
Raisonnement abductif en logique de description exploitant les domaines concrets spatiaux pour l'interprétation d'images

Yifan Yang¹, Jamal Atif², Isabelle Bloch¹

1. LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
[yifan.yang,isabelle.bloch}@telecom-paristech.fr](mailto:{yifan.yang,isabelle.bloch}@telecom-paristech.fr)

2. Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243,
LAMSADE, 75016 Paris, France
jamal.atif@dauphine.fr

RÉSUMÉ. L'interprétation d'images a pour objectif non seulement de détecter et reconnaître des objets dans une scène mais aussi de fournir une description sémantique tenant compte des informations contextuelles dans toute la scène. Le problème de l'interprétation d'images peut être formalisé comme un problème de raisonnement abductif, c'est-à-dire comme la recherche de la meilleure explication en utilisant une base de connaissances. Dans ce travail, nous présentons une nouvelle approche utilisant une méthode par tableau pour la génération et la sélection d'explications possibles d'une image donnée lorsque les connaissances, exprimées dans une logique de description, comportent des concepts décrivant les objets mais aussi les relations spatiales entre ces objets. La meilleure explication est sélectionnée en exploitant les domaines concrets pour évaluer le degré de satisfaction des relations spatiales entre les objets.

ABSTRACT. Image interpretation aims not only at detecting and recognizing objects in a scene but also at deriving a semantic description considering contextual information in the whole scene. Image interpretation can be formalized as an abductive reasoning problem, i.e. an inference to the best explanation using a background knowledge. In this work, we present a framework using a tableau method for generating and selecting potential explanations of the given image when the background knowledge is encoded in description logics, and includes concepts describing objects and their spatial relations. The best explanation is selected according to a minimality criterion based on the satisfaction degree of spatial relations between the objects, computed in concrete domains.

MOTS-CLÉS : interprétation d'images, abduction, logiques de description, tableau sémantique, relations spatiales, représentations floues, domaines concrets.

KEYWORDS: image interpretation, abduction, description logics, semantic tableau, spatial relations, fuzzy representations, concrete domains.

DOI:10.3166/RIA.31.11-39 © 2017 Lavoisier

1. Introduction

L'une des questions centrales en intelligence artificielle, qui reste ouverte aujourd'hui malgré les progrès récents du domaine, notamment en reconnaissance des formes (Krizhevsky *et al.*, 2012), consiste à associer des descriptions linguistiques de haut niveau à des données numériques telles que des percepts visuels. Ce problème est connu sous le nom de fossé sémantique (Smeulders *et al.*, 2000), ou ancrage de symboles (Harnad, 1990; Coradeschi, Saffiotti, 2000).

Dans notre travail, au-delà de la reconnaissance d'objets isolés dans une scène, nous nous intéressons à l'inférence de descriptions complexes permettant d'expliquer ce que l'on observe en exploitant les informations de contexte, les relations spatiales entre les objets, et une connaissance du domaine. A cette fin, nous formalisons le problème d'interprétation ainsi défini comme une tâche de raisonnement abductif logique.

L'abduction, introduite dans (Peirce, 1958), est un mode de raisonnement qui permet d'inférer la *meilleure* explication d'un phénomène observé en prenant en compte les connaissances expertes du domaine. Elle comporte en général deux phases : la génération d'hypothèses et la sélection de la meilleure hypothèse selon des critères prédéfinis. A l'aune de cette définition, il apparaît naturel de formaliser l'interprétation sémantique comme un raisonnement abductif, à l'instar des travaux de (Hobbs, 2004) pour le traitement naturel du texte, (Shanahan, 2005) pour des applications en robotique autonome, ou (Neumann, Möller, 2008; Peraldi *et al.*, 2007; Atif *et al.*, 2014) pour la compréhension de scènes visuelles. Nos travaux s'inscrivent dans la lignée de ces derniers en allant plus loin dans l'exploitation du formalisme d'abduction en logique de description et des domaines concrets.

Une image est une représentation numérique qui ne contient pas de sémantique explicite. En extraire une description de haut niveau ne peut donc se faire qu'en recourant à une connaissance experte du domaine. Cette connaissance peut être représentée sous forme de modèles structurels à base de graphes (Chein, Mugnier, 2008), de grammaires (Zhu, Mumford, 2006) ou d'ontologies (Town, 2006; Hudelot *et al.*, 2008). L'intérêt des ontologies est de permettre d'exploiter les capacités de raisonnement offertes par les logiques de description qui servent de fondement à ce type de représentation. En effet, ces logiques (Baader *et al.*, 2003) sont dédiées à la représentation structurée des connaissances. Initialement utilisées comme langage formel de représentation du web sémantique, leur champ d'application dépasse largement ce cadre aujourd'hui, et s'étend par exemple à l'interprétation d'images (Atif *et al.*, 2014; Elsenbroich *et al.*, 2006; Neumann, Möller, 2008). Beaucoup de travaux s'appuient aujourd'hui sur des modèles d'apprentissage supervisé (notamment par réseaux de neurones profonds) pour la compréhension de scènes. Ceux-ci se limitent dans leur majorité à affecter une étiquette à la scène. Des travaux récents, combinant données textuelles et données d'images ou vidéos, que l'on peut catégoriser sous l'appellation approches *show and tell* (Vinyals *et al.*, 2015), tentent d'inférer une description textuelle riche pour les images à annoter. Ces descriptions textuelles ont pour but

de décrire de façon précise la scène, mais ne permettent pas d'expliquer ce que l'on observe dans celle-ci. C'est en ce sens que notre approche est différente. De plus, ces approches par apprentissage, utilisant des réseaux de neurones profonds, sont très gourmandes en exemples étiquetés, ce qui freine leur application à des domaines d'application où l'on manque de ces données étiquetées.

Notre objectif est d'extraire une information de haut niveau à partir d'une image, qui explique pourquoi l'on observe telle ou telle situation dans l'image. Pour appuyer notre propos, considérons l'exemple de la figure 1, qui illustre une tâche d'interprétation d'une image d'un cerveau comportant une pathologie tumorale, image acquise par résonance magnétique (IRM). L'information sémantique de haut niveau que l'on cherche à extraire dans ce cas consiste en la description de la pathologie, des structures cérébrales affectées par celle-ci, ainsi que de l'agencement structurel entre ces structures et la pathologie, afin de fournir une aide au diagnostic. Selon le niveau de granularité de la connaissance décrite, du but de l'interprétation, et de la personne à qui s'adresse le résultat de celle-ci, deux exemples d'hypothèses possibles (explications au sens de l'abduction) pourraient être :

- une structure anormale est présente dans le cerveau,
- une tumeur périphérique de contraste faible est présente dans l'hémisphère droit¹.

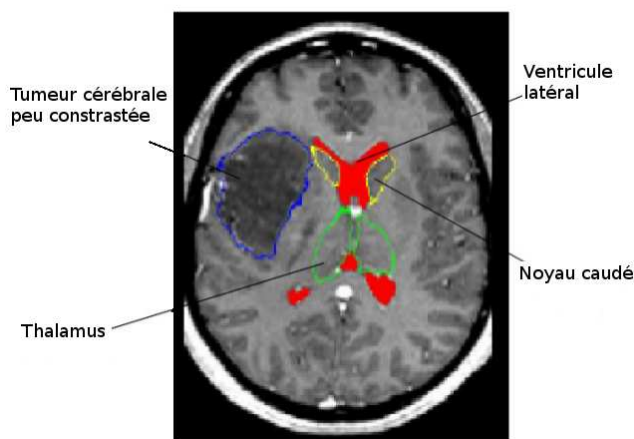


Figure 1. Une coupe d'une image IRM 3D représentant un cerveau avec une pathologie. Quelques structures sont annotées : les ventricules latéraux en rouge, les noyaux caudés en jaune, le thalamus en vert, et une tumeur en bleu

1. On utilise la convention classique « la gauche est à droite » pour l'affichage . La structure dite à droite est du côté gauche dans la figure 1.

Pour aboutir à ce résultat, nous exploitons une connaissance anatomique structurée exprimée dans une logique de description dans sa partie terminologique (TBox), et formalisons le problème d'interprétation comme un raisonnement abductif. Nous partons d'une première reconnaissance des objets segmentés dans l'image, que nous représentons par la suite comme des instances des connaissances *a priori* (ABox). Nous introduisons une nouvelle approche d'abduction dans les logiques de description qui exploite les tableaux sémantiques (Baader *et al.*, 2003). Notre apport dans ce cadre est double. D'abord, nous proposons un algorithme de génération des hypothèses fondé sur les tableaux sémantiques. Cela vient combler un manque dans les travaux sur les logiques de description, puisque peu de recherches ont été consacrées à ce type de raisonnement malgré le vaste champ d'applications qu'il peut concerner. Puis, nous définissons un critère de sélection fondé sur les relations spatiales entre les objets et exploitant les domaines concrets. Cela vient enrichir les travaux sur les critères de sélection en abduction dans les logiques de description qui se sont intéressés principalement aux minimalités sémantique ou de cardinalité (Bienvenu, 2008; Colucci *et al.*, 2004; Elsenbroich *et al.*, 2006).

