Doctoral Dissertation

Doctoral Program in Computer and Control Engineering (33.th cycle)

Semantics-aware image understanding

Andrea Pasini

Prof. Elena Baralis, Supervisor

Doctoral Examination Committee:
Prof. Rosa Meo, Referee, Università degli studi di Torino
Prof. Elisa Quintarelli, Referee, Università degli studi di Verona
Prof. Genoveva Vargas Solar, Examination board, CNRS, France
Prof. Silvia Chiusano, Examination board, DAUIN - Politecnico di Torino
Prof. Marco Mellia, Examination board, DAUIN - Politecnico di Torino

Motivational ideas

Human **knowledge** is “general purpose”

Machine learning model

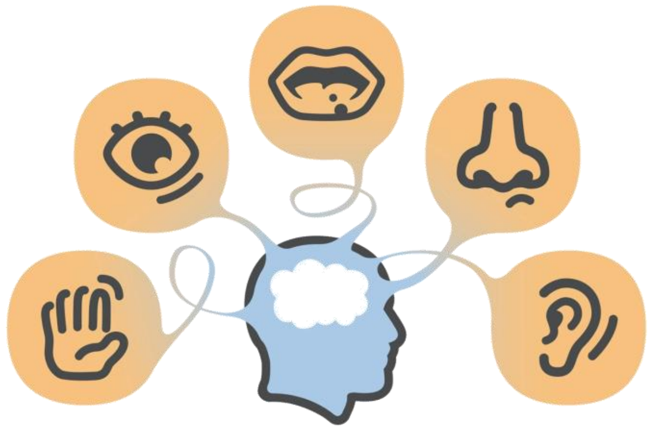
- Address a **single task** (or a limited amount)
  - Specialized on a specific **domain**

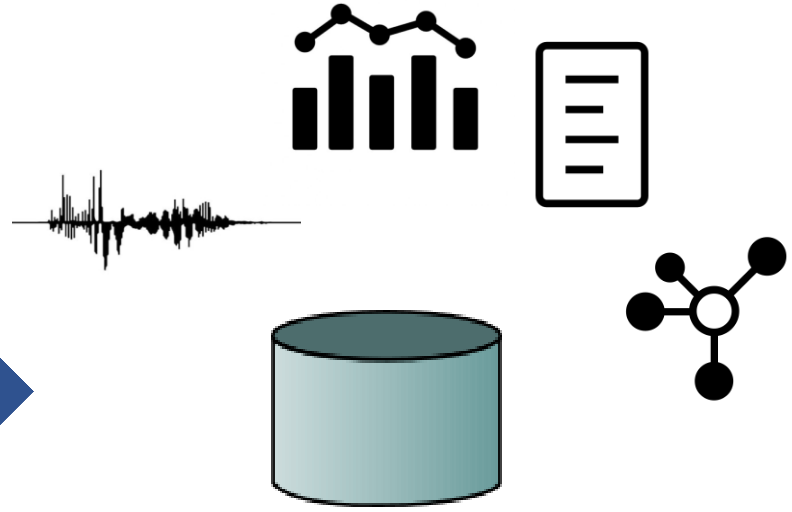
Human

- Address **different tasks in heterogeneous domains**
  - Knowledge is **shared** across domains

Motivational ideas

Human experience



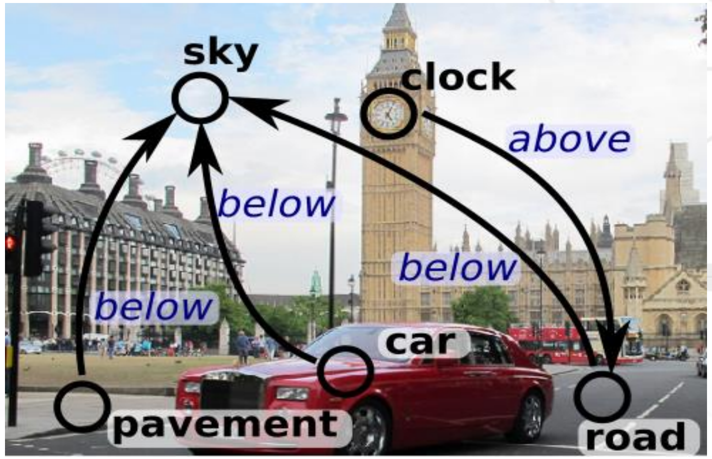


Semantic, heterogeneous data

Dissertation plan

- **Semantic** image understanding in this research
  - **Knowledge** extraction process from a set of labeled images
  - Learn common **object patterns** and relationships

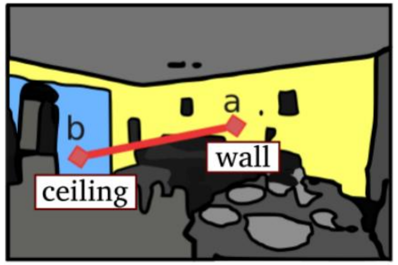
- With the following goals:
  1. Identify **anomalies** in labeled images
  2. Learning common object patterns to **summarize** the content of an image collection



Dissertation plan

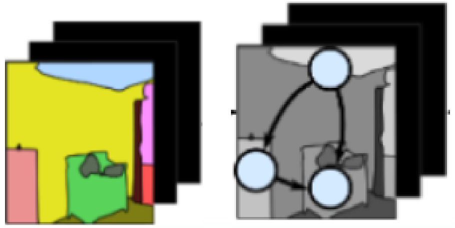
1. Identify **anomalies** in labeled images

SAD Semantic Anomaly Detection

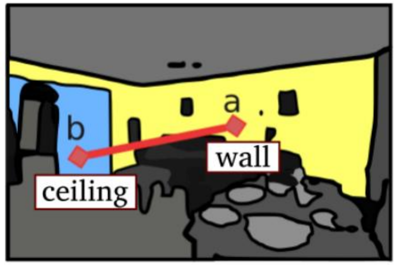


2. Learning common object patterns to **summarize** image collections

SImS Semantic Image Summarization



Dissertation plan



(1) SAD

(2) SImS

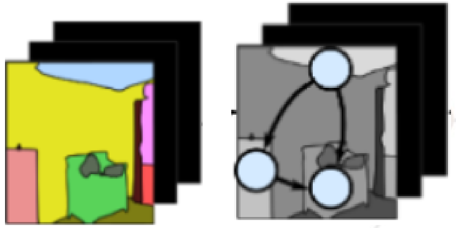
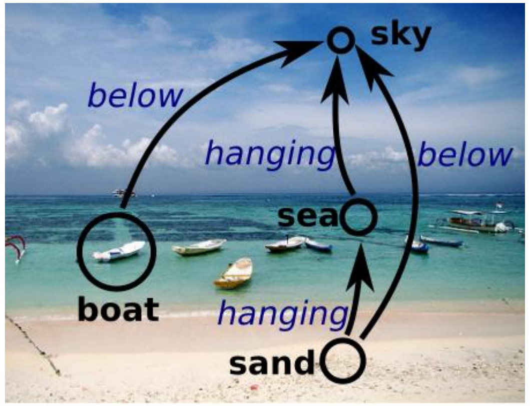


