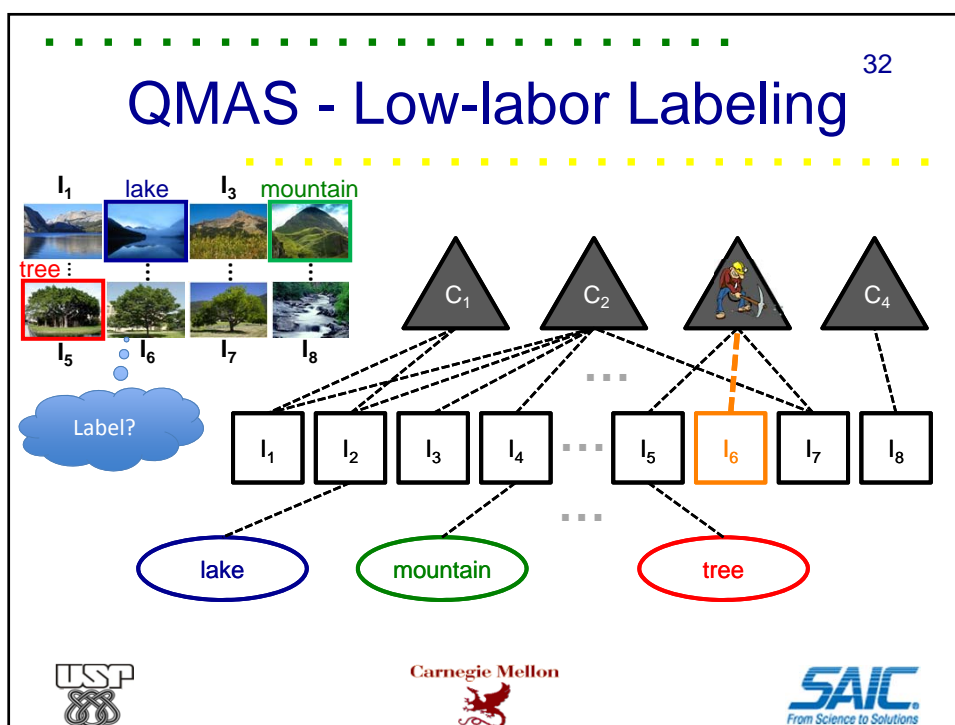
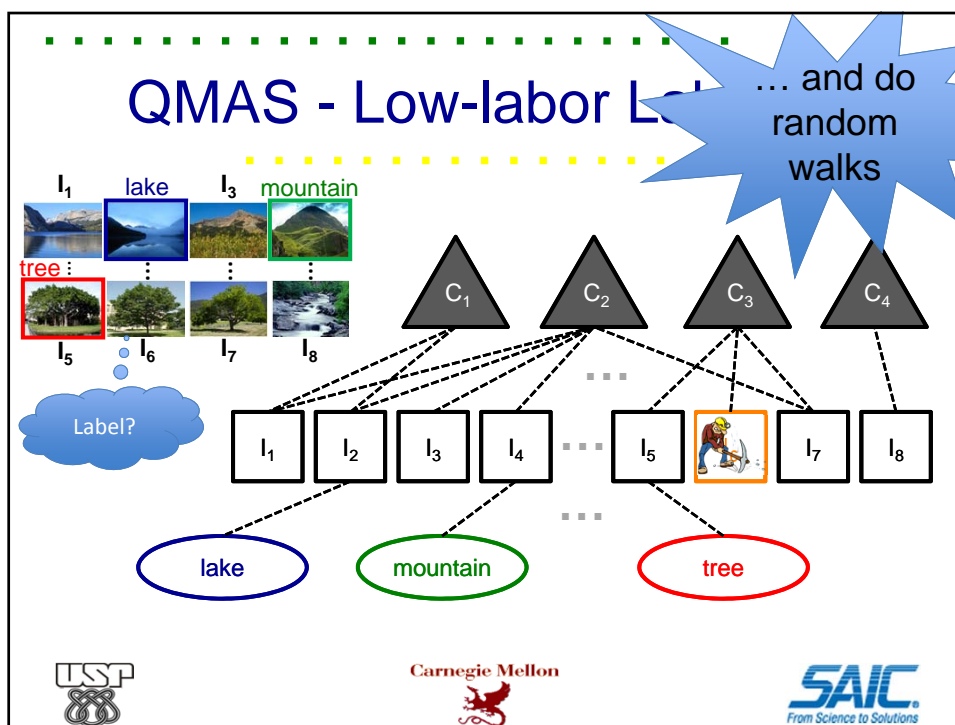
QMAS - Low-labor Labeling ²⁷

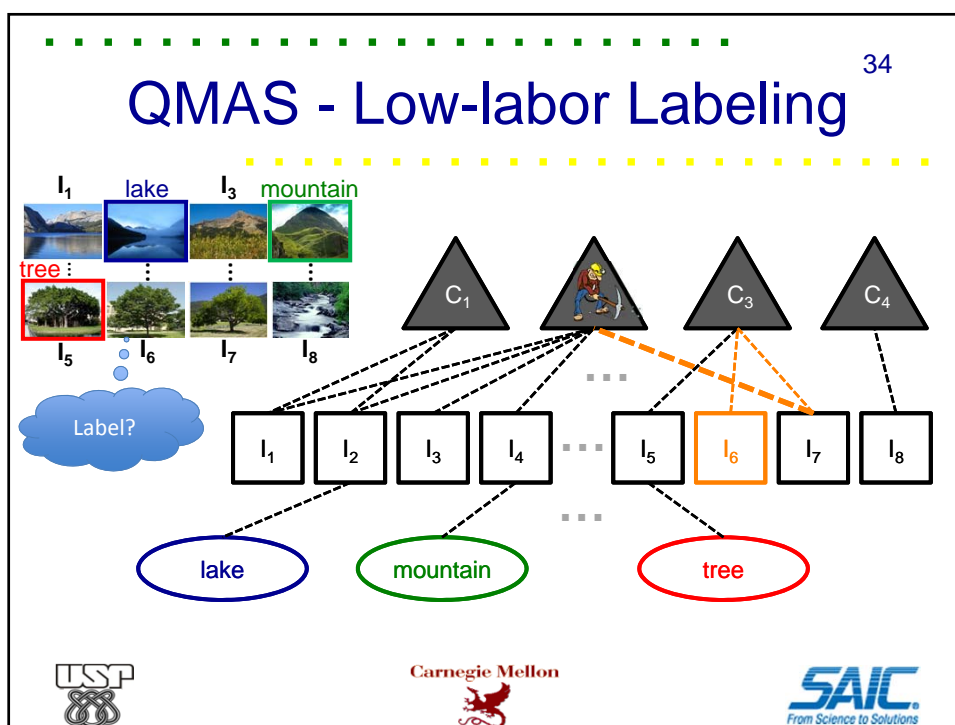
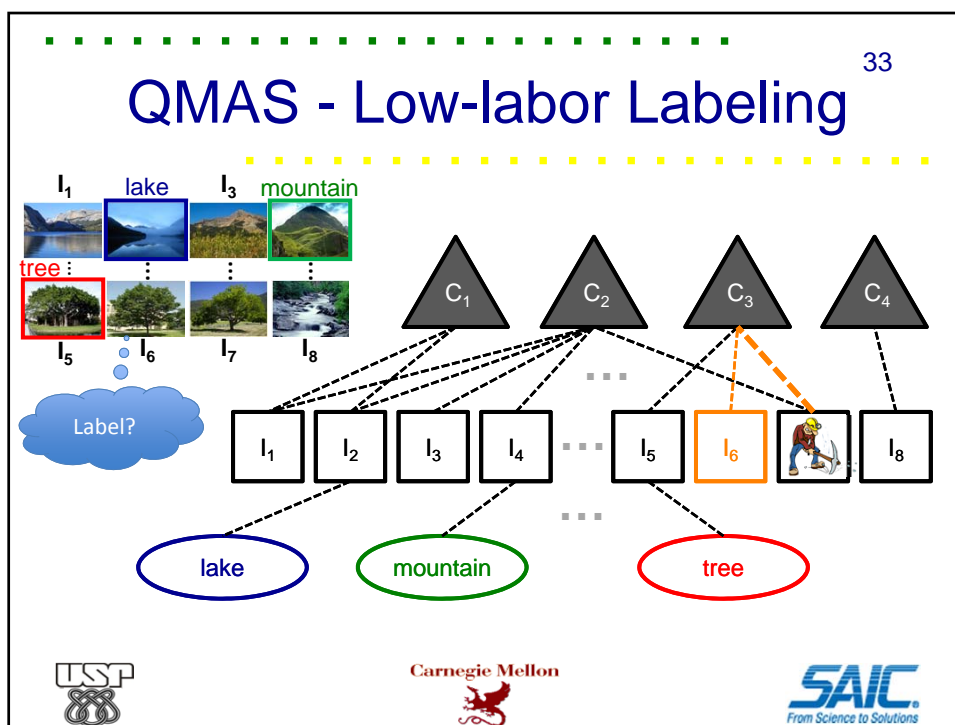