L'introduction d'un nouveau critère de sélection d'hypothèses fondé sur les relations spatiales trouve sa justification dans le fait que celles-ci sont plus robustes aux conditions d'acquisition, aux variations de forme et d'apparence, contrairement aux informations iconiques intrinsèques aux objets (Bloch, 2005). De plus, les domaines concrets ont été introduits pour enrichir le raisonnement en logique de description par des informations numériques. Ils permettent d'associer des prédicats numériques à des concepts abstraits. Ainsi, dans un contexte d'interprétation d'images, ils aident à réduire le fossé sémantique entre les percepts visuels extraits des images et leurs étiquettes symboliques ou linguistiques.

Un schéma récapitulant l'approche proposée est donné dans la figure 2. Il met en particulier en exergue les modules clés du système d'interprétation dans un contexte d'imagerie cérébrale.

Dans la section 2, nous introduisons le formalisme de représentation de connaissances et de raisonnement. Ce formalisme repose sur la logique de description $\mathcal{ALCH}\mathcal{I}\mathcal{R}_+$ enrichie de domaines concrets (section 2.1), les relations spatiales qualitatives et leurs représentations dans le domaine de l'image (section 2.2), et le raisonnement abductif (section 2.3). Les sections 3 et 4 constituent le cœur de notre contribution : la génération des hypothèses d'interprétation (explications) à partir d'une approche par tableaux sémantiques dans la section 3, et la sélection des meilleures explications en exploitant un nouveau critère fondé sur les relations spatiales et les domaines concrets flous dans la section 4. Nous illustrons dans la section 5 notre formalisme sur des tâches d'interprétation d'images cérébrales pathologiques et d'images illustrant un événement sportif.

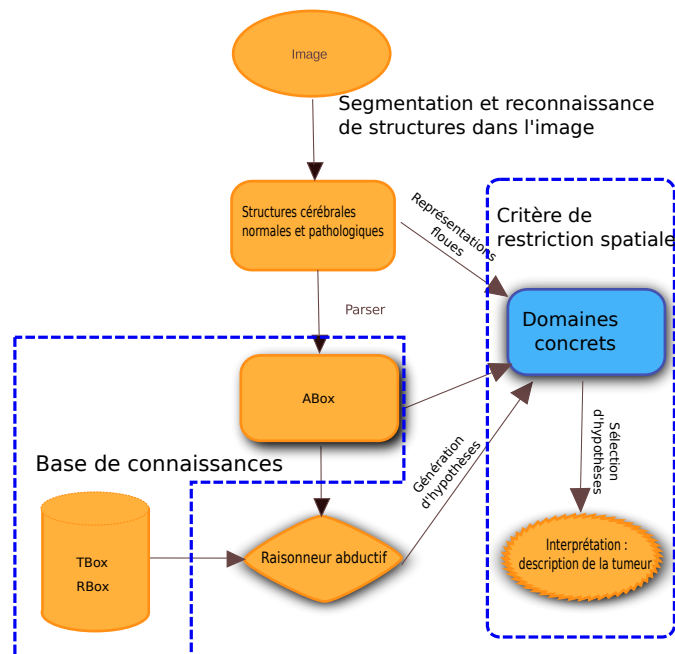


Figure 2. Schéma général du système d'interprétation proposé, illustré dans un contexte d'imagerie cérébrale

2. Représentation des connaissances

2.1. Logiques de description

Les logiques de description (LD) sont une famille de logiques servant de langage formel pour la représentation des connaissances et des ontologies (Baader *et al.*, 2003). Il existe plusieurs logiques de description qui diffèrent par leur pouvoir expressif et naturellement par la complexité des algorithmes de satisfiabilité associés. Dans ce qui suit, seront introduites la syntaxe et la sémantique de la logique de description $\mathcal{ALCH}\mathcal{I}_{\mathcal{R}^+}$ (Horrocks, Sattler, 1999), dont l'expressivité est assez riche pour permettre de représenter des relations spatiales en tant que rôles, et de raisonner sur celles-ci.

DÉFINITION 1 (Syntaxe de concepts). — Soient N_C et N_R deux ensembles disjoints dénombrables désignant les noms de concepts et les noms de rôles. Nous utilisons les lettres A et B pour les noms de concepts atomiques, les lettres r, s, u, v pour les noms de rôles, et les lettres C et D pour les noms de concepts. Les symboles \top et

\perp désignent les concepts universel et vide, respectivement. Les concepts du langage $\mathcal{ALCHL}_{\mathcal{R}^+}$ sont définis comme suit :

$$C, D \rightarrow A \mid \perp \mid \top \mid (C \sqcap D) \mid (C \sqcup D) \mid \neg C \mid \forall r.C \mid \exists r.C$$

DÉFINITION 2 (Sémantique des concepts). — Une interprétation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ se compose d'un ensemble $\Delta^{\mathcal{I}}$, le domaine de \mathcal{I} , et d'une fonction $\cdot^{\mathcal{I}}$ qui associe à tout concept C un sous-ensemble $C^{\mathcal{I}}$ de $\Delta^{\mathcal{I}}$ et à tout rôle r un sous-ensemble $r^{\mathcal{I}}$ de $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ tel que, pour tous les concepts C, D et les rôles r , les propriétés suivantes sont satisfaites :

- $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$ et $\perp^{\mathcal{I}} = \emptyset$
- $(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
- $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\exists r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} \text{ t.q. } (x, y) \in r^{\mathcal{I}} \text{ et } y \in C^{\mathcal{I}}\}$
- $(\forall r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}}, (x, y) \in r^{\mathcal{I}} \text{ implique } y \in C^{\mathcal{I}}\}$

DÉFINITION 3 (Satisfiabilité, subsumption et équivalence de concepts). —

- Un concept C est satisfiable s'il existe une interprétation \mathcal{I} telle que $C^{\mathcal{I}} \neq \emptyset$.
- Un concept C subsume un concept D , noté $D \sqsubseteq C$, si $D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ pour toute interprétation \mathcal{I} .
- Deux concepts C et D sont dits équivalents, noté $D \equiv C$, si $D^{\mathcal{I}} = C^{\mathcal{I}}$ pour toute interprétation \mathcal{I} .

DÉFINITION 4 (Syntaxe des rôles). — Soit $N_{\mathcal{R}}$ un ensemble de noms de rôles. Les rôles inverses et la négation de rôles sont notés respectivement r^- et $\neg r$. Les rôles complexes sont définis comme $r_1 \sqcap r_2$ et $r_1 \sqcup r_2$. L'inclusion de rôles ($r_1 \sqsubseteq r_2$) et leur composition ($r_1 \circ r_2$) définissent des axiomes de rôles. L'équivalence de rôles $r \equiv s$ peut s'écrire sous la forme ($r \sqsubseteq s$ et $s \sqsubseteq r$).

DÉFINITION 5 (RBox). — Une base de rôles, notée $RBox$, est un ensemble fini d'axiomes de rôles construits à partir d'éléments dans $\mathbf{R} = N_{\mathcal{R}} \cup \{r^- \mid r \in N_{\mathcal{R}}\}$, où r^- représente l'inverse d'un rôle r . Soit $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ une interprétation. La sémantique des constructeurs et des axiomes de rôles est donnée comme suit :

- Composition de rôles : $u \circ v \sqsubseteq r_1 \sqcup \dots \sqcup r_n$, où $n \geq 1$, est interprété comme $(u \circ v)^{\mathcal{I}} \subseteq r_1^{\mathcal{I}} \cup \dots \cup r_n^{\mathcal{I}}$. Pour tous éléments $x, y, z \in \Delta^{\mathcal{I}}$, $(\langle x, y \rangle \in u^{\mathcal{I}} \text{ et } \langle y, z \rangle \in v^{\mathcal{I}})$ implique $\langle x, z \rangle \in r_1^{\mathcal{I}} \cup \dots \cup r_n^{\mathcal{I}}$.
- Transitivité de rôles : $r \circ r \sqsubseteq r$ est interprété comme $(r \circ r)^{\mathcal{I}} \subseteq r^{\mathcal{I}}$. Pour tous éléments $x, y, z \in \Delta^{\mathcal{I}}$ tels que $\langle x, y \rangle \in r^{\mathcal{I}}$ et $\langle y, z \rangle \in r^{\mathcal{I}}$, alors $\langle x, z \rangle \in r^{\mathcal{I}}$.
- Rôles inverses : $u \equiv v^-$ est interprété comme $u^{\mathcal{I}} = v^{-\mathcal{I}}$. Pour tous éléments $x, y \in \Delta^{\mathcal{I}}$, alors $\langle x, y \rangle \in u^{\mathcal{I}}$ ssi $\langle y, x \rangle \in v^{\mathcal{I}}$.
- Rôles symétriques : $r \equiv r^-$ est interprété comme $r^{\mathcal{I}} = r^{-\mathcal{I}}$. Pour tous éléments $x, y \in \Delta^{\mathcal{I}}$, alors $\langle x, y \rangle \in r^{\mathcal{I}}$ ssi $\langle y, x \rangle \in r^{\mathcal{I}}$.

– *Disjonction de rôles* : $u \sqsubseteq \neg v$ est interprété comme $u^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \setminus v^{\mathcal{I}}$ où $u^{\mathcal{I}} \cap v^{\mathcal{I}} = \emptyset$. Pour tous éléments $x, y \in \Delta^{\mathcal{I}}$, $\langle x, y \rangle \in u^{\mathcal{I}}$ implique $\langle x, y \rangle \notin v^{\mathcal{I}}$.