Image content representation



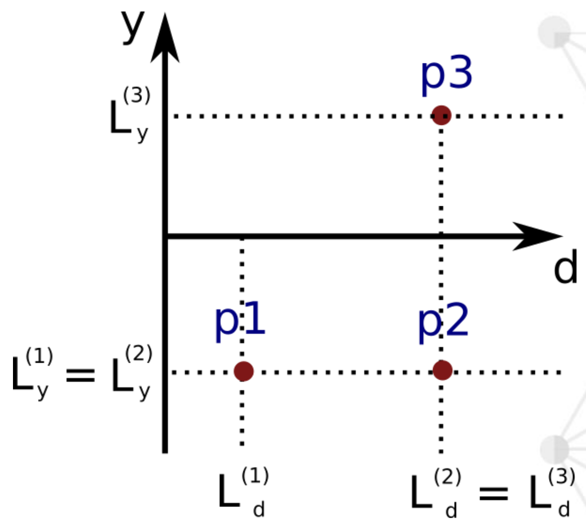
(3) Object relationships

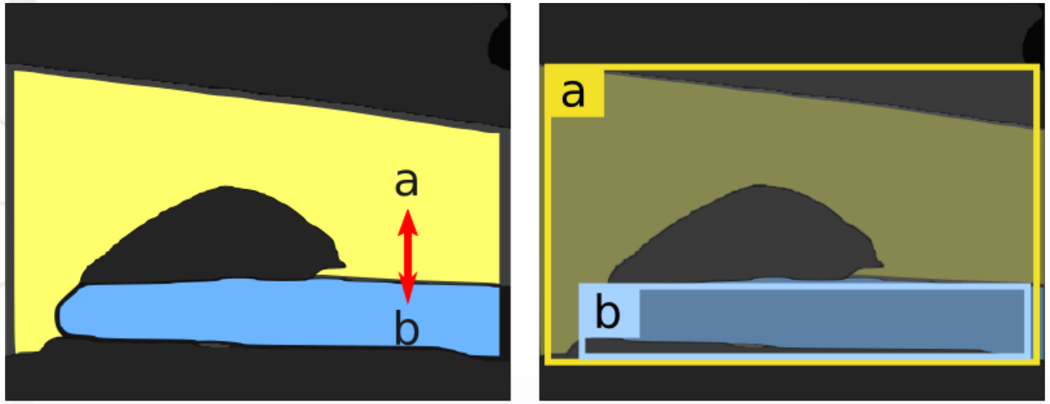
Object relative position

Object relative position

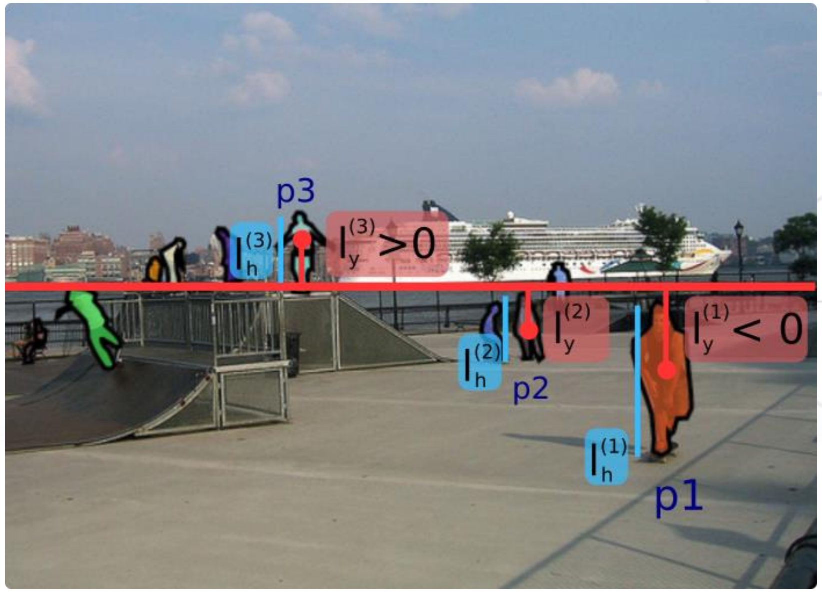
Previous methods

- Bounding boxes/centroids
- Restricted set of semantic relatioships
  - E.g. do not distinguish between ***on*** and ***above***





a) actual shape b) bounding boxes

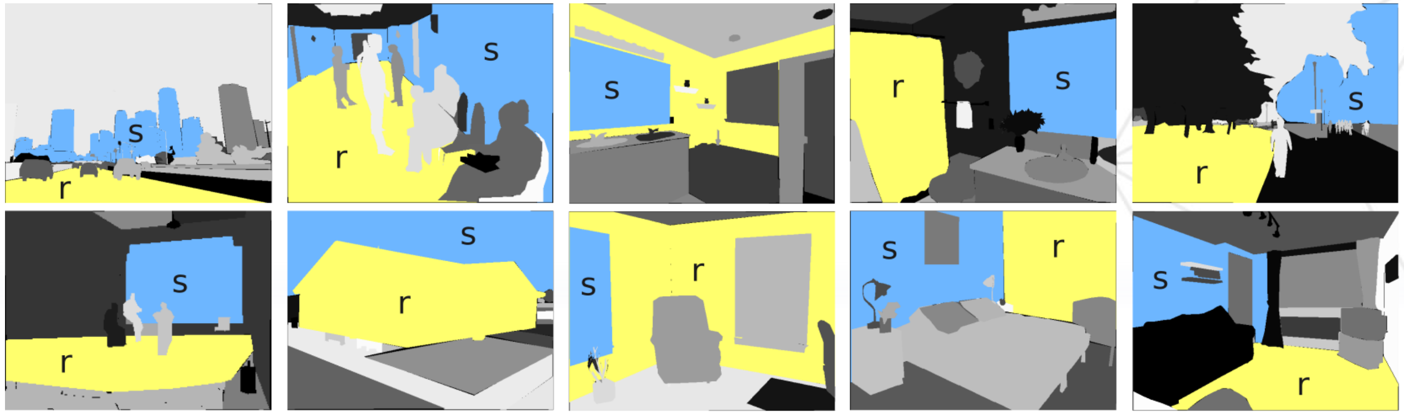


Object relative position

Our technique:

- Analyze **semantic/panoptic segmentation** for a given image
- Design a set of rule-based **features**
- Apply a random forest **classifier** on top of them

Label	Description
above	s is above r without contact
below	s is below r without contact
on	s is on top of r with contact
hanging	s is below r with contact
side	s and r are not vertically aligned
side-up	s and r are not vertically aligned, s is in a higher position
side-down	s and r are not vertically aligned, s is in a lower position
inside	s pixels are inside r shape
around	s pixels are around r shape

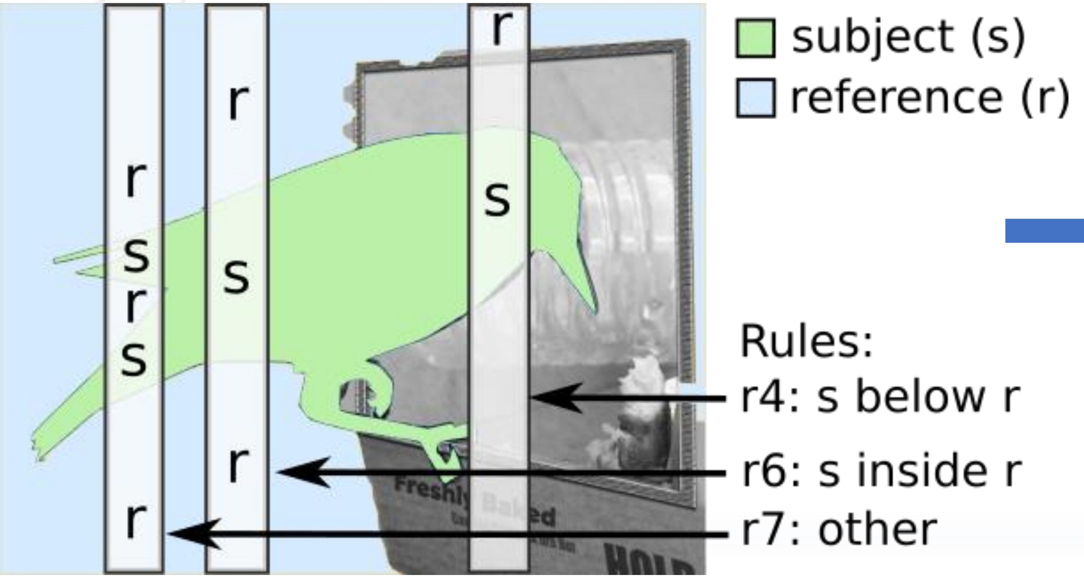


a) above b) on c) inside d) side e) side up

Object relative position

Our technique: string-based representation

- Inspect the **real object shapes**
- Analyze the image by **vertical** strings



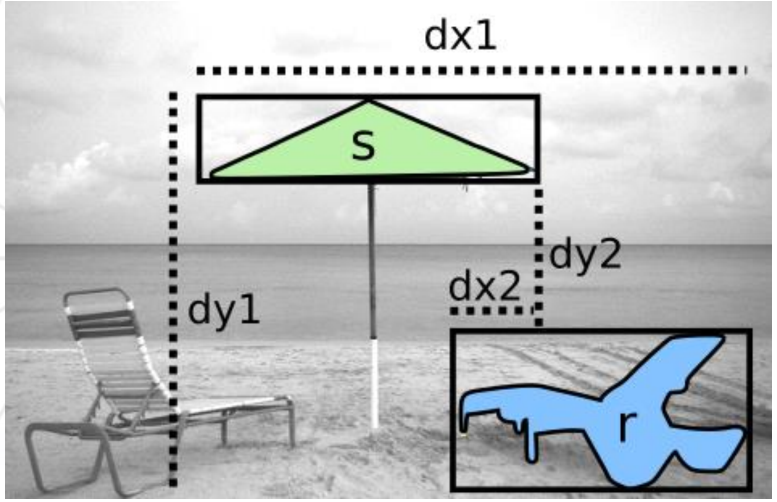
Feature vector (string-based rules)

Below	Above	Inside	Around	On	Hanging	Other
57	0	115	0	0	0	20

Object relative position

Bounding box based features

- Additional insights on the relative position
- Compensate string-based ones when the objects are not vertically aligned



Feature vector (string-based rules)

Below	Above	Inside	Around	On	Hanging	Other
57	0	115	0	0	0	20

+

Feature vector (bounding-box based)

dx1	dx2	dy1	dy2
57	0	115	0

Object relative position

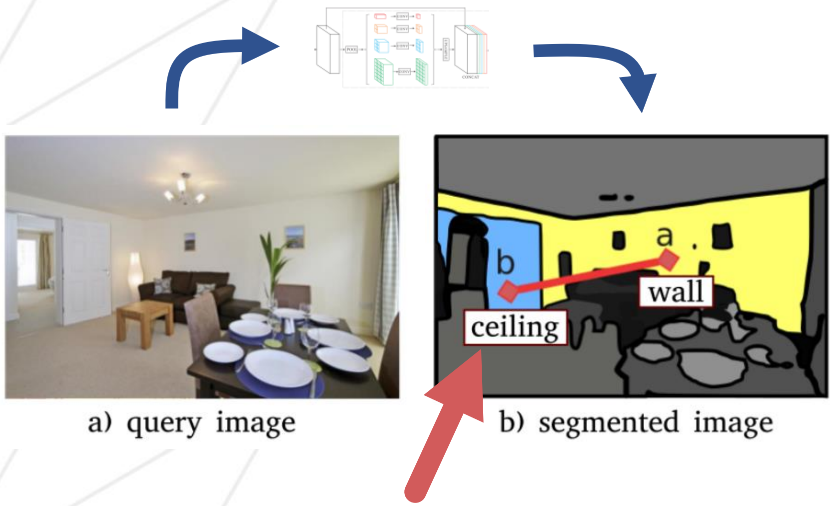
- Classification results (F1-score) on **our dataset**
  - COCO subset with 1000 labeled object pairs

