Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it
Bonn-Aachen
International Center for
Information Technology

Institute for AI and
Autonomous Systems

Spatial Relation Learning

An Overview

Dr. Alex Mitrevski
Master of Autonomous Systems

Structure

- Importance of spatial relations
- Spatial relation learning overview

Importance of Spatial Relations

Spatial Relation Preliminaries

- When considering an object-centric state space representation (recall our introductory lectures!), it is often useful to consider **the relation of one object with respect to another**

Spatial Relation Preliminaries

- When considering an object-centric state space representation (recall our introductory lectures!), it is often useful to consider **the relation of one object with respect to another**
- **Qualitative descriptions** of such relations are referred to as **spatial relations** (e.g. **on**, **next to**, **to the left of**, **in**, etc.)

Spatial Relation Preliminaries

- When considering an object-centric state space representation (recall our introductory lectures!), it is often useful to consider **the relation of one object with respect to another**
- **Qualitative descriptions** of such relations are referred to as **spatial relations** (e.g. **on**, **next to**, **to the left of**, **in**, etc.)
- Spatial relations are **particularly useful for describing manipulation activities in natural language** as well as **for explaining what a robot is attempting to do**
  - E.g. a robot may need to place a mug **on** a desk, **to the left of** a fork

Spatial Relation Preliminaries

- When considering an object-centric state space representation (recall our introductory lectures!), it is often useful to consider **the relation of one object with respect to another**
- **Qualitative descriptions** of such relations are referred to as **spatial relations** (e.g. **on**, **next to**, **to the left of**, **in**, etc.)
- Spatial relations are **particularly useful for describing manipulation activities in natural language** as well as **for explaining what a robot is attempting to do**
  - E.g. a robot may need to place a mug **on** a desk, **to the left of** a fork

A spatial relation is a qualitative representation of some aspect of space, typically defined for a pair of objects

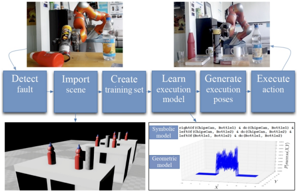
Spatial Relations and Planning

- A common use of spatial relations is in the **process of planning robot actions** — the action preconditions and effects are typically described by such relations

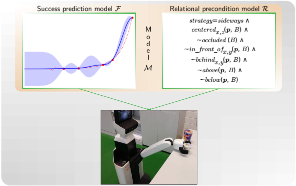
The snippet is based on an example used in the presentation of A. Mitrevski et al., "On the Diagnosability of Actions Performed by Contemporary Robotic Systems," in *31th Int. Workshop Principles of Diagnosis (DX)*, 2020.

```
(:action Pick
    :parameters (?Object - Object ?Table - Table ?Robot - Robot ?Waypoint -
          Waypoint)
    :precondition (and
        (robotAt ?Robot ?Waypoint)
        (tableAt ?Table ?Waypoint)
        (on ?Object ?Table)
        (emptyGripper ?Robot)
    )
    :effect (and
        (not (on ?Object ?Table))
        (not (emptyGripper ?Robot))
        (holding ?Robot ?Object)
    )
)
```

Spatial Relations and the Execution Process



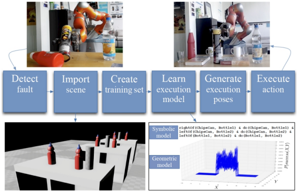
Object placing guided by spatial relations. A. Mitrevski et al, "Improving the Reliability of Service Robots in the Presence of External Faults by Learning Action Execution Models," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4256–4263, 2017.



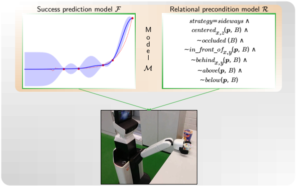
Grasping objects based on spatial relations. A. Mitrevski, "Skill Generalisation and Experience Acquisition for Predicting and Avoiding Execution Failures," *Ph.D. dissertation*, Department of Computer Science, RWTH Aachen University, 2023.