Une base de connaissances dans le langage $\mathcal{ALCHL}_{\mathcal{R}^+}$ est un triplet $(\mathcal{K} = \{\mathcal{T}, \mathcal{R}, \mathcal{A}\})$. La TBox \mathcal{T} décrit la terminologie en listant les concepts et leurs relations, et les relations de subsomption entre concepts. La Rbox \mathcal{R} décrit les axiomes de rôles. La ABox \mathcal{A} contient des informations sur les individus. Une Abox \mathcal{A} est un ensemble fini d'assertions $(a : C$ (appartenance à un concept), $(a, b) : r$ (rôle)).

Donnons un exemple simple de Rbox pour le raisonnement spatial :

$$\begin{aligned} RBox = \{ & rightOf \equiv leftOf^- \\ & above \equiv below^- \\ & closeTo \equiv closeTo^- \\ & farFrom \equiv farFrom^- \\ & isPartOf \circ isPartOf \sqsubseteq isPartOf \\ & hasPart \circ hasPart \sqsubseteq hasPart \\ & isPartOf \equiv hasPart^- \} \end{aligned}$$

Les domaines concrets des logiques de description proposés initialement dans (Baader, Hanschke, 1991) permettent d'introduire des informations numériques ou textuelles sur les concepts. Plus généralement, ils permettent de représenter les propriétés concrètes des objets réels tels que leur taille, leur apparence visuelle ou encore leur organisation spatiale. Par exemple, le concept $\text{Person} \sqcap \exists \text{age}.\leq_{20}$ représente l'ensemble des personnes dont l'âge est inférieur ou égal à 20. Dans cet exemple, \leq_{20} est un prédicat sur le domaine concret des entiers naturels \mathbb{N} .

Formellement les domaines concrets sont définis comme suit :

DÉFINITION 6 (Domaines concrets). — *Un domaine concret D est une paire (Δ_D, Φ_D) où Δ_D est un ensemble non vide et Φ_D un ensemble non vide de noms de prédicats définis sur Δ_D . À chaque nom de prédicat $P \in \Phi_D$ sont associés une arité n et un prédicat n -aire $P^D \subseteq \Delta_D^n$. Un domaine concret est dit admissible si (i) l'ensemble de ses noms de prédicat est fermé par négation, c'est-à-dire pour tout $P \in \Phi_D$, il existe $\bar{P} \in \Phi_D$ interprété comme $\bar{P}^D = \Delta_D \setminus P^D$, (ii) il contient le prédicat \top_D interprété comme Δ_D , et (iii) le problème de satisfiabilité d'une conjonction de prédicats $\bigwedge_{i=1}^n P_i(u_i)$ est décidable (il existe une application δ telle que $\delta(u_i) \in P_i^D$, pour tout $1 \leq i \leq n$).*

2.2. Relations spatiales qualitatives

Dans notre approche, nous utilisons les relations spatiales entre objets dans l'image comme critère de sélection de la meilleure hypothèse à partir de l'ensemble donné par

l’algorithme d’abduction. Cette section traite donc du calcul des relations spatiales à partir d’une image segmentée. Plus précisément, nous introduisons des critères d’évaluation de degrés de satisfaction d’une relation spatiale donnée entre deux objets dans l’image.

Nous considérons l’espace de l’image comme le domaine concret dans lequel les relations spatiales sont évaluées. Nous exploitons les travaux de (Bloch, 2005) qui se fondent sur la morphologie mathématique floue pour soit définir un degré de satisfaction d’une relation spatiale entre deux objets donnés, soit construire dans l’image une carte des degrés de satisfaction en chaque point d’une relation spatiale donnée par rapport à un objet de référence.

Soit S l’espace de l’image, typiquement \mathbb{Z}^3 pour des images 3D. Un ensemble flou dans S est défini par une fonction d’appartenance $\mu : S \rightarrow [0, 1]$ où pour tout x dans S , $\mu(x)$ représente le degré avec lequel x appartient à l’ensemble flou. Pour un ensemble net (non flou, binaire), les degrés d’appartenance prennent leurs valeurs dans $\{0, 1\}$. Dans un cadre de modélisation des relations spatiales, $\mu(x)$ peut représenter le degré de satisfaction d’une relation spatiale donnée en un point x (pixel ou voxel) de l’image par rapport à un objet de référence.

2.2.1. Inclusion

Des descriptions de rôles telles que *partie de* et *contient* sont souvent utilisées dans les représentations structurelles de l’anatomie, d’où l’intérêt de fonder le raisonnement spatial dans l’image sur ces relations.

Dans le cas binaire, le degré de satisfaction de la relation « X est inclus dans Y » est lui-même binaire. Un degré avec lequel un objet est inclus dans un autre peut aussi être défini par une valeur dans $[0, 1]$, par exemple $\frac{|X \cap Y|}{|X|}$, où $|\cdot|$ représente le cardinal d’un ensemble. Dans le cas où les objets sont flous, plusieurs définitions de l’inclusion ont été introduites dans la littérature (Dubois, Prade, 1980). Un exemple s’écrit $\inf_{x \in S} I(u(x), v(x))$, où u et v sont les fonctions d’appartenance des ensembles flous, et I est une implication floue. L’extension de la définition par la cardinalité donnée ci-dessus, et que nous utilisons dans nos expériences, peut être étendue selon :

$$\frac{\sum_{x \in S} \min(u(x), v(x))}{\sum_{x \in S} u(x)}, \quad (1)$$

l’intersection floue étant définie ici par le minimum et le cardinal par la somme.

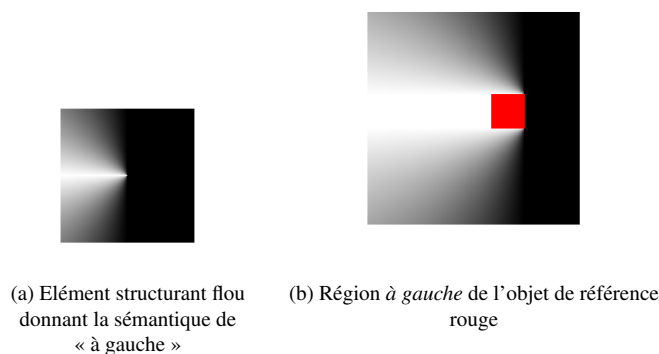
2.2.2. Relations directionnelles

A l’instar des relations d’inclusion, les relations directionnelles (par exemple à droite, en-dessous, au-dessus, etc.) sont abondamment utilisées par les experts pour décrire l’anatomie cérébrale (Neuranat, 2006 ; Waxman, 2000 ; Rosse *et al.*, 2003)). De telles relations sont intrinsèquement imprécises même quand elles impliquent des objets nets, d’où l’intérêt de représentations floues. Dans (Bloch, 2005), l’auteure a introduit des modèles de calcul de relations directionnelles floues dans l’espace de

l'image en exploitant la morphologie mathématique floue. A titre d'exemple, la relation *à gauche* (dans l'image) est calculée par une dilatation floue de l'objet de référence R (figure 3a et carré rouge dans la figure 4b) par un élément structurant flou (figure 4a). Le résultat (figure 4b) est une carte floue dans laquelle chaque pixel (ou voxel) a pour valeur le degré de satisfaction de la relation spatiale. Un degré d'inclusion peut alors être calculé entre l'objet cible A (figure 3b par exemple) et cette carte floue, donnant alors le degré de satisfaction d'une relation spatiale donnée.

(a) Objet de référence R .(b) Objet cible A .

Figure 3. Un exemple d'objets cible et de référence pour illustrer le calcul de la relation « à gauche »



(a) Élément structurant flou donnant la sémantique de « à gauche »

(b) Région à gauche de l'objet de référence rouge

Figure 4. Calcul des relations directionnelles par morphologie mathématique floue

Une autre approche repose sur l'histogramme d'angles $H^R(A)$, qui calcule la fréquence des angles entre tous les segments joignant un point de R et un point de A et l'axe horizontal. En 3D, deux angles (α_1, α_2) sont considérés : α_1 prend ses valeurs dans $[-\pi, \pi]$ et mesure l'angle entre les projections des segments sur le plan x-y et l'axe des x ; α_2 prend ses valeurs dans $[-\pi/2, \pi/2]$ et mesure l'angle entre le segment et la projection sur le plan x-y.

Le degré de satisfaction d'une relation directionnelle donnée est calculé en mesurant l'intersection entre l'histogramme (figure 5 pour les objets A et R de la figure 3) et une fonction d'appartenance donnant la sémantique de la relation (figure 6).

