# Representation and Retrieval of Images using Spatial Relations

Paulo E. Santos

joint work with Danilo Nunes, Leonardo Anjoletto and Adam Pease

March 24, 2019

# Introduction

- Integrating low-level information (NN) with high-level knowledge (Spatial Relations)
- Image Retrieval using Region Analysis (IRRA) method ensembles a stack of distinct neural networks in order to estimate spatial relations/prepositions between pairs of objects.
- permits a representation of an image by the objects depicted and the relations holding between them.
- Suggested Upper-Level Ontology (SUMO) is used to infer new relations beyond the original binary relations and allows retrieval of images based on queries with respect to spatial arrangements.
- Results for an indoor/outdoor classifier shows that neural networks alone are capable of achieving 88% in precision and recall, but when combined with ontology this result increases in 10 percentage points, reaching 98% of precision and recall.

# Introduction

- Integrating low-level information (NN) with high-level knowledge (Spatial Relations)
- Image Retrieval using Region Analysis (IRRA) method ensembles a stack of distinct neural networks in order to estimate spatial relations/prepositions between pairs of objects.
- permits a representation of an image by the objects depicted and the relations holding between them.
- Suggested Upper-Level Ontology (SUMO) is used to infer new relations beyond the original binary relations and allows retrieval of images based on queries with respect to spatial arrangements.
- Results for an indoor/outdoor classifier shows that neural networks alone are capable of achieving 88% in precision and recall, but when combined with ontology this result increases in 10 percentage points, reaching 98% of precision and recall.

# Introduction

- Integrating low-level information (NN) with high-level knowledge (Spatial Relations)
- Image Retrieval using Region Analysis (IRRA) method ensembles a stack of distinct neural networks in order to estimate spatial relations/prepositions between pairs of objects.
- permits a representation of an image by the objects depicted and the relations holding between them.
- Suggested Upper-Level Ontology (SUMO) is used to infer new relations beyond the original binary relations and allows retrieval of images based on queries with respect to spatial arrangements.
- Results for an indoor/outdoor classifier shows that neural networks alone are capable of achieving 88% in precision and recall, but when combined with ontology this result increases in 10 percentage points, reaching 98% of precision and recall.

# Introduction

- Integrating low-level information (NN) with high-level knowledge (Spatial Relations)
- Image Retrieval using Region Analysis (IRRA) method ensembles a stack of distinct neural networks in order to estimate spatial relations/prepositions between pairs of objects.
- permits a representation of an image by the objects depicted and the relations holding between them.
- Suggested Upper-Level Ontology (SUMO) is used to infer new relations beyond the original binary relations and allows retrieval of images based on queries with respect to spatial arrangements.
- Results for an indoor/outdoor classifier shows that neural networks alone are capable of achieving 88% in precision and recall, but when combined with ontology this result increases in 10 percentage points, reaching 98% of precision and recall.

# Introduction

- Integrating low-level information (NN) with high-level knowledge (Spatial Relations)
- Image Retrieval using Region Analysis (IRRA) method ensembles a stack of distinct neural networks in order to estimate spatial relations/prepositions between pairs of objects.
- permits a representation of an image by the objects depicted and the relations holding between them.
- Suggested Upper-Level Ontology (SUMO) is used to infer new relations beyond the original binary relations and allows retrieval of images based on queries with respect to spatial arrangements.
- Results for an indoor/outdoor classifier shows that neural networks alone are capable of achieving 88% in precision and recall, but when combined with ontology this result increases in 10 percentage points, reaching 98% of precision and recall.

# Motivation

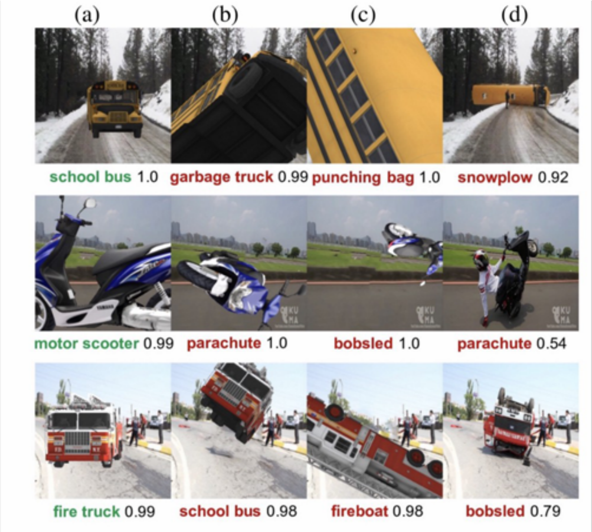



Figure: [Michael Alcorn et al., 2019]