- Using spatial relations in the preconditions and effects has implications for the execution process:
  - A robot should be able to **generate execution parameters that satisfy the spatial relations**
  - The robot should also be able to **verify that the relations are indeed satisfied after the execution** in order to identify execution anomalies

Spatial Relations and the Execution Process



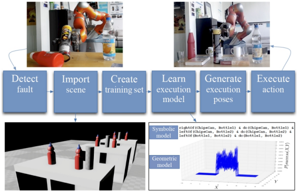
Object placing guided by spatial relations. A. Mitrevski et al, "Improving the Reliability of Service Robots in the Presence of External Faults by Learning Action Execution Models," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4256–4263, 2017.



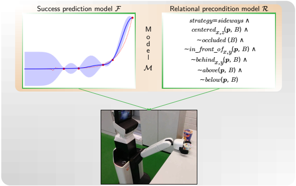
Grasping objects based on spatial relations. A. Mitrevski, "Skill Generalisation and Experience Acquisition for Predicting and Avoiding Execution Failures," *Ph.D. dissertation*, Department of Computer Science, RWTH Aachen University, 2023.

- Using spatial relations in the preconditions and effects has implications for the execution process:
  - A robot should be able to **generate execution parameters that satisfy the spatial relations**
  - The robot should also be able to **verify that the relations are indeed satisfied after the execution** in order to identify execution anomalies
- Spatial relations can be observed as **execution constraints** that should be satisfied during the execution

Spatial Relations and the Execution Process



Object placing guided by spatial relations. A. Mitrevski et al, "Improving the Reliability of Service Robots in the Presence of External Faults by Learning Action Execution Models," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4256–4263, 2017.



Grasping objects based on spatial relations. A. Mitrevski, "Skill Generalisation and Experience Acquisition for Predicting and Avoiding Execution Failures," *Ph.D. dissertation*, Department of Computer Science, RWTH Aachen University, 2023.

- Using spatial relations in the preconditions and effects has implications for the execution process:
  - A robot should be able to **generate execution parameters that satisfy the spatial relations**
  - The robot should also be able to **verify that the relations are indeed satisfied after the execution** in order to identify execution anomalies
- Spatial relations can be observed as **execution constraints** that should be satisfied during the execution
- The use of spatial relations during execution may potentially also contribute towards **increasing the explainability of robots**
  - As already mentioned, spatial relations and natural language are closely related

Defining (Spatial) Relations

- We can define relations (spatial, but also relations in general) in multiple ways:
  - In a first-order logic sense, a relation can be seen as **a relation $r : \mathbb{O}^n \rightarrow \{\textbf{true}, \textbf{false}\}$, where $O \in \mathbb{O}$ represents an object**

Defining (Spatial) Relations

- We can define relations (spatial, but also relations in general) in multiple ways:
  - In a first-order logic sense, a relation can be seen as **a relation $r : \mathbb{O}^n \rightarrow \{\textbf{true}, \textbf{false}\}$, where $O \in \mathbb{O}$ represents an object**
  - A more general way is to model a relation as **a mapping $r : \mathbb{R}^n \rightarrow \{0, 1\}$, where $\mathbb{R}^n$ is an appropriate object feature space**

Defining (Spatial) Relations

- We can define relations (spatial, but also relations in general) in multiple ways:
  - In a first-order logic sense, a relation can be seen as **a relation $r : \mathbb{O}^n \rightarrow \{\textbf{true}, \textbf{false}\}$, where $O \in \mathbb{O}$ represents an object**
  - A more general way is to model a relation as **a mapping $r : \mathbb{R}^n \rightarrow \{0, 1\}$, where $\mathbb{R}^n$ is an appropriate object feature space**
  - An even more general definition is to consider a relation as **a mapping from a feature space to a probability value $r : \mathbb{R}^n \rightarrow \mathbb{R}$**

