

Generation of Textual Explanations in XAI: the Case of Semantic Annotation *

Jean-Philippe Poli

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

jean-philippe.poli@cea.fr

Wassila Ouerdane

Université Paris-Saclay, CentraleSupélec, MICS, 91190, Gif-sur-Yvette, France

wassila.ouerdane@centralesupelec.fr

Régis Pierrard

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

regis.pierrard@cea.fr

Abstract

Semantic image annotation is a field of paramount importance in which deep learning excels. However, some application domains, like security or medicine, may need an explanation of this annotation. Explainable Artificial Intelligence is an answer to this need. In this work, an explanation is a sentence in natural language that is dedicated to human users to provide them clues about the process that leads to the decision: the labels assignment to image parts. We focus on semantic image annotation with fuzzy logic that has proven to be a useful framework that captures both image segmentation imprecision and the vagueness of human spatial knowledge and vocabulary. In this paper, we present an algorithm for textual explanation generation of the semantic annotation of image regions.

1 Introduction

Semantic image annotation is the ability for a computer to label images or image regions. It is a task of paramount importance with the daily production of images in all the domains (e.g. medicine, surveillance).

In this field, deep learning has enabled to build models that can efficiently classify images and recognize objects. Sometimes, these models can even top

*This work has been partly funded by the DeepHealth project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

human capabilities on several specific tasks [1]. For some critical applications of Artificial Intelligence (AI), performance is not the only criterion to optimize [2]. Such applications may require a relative understanding of the logic performed by the AI. In other words, the end-user would like to get a response to the question “Why ?” [3]

For semantic annotation, Constraint Satisfaction Problems (CSP) have been successfully applied to geometrical figure annotation [4] and region labelling from a model [5]. Vanegas et al. extended these previous works to fuzzy constraint satisfaction problems (FCSP) to involve fuzzy spatial relations and illustrate their approach with an automatic interpretation of Earth observation images [6]. Since CSP and FCSP are interpretable models and the process of solving is also interpretable and explainable, this kind of approaches are good candidates for explainable semantic annotation of images. Pierrard et al. [7] propose algorithms to extract automatically relevant fuzzy spatial relations for image annotation from a few learning images whose regions are segmented and labelled. The appropriate relations are then used to constitute a FCSP for annotating areas of an image or a rule base to classify the image.

In this paper, we focus on the generation of a textual explanation of the semantic annotation in the context of [7]. Given a solution of such FCSPs and the degree of satisfaction of all the involved constraints, we propose and evaluate two algorithms to extract clues of the reasoning and to order the pieces of the explanation efficiently.

The paper is structured as follows. In section 2, fuzzy spatial relations, constraint satisfaction problems and their solving are described. Next, sections 3 and 4 are devoted to describe the methods for generating explanation of semantic annotation. Then, the two approaches are evaluated and compared in section 5. Finally, we draw some conclusions and perspectives in section 6 .

2 Background

2.1 Fuzzy Spatial Relations

The fuzzy logic framework allows using words instead of numbers during computations and also during problem formalization. Indeed, relations are represented by a linguistic description that can be directly used in the explanation [7].

Many fuzzy spatial relations have been studied in the literature [8]. For instance, Vanegas considers three types of spatial relations: topological, metric and structural relations [6]. The two first types are often used in computer vision. We can cite for instance the RCC8 framework that defines relations between regions and their fuzzy counterparts that have been introduced in [9,10]. Bloch introduced a framework based on fuzzy morpho-mathematics to evaluate fuzzy spatial relations [8]. In particular, metric directional relations can be expressed based on the fuzzy dilation operator.

Without loss of generality, in the remainder of this paper, we use specifically directional, distance and symmetry relations. Directional and distance rela-

tions [8] are computed as a fuzzy landscape and assessed using a fuzzy pattern matching approach [11]. The symmetry relation [12] we use consists in finding the line that maximizes a symmetry measure between two objects (regions). Since this measure is not differentiable, a direct search method is used to solve this optimization problem, such as the downhill simplex method.

2.2 Fuzzy Constraint Satisfaction Problems

A constraint satisfaction problem (CSP) consists in assigning some values to a set of variables that must respect a set of constraints.