Classifier	above	around	below	hanging	inside	on	side	side-down	side-up	macro-avg
KNN	0.89	0.93	0.95	0.90	0.92	0.78	0.82	0.89	**0.92**	0.89
RBF-SVC	0.89	0.94	0.92	0.84	0.94	0.82	0.75	0.88	0.85	0.87
Decision tree	0.92	0.95	0.94	0.91	0.93	0.86	0.78	0.92	0.82	0.89
Random forest	**0.93**	**0.96**	**0.98**	**0.92**	**0.95**	**0.93**	**0.85**	**0.94**	0.88	**0.93**
Average	0.91	0.94	0.95	0.89	0.93	0.85	0.80	0.91	0.87	0.89

SAD

Semantic Anomaly Detection

Semantic segmentation



There is no perfect model

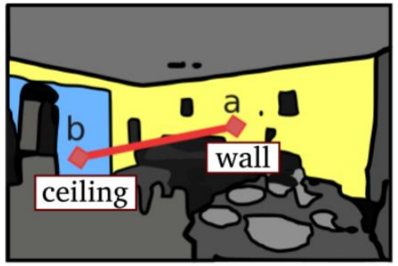
At inference time **ground truth is not available**

Task: find possibly misclassified objects

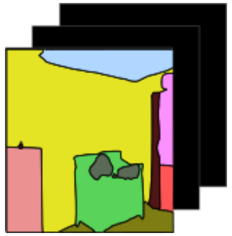
Semantic Anomaly Detection

(**SAD**)

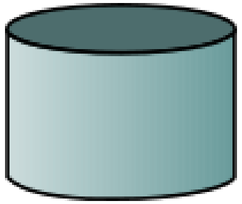
The **SAD (Semantic Anomaly Detection)** process



SAD output: **Anomaly:** wall (a) on ceiling (b) has likelihood <0.01



learn from labeled images



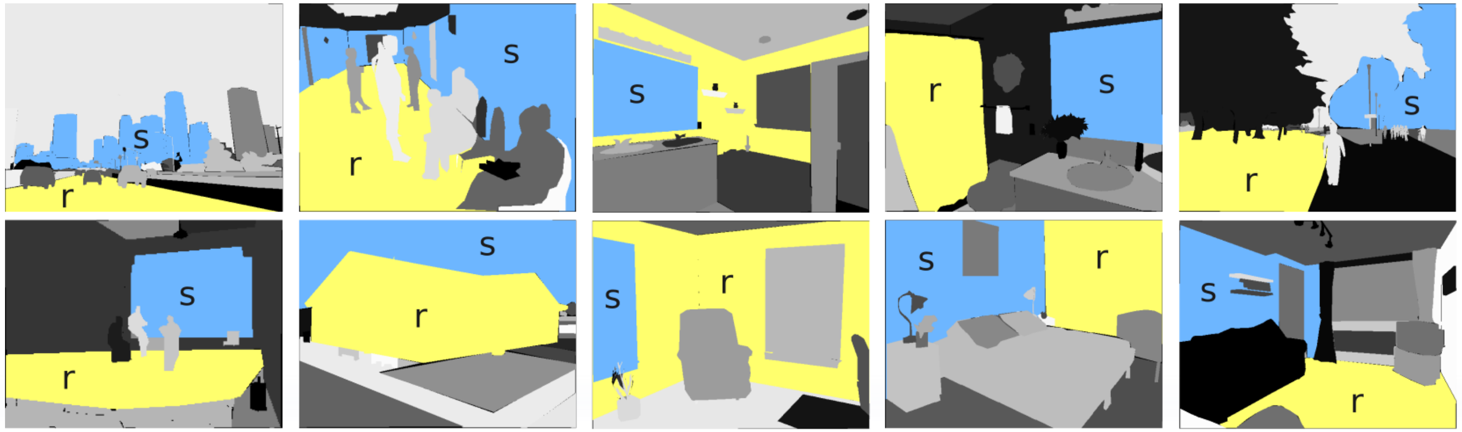
knowledge base

common object relationships:

- relative position
- relative size
- co-occurrence

SAD - Relative position between objects

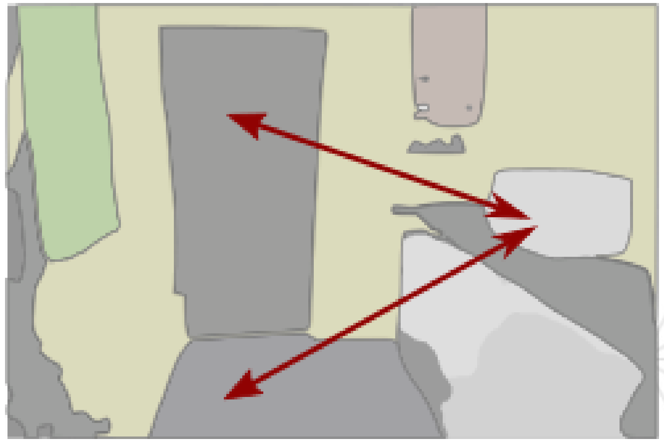
- Identify common relationships
  - E.g., **chair** is typically **on** the **floor**



a) above b) on c) inside d) side e) side up

SAD - Relative size and co-occurrence

- E.g., **cup** is typically **smaller** than **table**
- E.g., **bed** and **pillow** typically co-occur in the same image



SAD - Knowledge base definition

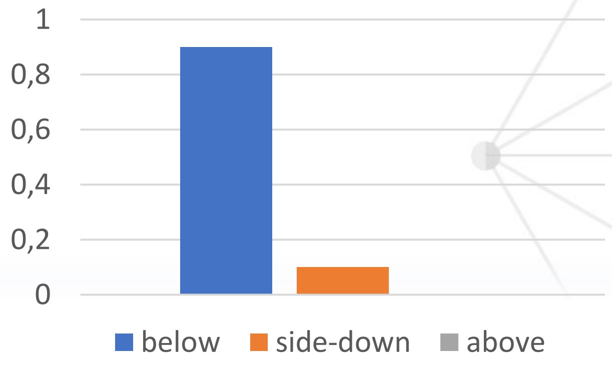
Relationships

Category	Properties
position	*above, below, on, hanging, inside, around, side-up, side, side-down*
width	*bigger, same, smaller*
height	*bigger, same, smaller*
area	*bigger, same, smaller*
co-occurrence	*co-occurs, ¬co-occurs*

1. Consider a pair of classes: s, r
   - s = lamp, r = ceiling
2. Learn from the training images a **histogram** for each category c (e.g., position):

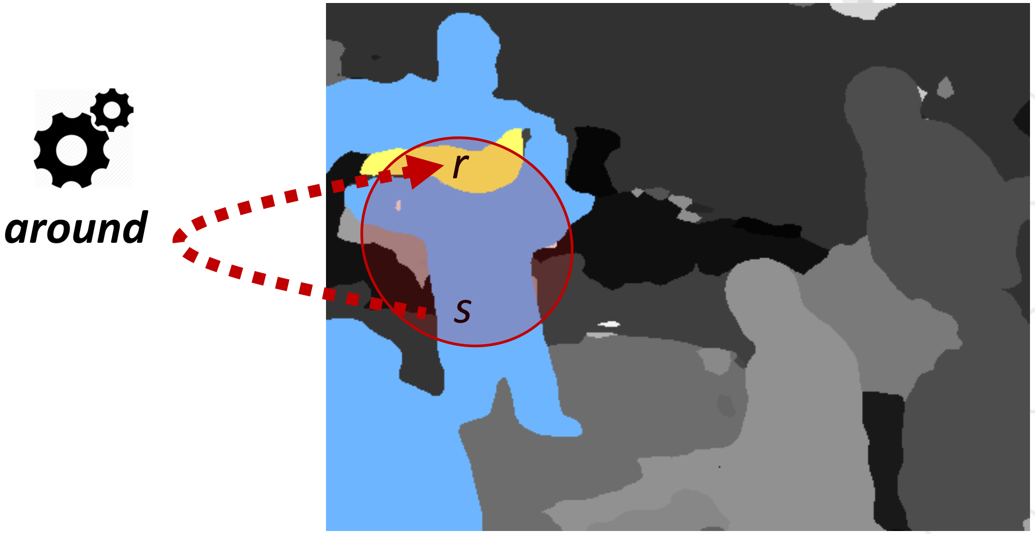
$$H(\mathcal{T}) = [I(p_0) \dots I(p_i) \dots I(p_{n-1})]$$

- $H(\mathcal{T}) = [I(below)=0.9, I(side\text{-}down)=0.1, I(above)=0.0, \dots]$
- 90% of the times lamps are below ceiling...