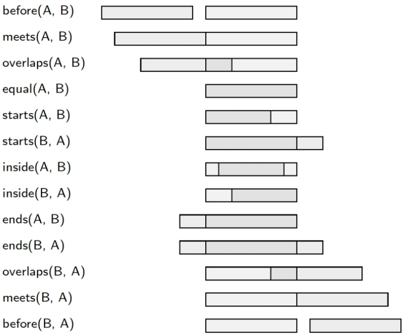
Defining (Spatial) Relations

- We can define relations (spatial, but also relations in general) in multiple ways:
  - In a first-order logic sense, a relation can be seen as **a relation $r : \mathbb{O}^n \rightarrow \{\textbf{true}, \textbf{false}\}$, where $O \in \mathbb{O}$ represents an object**
  - A more general way is to model a relation as **a mapping $r : \mathbb{R}^n \rightarrow \{0, 1\}$, where $\mathbb{R}^n$ is an appropriate object feature space**
  - An even more general definition is to consider a relation as **a mapping from a feature space to a probability value $r : \mathbb{R}^n \rightarrow \mathbb{R}$**
  - In the above cases, a relation can really only take two possible values (true and false), but **it may make sense for a spatial relation to take an arbitrary number of qualitative values, so we can define $r : \mathbb{R}^n \rightarrow \mathbb{Z}$**

Qualitative Calculi

- When defining spatial relations, a common question is **which relations to use**, or a related question, **is there a minimal set of relations that can be defined**?
- In most robotics applications, commonly used natural language relations are used, such as **on**, **behind**, **in**, or **to the left of**
- In the context of spatial relations, there are also various **qualitative spatial calculi**, which specify a nominal set of well-defined spatial relations
  - Using a qualitative spatial calculus, common spatial relations can be defined through combinations of the nominal relations
- Let us briefly take a look at a few such calculi

Rectangle Algebra

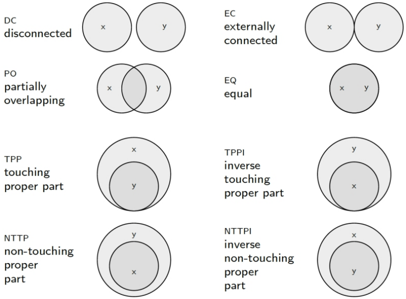


C. Landsiedel et al., "A review of spatial reasoning and interaction for real-world robotics," *Advanced Robotics*, vol. 31, no. 5, pp. 222–242, 2017.

- One qualitative calculus is the **rectangle algebra** system, which is based on **Allen's interval relations***
    - These relations were originally introduced in the context of temporal intervals
- According to this system, there are **13 relations** that are defined for pairs of rectangles
    - Seven of these are unique, while the remaining six are inverse relations
- The use of rectangle algebra requires **two-dimensional projections of axis-aligned object bounding boxes**; the relations can then be computed over these projections

* James F. Allen "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, pp. 832–843, 1983.

Region Connection Calculus (RCC8)

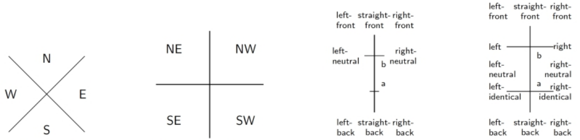


C. Landsiedel et al., "A review of spatial reasoning and interaction for real-world robotics," *Advanced Robotics*, vol. 31, no. 5, pp. 222–242, 2017.

- Another spatial relation calculus is the **region connection calculus (RCC8)**
- The RCC8 system specifies **8 relations** that are **defined for two-dimensional spatial regions**
  - Six of these are unique; the other two relations are inverses
- Just as in the case of the rectangle algebra, the use of RCC8 requires **two-dimensional projections of object bounding boxes**, and the relations are computed over the projections

Qualitative Orientation Representations

- Both the rectangle calculus and RCC8 define relations without considering orientations, but there are also qualitative calculi that can be used to define orientation relations



(a) Cone-based direction calculus (b) Projection-based direction calculus (c) Single Cross Calculus (d) Double Cross Calculus

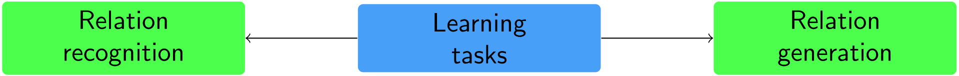
C. Landsiedel et al., "A review of spatial reasoning and interaction for real-world robotics," *Advanced Robotics*, vol. 31, no. 5, pp. 222–242, 2017.