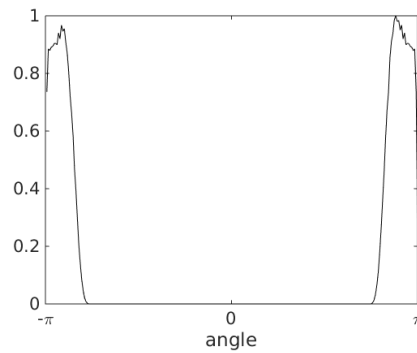


Figure 5. Histogramme d'angles entre les deux objets de la figure 3

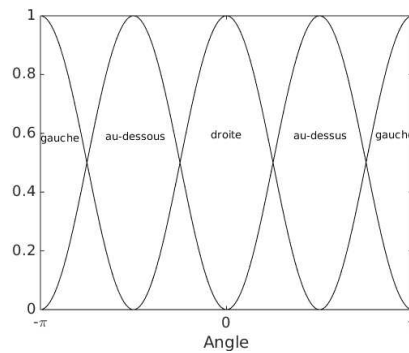


Figure 6. Fonctions d'appartenance définissant la sémantique des relations de direction

2.2.3. Relations de distance

Les relations de distance telles que *proche de* ou *loin de* peuvent aussi bénéficier d'une modélisation par ensembles flous, comme illustré dans la figure 9. A titre d'exemple, la relation *loin de* peut être calculée en appliquant une fonction d'appartenance croissante (figure 7) à la carte de distance (figure 8) associée à un objet de référence. Cela permet d'associer à chaque point de l'image un degré de satisfaction de la relation dans $[0, 1]$. Dans la figure 9, l'objet de référence est le ventricule latéral droit, et les pixels clairs indiquent une forte satisfaction de la relation *loin du ventricule latéral*. La satisfaction de la relation entre deux objets peut alors être calculée en quantifiant le degré d'inclusion de l'objet cible dans la région associée à la relation.

Un autre mode de calcul du degré de satisfaction d'une relation de distance consiste à appliquer une fonction d'appartenance donnant la sémantique de la relation (figure 7) à une valeur de distance d entre l'objet cible et l'objet de référence (distance minimum, moyenne ou de Hausdorff par exemple, figure 10). Les relations de distance dépendent du système de référence et de l'échelle des images. Les paramètres de la fonction d'appartenance sont modifiables selon les applications et peuvent être appris à partir d'une base de données annotées (Atif *et al.*, 2007). Nous utilisons cette approche et la fonction d'appartenance de la figure 7 dans nos expériences.

Notons aussi que, à l'instar des histogrammes d'angles, des histogrammes de distances peuvent être calculés. On peut donc mesurer le degré de satisfaction d'une relation de distance par mise en correspondance floue entre l'histogramme de distance et la fonction d'appartenance donnant la sémantique de la relation.

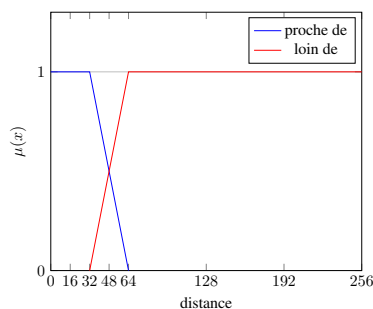


Figure 7. Fonctions d'appartenance de relations de distance (proche de et loin de)

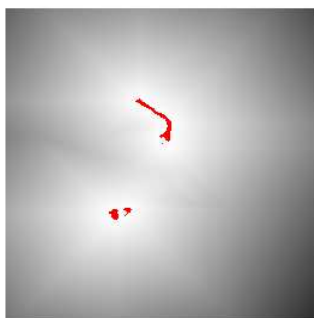


Figure 8. Carte de distance au ventricule latéral (en rouge) : les niveaux de gris élevés correspondent à de petites valeurs de la distance



Figure 9. Région floue loin du ventricule latéral (contour rouge)

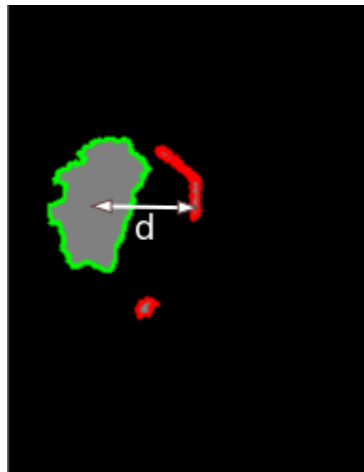


Figure 10. Evaluation de la distance entre deux structures cérébrales

2.3. Raisonnement abductif en logique de description

Dans le cadre des LD, l'abduction peut être considérée selon différents angles (Elsenbroich *et al.*, 2006; Klarman *et al.*, 2011) : abduction de concepts, abduction de TBox, abduction de ABox et abduction de base de connaissances. Nous détaillons dans la suite le cas général de l'abduction de concepts. En effet des assertions de la ABox peuvent être représentées par des concepts, et l'abduction d'une base

de connaissances peut être vue comme un ensemble d'abductions de concepts. Dans un contexte d'interprétation d'images, le résultat de la première étape de segmentation et de reconnaissance est traduit sous la forme d'assertions dans la ABox, puis en un *concept spécifique*, défini dans la suite.

DÉFINITION 7 (Abduction de concepts). — *Soit \mathcal{L} une LD quelconque, $\mathcal{K} = \{\mathcal{T}, \mathcal{R}, \mathcal{A}\}$ une base de connaissances dans \mathcal{L} , et C, D deux concepts dans \mathcal{L} , supposés satisfiables par rapport à \mathcal{K} . Le problème d'abduction de concepts \mathcal{P} est défini comme suit : étant donnée une observation sous forme d'un concept \mathcal{O} , un concept \mathcal{H} satisfiable dans \mathcal{K} est une explication qui porte l'information au niveau abstrait pour impliquer l'observation si $\mathcal{K} \models \mathcal{H} \sqsubseteq \mathcal{O}$ et $\mathcal{H} \not\models \mathcal{O}$.*

DÉFINITION 8 (Concept le plus spécifique). — *(Baader, 2003) Soient \mathcal{T} une TBox dans \mathcal{L} , $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ une interprétation, $X \subseteq \Delta^{\mathcal{I}}$, et E un concept dans \mathcal{L} . Le concept E est un concept le plus spécifique de X dans l'interprétation \mathcal{I} si :*

- $X \subseteq E^{\mathcal{I}}$,
- pour tout concept $F \in \mathcal{L}$ tel que $X \subseteq F^{\mathcal{I}}$, $\mathcal{T} \models E \sqsubseteq F$.

Par exemple, en considérant la ABox

$$\begin{aligned} ABox = \{ & a : BrainTumor \\ & b : NonEnhanced \\ & \langle a, b \rangle : hasEnhancement \}, \end{aligned}$$

le concept le plus spécifique de $a^{\mathcal{I}}$ s'écrit :

$$BrainTumor \sqcap \exists hasEnhancement.NonEnhanced$$

DÉFINITION 9 (Sous-concept). — *L'ensemble $sub(D)$ des sous-concepts d'un concept D comporte tous les concepts apparaissant dans D , et est défini récursivement comme suit :*

$$\begin{aligned} sub(A) &= \{A\} \text{ pour les noms de concepts } A \in N_C \\ sub(C \sqcap E) &= \{C \sqcap E\} \cup sub(C) \cup sub(E) \\ sub(C \sqcup E) &= \{C \sqcup E\} \cup sub(C) \cup sub(E) \\ sub(\exists r.C) &= \{\exists r.C\} \cup sub(C) \\ sub(\forall r.C) &= \{\forall r.C\} \cup sub(C) \end{aligned}$$

Dans le cas de la base de connaissances sur le cerveau (cf. section 5), nous avons par exemple :

$$\begin{aligned} sub(\exists leftOf.CNI \sqcap \exists closeTo.CNI) &= \{\exists leftOf.CNI \sqcap \exists closeTo.CNI, \\ & \exists leftOf.CNI, \\ & \exists closeTo.CNI, \\ & CNI\} \end{aligned}$$

3. Un nouvel algorithme d'abduction reposant sur les tableaux sémantiques

Nous présentons dans cette section une partie de notre contribution, à savoir un nouvel algorithme d'abduction de concepts fondé sur les tableaux sémantiques pour la logique $\mathcal{ALCHL}_{\mathcal{R}^+}$. Notre approche s'inspire de l'approche introduite dans (Colucci *et al.*, 2004) et l'enrichit. L'abduction donnée dans (Colucci *et al.*, 2004) répond au but d'établir une relation entre un demandeur et un fournisseur (*matchmaking decision* en anglais) qui partagent les mêmes concepts. Dans notre formalisation, puisque nous cherchons une description de haut niveau qui peut être quelconque, nous nous affranchissons de cette contrainte, et généralisons leur approche. Nous détaillons en particulier ici la phase de génération d'hypothèses et la phase de sélection est décrite dans la section suivante.

Avant de détailler notre approche par tableaux, rappelons quelques définitions préliminaires.

DÉFINITION 10 (Forme normale négative). — Une forme normale négative (FNN) est une expression de concept telle que le constructeur de négation apparaît seulement devant les concepts atomiques. Les règles de réécriture pour construire une telle forme sont :

- $\neg(\neg C) \equiv C$,
- $\neg(C \sqcup D) \equiv \neg C \sqcap \neg D$,
- $\neg(C \sqcap D) \equiv \neg C \sqcup \neg D$,
- $\neg(\exists r.C) \equiv \forall r.\neg C$,
- $\neg(\forall r.C) \equiv \exists r.\neg C$.

Par exemple, la forme normale négative du concept $\neg(\text{BrainStructure} \sqcap \exists \text{leftOf}. \text{CNl})$ s'écrit $\neg \text{BrainStructure} \sqcup \forall \text{leftOf}.\neg \text{CNl}$.