SAD - The anomaly detection process

Input: segmented image



1. Consider 1 relationship **category**: *c = position*
2. Retrieve **histogram** in the knowledge base:

around

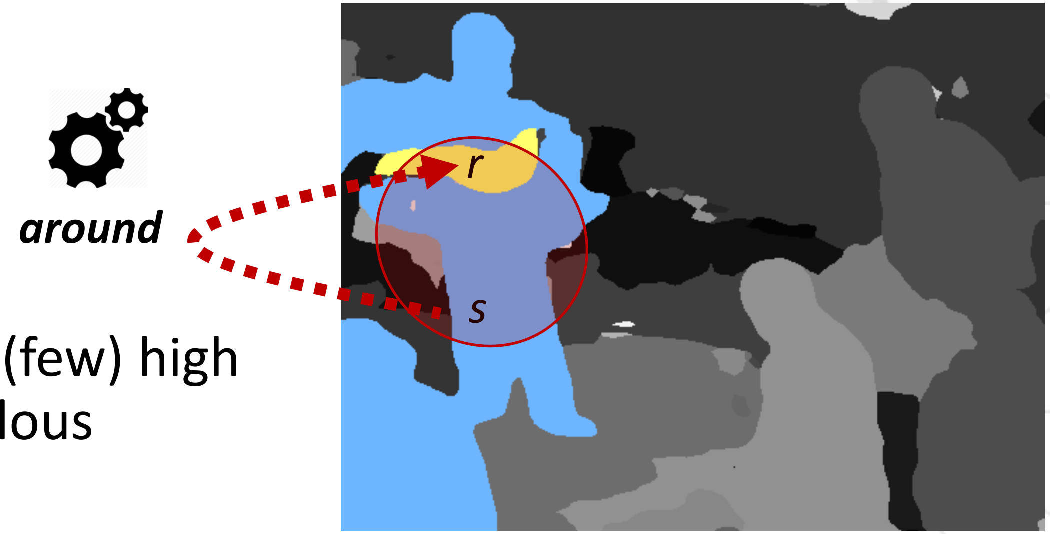
$$H(\mathcal{T}) = [l(p_0) \ldots \mathbf{l(p_i)} \ldots l(p_{n-1})]$$

- If $\mathbf{l(p_i) < thr}$ then an ***anomaly*** is detected between S and R

s = person, r = wall

$H(\mathcal{T}) \rightarrow l(around) = 0.001$

SAD - The anomaly detection process



False positives

- Since anomalies involve **object pairs**, also (few) high accuracy objects can be labeled as anomalous
- E.g., between s (person) and r (wall), only s is misclassified

s = person, r = wall

Anomaly between s and r

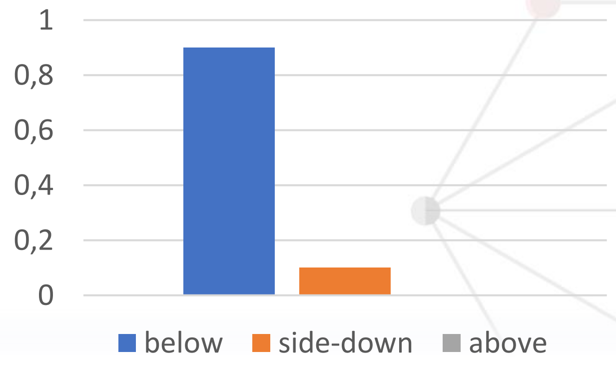
We propose the **Delta method** to tackle this issue

SAD - The Delta method

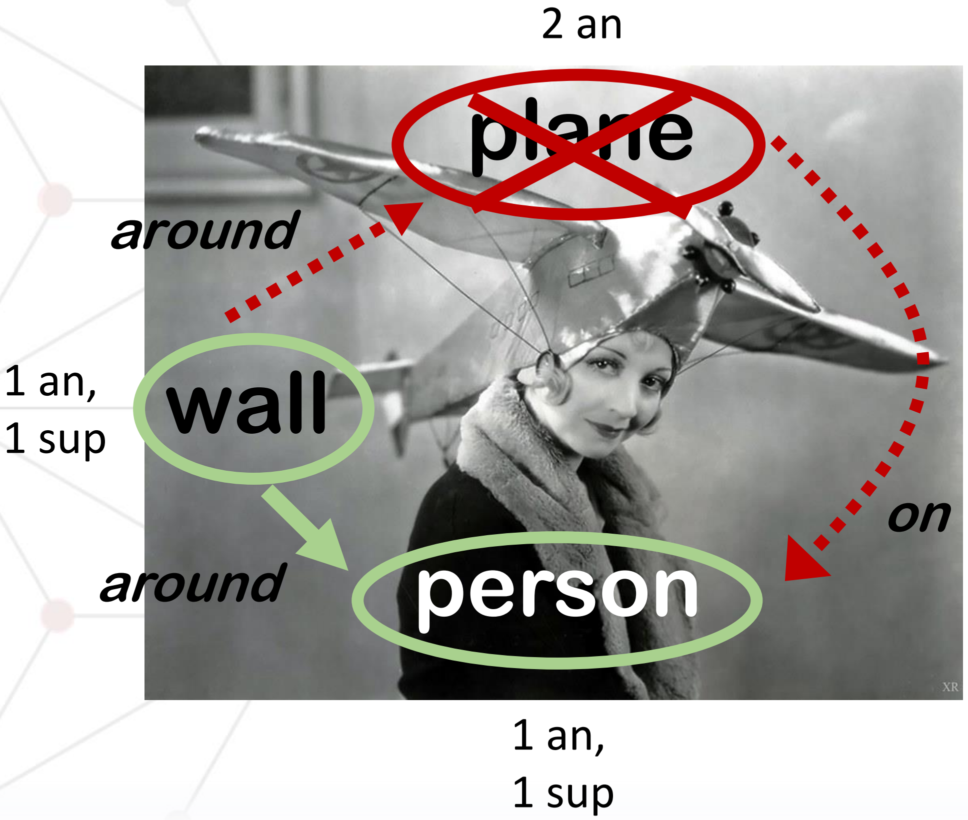
Definitions

- Anomaly: If $I(\neg p_i) > thr$

- Supporter: If $I(p_i) > thr$

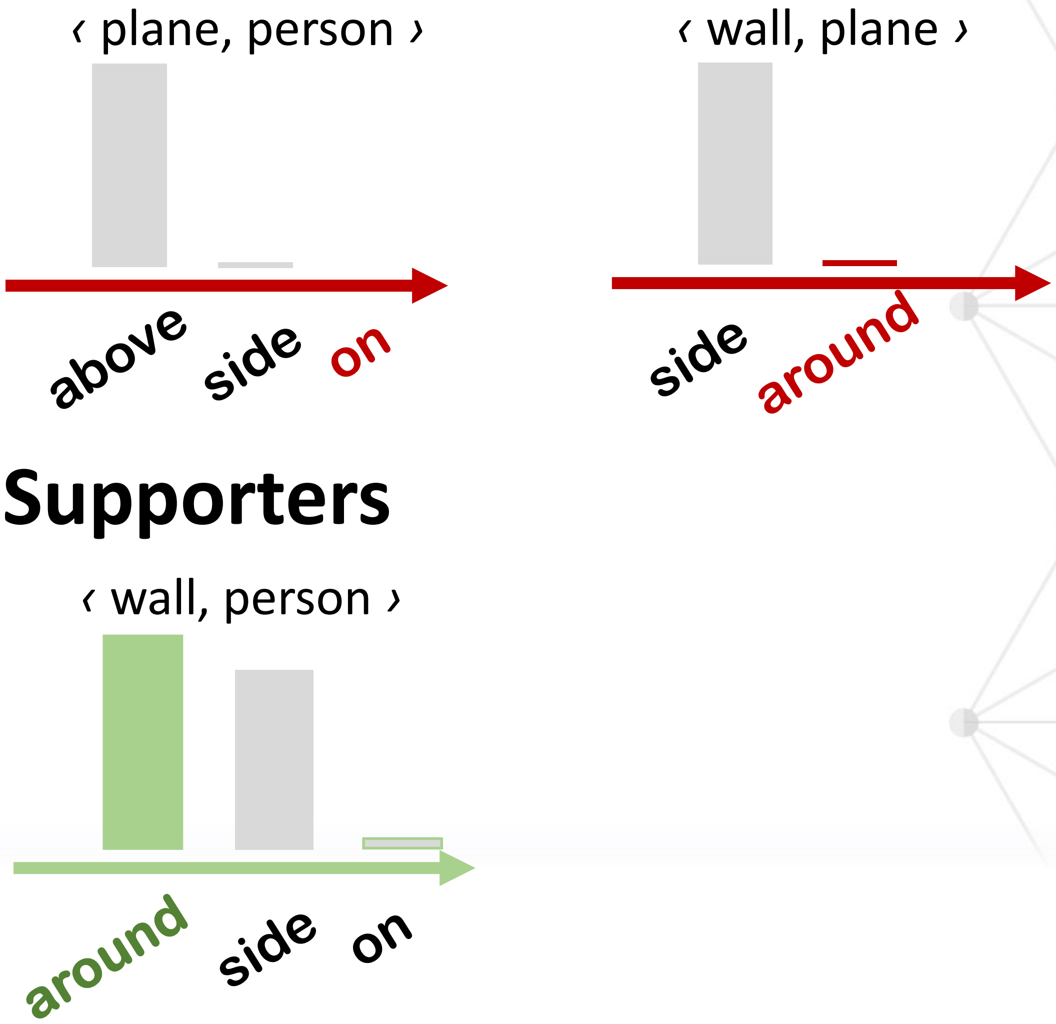


SAD - The Delta method

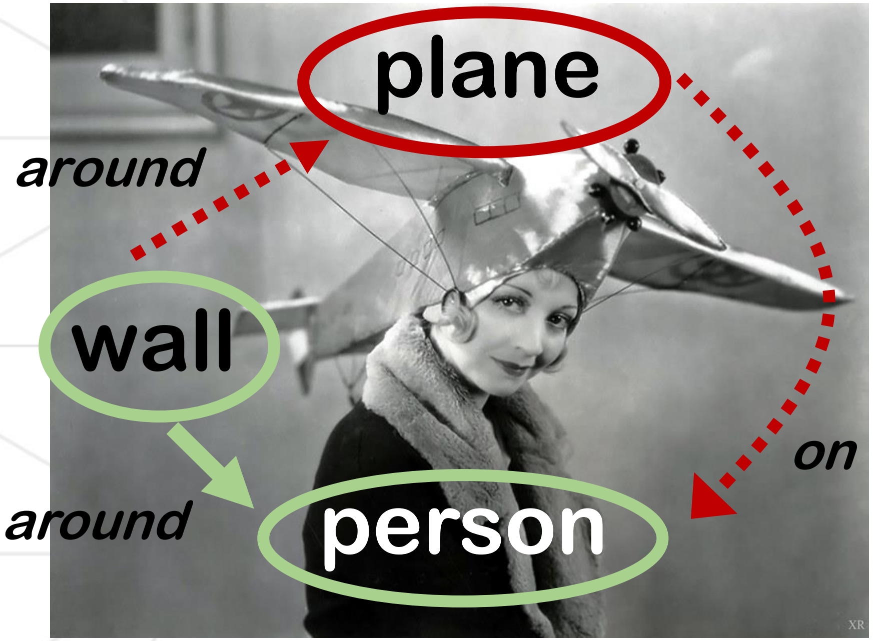


Anomalies

Supporters



SAD - The Delta method



Classification confidence score

(lower = misclassified)