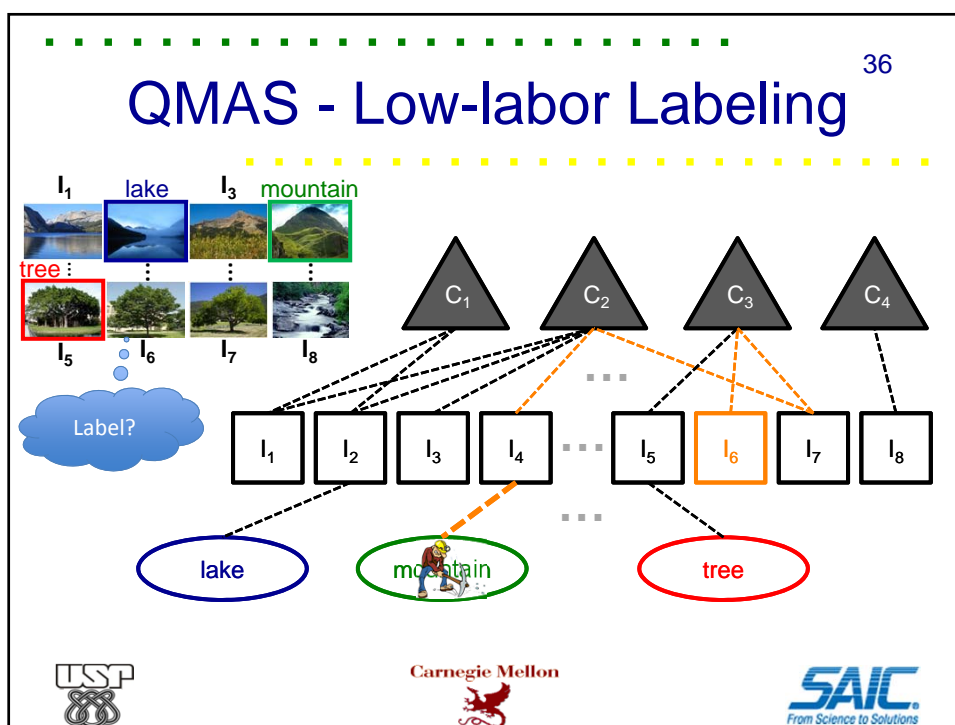
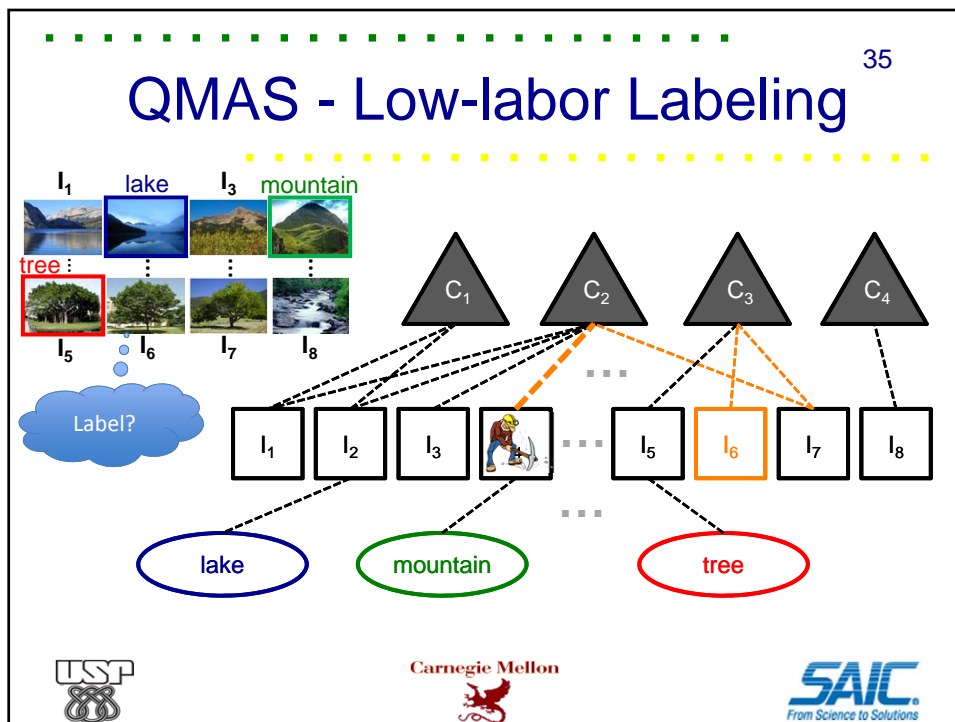
QMAS - Low-labor Labeling ²⁸