# Image Retrieval Using Region Analysis (IRRA)

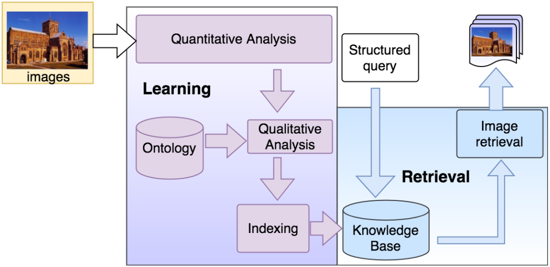



Figure: Overview of IRRA method numerical analysis phase.

# Connectionist Processing

1. Identify the context in which objects are inserted;
2. Detect regions in the images occupied by each object;
3. Combine the detected objects in pairs;
4. Estimate a spatial relation for every detected object pair.

In order to perform all these tasks, a stack of neural networks is applied. We apply three distinct neural networks to hierarchically detect context, objects and spatial relations.

# Connectionist Processing

1. Identify the context in which objects are inserted;
2. Detect regions in the images occupied by each object;
3. Combine the detected objects in pairs;
4. Estimate a spatial relation for every detected object pair.

In order to perform all these tasks, a stack of neural networks is applied. We apply three distinct neural networks to hierarchically detect context, objects and spatial relations.

# Connectionist Processing

1. Identify the context in which objects are inserted;
2. Detect regions in the images occupied by each object;
3. Combine the detected objects in pairs;
4. Estimate a spatial relation for every detected object pair.

In order to perform all these tasks, a stack of neural networks is applied. We apply three distinct neural networks to hierarchically detect context, objects and spatial relations.

# Connectionist Processing

1. Identify the context in which objects are inserted;
2. Detect regions in the images occupied by each object;
3. Combine the detected objects in pairs;
4. Estimate a spatial relation for every detected object pair.

In order to perform all these tasks, a stack of neural networks is applied. We apply three distinct neural networks to hierarchically detect context, objects and spatial relations.

# Connectionist Processing

1. Identify the context in which objects are inserted;
2. Detect regions in the images occupied by each object;
3. Combine the detected objects in pairs;
4. Estimate a spatial relation for every detected object pair.

In order to perform all these tasks, a stack of neural networks is applied. We apply three distinct neural networks to hierarchically detect context, objects and spatial relations.

# Context identification

- A first neural network is applied in order to estimate context.
- *indoor* and *outdoor*.
- AlexNet (Krizhevsky et al., 2012) in order to perform this classification.
- The purpose of this step is to reduce the number of target objects, dismissing inconsistent objects with respect to the scene;
  - objects such as buildings and cars will not be part of the segmentation of an indoor model, for instance.

# Context identification

- A first neural network is applied in order to estimate context.
- *indoor* and *outdoor*.
- AlexNet (Krizhevsky et al., 2012) in order to perform this classification.
- The purpose of this step is to reduce the number of target objects, dismissing inconsistent objects with respect to the scene;
    - objects such as buildings and cars will not be part of the segmentation of an indoor model, for instance.

# Context identification

- A first neural network is applied in order to estimate context.
- *indoor* and *outdoor*.
- AlexNet (Krizhevsky et al., 2012) in order to perform this classification.
- The purpose of this step is to reduce the number of target objects, dismissing inconsistent objects with respect to the scene;
  - objects such as buildings and cars will not be part of the segmentation of an indoor model, for instance.

# Context identification

- A first neural network is applied in order to estimate context.
- *indoor* and *outdoor*.
- AlexNet (Krizhevsky et al., 2012) in order to perform this classification.
- The purpose of this step is to reduce the number of target objects, dismissing inconsistent objects with respect to the scene;
  - objects such as buildings and cars will not be part of the segmentation of an indoor model, for instance.

# Context identification

- A first neural network is applied in order to estimate context.
- *indoor* and *outdoor*.
- AlexNet (Krizhevsky et al., 2012) in order to perform this classification.
- The purpose of this step is to reduce the number of target objects, dismissing inconsistent objects with respect to the scene;
  - objects such as buildings and cars will not be part of the segmentation of an indoor model, for instance.