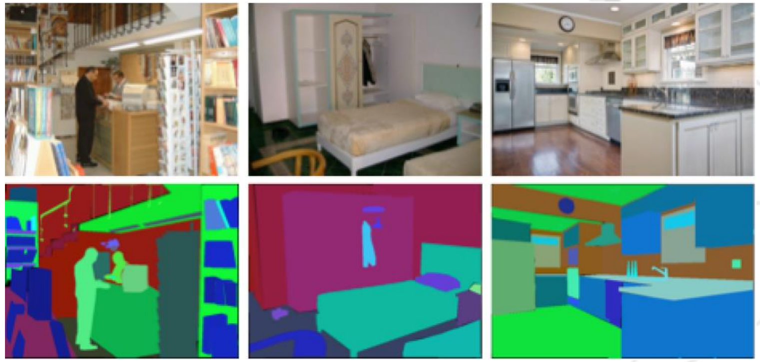
$$score(z) = \sum_{sup \in \mathcal{S}up_z} conf(sup) - \sum_{an \in \mathcal{A}n_z} conf(an)$$

SAD - Experiments on ADE20K dataset

POLITECNICO DI TORINO

SmartData

- 20000 training images, 2000 test images, 150 class labels



Knowledge base examples

co-occurrence

Class Pair	CF	Class Pair	CF
wall, oven	1.00	sky, microwave	-0.99
wall, sink	0.99	cabinet, road	-0.99
floor, sofa	0.96	sofa, car	-0.99
bed, pillow	0.94	sky, countertop	-0.99
building, sidewalk	0.93	floor, hill	-0.98
sky, mountain	0.91	lamp, river	-0.98

area

Class Pair	Sup	Histogram
plate, swivel chair	25	bi=0.00 sa=0.00 sm=1.00
light, microwave	378	bi=0.02 sa=0.06 sm=0.92
runway, van	20	bi=0.95 sa=0.05 sm=0.00
painting, pool table	271	bi=0,03 sa=0,04 sm=0,94

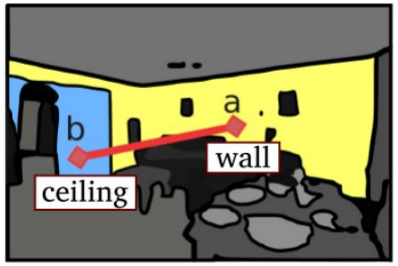
position

Class Pair	Sup	Histogram
runway, sky	151	*below*=0.87 *side-down*=0.1
ball, pool table	33	*inside*=0.91 *above*=0.03
light, sink	1321	*side-up*=0.83 *above*=0.17

SAD - Experiments on ADE20K dataset

Anomaly detection

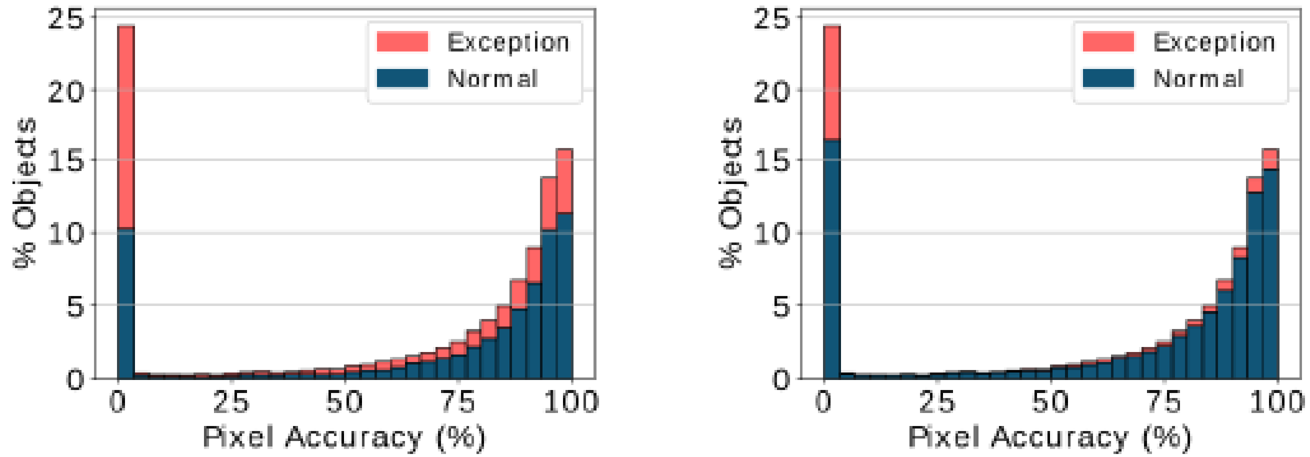
- Segmentation network under evaluation: **PSPnet**
  - **Misclassified** objects (low pixel accuracy) should be labeled as **anomalies**
    - Misclassified object have pixel accuracy <75%



Anomaly: wall (a) on ceiling (b) has likelihood <0.01

SAD - Experiments on ADE20K dataset

Anomaly detection



(a) Anomaly only method.

(b) Delta method.

Method	Precision (Ex)	Recall (Ex)	Precision (Norm)	Recall (Norm)	Macro avg. precision	Macro avg recall
Anomaly only	0.50	**0.56**	**0.74**	0.70	0.62	**0.63**
Delta	**0.63**	0.29	0.70	**0.91**	**0.67**	0.60

SAD - Points of strength

- Provides a human **understandable** description of the anomaly
- Highlights **potentially misclassified objects**
- **Semantic enrichment** of the image segmentation even when the classification is correct

SImS

Semantic Image Summarization

Visual summarization

Big image collection

What are the **main themes** of the images inside this collection?

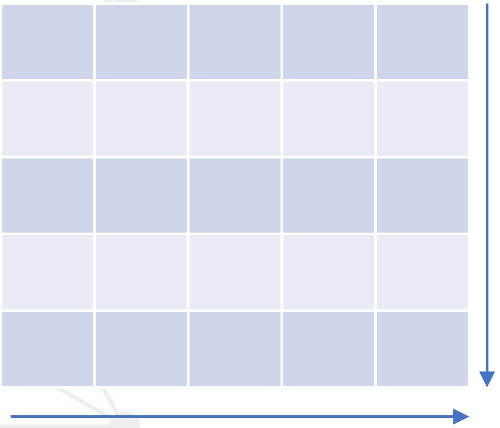
Where do you typically find people in the collection?

What can be typically found **below** sky in the images?

Visual summarization

Previous approaches to image **collection** summarization

Features matrix

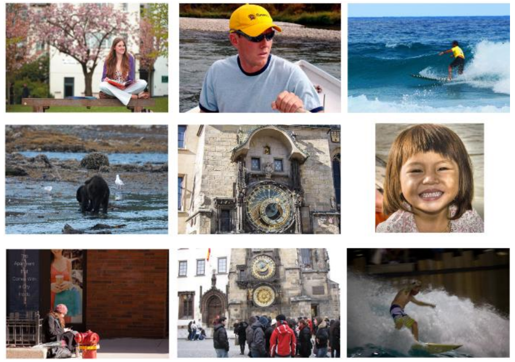


images

features

KMedoids clustering

Medoids = summary images

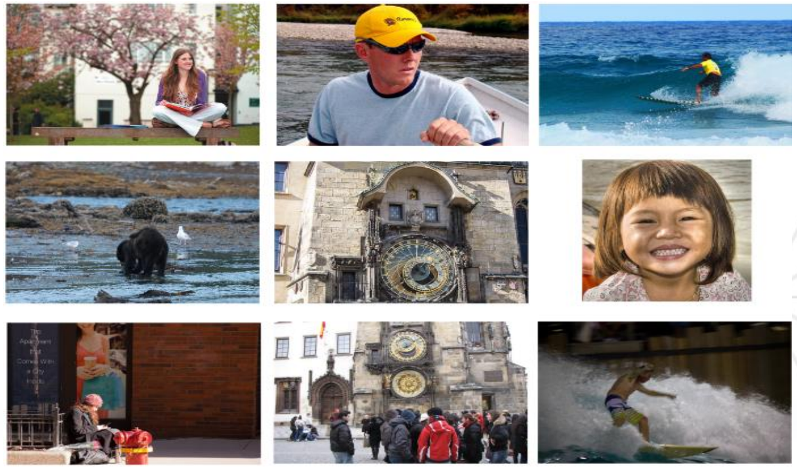


"Video summarization by k-medoid clustering", 2006, ACM symposium on Applied computing, Hadi, F. Essannouni, and R. O. H. Thami.