An extension of CSP to the fuzzy logic framework to deal with imprecise parameters and flexible constraints is presented in [13]. This is called a fuzzy constraint satisfaction problem (FCSP). A FCSP is defined by:

- A set of variables $X = \{x_1, \dots, x_n\}$,
- A set of domains $D = \{D_1, \dots, D_n\}$ such as D_i is the range of values that can be assigned to x_i ,
- A set of flexible constraints $C = \{c_1, \dots, c_p\}$. Each constraint c_k is defined by a fuzzy relation R_k and by the set of variables V_k that are involved in it.

To solve a FCSP, the backtracking algorithm is applied. It starts with an empty set of instantiations and selects a variable $x \in X$ to instantiate. Then, it finds a value in the domain of X that maintains the consistency of the current instantiation, regarding the set of constraints C . The steps are repeated until all the variables are instantiated. When a variable x has no more value to test, the algorithm backtracks and tries the next value of the previously instantiated variable.

An instantiation that is consistent and complete is a solution. One solution of the FCSP is evaluated by its degree of consistency. Given a solution γ , its degree of consistency [6] is:

$$\text{cons}(\gamma) = \min_{c_k \in C} \mu_{R_k}(\gamma|_{V_k}) \quad (1)$$

where $\gamma|_{V_k}$ is the projection of γ on V_k and μ_{R_k} the membership function representing R_k .

This consistency degree also enables to compare different solutions so that the best one can be extracted.

To improve the performance of the backtracking algorithm, [6, 13] have adapted the AC-3 algorithm of crisp CSP that prunes the domains, discarding values that are inconsistent with the current instantiation.

2.3 Image Annotation with FCSP

When dealing with image annotation, the set of variables X corresponds to the objects we would like to instantiate. The variables share the same domain D

that represents the regions in the image that we get after segmentation. Thus, $|X| \leq |D|$. The constraints in C are defined by fuzzy relations: some of them can deal with groups of objects [6].

This can solve specific annotation problems in which the objects to annotate and the labels are known (even if they are automatically detected, by a segmentation for instance). The intuition behind is that such annotation problem can be combinatorial and the labels are affected accordingly to each other, by opposition with individually like in classical approaches.

In [7], this approach was applied to organ annotation in medical images, with a focus on automatically generating the FCSP from few data. In the remainder of this paper, we will take this work as an illustration with an automatic generated FCSP.

To generate our explanations, the algorithms we propose in this work (Algo 1 and Algo 2) take as input a trace $T = \langle P, s, \bar{C} \rangle$ of the execution of the solving algorithms. T is composed of:

- $P = \langle X, D, C \rangle$ is a FCSP.
- s , a chosen solution among all the solutions of P , for instance the best one regarding the degree of consistency. s contains the assignment for each variable in X .
- \bar{C} , the set of degrees of satisfaction of each $c \in C$.

2.4 Surface Realization

In linguistics, a realization consists in generating a *surface form*, which is a correct sentence in a given natural language, from a more abstract representation, in which the different components such as the subject or the verb are specified. Therefore, a *surface realizer* is a system that is able to take an abstract semantic representation as an input to generate a syntactically-correct sentence.

In this work, we rely on SimpleNLG [14] for performing this task. This realization engine provides an API that is easy to use and complete enough for the kind of explanation we would like to generate. We do not explain here how we use it (e.g. the function calls). We will just describe the form of the sentences.

3 Complete Textual Explanation Generation

In this section, we present a first algorithm for explanation generation in natural language.

3.1 Algorithm

Algo1 uses all the constraints of the FCSP and turns them into sentences. The vocabulary of relations contains: *to the left of*, *to the right of*, *below*, *above*, *close*

to, *symmetrical to* and *stretched*. That makes 6 binary and one unary relations. We note that \bar{c}_x is the complement of x in the scope of c , and the *moderator* is selected among those cited in table 1 according to the satisfaction of c . This idea is inspired from [15].