# Object segmentation

- The segmentation neural networks generate a set of class proposals for each pixel in the image.
- similar pixels are grouped together in order to represent objects.
- spatial relations are assigned between pairs of these segmented objects.
  - mapping pairs of detected objects: topological terms.
  - another neural network : estimate a spatial preposition from the previous (topological) classification.

# Object segmentation

- The segmentation neural networks generate a set of class proposals for each pixel in the image.
- similar pixels are grouped together in order to represent objects.
- spatial relations are assigned between pairs of these segmented objects.
  - mapping pairs of detected objects: topological terms.
  - another neural network : estimate a spatial preposition from the previous (topological) classification.

# Object segmentation

- The segmentation neural networks generate a set of class proposals for each pixel in the image.
- similar pixels are grouped together in order to represent objects.
- spatial relations are assigned between pairs of these segmented objects.
  - mapping pairs of detected objects: topological terms.
  - another neural network : estimate a spatial preposition from the previous (topological) classification.

# Object segmentation

- The segmentation neural networks generate a set of class proposals for each pixel in the image.
- similar pixels are grouped together in order to represent objects.
- spatial relations are assigned between pairs of these segmented objects.
  - mapping pairs of detected objects: topological terms.
  - another neural network : estimate a spatial preposition from the previous (topological) classification.

# Symbolic Processing

We use SUMO (Niles and Pease, 2001) to represent some aspects of the knowledge acquired through the Connectionist Processing.

# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

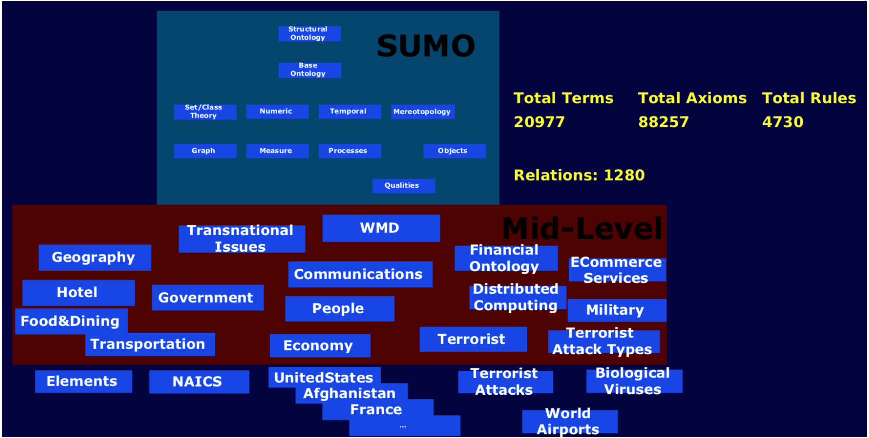
# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

# Suggested Upper Merged Ontology - SUMO

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
  - Now linked with factbases including YAGO for millions of facts
- Free - GNU GPL for ontology and tools
- Reuse of the SUMO library means not having to recreate and define spatial labels - and not have to redo our work when scope grows to more objects and relations
- www.ontologyportal.org

# Suggested Upper Merged Ontology - SUMO



# Symbolic Processing

- create instances of the domain.
- each detected object is considered as an unique and independent instance.
- For example: image $I_x$ with the detected objects *Building, Floor* and *Sky*:
  (instance $Building_x$ Building)
  (instance $Floor_x$ Floor)
  (instance $Sky_x$ Sky)

# Symbolic Processing

- create instances of the domain.
- each detected object is considered as an unique and independent instance.
- For example: image $I_x$ with the detected objects *Building, Floor* and *Sky*:
  (instance $Building_x$ Building)
  (instance $Floor_x$ Floor)
  (instance $Sky_x$ Sky)

# Symbolic Processing

- create instances of the domain.
- each detected object is considered as an unique and independent instance.
- For example: image $I_x$ with the detected objects *Building, Floor* and *Sky*:
  (instance $Building_x$ Building)
  (instance $Floor_x$ Floor)
  (instance $Sky_x$ Sky)

# Experiments- Datasets used

SUN09 (Choi et al. 2010): over 12,000 images containing various classes of objects in distinct scenes.

Two sets of annotations for SUN09 were used:

- data set 1: annotations in the form of structured queries ($\langle$*noun*, *preposition*, *noun*$\rangle$) representing two relations *below* and *above*.
- data set 2: the annotations that includes eleven (11) distinct preposition classes: *above, across from, behind, below, in, in front of, inside, left, right, on* and *under*.