DÉFINITION 11 (Forme normale conjonctive). — Une forme normale conjonctive (FNC) est une expression de concepts telle que les concepts complexes sont remplacés par la conjonction de leurs super-concepts (c'est-à-dire les concepts dont ils sont des sous-concepts) en prenant en compte la TBox. Pour un concept C et une TBox $\mathcal{T} = \{C \sqsubseteq D\}$, $FNC(C, \mathcal{T}) = C \sqcap D$.

DÉFINITION 12 (Concept internalisé). — Soit \mathcal{T} une TBox comportant un ensemble d'axiomes de la forme $C_i \sqsubseteq D_i$. Le concept internalisé de la TBox \mathcal{T} s'écrit :

$$C_{\mathcal{T}} \equiv \sqcap_{(C_i \sqsubseteq D_i \in \mathcal{T})} (\neg C_i \sqcup D_i)$$

Par exemple, le concept internalisé de l'axiome $LVI \sqsubseteq \text{BrainStructure} \sqcap \exists(\text{rightOf} \sqcap \text{closeTo}). \text{CNl}$ s'écrit $\neg LVI \sqcup (\text{BrainStructure} \sqcap \exists(\text{rightOf} \sqcap \text{closeTo}). \text{CNl})$.

La proposition suivante donne notre stratégie de résolution du problème d'abduction.

PROPOSITION 13 (Stratégie pour l'abduction de concepts dans les LD). — *Le problème d'abduction de concepts \mathcal{P} dans la définition 7 (étant donnée une observation sous forme d'un concept \mathcal{O} , un concept \mathcal{H} satisfiable dans \mathcal{K} est une explication si $\mathcal{K} \models \mathcal{H} \sqsubseteq \mathcal{O}$ et $\mathcal{H} \not\models \mathcal{O}$) revient à trouver un concept \mathcal{H} tel que $\mathcal{H} \sqcap \neg\mathcal{O}$ est insatisfiable dans \mathcal{K} .*

Si l'on ne considère que la partie terminologique, le problème revient à trouver l'ensemble des concepts \mathcal{H} qui rendent le concept $\mathcal{H} \sqcap \neg\mathcal{O} \sqcap C_{\mathcal{T}}$ insatisfiable.

Preuve : Soit \mathcal{T} une TBox dans la base de connaissances \mathcal{K} , \mathcal{H} et \mathcal{O} deux concepts qui satisfont $\mathcal{K} \models \mathcal{H} \sqsubseteq \mathcal{O}$. Pour un ensemble $\{a, b, c, \dots\}$ d'éléments d'interprétation de \mathcal{K} , si $a \in \mathcal{H}^{\mathcal{I}}$, on a aussi $a \in \mathcal{O}^{\mathcal{I}}$. Donc $a \notin \neg\mathcal{O}^{\mathcal{I}}$ et $\mathcal{K} \models \mathcal{H} \sqcap \neg\mathcal{O} \sqsubseteq \perp$.

A la lumière de la proposition ci-dessus, il est alors possible d'utiliser une méthode de vérification de satisfiabilité pour générer l'ensemble des hypothèses. Dans les logiques de description, les méthodes par tableaux sémantiques permettent de faire cela efficacement (Baader *et al.*, 2003). L'approche par tableaux sémantiques permet en effet de construire pour un concept C donné un modèle (une interprétation non vide) relativement à la base de connaissances, en construisant de façon récursive l'arbre d'interprétation, par l'application d'un ensemble de règles d'expansion. Pour les besoins de l'abduction, cela revient à trouver les concepts \mathcal{H} qui conduisent à la *fermeture* de l'arbre associé au concept $\mathcal{H} \sqcap \neg\mathcal{O} \sqcap C_{\mathcal{T}}$. Cette notion de fermeture sera précisée dans la suite.

DÉFINITION 14 (Tableau sémantique pour $\mathcal{ALCHIT}_{\mathcal{R}^+}$). — *Soit D un concept $\mathcal{ALCHIT}_{\mathcal{R}^+}$ en FNN, R_D l'ensemble des rôles dans $\mathcal{ALCHIT}_{\mathcal{R}^+}$, et $\mathcal{I} = (\mathcal{I}, \Delta^{\mathcal{I}})$ une interprétation. Le tableau sémantique T pour D est défini comme un triplet $(\mathbf{S}, \mathcal{L}, \mathcal{E})$, où \mathbf{S} est un ensemble d'éléments dans $\Delta^{\mathcal{I}}$, \mathcal{L} associe à chaque élément dans \mathbf{S} un ensemble de concepts dans D ($\mathcal{L} : \mathbf{S} \rightarrow \mathbb{P}(\text{sub}(D))$), où $\mathbb{P}(\text{sub}(D))$ est l'ensemble des parties de $\text{sub}(D)$, \mathcal{E} associe à chaque élément dans \mathbf{S} un ensemble de rôles dans R_D ($\mathcal{E} : \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{P}(R_D)$).*

Soient x et y deux éléments de \mathbf{S} , C, E deux sous-concepts de D et r un rôle dans R_D . Le tableau sémantique permettant de vérifier la satisfiabilité du concept D est une structure d'arbre où chaque nœud correspond à un élément x de $\Delta^{\mathcal{I}}$. Chaque nœud est étiqueté par un ensemble de concepts $\mathcal{L}(x)$. Chaque arc entre deux nœuds x et y est étiqueté par les rôles associant ces derniers ($r \in \mathcal{E}(\langle x, y \rangle)$). Les règles d'expansion permettant le développement de l'arbre sont les suivantes :

1. *si $C \in \mathcal{L}(x)$, alors $\neg C \notin \mathcal{L}(x)$;*
2. *si $C \sqcap E \in \mathcal{L}(x)$, alors $C \in \mathcal{L}(x)$ et $E \in \mathcal{L}(x)$;*
3. *si $C \sqcup E \in \mathcal{L}(x)$, alors $C \in \mathcal{L}(x)$ ou $E \in \mathcal{L}(x)$;*
4. *si $\exists r.C \in \mathcal{L}(x)$, alors il existe un élément $y \in \mathbf{S}$ tel que $r \in \mathcal{E}(\langle x, y \rangle)$ et $C \in \mathcal{L}(y)$;*
5. *si $\forall r.C \in \mathcal{L}(x)$, alors pour tout $y \in \mathbf{S}$ tel que $r \in \mathcal{E}(\langle x, y \rangle)$, $C \in \mathcal{L}(y)$;*
6. *si $\forall r.C \in \mathcal{L}(x)$ et r est un rôle transitif, alors pour tout $y \in \mathbf{S}$ tel que $r \in \mathcal{E}(\langle x, y \rangle)$, $\forall r'.C \in \mathcal{L}(y)$;*

7. $r \in \mathcal{E}(\langle x, y \rangle)$ ssi $r^- \in \mathcal{E}(\langle y, x \rangle)$;
8. si $r \in \mathcal{E}(\langle x, y \rangle)$ et $r \sqsubseteq v$ (ou $r^- \sqsubseteq v^-$) alors $v \in \mathcal{E}(\langle x, y \rangle)$.

DÉFINITION 15 (Conflit). — Une branche contient un conflit (la branche est fermée) quand $\{C, \neg C\} \subseteq \mathcal{L}(x)$ pour un nœud x et un concept C .

$$\mathcal{L}(x) = \{C, \neg C\}$$

$$\quad \quad \quad \downarrow$$

$$\quad \quad \quad \boxtimes$$

Une branche est dite *complète* quand il existe un conflit dans un nœud x ou aucune des règles d'expansion de l'arbre ne peut être appliquée. Pour un concept D donné, D est satisfiable si et seulement si toutes les branches de l'arbre sauf une sont fermées. Autrement le concept D est dit insatisfiable.

Le tableau sémantique est construit en appliquant les règles d'expansion jusqu'à ce que toutes les règles soient appliquées ou un conflit apparaît dans un nœud. Pour les besoins de l'abduction, l'ensemble des hypothèses est constitué de concepts construits par conjonction avec au moins la négation d'un concept dans les branches ouvertes. La TBox sert à modéliser la connaissance terminologique, qui est ensuite traduite en un concept équivalent pour être intégrée dans la construction du tableau. La RBox sert à modéliser les propriétés de rôles (ici, les relations spatiales), et est ensuite utilisée dans les règles d'expansion. La RBox est utile pour inférer les relations spatiales. Par exemple, prenons deux éléments d'interprétation x, y tels que $x \in C, y \in D$ dans le tableau sémantique et une relation $L(x, y) = \{rightOf\}$. En appliquant la règle d'inversion des rôles, on a aussi $L(y, x) = \{leftOf\}$. On peut alors vérifier le degré de satisfaction du concept D avec une description $\exists leftOf.C$ par exemple. La ABox est utilisée pour représenter les objets dans l'image observée, et sert à construire le concept correspondant à l'observation \mathcal{O} . La RBox et la TBox peuvent bien sûr être confondues. L'intérêt de les maintenir séparées est d'ordre pratique puisque l'on peut appliquer un raisonnement particulier sur les RBox (raisonnement spatial dans notre cas) sans définir et porter celui-ci sur toute la TBox.