Moderator	Degree of satisfaction
very high	from 0.9
high	from 0.7 to 0.9
average	from 0.4 to 0.7
moderate	from 0.2 to 0.4
low	from 0 to 0.2

Table 1: Simplified confidence scale

Algorithm 1: Complete Explanations Generation

Input: a trace $T = \langle P, s, \bar{C} \rangle$
Output: a complete textual explanation

```

1 foreach unprocessed variable  $x \in X$  do
2    $v \leftarrow$  value of  $x$  in  $s$ 
3   Create a sentence of the form: “Region  $v$  is annotated as  $x$  with a
   moderator confidence because:”
4   foreach constraint  $c \in C$  involving  $x$  in its scope do
5     if  $x$  is the first variable in the scope of  $c$  then
6       Generate a sentence of the form: “it is  $c \bar{c}_x$ ” (eventually, for each
       variable  $x' \in \bar{c}_x$ , indicate the associated  $v' \in s$ )
7     else
8       Generate a sentence of the form: “ $\bar{c}_x$  is/are  $c x$ ”
9     end
10  end
11 end

```

3.2 Results

In this work, the FCSP has been extracted automatically from few images from the Visceral dataset¹. Figure 1 shows one of the image and different organs of interest.

The segmentation has been obtained automatically and the regions were given an identifier in an arbitrary order. Thus, in this first approach, items are not sorted. However, for the sake of comprehension of this article, we numerated ourselves the organs, from left to right and top to bottom.

We consider the solution of such a FCSP for Figure 1 with the highest degree of consistency.

¹<http://www.visceral.eu/>

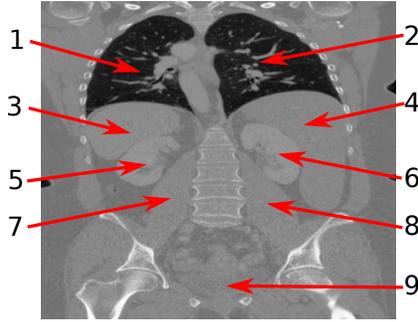


Figure 1: Backward MRI image with different regions to annotate

The result, as it can be seen in figure 2 is obviously a long but complete explanation.

Region 1 is annotated as the left lung with a high confidence **because:**

- it is *completely to the left of* region 2 (annotated as the right lung by the model),
- region 2 (right lung) is *completely to the right of* region 1,
- it is *above* region 3 (spleen),
- region 3 (spleen) is *completely below* region 1,
- it is *above* region 7 (left psoas),
- region 7 (left psoas) is *completely below* region 1,
- region 5 (left kidney) is *completely below* region 1.

Region 2 is annotated as the right lung with a very high confidence **because:**

- it is *completely to the right of* region 1 (left lung),
- region 1 (left lung) is *completely to the left of* region 2,
- region 3 (spleen) is *to the left of* region 2,
- region 4 (liver) is *below* region 2,
- it is *above* region 8 (right psoas),
- region 8 (right psoas) is *completely below* region 2,
- region 6 (right kidney) is *completely below* region 2,
- region 9 (bladder) is *below* region 2.

Figure 2: Extract of an explanation for an annotation with the complete approach. The complete one can be found in [7].

In the next, we investigate the possibility to shorten this explanation. Thus, the next section is dedicated to describe a second algorithm to generate a more concise explanation.

4 Concise Textual Explanation Generation

4.1 Cognitive Science Considerations

Cognitive science has largely studied the way Humans represent a scene or scan images. Thus, it seems natural to consider those insights to create an explanation.

Zwaan *et al.* present more than a decade of studies about situation model, i.e. a mental representation of affairs [16]. They highlight the difficulty to describe correctly a spatial scene with language, because of the difference between its dimensionality and the dimensionality of space. For instance, if one describes a room in a circular way, the first and the last objects are far from each other in the description but close in the room. This also shows the importance of the order in which the parts of the scene have to be described.

This leads us to the studies about image scanning [17], which is related to the mental representation of a scene or an image. Authors of [18] state that the visual images preserve the metric spatial information. This implies that starting from a focus point, subjects need more and more time to mentally visualize the information when going further to this focus point. Other works study the difficulties of subjects to represent a scene if the description is too long and if the description is too precise [19, 20]. Another difficulty is the direction of reading: [21] indicates that it affects the description of a scene.

The studies about image scan paths bring also good information. The attention of subjects is classically attracted by focus points. In image understanding, this is called salient objects and [22] gives a comprehensive review on their automatic detection. Nevertheless, cognitive science warns of the difficulty of defining saliency because it can be context-dependent, or due to the singularity of an object, of the user’s goal, etc. However, when a same subject watches the same picture, the scan paths may be different [23]: thus, the scan path does not depend only on the objects in the image. If several similar pictures are presented, the scan path can also be more and more efficient [23].