Spatial Relations and Frames of Reference

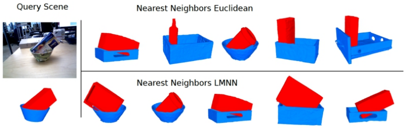
- The definition of spatial relations is **dependent on a well-defined reference frame**
  - In other words, relations that hold in one reference frame may not hold in another frame
- For instance, the cup in the example figure is to the left of the plate with respect to the right-handed coordinate frame of the robot's base, but what if we consider the frame of the camera on the robot's head?

Spatial Relations Learning Overview

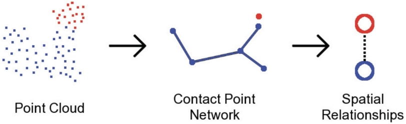
Learning Tasks



Relation Recognition



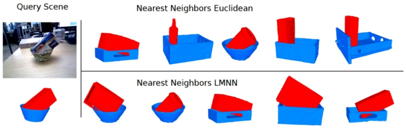
O. Mees et al., "Metric learning for generalizing spatial relations to new objects," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2017, pp. 3175–3182.



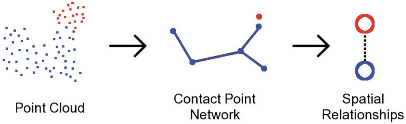
B. Rosman and S. Ramamoorthy, "Learning spatial relationships between objects," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1328–1342, 2011.

- The problem of relation recognition is that of **identifying which relations hold between a pair of objects**

Relation Recognition



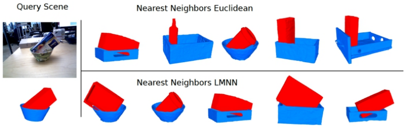
O. Mees et al., "Metric learning for generalizing spatial relations to new objects," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2017, pp. 3175–3182.



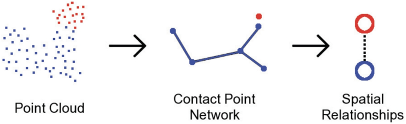
B. Rosman and S. Ramamoorthy, "Learning spatial relationships between objects," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1328–1342, 2011.

- The problem of relation recognition is that of **identifying which relations hold between a pair of objects**
- Typically defined as a **classification problem**, namely a classifier $f : (O, O) \to C$ is learned for objects O_1 and O_2 given in some representation (e.g. a point cloud), where $c \in C$ is a relation
  - Relation recognition is thus a visual classification task

Relation Recognition



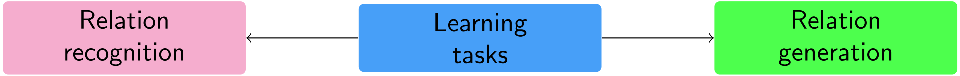
O. Mees et al., "Metric learning for generalizing spatial relations to new objects," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2017, pp. 3175–3182.



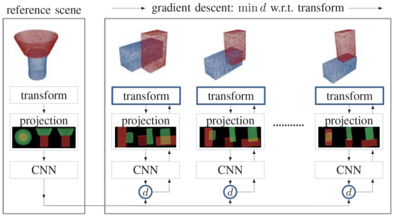
B. Rosman and S. Ramamoorthy, "Learning spatial relationships between objects," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1328–1342, 2011.