L'ensemble d'hypothèses est infini dans les logiques de description. Des contraintes sont donc souvent imposées pour limiter l'ensemble de candidats en utilisant un langage moins riche ou une évaluation quantitative définie selon l'application considérée (Bienvenu, 2008 ; Klarman *et al.*, 2011).

Un moyen de fermer le tableau est de l'augmenter par la conjonction des concepts qui ferment chacune des branches ouvertes. Plus précisément, pour chaque branche ouverte i , on peut générer un ensemble de concepts, notés H_i , composés des négations des concepts atomiques dans la branche i . Ainsi, une hypothèse \mathcal{H} est construite comme la conjonction des concepts dans chaque H_i . C'est un mode d'abduction où l'explication est conjonctive. D'autres modes existent, comme la disjonction des H_i (Eiter, Gottlob, 1995). La disjonction de toutes les hypothèses est alors aussi une

hypothèse. Cette hypothèse inclut toutes les possibilités, et apporte donc très peu d'information pour la prise de décision dans le contexte d'une application réelle.

L'ensemble des hypothèses ainsi généré peut contenir beaucoup de redondances, ou des concepts non satisfiables. Pour éviter les redondances, nous commençons par construire l'ensemble de couverture minimale (*minimum hitting set*).

DÉFINITION 16 (Couverture minimale). — Soit $\{S_1, \dots, S_n\}$ une collection d'ensembles. Une couverture T est un ensemble $T \subseteq \cup_{i=1}^n S_i$ tel que T contient au moins un élément de chaque ensemble de la collection $T \cap S_i \neq \emptyset$ ($1 \leq i \leq n$). La couverture minimale est une couverture T_m telle qu'il n'existe aucune couverture T' telle que $T' \subset T_m$.

Une telle couverture minimale de concepts garantit la validité syntaxique des hypothèses relativement à un langage donné (LD dans notre cas). Les hypothèses insatisfiables ($\mathcal{K} \cup \mathcal{H} \models \perp$) ou non pertinentes ($\mathcal{H} \models O$) sont aussi éliminées de l'ensemble des hypothèses. L'algorithme 1 détaille une procédure exhaustive adaptée à notre problème, et reposant sur le calcul de couverture minimale introduit dans (Reiter, 1987).

L'algorithme proposé est composé d'une méthode par tableau sémantique et d'un traitement a posteriori utilisant la couverture minimale. La complexité d'un test de satisfiabilité d'un concept dans un tableau sémantique¹ est EXPTIME dans $\mathcal{ALCHIT}_{\mathcal{R}^+}$ (Horrocks, Sattler, 1999). Dans notre algorithme, toutes les branches sont générées, ce qui est le pire cas pour le calcul. L'algorithme de la couverture minimale est un problème NP-complet. Nous démontrons aussi ci-dessous la cohérence de notre méthode. Les meilleures explications sont sélectionnées selon les évaluations des relations spatiales, avec un critère défini à la section suivante.

THÉORÈME 17 (Cohérence). — *Tous les résultats générés par la méthode par tableau sémantique et la couverture minimale sont cohérents.*

Preuve : Toutes les explications \mathcal{H} générées ferment toutes les branches ouvertes dans le tableau. Donc on a $\mathcal{K} \models \mathcal{H} \sqsubseteq \mathcal{O}$. Les explications incohérentes sont ensuite éliminées par l'algorithme 1. Les explications non pertinentes ($\mathcal{H} \models O$) sont éliminées dans le traitement a posteriori.

4. Critère de sélection d'hypothèses guidé par l'information de l'image

Cette section décrit l'autre contribution importante de ce travail, à savoir la définition de critères de sélection d'hypothèses cohérentes et pertinentes en prenant en compte l'information dans l'image (relations spatiales principalement), et en exploitant les domaines concrets et les représentations floues. Cela permet d'une part de réduire le fossé sémantique (avec le concours à la fois des domaines concrets et des représentations floues), et d'autre part, d'ordonner l'espace des hypothèses à l'aide d'une information concrète dans l'image.

1. Pour le test de satisfiabilité, la méthode par tableau se termine lorsqu'une branche ouverte est trouvée.

Données : Une collection d'ensembles $\{S_1, \dots, S_n\}$;

Résultat : Une collection de couvertures minimales \mathcal{H} ;

Initialisation : $\mathcal{H} = \emptyset$;

Initialisation de la racine de l'arbre ;

pour i allant de 1 à n **faire**

 Créer une feuille pour chaque S_i dans chaque branche ;

 Construire une hypothèse intermédiaire H_j à partir de la conjonction de tous les concepts dans la même branche de l'arbre ;

 Supprimer la branche j si H_j est insatisfiable avec la TBox ;

fin

La conjonction de tous les concepts dans chaque branche j définit une hypothèse potentielle H_j ;

$\mathcal{H} = \mathcal{H} \cup \{H_j\}$;

retourner \mathcal{H} ;

Algorithme 1 : Recherche exhaustive d'ensembles de couvertures minimales pour la génération d'hypothèses

Précisons les choses au travers de l'exemple sur le cerveau atteint d'une pathologie. Le but de l'interprétation pour cet exemple est de fournir une description plus ou moins détaillée selon la personne à qui s'adresse le résultat. Il s'agit alors de trouver l'explication (hypothèse) la plus spécifique et qui a un degré de satisfaction élevé dans l'image. Soit $\{\mathcal{H}_1, \dots, \mathcal{H}_n\}$ un ensemble de n concepts spécifiques candidats. Chaque \mathcal{H}_i peut être écrit en FNC modulo la théorie \mathcal{T} . Puisque chaque concept est ancré dans le domaine de l'image, un degré de satisfaction $SD(NCF(\mathcal{H}_i))$ impliquant les relations spatiales est calculé, permettant ainsi de sélectionner la « meilleure » explication.

4.1. Représentations floues dans les domaines concrets

L'information spatiale bénéficie d'une représentation floue permettant de prendre en compte l'imprécision dans l'image et l'aspect vague des relations elles-mêmes. Nous introduisons des domaines concrets spatiaux flous¹ pour représenter ces informations, et les associer à des concepts et des relations dans la partie terminologique.

DÉFINITION 18 (Domaine concret spatial flou). — *Un domaine concret spatial flou est une paire $\mathcal{D} = \{\Delta_{\mathcal{D}}, \Phi_{\mathcal{D}}\}$, où $\Delta_{\mathcal{D}}$ est un sous-ensemble de l'espace de l'image S (S étant \mathbb{Z}^3 pour une image en 3D par exemple), et $\Phi_{\mathcal{D}}$ est un ensemble de fonctions composé de :*

- une fonction f associant un individu dans la ABox à une région floue dans $\Delta_{\mathcal{D}}$;

1. Cette expression est un raccourci pour exprimer que les concepts sont représentés par des sous-ensembles flous du domaine concret, celui-ci n'étant en réalité pas flou (le domaine spatial de l'image dans les exemples).

– une fonction d'évaluation e_r^t assignant à une paire d'individus (r, t) un degré de satisfaction d'une relation spatiale, où r est l'objet de référence, et t l'objet cible.

Supposons par exemple que, dans la figure 1, une région ait été segmentée et reconnue comme étant le ventricule latéral droit (on peut utiliser par exemple les méthodes de (Fouquier *et al.*, 2012; Nempont *et al.*, 2013)). Un individu $c : LVr$ est alors ajouté à la ABox. L'individu c est donc ancré dans le domaine spatial par la fonction $f(c)$. Un degré de satisfaction de la relation à droite entre l'objet cible $a : BrainTumor$ et l'objet de référence $c : LVr$ peut donc être calculé. Nous avons donc un degré de satisfaction $e_c^a(rightOf)$ de la proposition « la région $f(a)$ est à droite de la région $f(c)$ » dans le domaine de l'image.

4.2. Critère spatial de sélection

Des critères de minimalité sont nécessaires pour sélectionner la « meilleure » explication dans un ensemble d'hypothèses. Différents critères sont discutés dans (Bienvenu, 2008) pour les logiques de description. Nous introduisons ici un nouveau critère qui trouve son intérêt dans le domaine de l'interprétation d'images.

Le critère spatial de sélection est une estimation quantitative de la satisfaction d'un concept C en prenant en compte les descriptions spatiales $\exists sr.D$ ou $\forall sr.D$, où sr est un rôle décrivant une relation spatiale telle que à droite, derrière, devant, proche, loin, etc. Le degré de satisfaction SD d'un concept complexe est calculé par minimum ou maximum des degrés de satisfaction de ses sous-concepts en fonction des connecteurs logiques utilisés. Par exemple, le degré de satisfaction d'un concept construit par conjonction de deux concepts est calculé comme le minimum entre les degrés de satisfactions des deux sous-concepts. Le degré de satisfaction du concept $\exists r.D$ correspond au degré de satisfaction du meilleur candidat qui satisfait la relation r . Le degré de satisfaction SD est calculé comme suit :

– soit une assertion $a : C, E = \exists sr.D$ et $E \in sub(C)$, et un ensemble d'individus I , où $i \in I$ est une instance du concept D dans la ABox ($i : D$), $SD(E) = \sup_{i \in I}(e_i^a(sr))$;

– soit une assertion $a : C, E = \forall sr.D$ et $E \in sub(C)$, et un ensemble d'individus I , où $i \in I$ est une instance du concept D dans la ABox ($i : D$), $SD(E) = \inf_{i \in I}(e_i^a(sr))$;

– pour la conjonction de descriptions spatiales $C = \prod_{j \in J} C_j$, $SD(C) = \inf_{j \in J}(SD(C_j))$;

– pour la disjonction de descriptions spatiales $C = \sqcup_{j \in J} C_j$, $SD(C) = \sup_{j \in J}(SD(C_j))$.