Finally, the Gestalt psychologists [24] studied the cognitive issues of visual perception, in particular the shape of objects. The 7 Gestalt principles concern figure-ground, similarity, proximity, common region, continuity, closure and focal point of images. They are particularly useful in design, but give some insight about how objects are perceived. In particular, they recommend to group objects that are similar or that share properties.

This short overview of cognitive science helped us to design our explanation strategy.

4.2 Principles

The previous subsection gives raw information from the cognitive science. The idea of our approach is to improve the previous version of the generation of explanation from a FCSP by considering cognitive science insights. We thus observe these principles:

- *Sorting*: the order of the results has an importance. It is important to start with regions in images that are salient, and then, regarding the recommendations of cognitive science papers, use diagonals and increasing distances to select the next results. The spiral order is not recommended.
- *Saliency*: the saliency is a difficult concept that can be context-dependent. *A minima*, one can select the biggest object or a group of objects as focus point.
- *Symmetry*: a pair of objects that are symmetrical must be grouped.
- *Priority*: we must select the most satisfied constraints first.
- *Associativity*: some relations are associative (e.g. “to the left of”) and explainees can immediately infer it, so we must use that to reduce the number of constraints involved in the explanations.
- *Locality*: if possible, we will use first the constraints with the closest regions in the image.

Moreover, an explanation must somehow indicate how the task has been achieved. In our case, the solving of a FCSP is quite simple to explain since the algorithm searches for the values of the variables such as the constraints are satisfied. However, it makes the explanation more complicated when constraints are not all unary, since these assignments are dependent from each other. Indeed, for instance, a binary constraint will force the assignment of two variables together. In the case of semantic annotation or classification, the constraints are relations so that it is a little bit simpler than, for instance, quadratic constraints.

Another point is that we are selecting a maximum number of constraints for each variable, such that there is no correlation between these constraints: for instance, the values of “to the left of” and “to the very left of” may be correlated and so we do not want to use them at the same time for the same variable because they are redundant. We use mutual information to detect this correlation.

In the next subsection, we introduce an algorithm that considers those different principles.

4.3 Algorithm

Algo 2 presents the algorithm to generate concise explanation for semantic annotation.

The explanation starts with a general sentence that indicates the global confidence about the annotation based on the degree of consistency of the solution (line 1). The algorithm then selects the region from the segmentation that is the most salient (line 2). Regarding this object, the image is divided into four quadrants. The explanation will start with the most salient region, then with the other objects in the same quadrant, then quadrant by quadrant, in the clockwise order. This order is materialized in an ordered set X' (lines 3-4).

For each variable in X' , the algorithm has to select at most N_{max} constraints to justify the explanation. The constraints are chosen regarding not only their level of satisfaction (that must be the highest as possible not to overload the text with moderators), but also their mutual link and the proximity with the other variables (lines 5-12).

The mutual link between relations is a tricky part. We use a knowledge graph about the relations as proposed in [7]. Such a graph emphasizes different links between two relations r_1 and r_2 , like $r_1 \implies r_2$, $\neg r_1 \implies r_2$, but also symmetry. Symmetry is important not to use twice the same constraint. Let o_1 and o_2 be two objects in the image, and r a symmetrical relation, if $o_1 r o_2$ is used in a sentence, we cannot use $o_2 r o_1$ anymore.

Then, the algorithm looks for grouping constraints such as “is symmetrical to” that constitutes a pair of variables (line 9). Indeed, the previous section highlights that groups of objects must be treated together. Thus, the other variables in the scope of this constraint must be processed just after (line 10).