- The problem of relation recognition is that of **identifying which relations hold between a pair of objects**
- Typically defined as a **classification problem**, namely a classifier $f : (O, O) \rightarrow C$ is learned for objects O_1 and O_2 given in some representation (e.g. a point cloud), where $c \in C$ is a relation
  - Relation recognition is thus a visual classification task
- **Generally a well-studied problem** in the literature, with a variety of solution approaches proposed

Learning Tasks



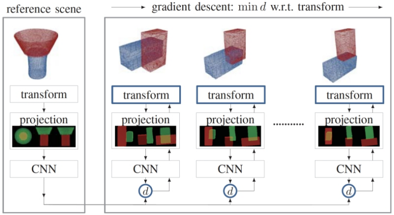
Relation Generation



P. Jund et al., “Optimization Beyond the Convolution: Generalizing Spatial Relations with End-to-End Metric Learning,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2018, pp. 4510–4516.

- The problem of relation generation is that of **finding a object configuration in which a set of objects will end up in a desired spatial relation**

Relation Generation



P. Jund et al., "Optimization Beyond the Convolution: Generalizing Spatial Relations with End-to-End Metric Learning," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2018, pp. 4510–4516.

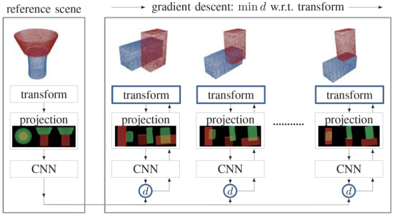
- The problem of relation generation is that of **finding a object configuration in which a set of objects will end up in a desired spatial relation**
- Given **a desired relation** r^*, as well as an **anchor object** O_a (which is static) and a **target object** O_t (which is manipulated by a robot), the problem can be expressed as an optimisation objective

  find a pose $P^*_{O_a}$

  such that
  $$\frac{1}{r^*_N}\sum_{i=1}^{r^*_N}\left(g(P^*_{O_a}, P_{O_t}) - g(P_{O_{i_1}}, P_{O_{i_2}})\right)^2 \geq \frac{1}{r^*_N}\sum_{i=1}^{r^*_N}\left(g(P^k_{O_a}, P_{O_t}) - g(P_{O_{i_1}}, P_{O_{i_2}})\right)^2, \forall P^k_{O_a} \neq P^*_{O_a}$$

  where g is some feature extractor

Relation Generation



P. Jund et al., "Optimization Beyond the Convolution: Generalizing Spatial Relations with End-to-End Metric Learning," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2018, pp. 4510–4516.

- The problem of relation generation is that of **finding a object configuration in which a set of objects will end up in a desired spatial relation**

- Given **a desired relation** r^*, as well as an **anchor object** O_a (which is static) and a **target object** O_t (which is manipulated by a robot), the problem can be expressed as an optimisation objective

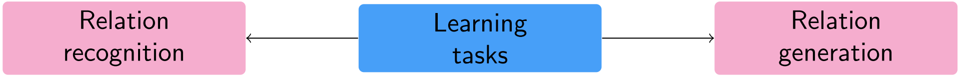
  find a pose $\quad P^*_{O_a}$

  such that
  $$\frac{1}{r^*_N}\sum_{i=1}^{r^*_N}\left(g(P^*_{O_a}, P_{O_t}) - g(P_{O_{i_1}}, P_{O_{i_2}})\right)^2 \geq \frac{1}{r^*_N}\sum_{i=1}^{r^*_N}\left(g(P^k_{O_a}, P_{O_t}) - g(P_{O_{i_1}}, P_{O_{i_2}})\right)^2, \; \forall P^k_{O_a} \neq P^*_{O_a}$$