Visual summarization

Main issues with KMedoids (and similar) approach

- Low **interpretability**
- Lack of **semantic** understanding of the image content

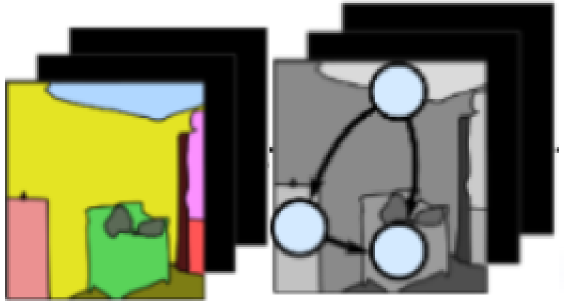


SImS: Semantic Image Summarization

Advantages of our approach

- **Semantics-aware** summarization patterns
- **Interpretable** results
    - Based on scene graphs

SImS

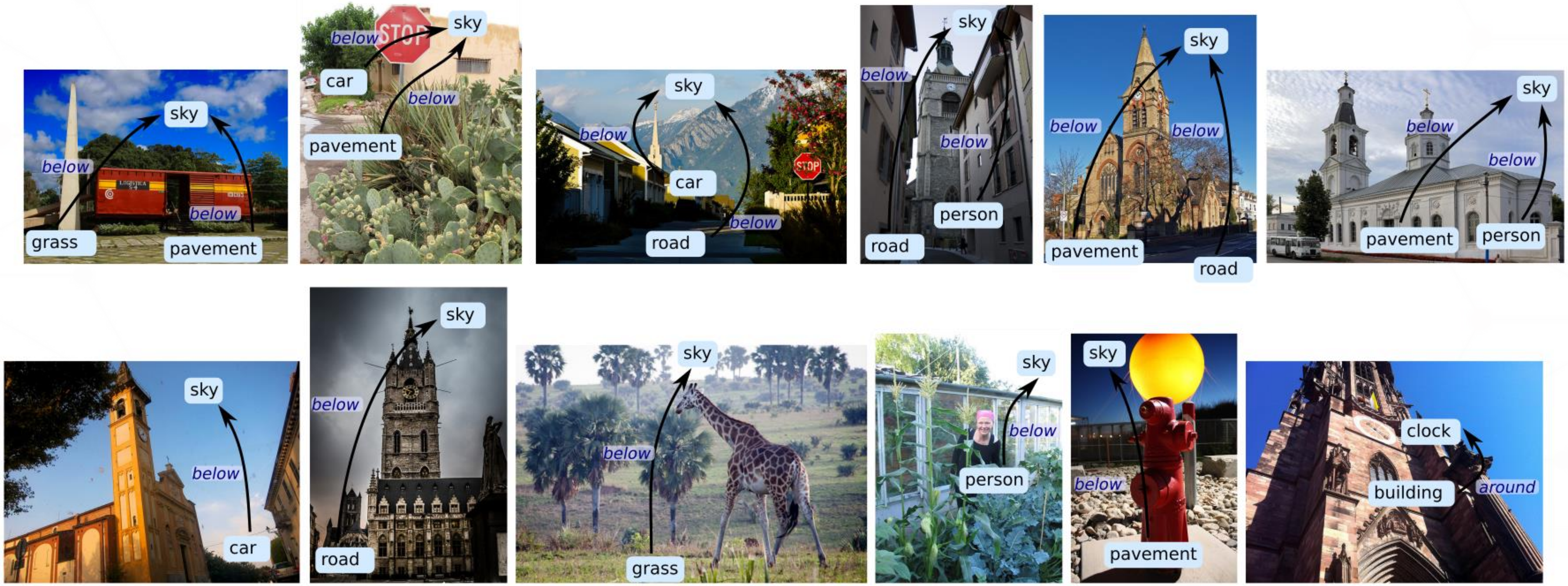


SImS: Semantic Image Summarization

POLITECNICO DI TORINO

SmartData

Example SGS (Scene Graph Summary): church-garden dataset

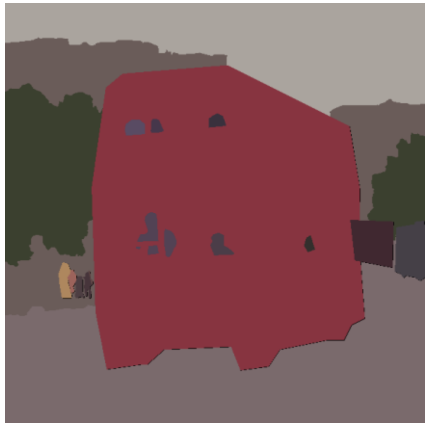


SImS - Scene Graphs

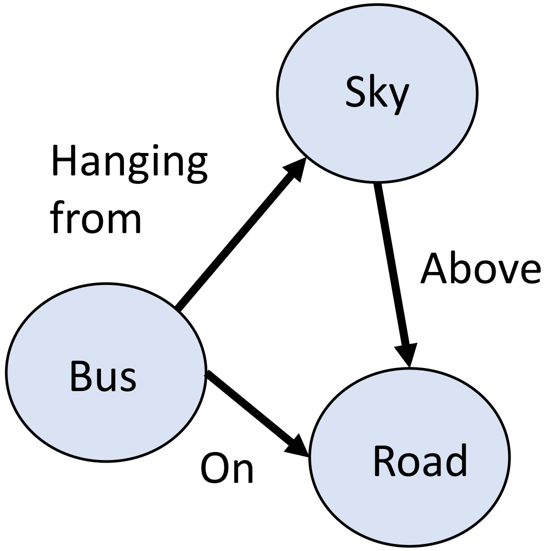
Panoptic segmentation

Relative position computation

Input image



Segmented image

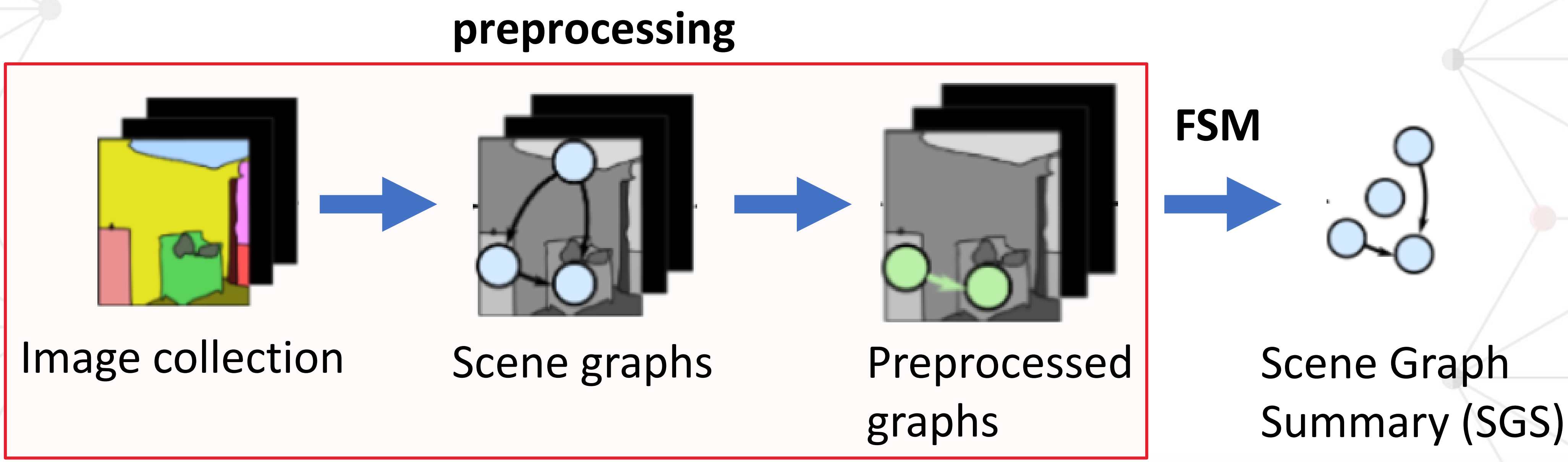


"Detecting Anomalies in Image Classification by Means of Semantic Relationships", 2019, IEEE AIKE, Andrea Pasini, Elena Baralis

SImS – Working Principles

Summarization based on **frequent subgraph mining (FSM)**

- Derive the Scene Graph Summary (SGS)



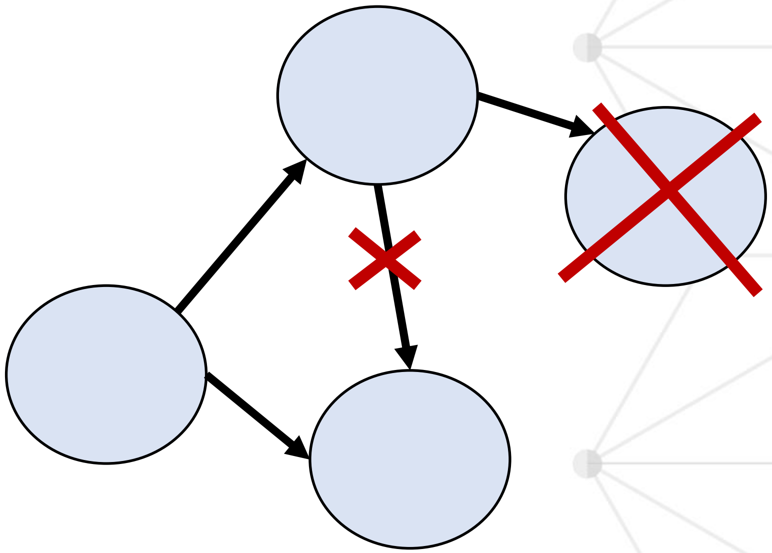
SImS – Working Principles

Frequent subgraph mining on the scene graphs:

- **SGS with redundancies**
- **Very slow, not scalable!**

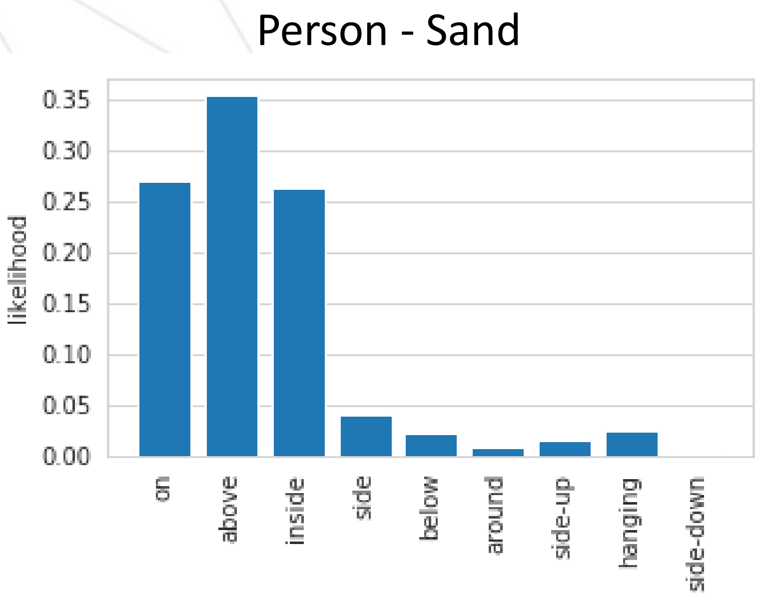
Solution: simplify input graphs

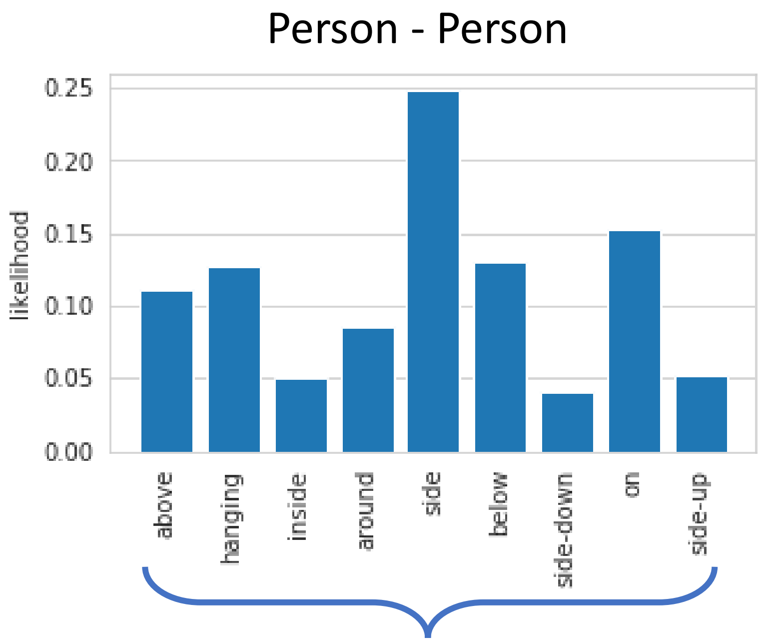
- Edge pruning
- Node pruning

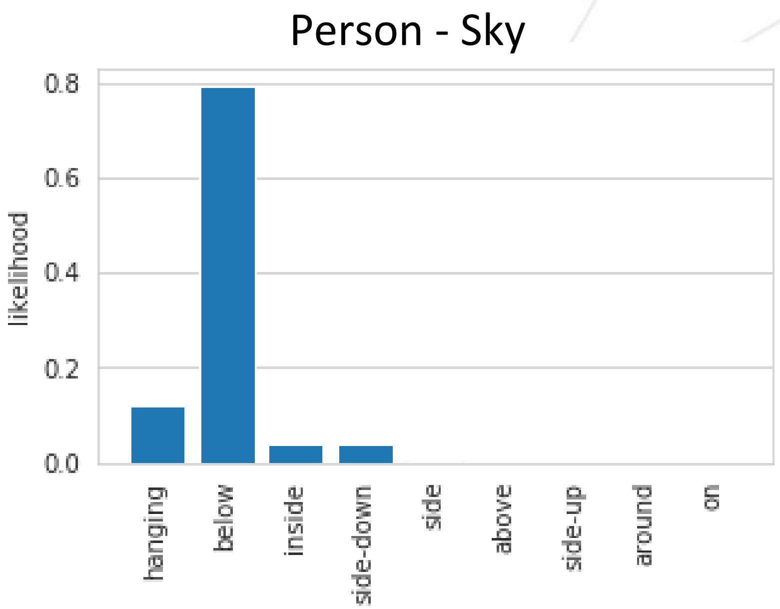


SImS – Edge pruning

1. Identify in the input collection **high-entropy relationships**
    - Build the **PRS** (Pairwise Relationship Summary)