Overall 4,367 images were used for training and 4,317 images for testing. In these datasets there are 186,299 pairs of objects for training and 173,111 for testing.

# Experiments- Datasets used

SUN09 (Choi et al. 2010): over 12,000 images containing various classes of objects in distinct scenes.
Two sets of annotations for SUN09 were used:

- data set 1: annotations in the form of structured queries ($\langle$*noun*, *preposition*, *noun*$\rangle$) representing two relations *below* and *above*.
- data set 2: the annotations that includes eleven (11) distinct preposition classes: *above, across from, behind, below, in, in front of, inside, left, right, on* and *under*.

Overall 4,367 images were used for training and 4,317 images for testing. In these datasets there are 186,299 pairs of objects for training and 173,111 for testing.

# Experiments- Datasets used

SUN09 (Choi et al. 2010): over 12,000 images containing various classes of objects in distinct scenes.
Two sets of annotations for SUN09 were used:

- data set 1: annotations in the form of structured queries ($\langle$*noun*, *preposition*, *noun*$\rangle$) representing two relations *below* and *above*.
- data set 2: the annotations that includes eleven (11) distinct preposition classes: *above, across from, behind, below, in, in front of, inside, left, right, on* and *under*.

Overall 4,367 images were used for training and 4,317 images for testing. In these datasets there are 186,299 pairs of objects for training and 173,111 for testing.

# Experiments- Datasets used

SUN09 (Choi et al. 2010): over 12,000 images containing various classes of objects in distinct scenes.

Two sets of annotations for SUN09 were used:

- data set 1: annotations in the form of structured queries ($\langle$*noun*, *preposition*, *noun*$\rangle$) representing two relations *below* and *above*.
- data set 2: the annotations that includes eleven (11) distinct preposition classes: *above, across from, behind, below, in, in front of, inside, left, right, on* and *under*.

Overall 4,367 images were used for training and 4,317 images for testing. In these datasets there are 186,299 pairs of objects for training and 173,111 for testing.

# Experiments- Datasets used

SUN09 (Choi et al. 2010): over 12,000 images containing various classes of objects in distinct scenes.
Two sets of annotations for SUN09 were used:

- data set 1: annotations in the form of structured queries ($\langle$*noun*, *preposition*, *noun*$\rangle$) representing two relations *below* and *above*.
- data set 2: the annotations that includes eleven (11) distinct preposition classes: *above, across from, behind, below, in, in front of, inside, left, right, on* and *under*.

Overall 4,367 images were used for training and 4,317 images for testing. In these datasets there are 186,299 pairs of objects for training and 173,111 for testing.

# Experiment 1- slowly adding high-level knowledge

1. results of this binary classification, using only a NN
2. whether the information provided by the scene identification improves the segmentation.
   - manually separated the objects as indoor and outdoor. According to each scene class detected we apply one or the other segmentation models (i.e. one trained with *indoor* objects or the other trained with *outdoor objects*).
3. a third *indoor*/*outdoor* classifier was tested, in which we combine low-level information of the scene with high-level information provided by the description of the objects in SUMO.

# Experiment 1- slowly adding high-level knowledge

1. results of this binary classification, using only a NN
2. whether the information provided by the scene identification improves the segmentation.
   - manually separated the objects as indoor and outdoor. According to each scene class detected we apply one or the other segmentation models (i.e. one trained with *indoor* objects or the other trained with *outdoor objects*).
3. a third *indoor*/*outdoor* classifier was tested, in which we combine low-level information of the scene with high-level information provided by the description of the objects in SUMO.

# Experiment 1- slowly adding high-level knowledge

1. results of this binary classification, using only a NN
2. whether the information provided by the scene identification improves the segmentation.
   - manually separated the objects as indoor and outdoor. According to each scene class detected we apply one or the other segmentation models (i.e. one trained with *indoor* objects or the other trained with *outdoor objects*).
3. a third *indoor*/*outdoor* classifier was tested, in which we combine low-level information of the scene with high-level information provided by the description of the objects in SUMO.