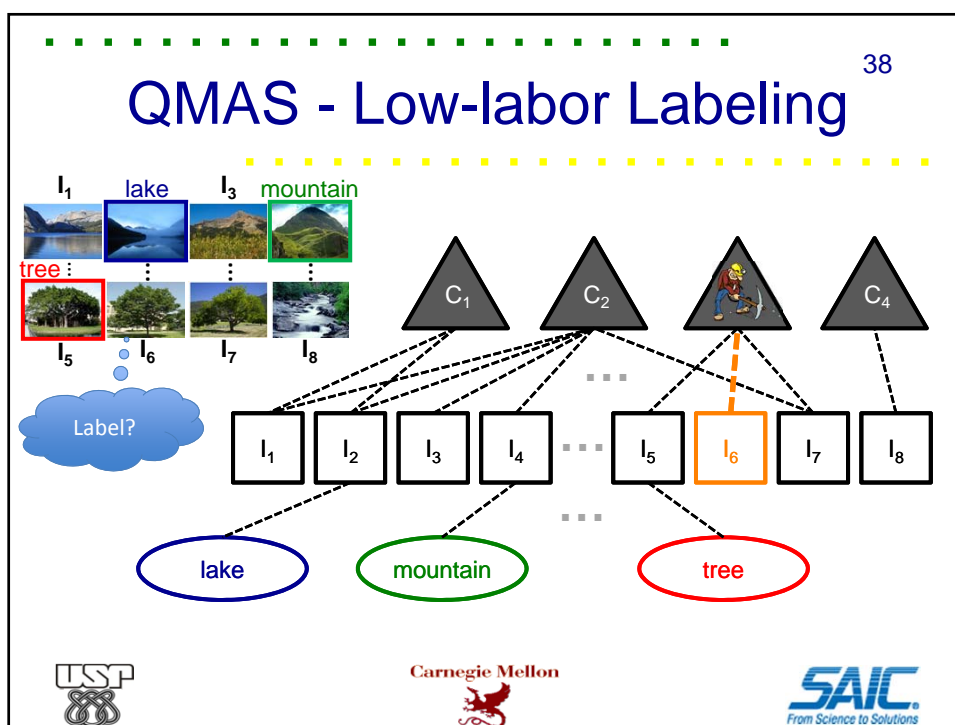
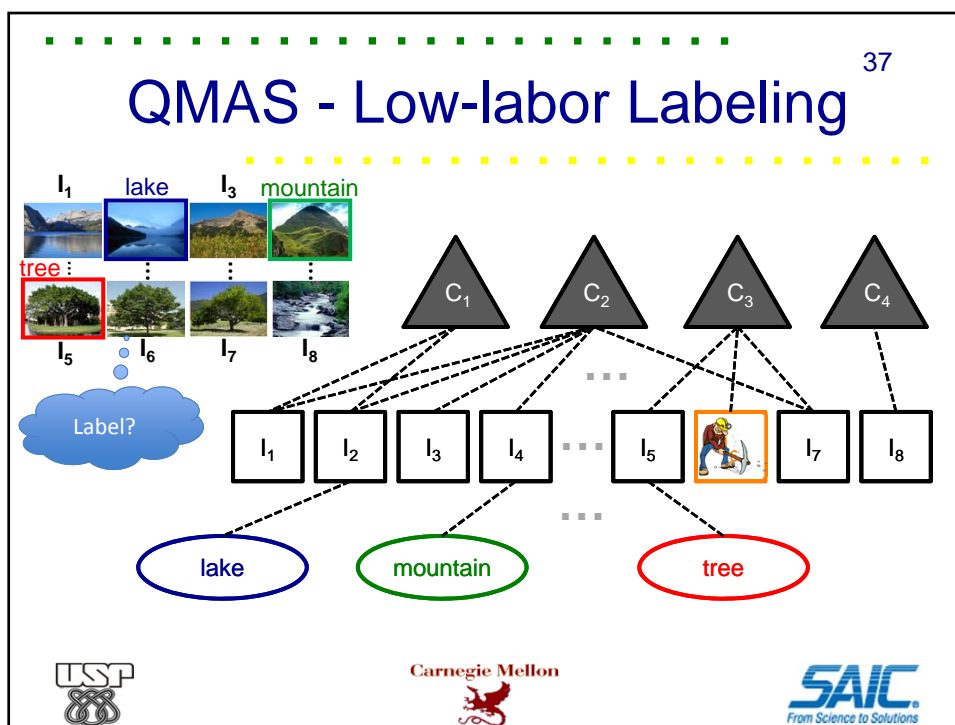


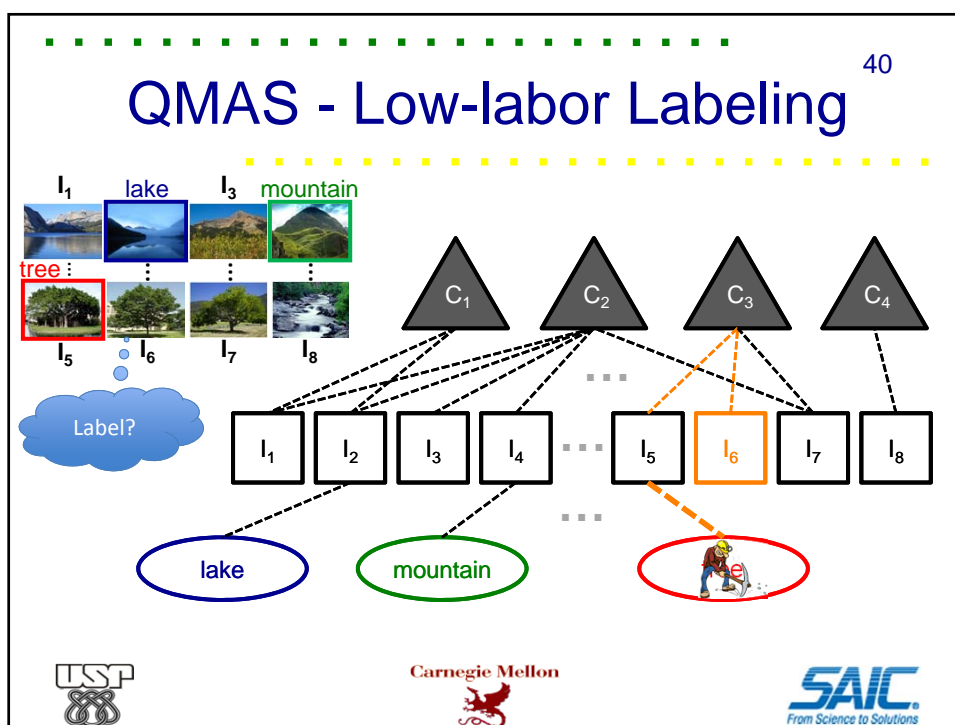
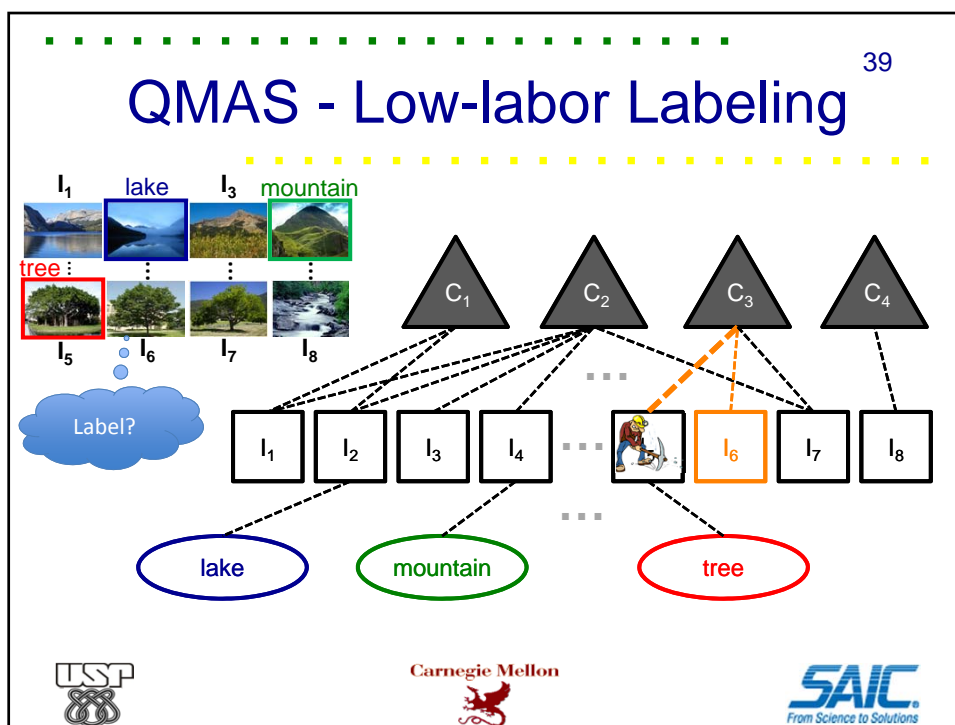