5. Exemples d'illustrations

5.1. Interprétation d'images IRM du cerveau

Pour illustrer l'approche proposée pour l'interprétation sémantique d'images, reprenons l'exemple de l'image IRM du cerveau pathologique.

La TBox suivante décrit les connaissances anatomiques et sur les pathologies, et la ABox représente les résultats sur les structures extraites d'une image IRM par des méthodes d'analyse d'images appropriées, que nous ne décrivons pas ici (voir (Fouquier *et al.*, 2012; Nempont *et al.*, 2013))¹. Nous reprenons ici la RBox de la section 2.1.

$$\begin{aligned}
 RBox = \{ & \text{rightOf} \equiv \text{leftOf}^- \\
 & \text{above} \equiv \text{below}^- \\
 & \text{closeTo} \equiv \text{closeTo}^- \\
 & \text{farFrom} \equiv \text{farFrom}^- \\
 & \text{isPartOf} \circ \text{isPartOf} \sqsubseteq \text{isPartOf} \\
 & \text{hasPart} \circ \text{hasPart} \sqsubseteq \text{hasPart} \\
 & \text{isPartOf} \equiv \text{hasPart}^- \}
 \end{aligned}$$

$$\begin{aligned}
 TBox = \{ & \text{Hemisphere} \sqsubseteq \exists \text{isPartOf}. \text{Brain} \\
 & \text{BrainStructure} \sqsubseteq \exists \text{isPartOf}. \text{Brain} \\
 & \text{BrainDisease} \sqsubseteq \exists \text{isPartOf}. \text{Brain} \sqcap \\
 & \quad \neg \text{BrainStructure} \\
 & \text{BrainTumor} \sqsubseteq \text{BrainDisease} \\
 & \text{LVl} \sqsubseteq \text{BrainStructure} \\
 & \text{LVr} \sqsubseteq \text{BrainStructure} \\
 & \text{CNl} \sqsubseteq \text{BrainStructure} \\
 & \text{CNr} \sqsubseteq \text{BrainStructure} \\
 & \text{PeripheralRegion} \sqsubseteq \exists \text{isPartOf}. \text{Brain} \\
 & \text{SubCorticalRegion} \sqsubseteq \exists \text{isPartOf}. \text{Brain}
 \end{aligned}$$

1. Dans les deux exemples de cette section, nous conservons les descriptions en anglais.

$$\begin{aligned}
& \textit{SmallDeformingTumor} \sqsubseteq \textit{BrainTumor} \sqcap \\
& \quad \exists \textit{hasEnhancement.NonEnhanced} \\
& \textit{PeripheralSmallDeformingTumor} \sqsubseteq \textit{SmallDeformingTumor} \sqcap \\
& \quad \exists \textit{isPartOf.PeripheralRegion} \sqcap \\
& \quad \exists \textit{farFrom.(LVl} \sqcup \textit{LVr)} \\
& \textit{SubCorticalSmallDeformingTumor} \sqsubseteq \textit{SmallDeformingTumor} \sqcap \\
& \quad \exists \textit{isPartOf.SubCorticalRegion} \sqcap \\
& \quad \exists \textit{rightOf.CNr} \}
\end{aligned}$$

$$\begin{aligned}
\textit{ABox} = \{ & a : \textit{BrainTumor} \\
& b : \textit{NonEnhanced} \\
& \langle a, b \rangle : \textit{hasEnhancement} \\
& c : \textit{CNr} \\
& d : \textit{CNl} \\
& k : \textit{LVr} \\
& m : \textit{LVl} \\
& p : \textit{PeripheralRegion} \\
& s : \textit{SubCorticalRegion} \}
\end{aligned}$$

La figure 11 illustre un résultat de segmentation sur une coupe. Les instances de la ABox, telles que a , c , d , k , m , p et s , sont associées à des sous-ensembles du domaine concret (l'espace de l'image ici).

L'objectif est d'exprimer une description de haut niveau de la tumeur dans le langage terminologique. Le concept le plus spécifique (Baader, 2003) de $a^{\mathcal{I}}$ est formalisé par $\textit{BrainTumor} \sqcap \exists \textit{hasEnhancement.NonEnhanced}$. La méthode du tableau sémantique est alors appliquée afin d'exhiber les hypothèses cohérentes. On présente une partie du tableau sémantique de cet exemple dans la figure 12, illustrant l'application de deux axiomes de la base de connaissances (ces axiomes sont notés en gras dans la TBox ci-dessus.). La racine du tableau sémantique est la négation de l'observation et les concepts internalisés de ces deux axiomes. Les opérateurs à décomposer et les conflits sont indiqués en gras dans la figure. Le tableau a deux branches ouvertes, qui peuvent être fermées avec les négations des concepts atomiques de ces deux branches, définissant ainsi les hypothèses. Les hypothèses possibles sont $\textit{BrainTumor} \sqcap \exists \textit{hasEnhancement.NonEnhanced}$, $\textit{SmallDeformingTumor}$, $\textit{PeripheralSmallDeformingTumor}$. La première hypothèse n'est pas une expli-

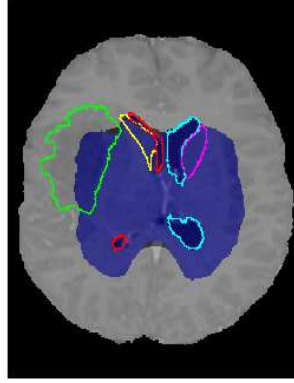


Figure 11. Segmentation de l'image IRM (sur une coupe). Les régions et contours en couleurs montrent les structures segmentées. La région bleue représente la région sous-corticale et la région grise la région périphérique. Les contours jaunes et violets sont ceux des noyaux caudés, les contours rouges et cyan ceux des ventricules latéraux, et les contours en vert sont ceux de la tumeur

cation car c'est l'observation elle-même ($\mathcal{H} \models \mathcal{O}$). Si l'on prend en compte tous les axiomes dans la base de connaissances, un ensemble d'hypothèses possibles est alors : $\{SmallDeformingTumor, PeripheralSmallDeformingTumor, SubCorticalSmallDeformingTumor\}$.

Les concepts *PeripheralSmallDeformingTumor* et *SubCorticalSmallDeformingTumor* sont des hypothèses plus spécifiques que *SmallDeformingTumor*.

Ensuite, le critère spatial de sélection est appliqué pour trouver la meilleure explication au sens de ce critère. Pour cela, les deux hypothèses les plus spécifiques sont exprimées sous forme normale conjonctive :

$$\begin{aligned}
 &PeripheralSmallDeformingTumor \sqcap SmallDeformingTumor \sqcap \\
 &\quad \exists isPartOf.PeripheralRegion \sqcap \\
 &\quad (\exists farFrom.LVl \sqcup \\
 &\quad \exists farFrom.LVr) \\
 &SubCorticalSmallDeformingTumor \sqcap SmallDeformingTumor \sqcap \\
 &\quad \exists isPartOf.SubCorticalRegion \sqcap \\
 &\quad \exists rightOf.CNr
 \end{aligned}$$

Par exemple, l'hypothèse *PeripheralSmallDeformingTumor* contient trois éléments de description spatiale : $\exists isPartOf.PeripheralRegion$, $\exists farFrom.LVl$ et $\exists farFrom.LVr$. Il faut donc calculer le degré de satisfaction des relations d'in-



Figure 12. Une partie du tableau sémantique construit pour expliquer l'image IRM du cerveau pathologique

clusion de la représentation spatiale floue de *BrainTumor* (en vert dans la figure 11) dans celle de *PeripheralRegion* (en gris), ainsi que des relations de distances entre les représentations spatiales floues de *BrainTumor* et *LVI* (en violet), *LVR* (en rouge). Ici nous prenons la définition du degré d'inclusion en fonction du volume (ou cardinal) des deux objets (équation 1), et la distance calculée est la distance euclidienne moyenne. Le degré global de satisfaction, après agrégation, de *PeripheralSmallDeformingTumor* est :

$$\begin{aligned} & \min(e_p^{\alpha}(isPartOf), \max(e_k^{\alpha}(farFrom), e_m^{\alpha}(farFrom))) \\ & = \min(0, 89, \max(0, 13, 0, 34)) = 0, 34. \end{aligned}$$

De même, le degré global de satisfaction de *SubCorticalSmallDeformingTumor* est :

$$\min(e_s^{\alpha}(isPartOf), e_c^{\alpha}(rightOf)) = \min(0, 81, 0, 11) = 0, 11.$$

Nous choisissons la description avec la valeur la plus élevée parmi tous les candidats. Dans cette illustration, la valeur de satisfaction de la distance a une valeur moins élevée, mais est très importante pour sélectionner le meilleur résultat. Comme les valeurs finales dépendent de l'application (et donc la définition des fonctions d'appartenance), les valeurs absolues de satisfaction des critères peuvent être différentes. Seules les valeurs relatives jouent un rôle dans la décision. Dans cette illustration, ces résultats conduisent à la conclusion que *PeripheralSmallDeformingTumor* est la meilleure explication de l'image observée, ce qui correspond bien à la description de haut niveau attendue. La tumeur est située dans la région périphérique et a peu d'impact sur la déformation des autres structures internes dans la région sous-corticale.

