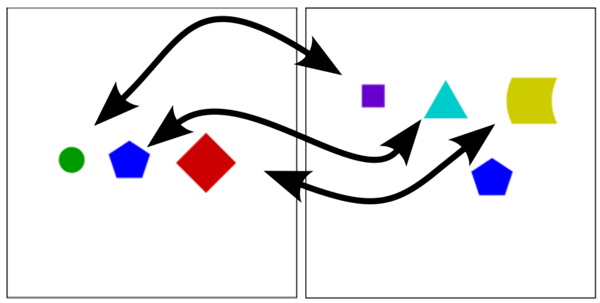
Visual analogical mapping

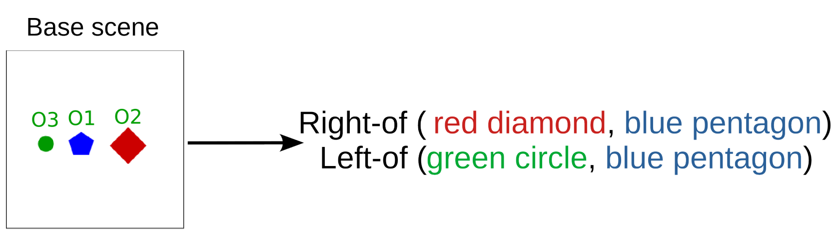
DFT summer School 2024

1 Project overview



The goal of the project is to build a DFT architecture that can make analogical mapping across two scenes. Concretely, this means that the architecture takes two images as inputs and builds a representation of which objects correspond to each other across the scenes based on relational similarities. The architecture can be built incrementally based on cognitive operations introduced in the lecture. We will divide the process of building the architecture into three parts: In the first part, we build an architecture that takes an image input and generates a conceptual *description* of that image by selectively activating a neural population representing the relational structure of the image. In the second part, we extend the architecture such that it can use the description to *guide the spatial selection* of objects in another image. The final part consists of changing the architecture to make true analogical mapping, i.e.) mapping based on relational similarities while disregarding featural dissimilarities.

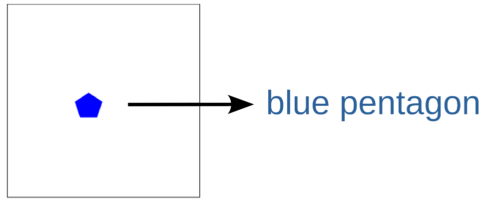
2 Describing a single scene



The goal of the first part of the project is to build an architecture generating a neural representation of the relational structure of the image input.

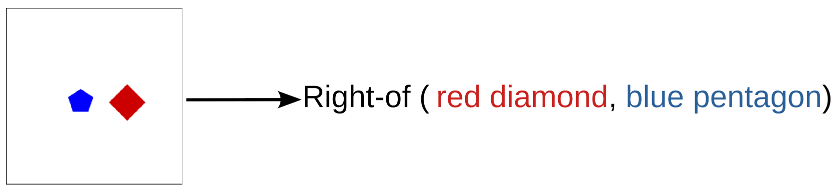
2.1 Describing a single object

First, build an architecture that activates the feature concepts of the spatially selected object. The given .json file already entails the visual front-end which pre-processes the image input and selects a location in the *spatial selection field* based on saliency. While the visual front end extracts the color,



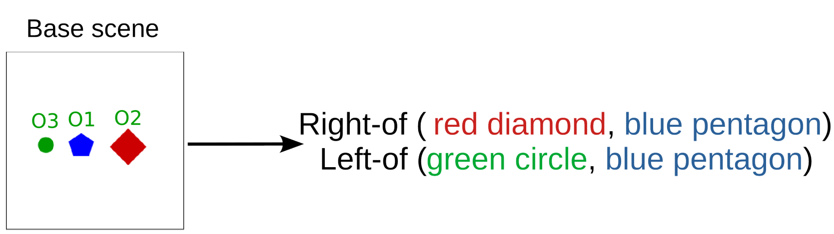
shape, and size of objects in the image, you can work with any number of features as you wish. Extend the architecture to extract the feature value of the object at the selected location. Once the feature value of the object is extracted, connect the field representing the feature value to feature concept nodes, which represent the extracted feature categorically. This requires setting up the connectivity in such a way that the node is a good prototype of the concept it is supposed to represent.

2.2 Describing the relation between a pair of objects



Next, extend the architecture to extract the relation between a pair of objects. First, change the architecture such that another object can be spatially selected once the currently selected object is described. The *IoR field* keeps sustained activation of previously selected locations and hinders the architecture from selecting the same location again. Selecting multiple objects sequentially creates a problem that the feature concepts of the first and the second objects cannot be distinguished, which we will address in the next step. To extract the spatial relation between two objects, the spatial location of the reference object and the target object needs to be represented in separate fields. Implement the coordinate transformation to represent the target object location in the coordinate centered at the reference object location. Finally, connect the field representing the centered target object to spatial relation concepts.