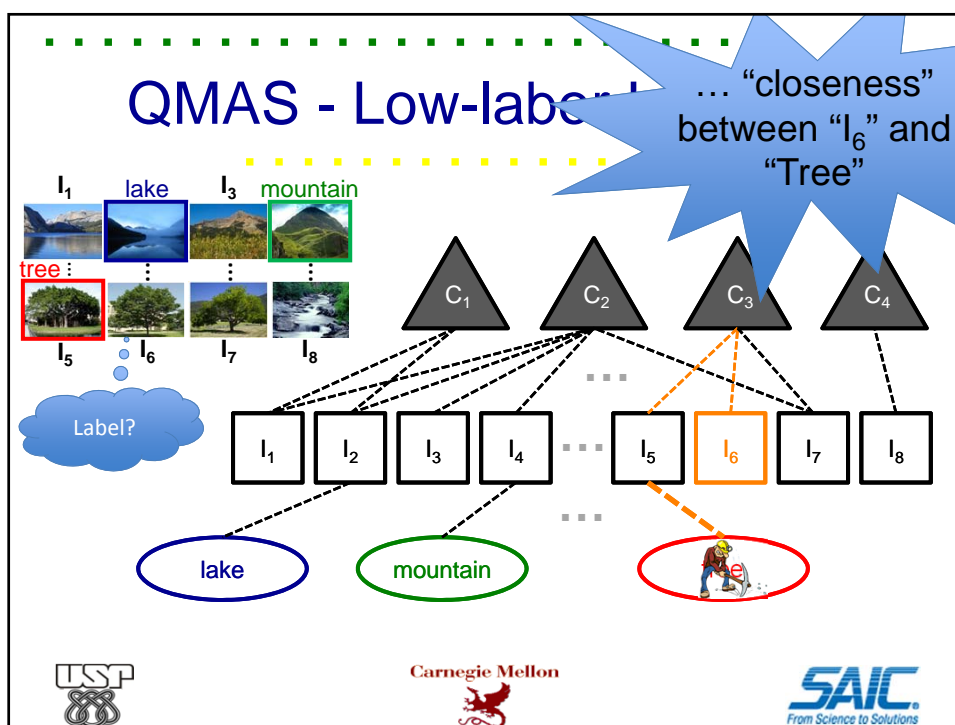
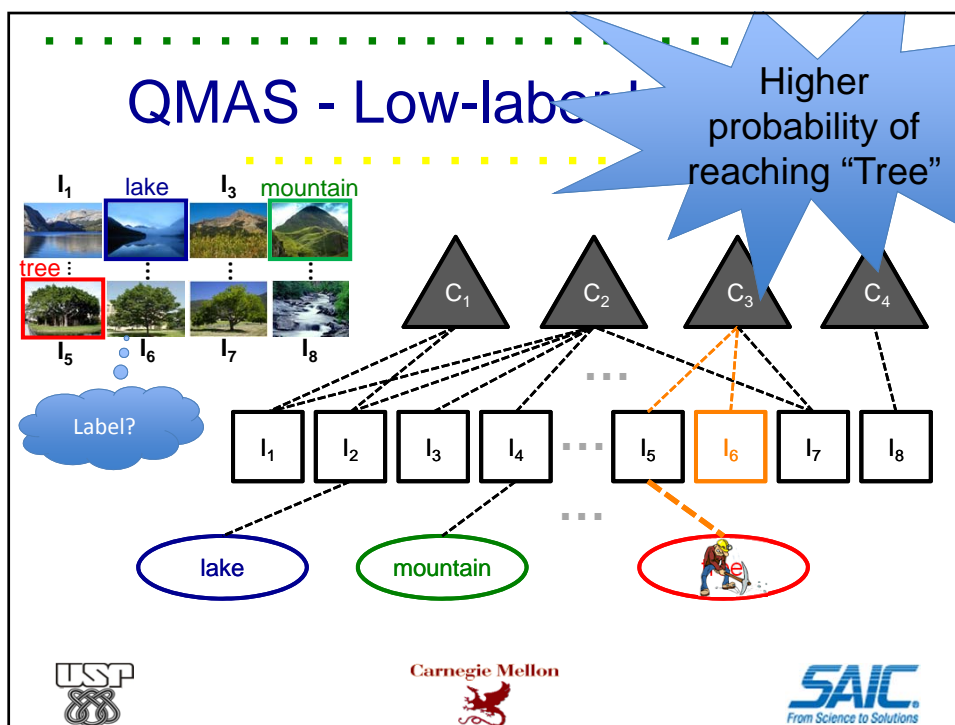