Algorithm 2: Concise Explanation Generation

Input: a trace $T = \langle P, s, \bar{C} \rangle$
Output: a concise textual explanation

- 1 Write a sentence to introduce the result and the global confidence
- 2 Select f the variable in s region that is the focus point in the image
- 3 From the center of f , divide the image into 4 quadrants Q_1, \dots, Q_4
- 4 X' = set of variables $x \in s$ sorted by quadrant
- 5 **while** $X' \neq \emptyset$ **do**
- 6 $x \leftarrow pop(X')$
- 7 $S \leftarrow$ Select N_{max} constraints c_i that are not linked in the knowledge graph and with maximal degrees of satisfaction
- 8 Write the sentence “ x is c_1, \dots , and is $c_{j \leq N_{max}}$ ”
- 9 **if** x involves a grouping constraint c **then**
- 10 | Move all variables in scope of c to the beginning of X'
- 11 **end**
- 12 **end**

4.4 Results

In this work, we define the focus point as the biggest object (in terms of area). We set $N_{max} = 2$.

For the same example (see Figure 1), and the same solution s , the result is shown in Figure 3.

Most of the constraints are linked in the knowledge graph, because we used mainly directional relations like “to the right of” and “to the left of”. This explains why we rarely reach N_{max} constraints.

The result is obviously shorter, and seems easier to read. The quadrant imposes an order for the description of each organ. The explanation seems less redundant thanks to the selection of the constraints.

“This is the annotation of the given image (with a very high confidence). The right lung (region 2) is *symmetrical to* the left lung (region 1) and *above* the liver (region 4). The liver (region 4) is *at the right of* the right kidney (region 6) and *at the right of* the right psoas (region 8). The right psoas (region 8) is *above of* the bladder (region 9) and is *symmetrical to* the left psoas (region 7). The left psoas (region 7) is *below* the left kidney (region 5). The spleen (region 3) is *above* the left kidney (region 5) and is *below* the left lung (region 1).”

Figure 3: Concise explanation produced by Algorithm 2

The next section is dedicated to the evaluation of both types of explanation.

5 Evaluation and Discussion

To compare the two approaches, we evaluated both of them. In this aim, we use the questionnaire presented in [25]: it is based on 17 questions organized in 3 categories: natural language, human-computer interaction and content and form. Each question is evaluated with a Likert scale (from 1 “strongly disagree” to 5 “strongly agree”). Our panel consists in 40 respondents, with 20 medical staff members (medical doctors, surgeons, nurses, radiologists), the other half being computer scientists (6) and other various non-medical professionals (14). To decrease the medical staff’s amount of time dedicated to the questionnaire, we selected only 14 questions out of the 17 initial ones that will allow comparing the both approaches. We removed the questions about the grammar and the one that indicates if the explanation made a respondent change his mind. Because of the lack of space, figure 4 highlights the answers to few questions.

Both explanations are comparable in terms of syntax correctness (87% for approach 1 and 95% for approach 2), of reasoning comprehension (67.5% agree for approach 1, 60% for approach 2), and of uncertainty communication (62.2% for approach 1, 65% for approach 2). “Reasoning comprehension” indicates if the respondents can infer about the reasoning process when they read the explanation. The “uncertainty communication” criterion evaluates the ability of the explanation to tell the user at which point the decision can be trusted. In our case, it is achieved by the translation of the constraints satisfaction into sentence parts like “with a very high confidence”. These facts show that not all the people understood how the algorithm annotates the organs and understood why the algorithm was not confident in all the cases.

For all other comparisons, the second approach outperforms the first approach. 19 persons found that the first explanation was too long whereas only 1 respondent was concerned by the length of the second explanation. Respondents found the first explanation repetitive (87.5%) and hard to read (72.5%), whereas only respectively 22.5% and 10% of the panel agree with these facts for the second one. Only 32.5% of the respondents found the order of the items in the