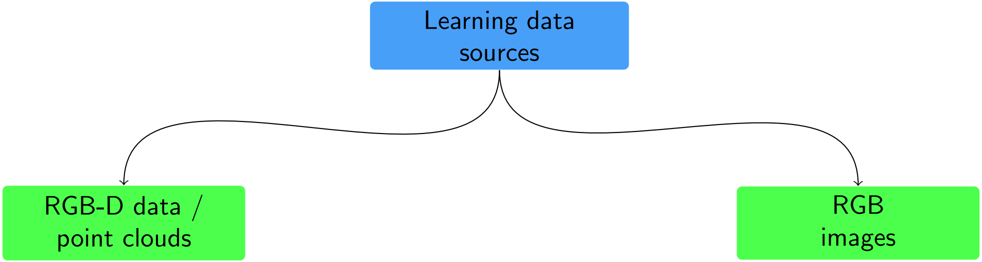
  where g is some feature extractor

- In words, we are looking for a configuration that minimises the similarity to known examples of r^*

Learning Tasks



Learning Data Sources



Learning Based on RGB-D Data

- Spatial relation learning based on RGB-D data / point clouds is very common in the literature — the examples of relation recognition and generation systems that we saw on the previous slides are all based on point cloud-based object representations

Learning Based on RGB-D Data

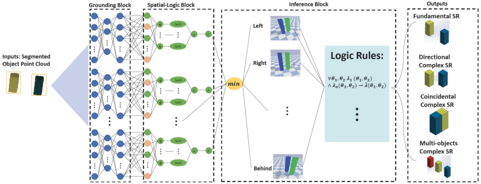
- Spatial relation learning based on RGB-D data / point clouds is very common in the literature — the examples of relation recognition and generation systems that we saw on the previous slides are all based on point cloud-based object representations
- Point cloud data is useful when considering spatial relations because **understanding spatial relations generally requires 3D information**
  - Due to perspective, pure 2D information can sometimes be misleading

Learning Based on RGB-D Data

- Spatial relation learning based on RGB-D data / point clouds is very common in the literature — the examples of relation recognition and generation systems that we saw on the previous slides are all based on point cloud-based object representations
- Point cloud data is useful when considering spatial relations because **understanding spatial relations generally requires 3D information**
  - Due to perspective, pure 2D information can sometimes be misleading
- In this case, **each object $O \in \mathbb{O}$ is represented by a point cloud**
  - When considering objects that are present in a scene, **object recognition typically precedes the point cloud extraction step**

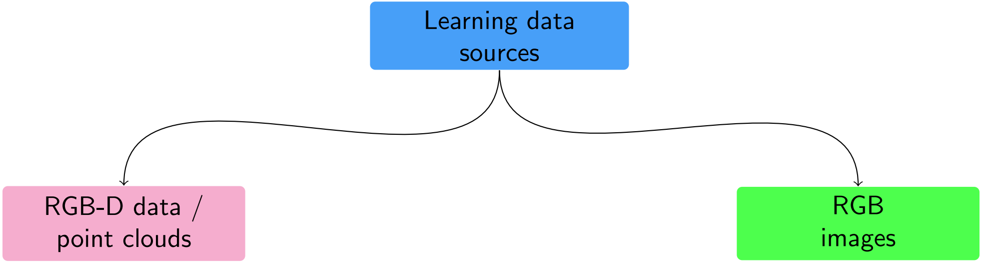
Learning Based on RGB-D Data

- Spatial relation learning based on RGB-D data / point clouds is very common in the literature — the examples of relation recognition and generation systems that we saw on the previous slides are all based on point cloud-based object representations
- Point cloud data is useful when considering spatial relations because **understanding spatial relations generally requires 3D information**
  - Due to perspective, pure 2D information can sometimes be misleading
- In this case, **each object $O \in \mathbb{O}$ is represented by a point cloud**
  - When considering objects that are present in a scene, **object recognition typically precedes the point cloud extraction step**
- In pre-deep learning-based systems, **point cloud features were extracted in a preprocessing step**; deep learning-based architectures tend to **process point clouds in an end-to-end fashion**

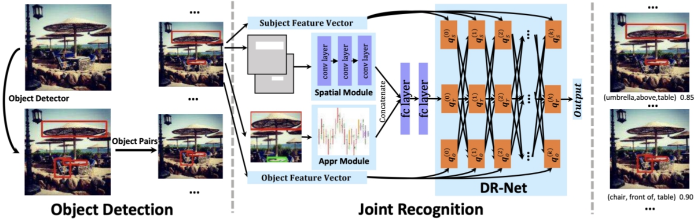


F. Yan, D. Wang and H. He, "Robotic Understanding of Spatial Relationships Using Neural-Logic Learning," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2020, pp. 8358–8365.