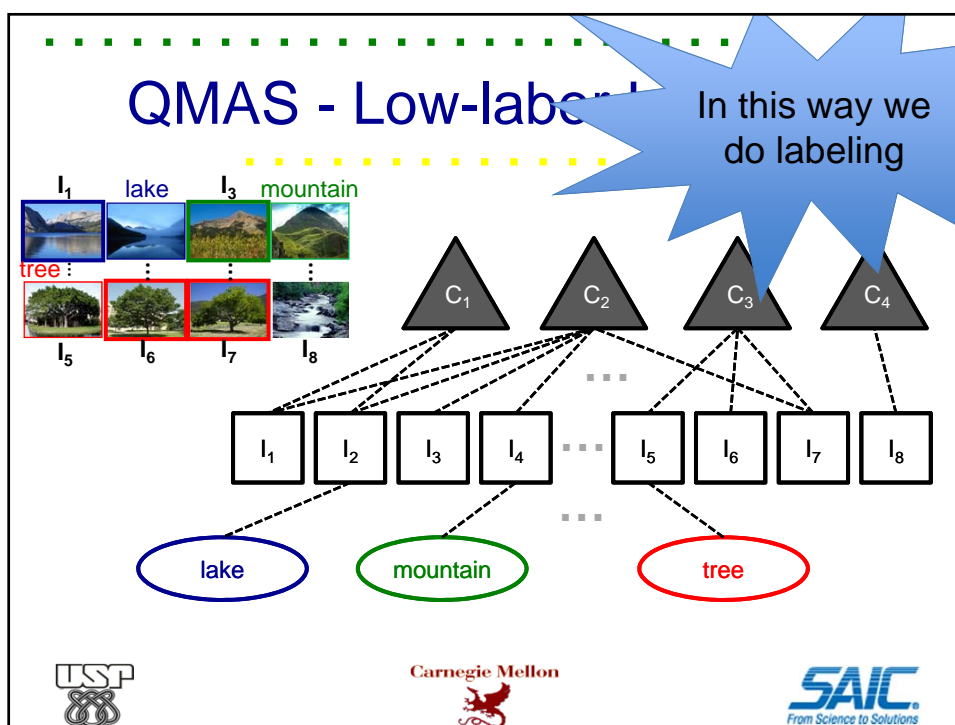
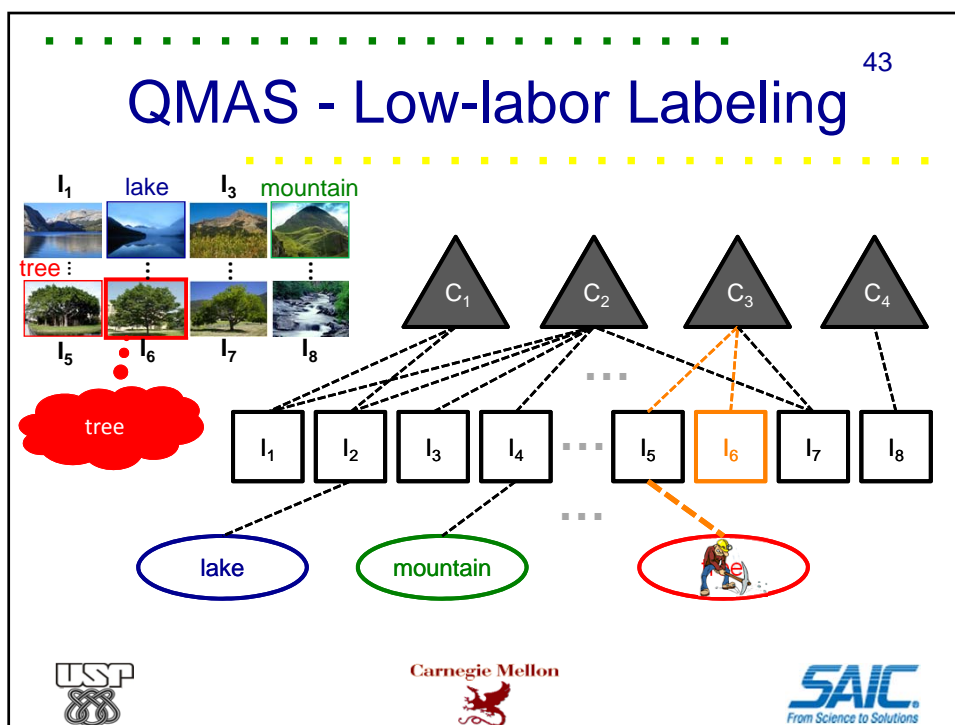


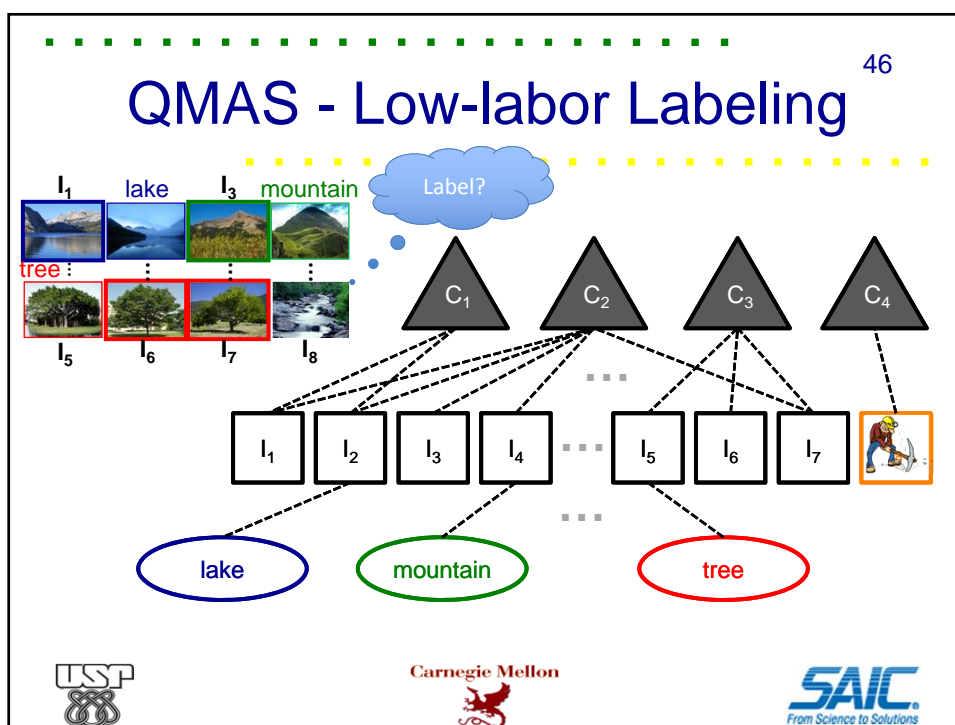
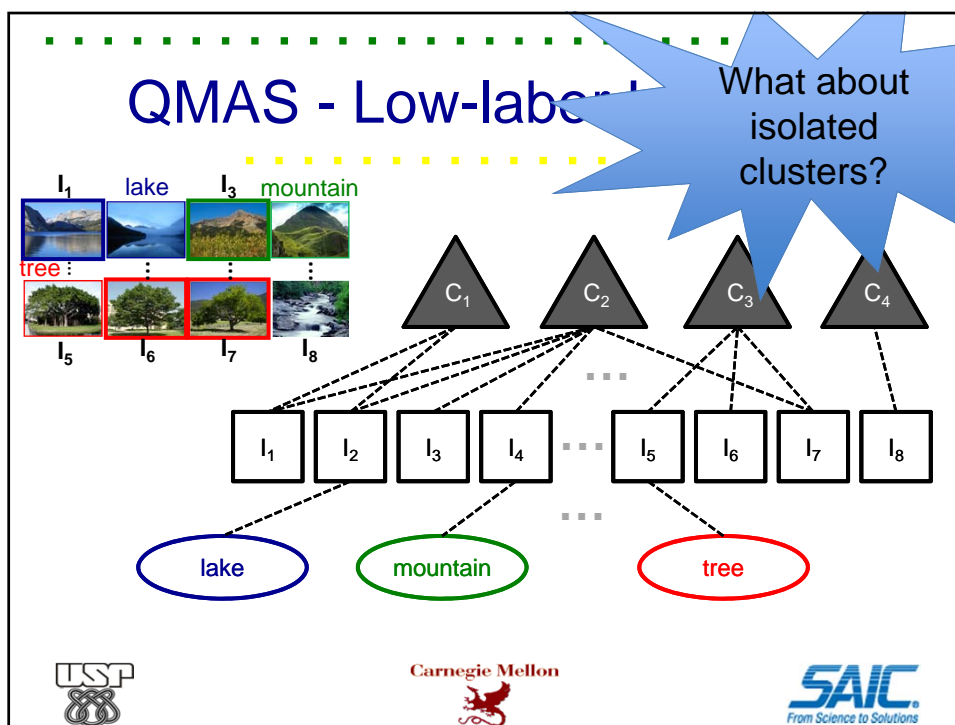