2.3 Describing multiple objects and relations



The final step is to represent the relational structure that keeps the identity of objects and relations distinct while allowing for flexible binding of objects to relations and roles.

In DFT, the core ingredient for such a representation is the ordinal binding dimension (first, second, third, ...). Take a look at the lecture explaining how sequences are generated and how this can assign ordinal values for each instance of objects and relations. There exist separate ordinal dimensions for objects and relations. The architecture needs to generate the ordinal dimension values autonomously. For the objects, the next order is given whenever the features of the currently spatially selected object are described and a new object is selected. For the relations, the next order is given whenever the relation between a different pair of objects is being described.

blue pentagon - O1	Right-of - R1	R1 - (O1 - Ref.) - (O2- Tar.)
red diamond - O2	Left-of - R2	R2 - (O1 - Ref.) - (O3- Tar.)
green circle - O3		

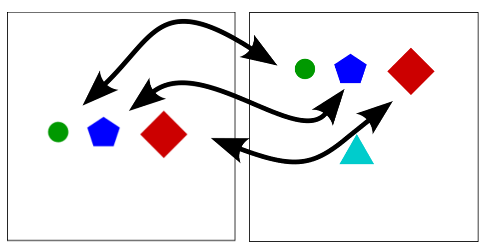
Once the architecture autonomously generates the ordinal dimension values, the joint representation of the concepts and the ordinal dimension can be built. Use two separate fields jointly representing the object feature concepts with the object ordinal dimension (blue pentagon - O1, ...) and the relation concepts with the relation ordinal dimension (Right-of - R1, ...).

Also, create another set of fields jointly representing the relation ordinal dimension, the object ordinal dimension, and the roles of the objects (R1 - (O1 - Ref.) - (O2 - Tar.), ...). As a heuristic, we posit that the currently spatially selected object always plays the target role. This means that the object ordinal value of the target object is simply the current ordinal value. The object ordinal value of the reference object must depend on which object is selected as the reference object. For the most simple case, you can just leave the reference object to be the first selected object (O1). Alternatively, the reference object can be selected on various factors such as its spatial proximity to the target object, and how recently it has been described.

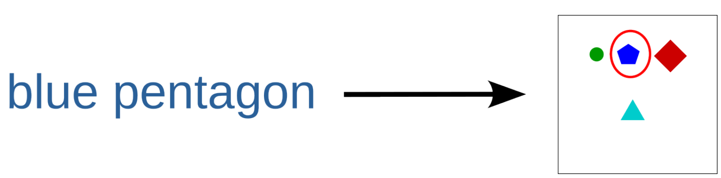
Finally, create a mental map jointly representing the spatial location of selected objects with the ordinal dimension. This mental map will ultimately represent the mapping between objects across scenes. It also contains information about how recent which object has been described.

3 Grounding the description in another scene

We have now built a cognitive architecture that takes an image input and generates a description of the input by selectively activating feature and relation concepts. The next step is to extend the architecture such that it can use the description to spatially select objects matching the description in another scene. For the current purpose, we use an image that is almost identical to the base scene, except that there is a distractor and that objects found in the base scene are not placed at different locations.



3.1 Featural guidance of single object search



The architecture can only attend to one object at a time due to the attentional bottleneck. This means that the architecture needs to select the description of a single object to search for. To keep the memory of the conceptual description and the currently selected description separate, create a copy of fields keeping the conceptual representation of the base scene. The selection of objects to be searched can simply follow the serial order of the description. Create another sequence generator to select one ordinal value at a time sequentially. The selected ordinal value should also selectively activate the feature concepts of the object with that ordinal value. Next, ensure that the selected feature concepts guide the search. This generally follows the opposite direction of the processing order of description. First, the selected feature concepts should create peaks in fields defined over featural dimensions. Next, these peaks representing feature values should create peaks in a field defined over spatial dimensions, ultimately providing input to the *spatial selection field*.

3.2 Relational guidance

We now want to extend the architecture so that the relation between a pair of objects can also guide the search. In the project, we limit our search to be structured such that the object that we are currently searching for always takes the target object role of a relation. From the description of the whole base scene, the architecture needs to select the relation that has the object that matches the currently searched object showing up as the target object. For example, when searching for the red diamond, the relation that the red diamond was to the right of the blue pentagon in the base scene needs to be selected. Also, the architecture needs to select the right reference object of the selected relation. This requires keeping a separate mental map of the target scene, representing the spatial location of all the objects that have already been selected. Once the relevant relational concept and the reference object have been selected, coordinate transformation can guide the spatial attention to the location at which the target object could be.

4 Advanced steps for analogical mapping

Consult the tutors to extend the architecture to make analogical mapping. This requires a process of deciding when to accept or reject the current mapping, organizing the processes to act based on the decision, and allowing featural mismatches between objects across scenes.