# Experiment 1- slowly adding high-level knowledge

1. results of this binary classification, using only a NN
2. whether the information provided by the scene identification improves the segmentation.
   - manually separated the objects as indoor and outdoor. According to each scene class detected we apply one or the other segmentation models (i.e. one trained with *indoor* objects or the other trained with *outdoor objects*).
3. a third *indoor*/*outdoor* classifier was tested, in which we combine low-level information of the scene with high-level information provided by the description of the objects in SUMO.

# Experiment 1- slowly adding high-level knowledge

Table: Precision-recall for scene classification.

| Scene | Precision | Recall | N |
|---|---|---|---|
| Indoor | 0.86 | 0.84 | 1,829 |
| Outdoor | 0.88 | 0.90 | 2,439 |
| Overall | 0.87 | 0.87 | 4,268 |

# Experiment 1- slowly adding high-level knowledge

**With Context**

Table: Intersection over union.

| Scene | Intersection Over Union % | |
|---|---|---|
| | Context | No context |
| Outdoor | 24.11 | 18.60 |
| Indoor | 16.57 | 14.07 |
| **Average** | **22.66** | **18.67** |

# Experiment 1- slowly adding high-level knowledge

**With SUMO Annotation**

Table: Precision-recall for classification with ontology.

| Scene | Precision | Recall | N |
|---|---|---|---|
| Indoor | 0.98 | 0.98 | 1,813 |
| Outdoor | 0.99 | 0.98 | 2,463 |
| Overall | 0.99 | 0.98 | 4,320 |

# Experiment 2- playing with spatial prepositions

- evaluate image retrieval using the annotations in [Lan et al., 2012]
- tested our method against all the structured query types proposed.
- The structured queries contain a *noun*, e.g. *pedestrians*, or a *relation* set expressed by a triple in the form *(noun, preposition, and noun)*, e. g. *"car on the road"*.
  - *Structure a (Sa)*, which contains only a relation, for instance, *"car on road"*;
  - *Structure b (Sb)* contains a relation and a noun, e.g, *"car on road, pedestrians"*;
  - *Structure c (Sc)* contains two relations, e.g., *"car on road, sky above building"*;
  - *Structure d (Sd)* contains two relations and a noun, e.g., *"car on road, sky above building, pedestrians"*;
  - *Structure e (Se)* contains three relations, e.g., *"car on road, sky above building, books inside of bookcase"*.

# Experiment 2- playing with spatial prepositions

- evaluate image retrieval using the annotations in [Lan et al., 2012]
- tested our method against all the structured query types proposed.
- The structured queries contain a *noun*, e.g. *pedestrians*, or a *relation* set expressed by a triple in the form *(noun, preposition, and noun)*, e. g. *"car on the road"*.
  - *Structure a (Sa)*, which contains only a relation, for instance, *"car on road"*;
  - *Structure b (Sb)* contains a relation and a noun, e.g, *"car on road, pedestrians"*;
  - *Structure c (Sc)* contains two relations, e.g., *"car on road, sky above building"*;
  - *Structure d (Sd)* contains two relations and a noun, e.g., *"car on road, sky above building, pedestrians"*;
  - *Structure e (Se)* contains three relations, e.g., *"car on road, sky above building, books inside of bookcase"*.

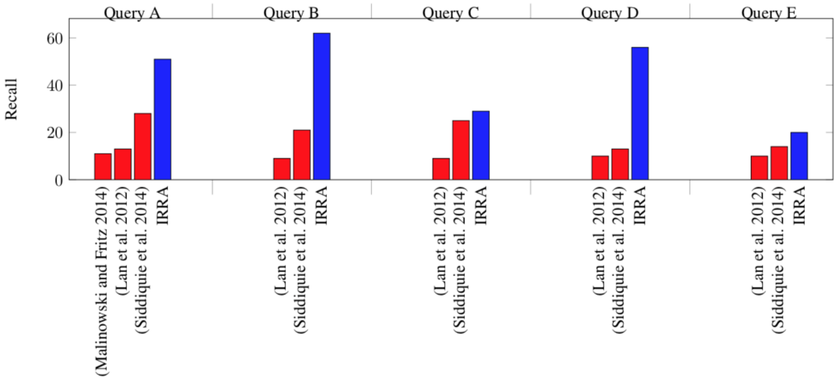
# Experiment 2- playing with spatial prepositions

- evaluate image retrieval using the annotations in [Lan et al., 2012]
- tested our method against all the structured query types proposed.
- The structured queries contain a *noun*, e.g. *pedestrians*, or a *relation* set expressed by a triple in the form *(noun, preposition, and noun)*, e. g. *“car on the road"*.
  - *Structure a (Sa)*, which contains only a relation, for instance, *“car on road"*;
  - *Structure b (Sb)* contains a relation and a noun, e.g, *“car on road, pedestrians"*;
  - *Structure c (Sc)* contains two relations, e.g., *“car on road, sky above building"*;
  - *Structure d (Sd)* contains two relations and a noun, e.g., *“car on road, sky above building, pedestrians"*;
  - *Structure e (Se)* contains three relations, e.g., *“car on road, sky above building, books inside of bookcase"*.