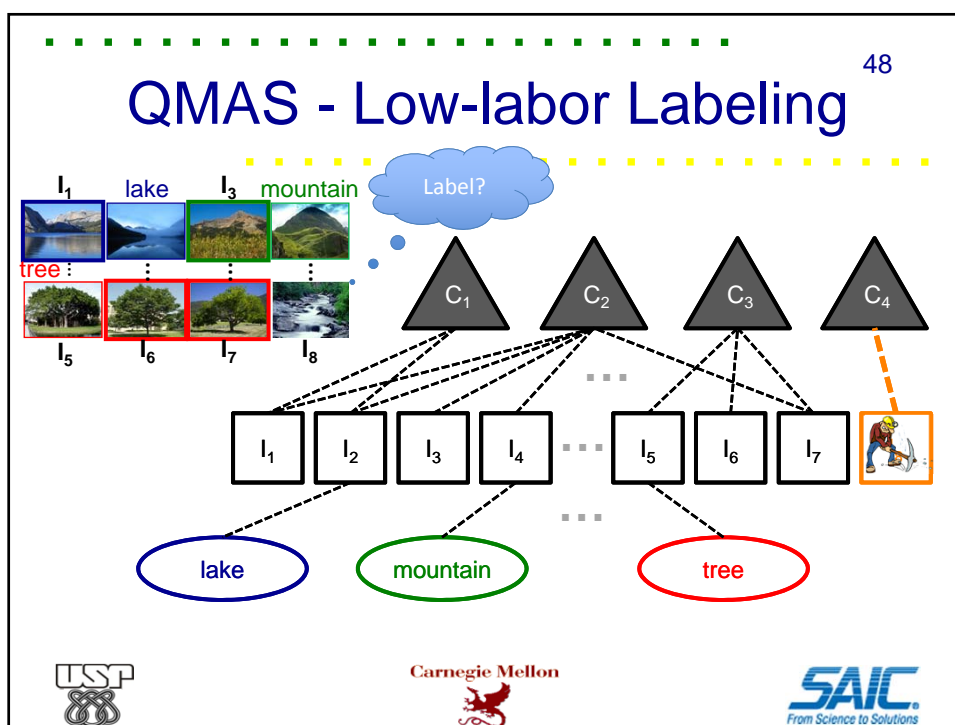
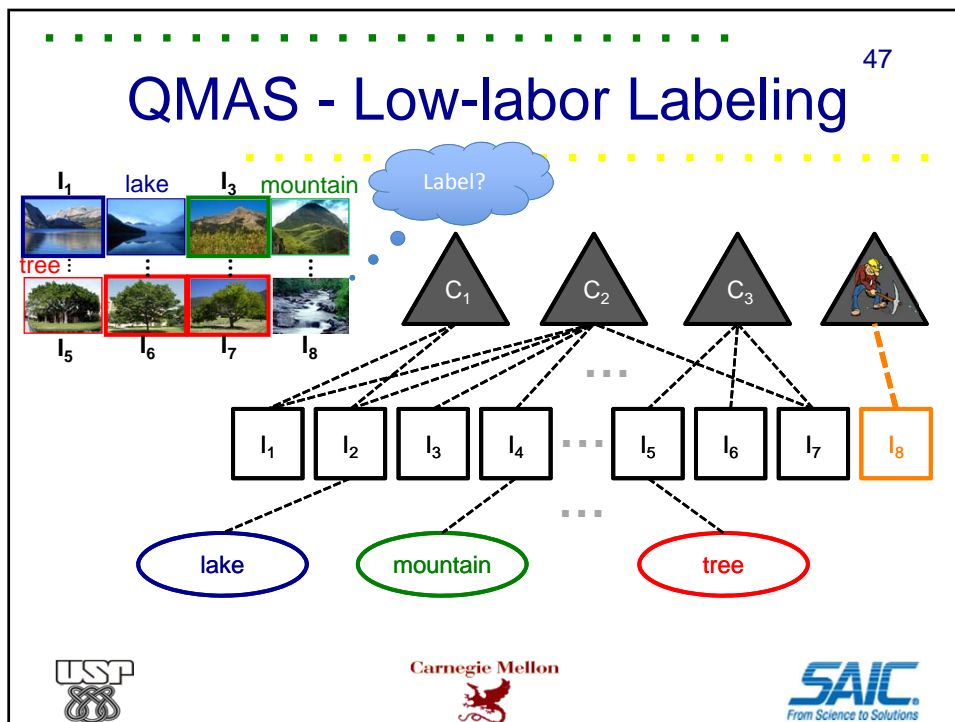


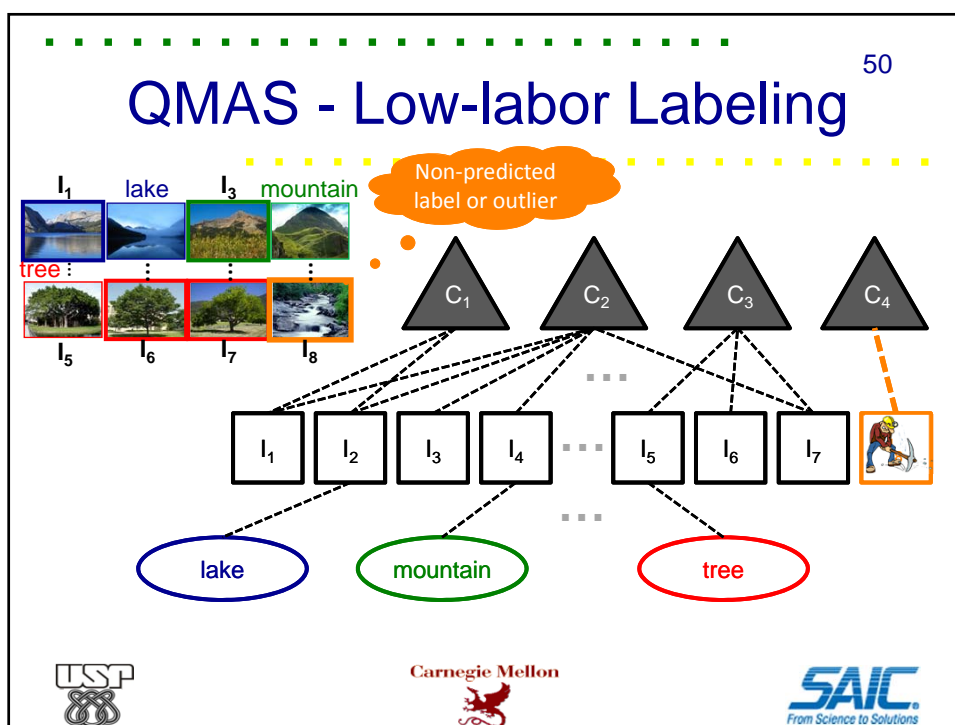
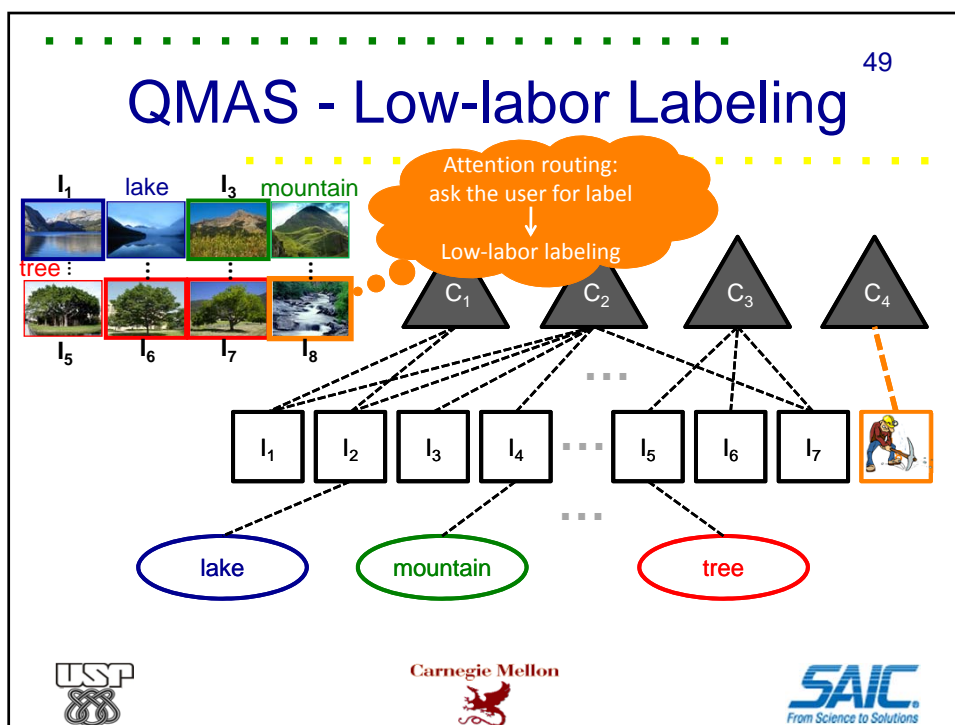