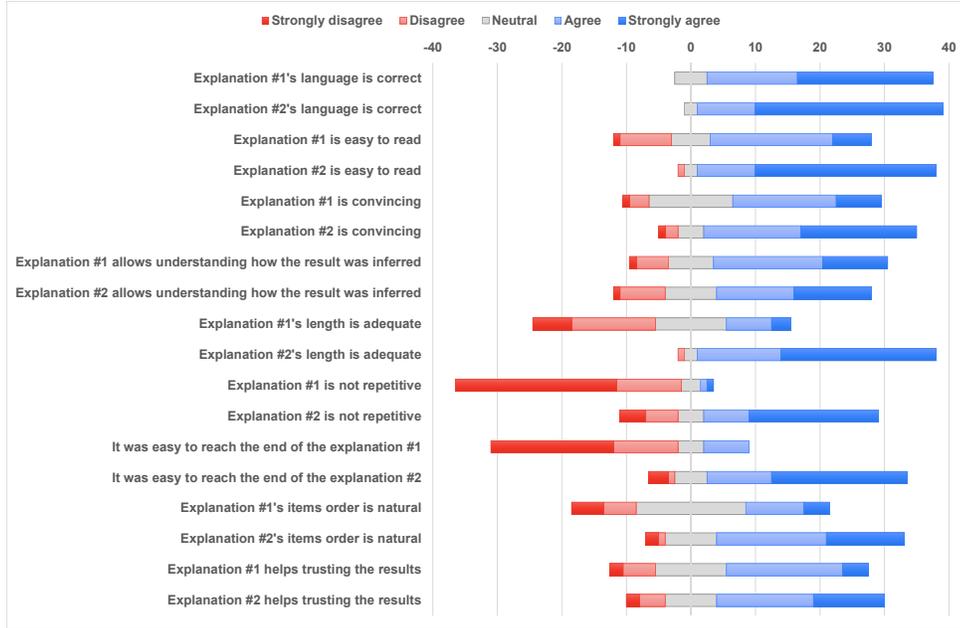


Figure 4: Highlights from the survey

explication suitable for explanation 1 versus 72.5% for the second explanation.

Both explanations make the respondents think they can trust the automatic labelling (55% for first approach and 65% for the second one).

These results confirm the advantages of the second algorithm.

First, it is important to note that these algorithms are not domain-specific. Indeed, the relations are generic in the sense that they could be used in another domain (such as satellite image annotation). They also manipulate image regions, and have no clue they represent organs. However, the labels that are used are organ names, because we want a semantic annotation. We do not use external domain knowledge, for instance to replace the word “region” by “organ” on the explanation, or to use a more technical vocabulary.

The results show that the order of the items inside an explanation are important for the end users. Conciseness seems to be a criterion of paramount importance too.

The questionnaire invited also the respondents to write comments after each type of explanation. Most of the medical staff felt uncomfortable with the fact that the MRI image was taken from the back. Nevertheless, no one declared the explanation was wrong: maybe it can have an impact on the confidence of the users in the AI.

One of the medical respondent said it could be useful to use the spine as main

region and use it for the labelling of the other regions. This idea emphasizes the importance of saliency: indeed, in such an image, we can see the spine first because it is whiter and central. Unfortunately, in the segmentation we use, bones are not considered.

Finally, we also made a comparison between the medical respondents and the others, but the results do not differ significantly.

6 Conclusion and Perspectives

In this paper, we presented our work on the generation of textual explanations of image annotation. The first part provides a form of explanation that was not pertinent for humans. The second part is an improvement of the first one that generates a more concise explanation. It relies on a more sophisticated selection of the constraints that are used in the explanation, based on cognitive science principles.

This work also shows the importance of realizers for explainable AI: although it is not the goal of this work, using synonyms or different sentence structures to break the monotony of the explanations can help. However, the survey we presented shows that most participants are convinced by the explanations and they understand the logic of the model.

What we observe is that to develop a model, then an algorithm to extract relevant clues and finally improve realizers involve too many fields and is difficult to manage. In our future work, we are thinking of the separation of these tasks.

Acknowledgement

The authors would like to thank the survey panel, in particular the medical staff who accepted to participate despite the pandemic.

References

- [1] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. C. Corrado, A. Darzi, M. Etemadi, F. Garcia-Vicente, F. J. Gilbert, M. Halling-Brown, D. Hassabis, S. Jansen, A. Karthikesalingam, C. J. Kelly, D. King, J. R. Ledsam, D. Melnick, H. Mostofi, L. Peng, J. J. Reicher, B. Romera-Paredes, R. Sidebottom, M. Suleyman, D. Tse, K. C. Young, J. D. Fauw, and S. Shetty, “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, no. 7788, pp. 89–94, 1 2020.
- [2] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” in *eprint arXiv:1702.08608*, 2017.
- [3] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial Intelligence*, vol. 2679, pp. 1–38, 2019.