# Experiment 2- playing with spatial prepositions

| Preposition | precision | recall | f-measure | $n$ |
|---|---|---|---|---|
| *Above* | 0.76 | 0.58 | 0.66 | 166 |
| *Across from* | 1.00 | 0.03 | 0.06 | 387 |
| *Behind* | 0.65 | 0.62 | 0.63 | 329 |
| *Below* | 0.84 | 0.79 | 0.81 | 361 |
| *In* | 0.59 | 0.84 | 0.69 | 475 |
| *In front of* | 0.55 | 0.69 | 0.61 | 317 |
| *Inside of* | 0.00 | 0.00 | 0.00 | 65 |
| *Left of* | 0.01 | 0.01 | 0.01 | 187 |
| *On* | 0.60 | 0.77 | 0.67 | 208 |
| *Right of* | 0.00 | 0.00 | 0.00 | 59 |
| *Under* | 0.87 | 0.98 | 0.93 | 2,399 |
| Overall | 0.75 | 0.75 | 0.71 | 4,953 |

# Experiment 2- playing with spatial prepositions



# Experiment 3- Ontology Expansion

- We extended the annotations to infer new relations derived from the original relations that were manually annotated.
- SUMO was used to extend the system's knowledge about the relations in order to evaluate spatial prepositions that were not obtained by the quantitative analysis processes.
- distinct queries were proposed, but the same information to be retrieved from the set of images. For instance, for two objects ($x$, $y$) and the query ($x$-above-$y$), we also evaluated the retrieval for ($y$-below-$x$) against the same annotation for above as used in the original query.

# Experiment 3- Ontology Expansion

- We extended the annotations to infer new relations derived from the original relations that were manually annotated.
- SUMO was used to extend the system's knowledge about the relations in order to evaluate spatial prepositions that were not obtained by the quantitative analysis processes.
- distinct queries were proposed, but the same information to be retrieved from the set of images. For instance, for two objects ($x$, $y$) and the query ($x$-above-$y$), we also evaluated the retrieval for ($y$-below-$x$) against the same annotation for above as used in the original query.

# Experiment 3- Ontology Expansion

- We extended the annotations to infer new relations derived from the original relations that were manually annotated.
- SUMO was used to extend the system's knowledge about the relations in order to evaluate spatial prepositions that were not obtained by the quantitative analysis processes.
- distinct queries were proposed, but the same information to be retrieved from the set of images. For instance, for two objects ($x$, $y$) and the query ($x$-above-$y$), we also evaluated the retrieval for ($y$-below-$x$) against the same annotation for above as used in the original query.

# Experiment 3- Ontology Expansion

- We extended the annotations to infer new relations derived from the original relations that were manually annotated.
- SUMO was used to extend the system's knowledge about the relations in order to evaluate spatial prepositions that were not obtained by the quantitative analysis processes.
- distinct queries were proposed, but the same information to be retrieved from the set of images. For instance, for two objects ($x$, $y$) and the query ($x$-above-$y$), we also evaluated the retrieval for ($y$-below-$x$) against the same annotation for above as used in the original query.

# Experiment 3- Ontology Expansion

- The instanciation of the whole ontology with SUN09 data generated more than $13,000$ terms with respect to the images and objects, and more than $18,000$ formulas referencing the created relations.
- The retrieval task based on the new set of queries was performed by evaluating every image using the E first-order logic theorem prover
- Every image was tested using the new annotated queries, the mean Averate Precision (mAp) achieved in this case was 41.60, whereas he original set for these prepositions obtained a mAp of 51.
- The decrease in performance observed with the extended set of relations (in contrast with the original) was due to the fact that, by increasing the size of the knowledge base, errors were possibly included in the process