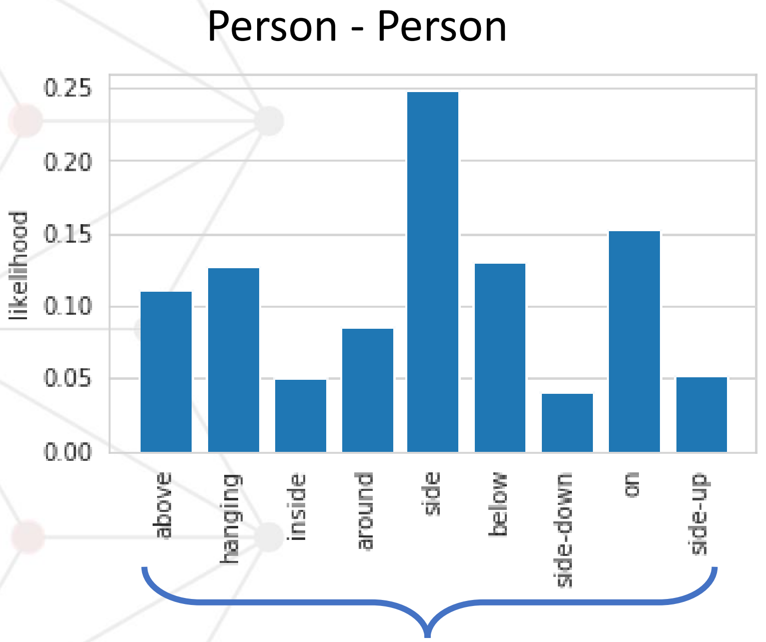




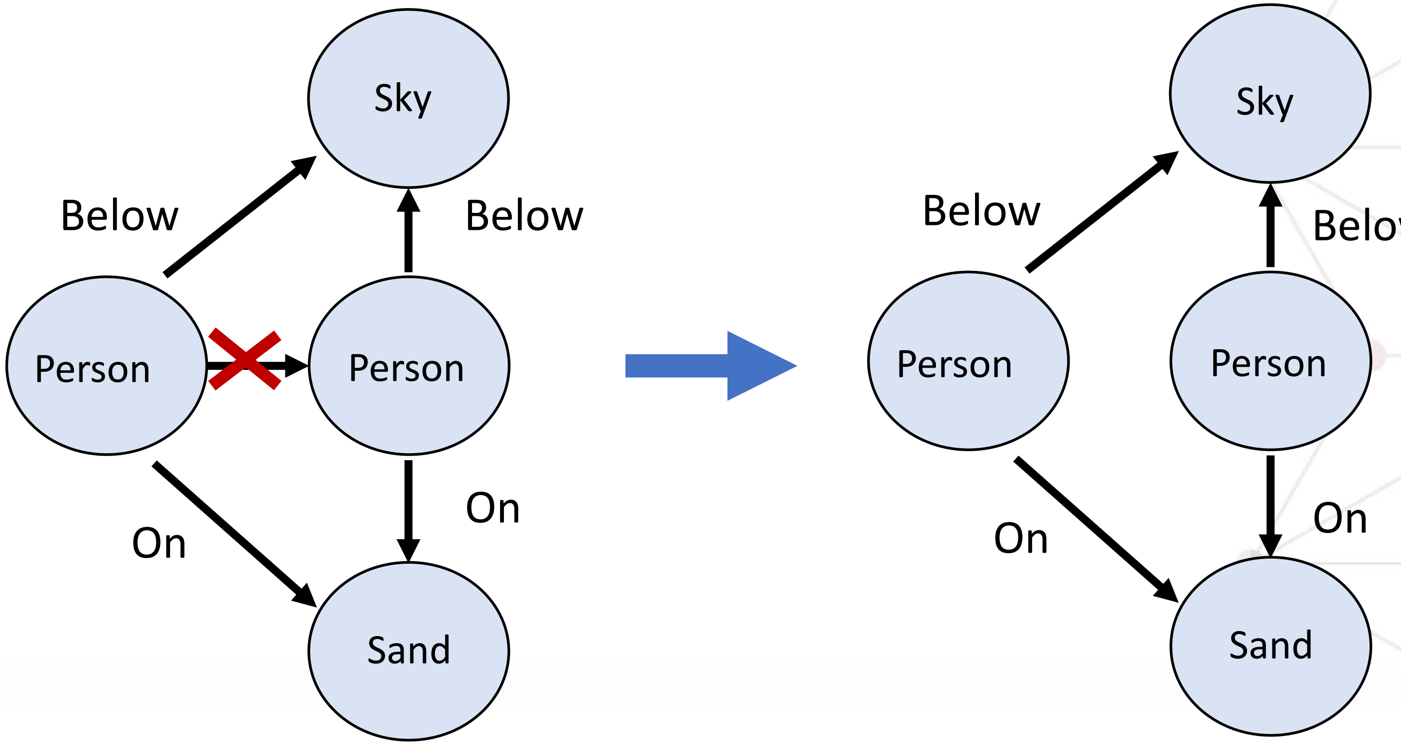
High entropy = fewer information

SImS – Edge pruning

2. Remove **high-entropy relationships** (for all input graphs)

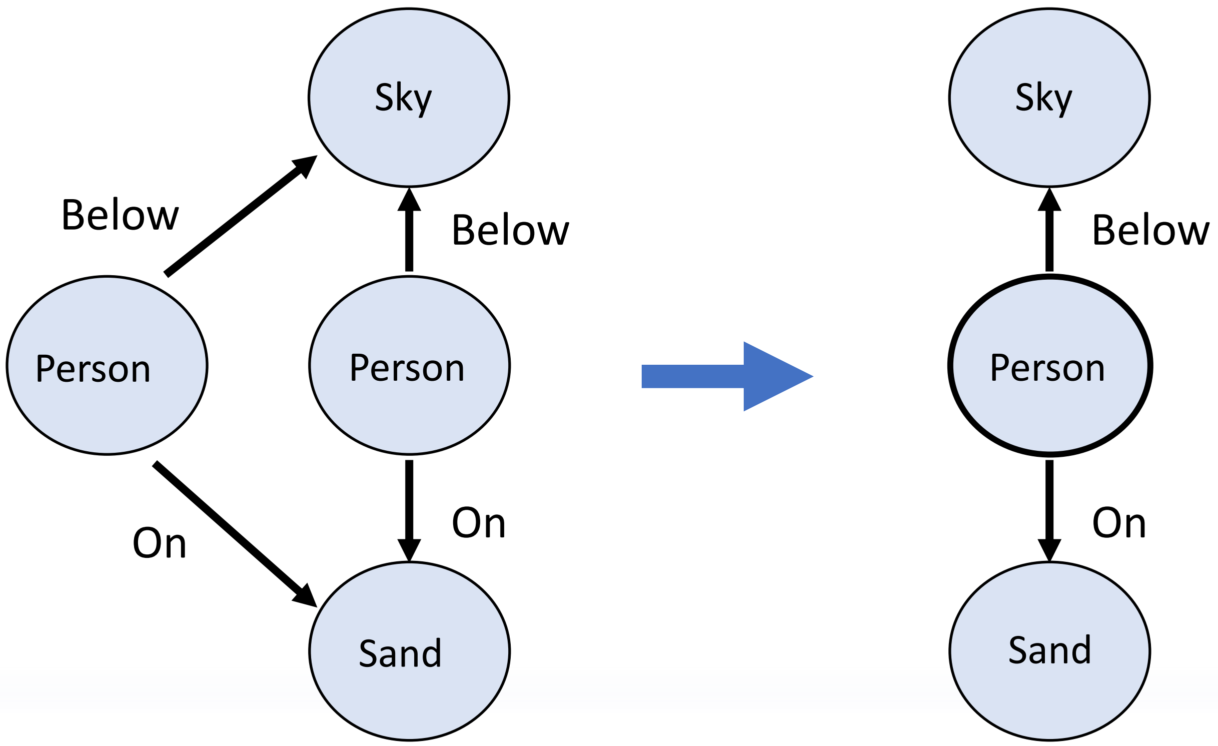


High entropy = fewer information



SImS – Node pruning

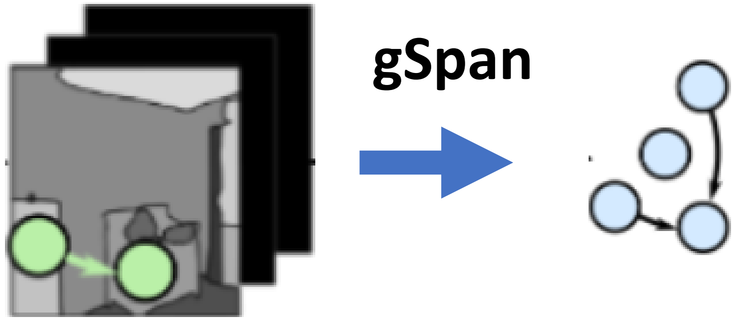
Remove **equivalent nodes** (same type of relationships)



SImS – FSM

Apply a Frequent Subgraph Mining algorithm: gSpan

- Find common **frequent subgraphs** in the collection



Preprocessed graphs

Scene Graph Summary (SGS)

Minsup = min % of images where a subgraph should occur to be considered as **frequent**

“gspan: Graph-based substructure pattern mining", 2002, IEEE International Conference on Data Mining, X. Yan and J. Han.

SImS – Summary Evaluation

Node and **edge diversity**

- Average **dissimilarity** among SGS graphs

Coverage

- Percentage of input collection images **represented** by the SGS

SImS – Preprocessing Evaluation

Summarization of Microsoft COCO dataset (118K images)

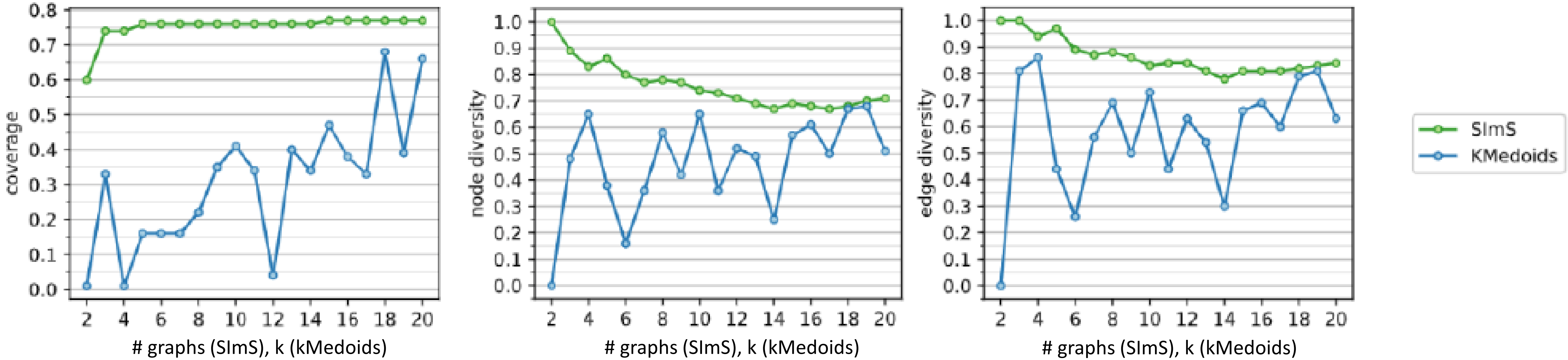
Focus on running **time (graph-mining step)**

Configuration			*Statistics*			
Minsup	Edge pruning	Node pruning	**Time**	N. graphs	Coverage	Diversity
0.010	N	N	15h 55m	6111	0.43	0.60
0.010	Y	N	4h 30m	237	0.43	0.69
0.010	**Y**	**Y**	**2 s**	**144**	**0.43**	**0.81**
0.001	Y	N	Doesn't finish	/	/	/
0.001	**Y**	**Y**	**7 s**	**3345**	**0.48**	**0.75**

"Panoptic segmentation", 2019, CVPR,
A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dolľar.

SImS – Comparison with KMedoids

Summarization of Microsoft COCO subset (4865 images): skiing, driving **topics**



* For kMedoids, graph coverage/diversity is computed by extracting scene graphs from the output images

SImS highlights

SImS - Semantic Image Summarization

- Based on **Frequent Subgraph Mining** on scene graphs
- Interpretable, **semantic** aware results
- Higher coverage and diversity

Thank you for your attention
Any questions?

"Additional reviewer assignment by means of weighted association rules", 2018, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, Cagliero, L.; Garza, P.; Pasini, A.; Baralis, E.

"Adaptive hierarchical clustering for petrographic image analysis", Data Analytics solutions for Real-LIfe APplications (DARLI-AP). 2019 Workshops of the EDBT/ICDT Joint Conference, EDBT/ICDT-WS 2019, Pasini, Andrea; Baralis, Elena; Garza, Paolo; Floriello, Davide; Idiomi, Michela; Ortenzi, Andrea; Ricci, Simone

"Detecting Anomalies in Image Classification by Means of Semantic Relationships", 2019, IEEE, Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Pasini, Andrea; Baralis, Elena

"Automatic pore typing classification from 2D images", OMC 2019, Floriello, D.; Ortenzi, A.; Idiomi, M.; Ricci, S.; Amendola, A.; Carminati, S.; Baralis, E.; Garza, P.; Pasini, A.

"DSLE: A Smart Platform for Designing Data Science Competitions", 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) Attanasio, Giuseppe; Giobergia, Flavio; Pasini, Andrea; Ventura, Francesco; Baralis, Elena Maria; Cagliero, Luca; Garza, Paolo; Apiletti, Daniele; Cerquitelli, Tania; Chiusano, Silvia Anna

“Severity Classification of Deep Learning U-Nets from Satellite Images”, 2020 IEEE International Conference on Big Data Monaco, Simone; Pasini, Andrea; Apiletti, Daniele; Colomba, Luca; Garza, Paolo; Baralis, Elena Maria