51

Agenda

- Introduction
- QMAS
 - Feature Extraction
 - Low-labor Labeling
- Experiments
- Conclusions

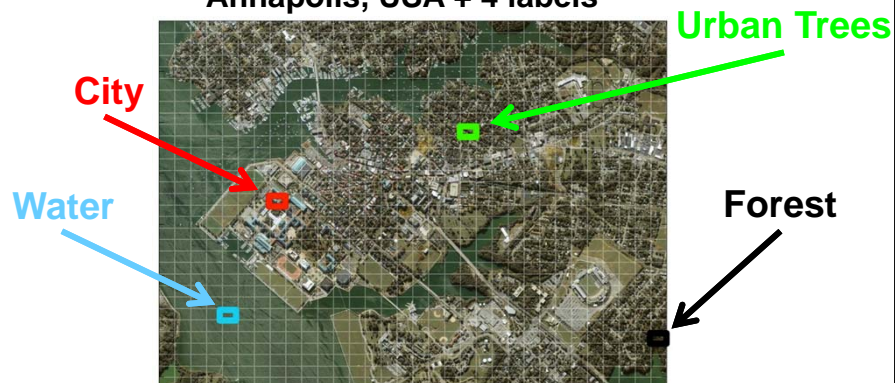


52

Experiments

- Application on Satellite Imagery

Annapolis, USA + 4 labels*



* geoeye.com



53

Experiments

- Public Data (geoeye.com)
 - **GeoEye**: ~17MB, ~14k tiles

Annapolis, USA



Venice, Italy



Athens, Greece



54

Experiments

- Proprietary Data
 - **SAT1.5GB**: ~1.5GB, ~700k tiles
 - **SATLARGE**: ~2.25GB, ~2.5 *million* tiles



Experiments

55

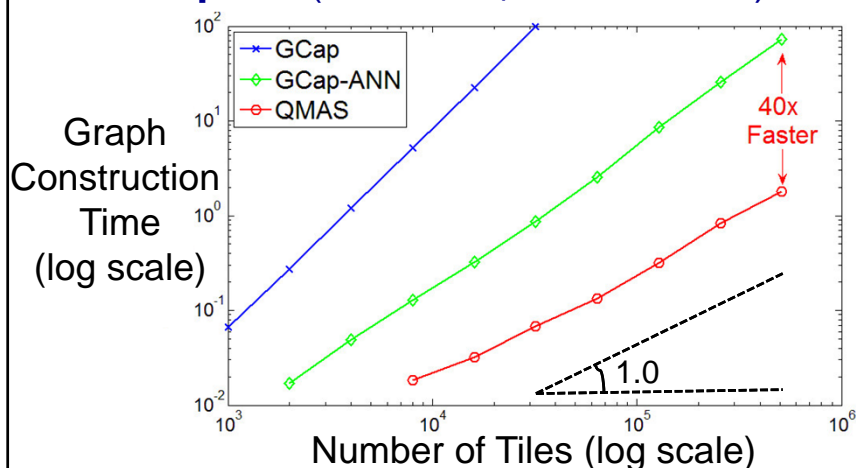
- **System Configuration**
 - Linux + 2.8 GHz core + 4GB RAM
- **Compared Related Works**
 - **GCap** (Pan, J. -Y. et al., CVPRW 2004) with 7NN
 - **GCap-ANN** with 7NN
- **All methods:**
 - RWR: *power iteration method*, $c = 0.15$



Experiments

56

G1 – Speed (random samples of SAT1.5GB)



57

Experiments

G2 - Quality and **G3 - Low-labor labeling**

(GeoEye)



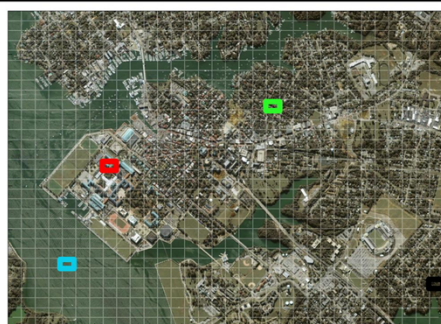
a) Annapolis-USA + 4 labels

58

Experiments

G2 - Quality and **G3 - Low-labor labeling**

(GeoEye)



a) Annapolis-USA + 4 labels

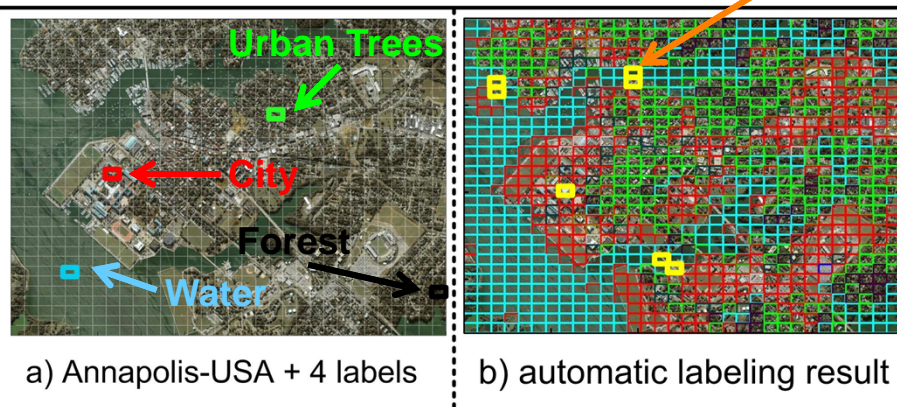
59

Experiments

G2 - Quality and **G3 - Low-labor labeling**

(GeoEye)