# Experiment 3- Ontology Expansion

- The instanciation of the whole ontology with SUN09 data generated more than $13,000$ terms with respect to the images and objects, and more than $18,000$ formulas referencing the created relations.
- The retrieval task based on the new set of queries was performed by evaluating every image using the E first-order logic theorem prover
- Every image was tested using the new annotated queries, the mean Averate Precision (mAp) achieved in this case was 41.60, whereas he original set for these prepositions obtained a mAp of 51.
- The decrease in performance observed with the extended set of relations (in contrast with the original) was due to the fact that, by increasing the size of the knowledge base, errors were possibly included in the process

# Experiment 3- Ontology Expansion

- The instanciation of the whole ontology with SUN09 data generated more than $13,000$ terms with respect to the images and objects, and more than $18,000$ formulas referencing the created relations.
- The retrieval task based on the new set of queries was performed by evaluating every image using the E first-order logic theorem prover
- Every image was tested using the new annotated queries, the mean Averate Precision (mAp) achieved in this case was 41.60, whereas he original set for these prepositions obtained a mAp of 51.
- The decrease in performance observed with the extended set of relations (in contrast with the original) was due to the fact that, by increasing the size of the knowledge base, errors were possibly included in the process

# Experiment 3- Ontology Expansion

- The instanciation of the whole ontology with SUN09 data generated more than $13,000$ terms with respect to the images and objects, and more than $18,000$ formulas referencing the created relations.
- The retrieval task based on the new set of queries was performed by evaluating every image using the E first-order logic theorem prover
- Every image was tested using the new annotated queries, the mean Averate Precision (mAp) achieved in this case was 41.60, whereas he original set for these prepositions obtained a mAp of 51.
- The decrease in performance observed with the extended set of relations (in contrast with the original) was due to the fact that, by increasing the size of the knowledge base, errors were possibly included in the process

# Conclusion

- an approach that establishes relations between objects in images by means of spatial arrangements.
- starts by decomposing images with respect to pairs of objects, where each pair is also combined with a spatial relation.
- Each spatial relation is related to a spatial preposition
- was evaluated on a public data set, whose results show that our approach outperforms previous work in the retrieval of images using spatial relations.
- Results showed that by combining SUMO's high-level description of objects with the output of a machine learning classifier, it is possible to increase the precision and recall of classification procedures.

# Conclusion

- an approach that establishes relations between objects in images by means of spatial arrangements.
- starts by decomposing images with respect to pairs of objects, where each pair is also combined with a spatial relation.
- Each spatial relation is related to a spatial preposition
- was evaluated on a public data set, whose results show that our approach outperforms previous work in the retrieval of images using spatial relations.
- Results showed that by combining SUMO's high-level description of objects with the output of a machine learning classifier, it is possible to increase the precision and recall of classification procedures.

# Conclusion

- an approach that establishes relations between objects in images by means of spatial arrangements.
- starts by decomposing images with respect to pairs of objects, where each pair is also combined with a spatial relation.
- Each spatial relation is related to a spatial preposition
- was evaluated on a public data set, whose results show that our approach outperforms previous work in the retrieval of images using spatial relations.
- Results showed that by combining SUMO's high-level description of objects with the output of a machine learning classifier, it is possible to increase the precision and recall of classification procedures.

# Conclusion

- an approach that establishes relations between objects in images by means of spatial arrangements.
- starts by decomposing images with respect to pairs of objects, where each pair is also combined with a spatial relation.
- Each spatial relation is related to a spatial preposition
- was evaluated on a public data set, whose results show that our approach outperforms previous work in the retrieval of images using spatial relations.
- Results showed that by combining SUMO's high-level description of objects with the output of a machine learning classifier, it is possible to increase the precision and recall of classification procedures.

# Conclusion

- an approach that establishes relations between objects in images by means of spatial arrangements.
- starts by decomposing images with respect to pairs of objects, where each pair is also combined with a spatial relation.
- Each spatial relation is related to a spatial preposition
- was evaluated on a public data set, whose results show that our approach outperforms previous work in the retrieval of images using spatial relations.
- Results showed that by combining SUMO's high-level description of objects with the output of a machine learning classifier, it is possible to increase the precision and recall of classification procedures.

# Future Work

- improve statistical classifiers in general
- include abstractions (in terms of high-level relations, spatial or not) to static data sets in order to enhance image retrieval tasks.

# Thanks!

This author was partially supported by:

- INFOSYS, Palo Alto, CA
- FAPESP grant: 2016/18792-9
- CAPES Finance Code 001