Learning Data Sources



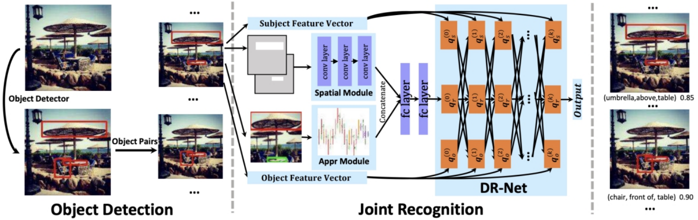
Learning Based on RGB Images



B. Dai, Y. Zhang and D. Lin, "Detecting Visual Relationships with Deep Relational Networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3298–3308.

- RGB images have also occasionally been used for spatial relation learning, but so far only for recognising relations
  - Generation only based on images is an ill-defined problem — again, due to perspective

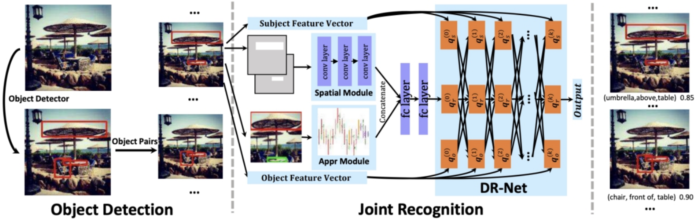
Learning Based on RGB Images



B. Dai, Y. Zhang and D. Lin, "Detecting Visual Relationships with Deep Relational Networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3298–3308.

- RGB images have also occasionally been used for spatial relation learning, but so far only for recognising relations
  - Generation only based on images is an ill-defined problem — again, due to perspective
- Most network architectures for relation recognition use **customised relation encoders** (in the example illustrated above, based on conditional random fields)

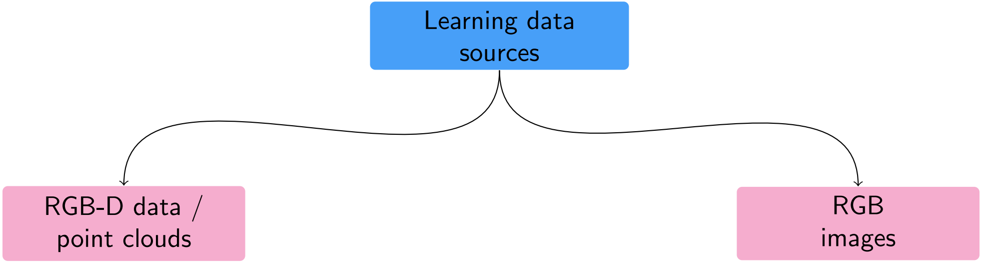
Learning Based on RGB Images



B. Dai, Y. Zhang and D. Lin, "Detecting Visual Relationships with Deep Relational Networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3298–3308.

- RGB images have also occasionally been used for spatial relation learning, but so far only for recognising relations
  - Generation only based on images is an ill-defined problem — again, due to perspective
- Most network architectures for relation recognition use **customised relation encoders** (in the example illustrated above, based on conditional random fields)
- Image-based learning of relations is often considered in the context of **visual question answering**

Learning Data Sources



Summary

- Spatial relations are qualitative representations, typically defined for pairs of objects, of some aspects of space
- There are various qualitative spatial calculi that define a nominal set of relations based on which more complex relations can be defined
- Spatial relation learning can be performed for the purpose of recognition (a classification problem) or generation (an optimisation problem of reproducing a desired relation)
- Point clouds are typically used for learning how to recognise and generate spatial relations; learning based on RGB images is done as well, but only for the recognition problem