Outliers



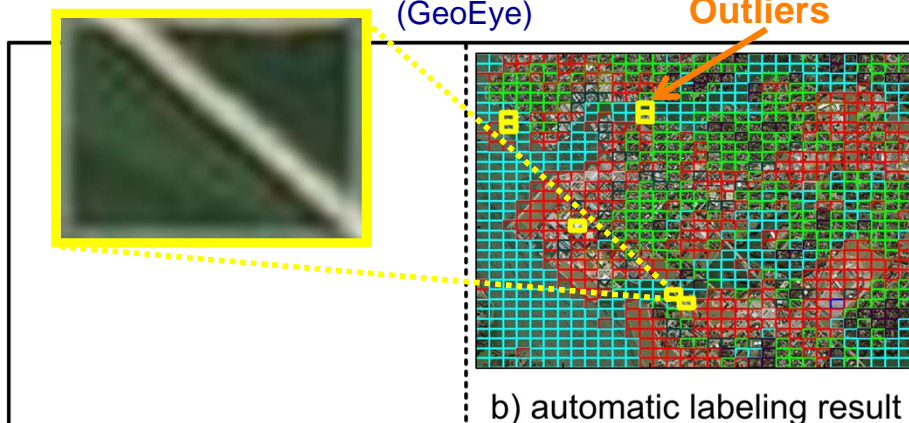
60

Experiments

G2 - Quality and **G3 - Low-labor labeling**

(GeoEye)

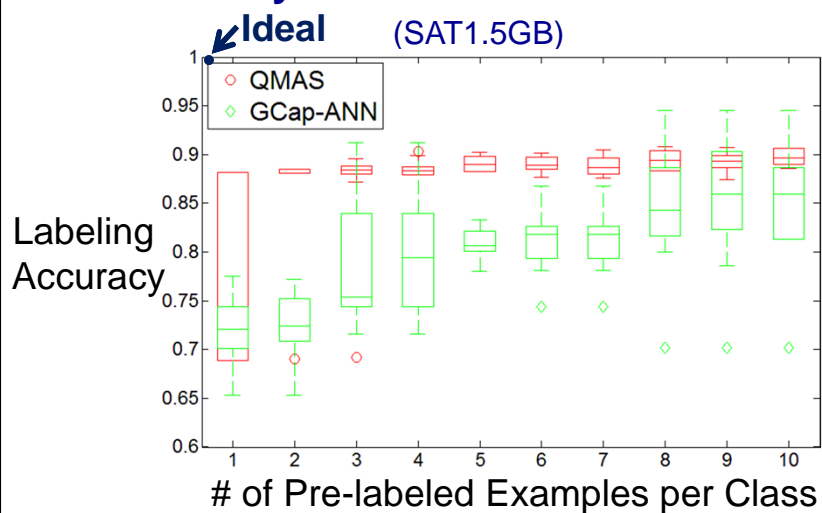
Outliers



61

Experiments

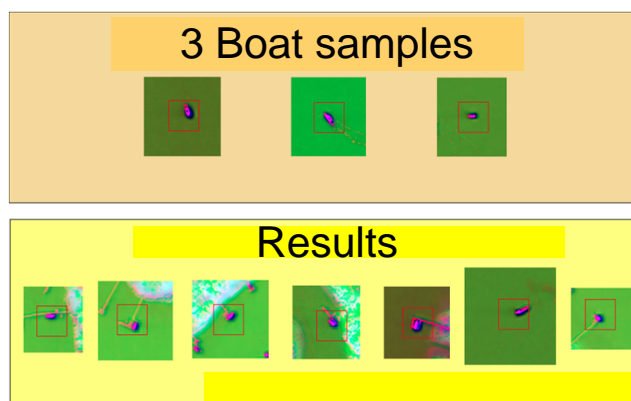
G2 - Quality and G3 - Low-labor labeling



62

Experiments

• Queries (SATLARGE)



63

Agenda

- Introduction
- QMAS
 - Feature Extraction
 - Low-labor Labeling
- Experiments
- Conclusions



64

Conclusions

- ✓ **G1 – Speed**
 - > 40x faster than others
- ✓ **G2 – Quality**
 - Better or equal quality compared to top related works
- ✓ **G3 – Low-labor labeling**
 - Less than *five* examples still leads to accurate results



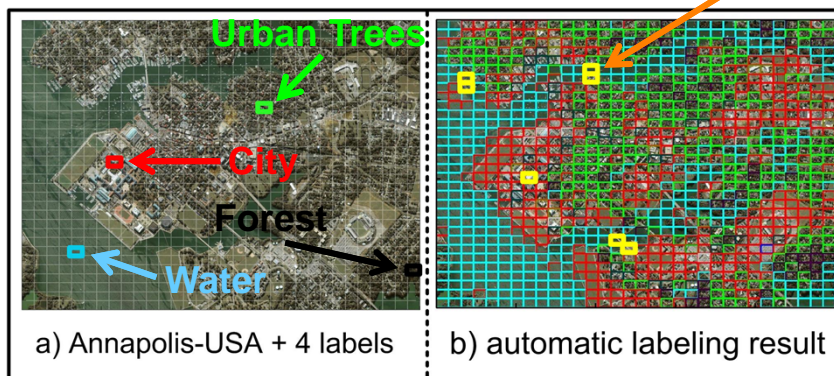
65

Conclusions (cnt'd)

- **Spots outliers**
 - tiles that may need new label-types



Thank you!



Contact e-mail: robson@icmc.usp.br



Microsoft

Research


Google research

67


Introduction

Problem 2: mining and attention routing


- **Given:**
 - Images
 - A few labels



⋮





tree




lake


⋮






⋮





mountain

⋮




68

Omitted for space


Introduction

Problem 2: mining and attention routing


- **Given:**
 - Images
 - A few labels
- **Output:**
 - Clusters → new labels / group labeling
 - A summary of the dataset → comprehension



⋮





tree




lake


⋮






⋮





mountain

⋮



... but in a
Technical
Report

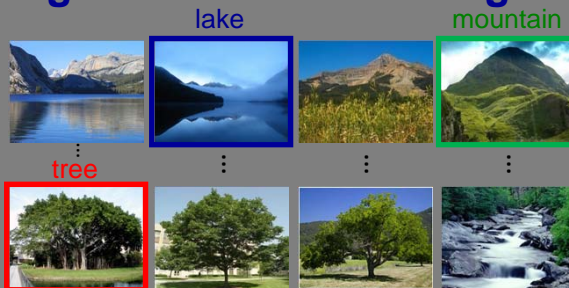
69

Introduction

Problem 2: mining and attention routing

- **Given:**

- Images
- A few labels



- **Output:**

- Clusters → new labels / group labeling
- A summary of the dataset → comprehension

70

Experiments

- **Queries (SATLARGE)**

Example Data ~50 Water samples