- [4] D. Waltz, “Understanding line drawings of scenes with shadows,” in *The Psychology of Computer Vision*. McGraw-Hill, 1975.
- [5] A. Deruyver, Y. Hodé, and L. Brun, “Image interpretation with a conceptual graph: Labeling over-segmented images and detection of unexpected objects,” *Artificial Intelligence*, vol. 173, no. 14, pp. 1245 – 1265, 2009.
- [6] M. C. Vanegas Orozco, “Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model.” Theses, Télécom ParisTech, Jan. 2011.
- [7] R. Pierrard, J.-P. Poli, and C. Hudelot, “Spatial relation learning for explainable image classification and annotation in critical applications,” *Artificial Intelligence*, vol. 292, p. 103434, 2021.
- [8] I. Bloch, *Fuzzy Models of Spatial Relations, Application to Spatial Reasoning*. Springer Berlin Heidelberg, 2013, pp. 51–58.
- [9] Yongming Li and Sanjiang Li, “A fuzzy sets theoretic approach to approximate spatial reasoning,” *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 6, pp. 745–754, Dec 2004.
- [10] S. Schockaert, M. D. Cock, C. Cornelis, and E. E. Kerre, “Fuzzy region connection calculus: Representing vague topological information,” *International Journal of Approximate Reasoning*, vol. 48, no. 1, pp. 314 – 331, 2008, special Section: Perception Based Data Mining and Decision Support Systems.
- [11] M. Cayrol, H. Farreny, and H. Prade, “Fuzzy pattern matching,” *Kybernetes*, vol. 11, no. 2, pp. 103–116, 1982.
- [12] O. Colliot, “Représentation, évaluation et utilisation de relations spatiales pour l’interprétation d’images. application à la reconnaissance de structures anatomiques en imagerie médicale,” Ph.D. dissertation, Télécom ParisTech, 2003.
- [13] D. Dubois, H. Fargier, and H. Prade, “Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty,” *Applied Intelligence*, vol. 6, no. 4, pp. 287–309, 1996.
- [14] A. Gatt and E. Reiter, “Simplenlg: A realisation engine for practical applications,” in *Proceedings of the 12th European Workshop on Natural Language Generation*, 2009, pp. 90–93.
- [15] D. V. Budescu, H.-H. Por, and S. B. Broomell, “Effective communication of uncertainty in the IPCC reports,” *Climatic change*, vol. 113, no. 2, pp. 181–200, 2012.

- [16] R. A. Zwaan and G. A. Radvansky, “Situation models in language comprehension and memory.” *Psychological Bulletin*, vol. 123, no. 2, p. 162, 1998. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1041143270>
- [17] G. Borst, S. Kosslyn, and M. Denis, “Different cognitive processes in two image-scanning paradigms,” *Memory & cognition*, vol. 34, pp. 475–90, 05 2006.
- [18] S. Kosslyn, T. Ball, and B. Reiser, “Visual images preserve metric spatial information: Evidence from studies of image scanning,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 1, pp. 47–60, 2 1978.
- [19] M. J. Farah and S. M. Kosslyn, “Structure and strategy in image generation,” *Cognitive Science*, vol. 5, no. 4, pp. 371 – 383, 1981.
- [20] M. Denis, M.-R. Goncalves, and D. Memmi, “Mental scanning of visual images generated from verbal descriptions: Towards a model of image accuracy,” *Neuropsychologia*, vol. 33, no. 11, pp. 1511 – 1530, 1995, the Neuropsychology of Mental Imagery.
- [21] A. Román, A. Fathi, and J. Santiago, “Spatial biases in understanding descriptions of static scenes: The role of reading and writing direction,” *Memory & cognition*, vol. 41, 01 2013.
- [22] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, “Salient object detection: A survey,” *Computational Visual Media*, vol. 5, pp. 117–150, 2014.
- [23] D. Noton and L. Stark, “Scanpaths in eye movements during pattern perception,” *Science*, vol. 171, no. 3968, pp. 308–311, 1971.
- [24] R. Luccio, *Gestalt problems in cognitive psychology: Field theory, invariance and auto-organisation*. Springer Berlin Heidelberg, 1993, pp. 1–19.
- [25] I. Baaj and J. Poli, “Natural language generation of explanations of fuzzy inference decisions,” in *2019 IEEE International Conference on Fuzzy Systems*. IEEE, 2019, pp. 1–6.