5.2. Saut à la perche

Dans cette section, nous illustrons notre approche sur un problème d'interprétation d'une activité sportive. Nous prenons l'exemple du saut à la perche, visant à franchir une barre (*bar*) à l'aide d'une perche (*pole*). L'objectif est d'interpréter les différentes phases d'un saut, telles que l'approche, la course, l'appui sur la perche, le saut et le franchissement (*approach*, *run*, *plant*, *swing* et *cross*), observées dans une séquence d'images.

Le saut à la perche est décrit par une base de connaissances, dont une partie est donnée ici. Elle contient quatre concepts correspondant à des activités, et des concepts représentant les objets élémentaires impliqués (*player*, *pole*, *bar* et *ground*).

$$\begin{aligned} TBox = \{ & PlayerFarFromBar \sqsubseteq \exists Player \sqcap \exists farFrom.Bar \\ & PoleCloseToGround \sqsubseteq \exists Pole \sqcap \exists closeTo.Ground \\ & PlayerRightOfBar \sqsubseteq \exists Player \sqcap \exists rightOf.Bar \\ & PlayerAboveBar \sqsubseteq \exists Player \sqcap \exists above.Bar \\ & PlayerCloseToBar \sqsubseteq \exists Player \sqcap \exists closeTo.Bar \end{aligned}$$

$$\begin{aligned}
PoleVault &\sqsubseteq \exists hasPart.Bar \sqcap \\
&\quad \exists hasPart.Player \sqcap \\
&\quad \exists hasPart.Pole \\
PoleVaultApproach &\sqsubseteq PoleVault \sqcap \\
&\quad \exists hasPart.PlayerFarFromBar \\
PoleVaultPlant &\sqsubseteq PoleVault \sqcap \\
&\quad \exists hasPart.PoleCloseToGround \\
PoleVaultSwing &\sqsubseteq PoleVault \sqcap \\
&\quad \exists hasPart.PlayerRightOfBar \sqcap \\
&\quad \exists hasPart.PlayerCloseToBar \\
PoleVaultCross &\sqsubseteq PoleVault \sqcap \\
&\quad \exists hasPart.PlayerAboveBar \sqcap \\
&\quad \exists hasPart.PlayerCloseToBar \}
\end{aligned}$$

Les données utilisées dans l'exemple proviennent de la base de vidéos UCF-101 (Soomro *et al.*, 2012). Nous supposons que les régions principales dans l'image sont segmentées et reconnues. Un exemple d'image est illustré à la figure 13. Les objets sont segmentés et annotés manuellement avec l'outil *LabelMe* (Russell *et al.*, 2008).

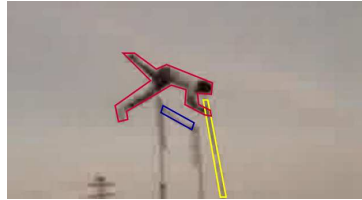


Figure 13. Phase de franchissement de la barre pendant un saut à la perche

La figure 13 montre trois régions reconnues et annotées. La région en rouge représente le sportif, celle en bleu représente la barre, et la perche correspond à la région jaune. Un individu d désigne la scène, représentant une phase inconnue pendant l'activité de saut à la perche. La scène observée est représentée par une ABox avec quatre individus a , b , c et d , et d contient d'autres individus. La ABox a ainsi l'expression suivante :

$$\begin{aligned}
ABox &= \{a : Player \\
&\quad b : Pole \\
&\quad c : Bar
\end{aligned}$$

$$\begin{aligned}
& d : \text{unknown activity} \\
& \langle d, a \rangle : \text{hasPart} \\
& \langle d, b \rangle : \text{hasPart} \\
& \langle d, c \rangle : \text{hasPart}
\end{aligned}$$

Pour interpréter globalement la scène, représentée par une instance d , le concept le plus spécifique de $d^{\mathcal{I}}$ est $\exists \text{hasPart.}Player \sqcap \exists \text{hasPart.}Bar \sqcap \exists \text{hasPart.}Pole$. La méthode par tableau sémantique est alors appliquée pour déterminer des hypothèses cohérentes. Un ensemble possible d'hypothèses est : $\{PoleVault, PoleVaultApproach, PoleVaultPlant, PoleVaultSwing, PoleVaultCross\}$. Il est clair que $PoleVault$ est une hypothèse plus générale que les autres.

Les explications préférées sont exprimées sous forme normale conjonctive et évaluées selon le critère spatial de sélection. Par exemple les hypothèses $PoleVaultApproach$ et $PoleVaultCross$ sont exprimées sous la forme suivante :

$$\begin{aligned}
& PoleVaultApproach \sqcap PoleVault \sqcap \exists \text{hasPart.}Bar \sqcap \\
& \quad \exists \text{hasPart.}Player \sqcap \\
& \quad \exists \text{hasPart.}Pole \sqcap \\
& \quad \exists \text{hasPart.}(PlayerFarFromBar \sqcap \\
& \quad \quad Player \sqcap \exists \text{farFrom.}Bar)
\end{aligned}$$

$$\begin{aligned}
& PoleVaultCross \sqcap PoleVault \sqcap \exists \text{hasPart.}Bar \sqcap \\
& \quad \exists \text{hasPart.}Player \sqcap \\
& \quad \exists \text{hasPart.}Pole \sqcap \\
& \quad \exists \text{hasPart.}(PlayerCloseBar \sqcap \\
& \quad \quad Player \sqcap \exists \text{closeTo.}Bar) \sqcap \\
& \quad \exists \text{hasPart.}(PlayerAboveBar \sqcap \\
& \quad \quad Player \sqcap \exists \text{above.}Bar)
\end{aligned}$$

Le degré de satisfaction global de $PoleVaultApproach$ est :

$$e_a^c(\text{farFrom}) = 0.$$

Le degré de satisfaction global de $PoleVaultCross$ est :

$$\min(e_a^c(\text{closeTo}), e_a^c(\text{above})) = \min(1, 0,75) = 0,75.$$

Dans cet exemple, le franchissement de la barre (*PoleVaultCross*) obtient le plus haut degré de satisfaction dans le domaine concret. Il est donc considéré comme la meilleure explication de la figure 13, et correspond aussi à l'interprétation de l'image qu'aurait donnée un observateur humain.

Cependant, cet exemple illustratif a des limites. En particulier une hypothèse sous-jacente est que le système de référence dans l'image correspond à celui qui est implicite dans la base de connaissances. La position de la caméra doit être constante, afin de pouvoir négliger le problème d'échelle et les directions relatives par rapport au point d'observation. Dans les deux illustrations, les relations spatiales *farFrom* et *closeTo* dépendent du système de référence et de l'échelle des images. Les paramètres de la fonction d'appartenance sont modifiables selon les applications et peuvent être appris à partir d'une base de données annotées, comme mentionné précédemment.

6. Conclusion

La contribution principale de cet article est de raisonner par abduction avec des représentations des connaissances et des informations dans des domaines concrets pour l'interprétation d'images. Le cadre proposé repose sur une logique de description et une méthode de tableau sémantique pour déterminer des hypothèses cohérentes pour l'abduction. Un critère spatial de sélection, calculé dans le domaine de l'image, permet d'identifier les explications préférées, ce qui constitue également un apport original. L'estimation quantitative exploite des représentations floues, dans le domaine concret de l'image, des relations spatiales exprimées dans la base de connaissances. Cette contribution constitue un critère concret de décision pour l'interprétation d'images, grâce aux liens établis entre deux modalités : les images d'une part, et l'inférence logique d'autre part. Pour les travaux futurs, des techniques d'optimisation de la méthode par tableaux et de recherche heuristique pourraient être intégrées pour améliorer les performances de calcul dans les applications réelles.

Remerciements

Ce travail a été financé par l'ANR (projet LOGIMA).

Bibliographie

- Atif J., Hudelot C., Bloch I. (2014). Explanatory reasoning for image understanding using formal concept analysis and description logics. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, n° 5, p. 552–570.
- Atif J., Hudelot C., Fouquier G., Bloch I., Angelini E. (2007). From Generic Knowledge to Specific Reasoning for Medical Image Interpretation using Graph-based Representations. In *International Joint Conference on Artificial Intelligence IJCAI'07*, p. 224–229. Hyderabad, India.
- Baader F. (2003). Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In *International Joint Conference on Artificial Intelligence IJCAI'03*, vol. 3, p. 319–324.

- Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F. (2003). *The description logic handbook: theory, implementation, and applications*. Cambridge University Press.
- Baader F., Hanschke P. (1991). A schema for integrating concrete domains into concept languages. In *International Joint Conference on Artificial Intelligence IJCAI'91*, p. 452–457.
- Bienvenu M. (2008). Complexity of abduction in the \mathcal{EL} family of lightweight description logics. In *11th International Conference on Principles of Knowledge Representation and Reasoning (KR08)*, p. 220–230.
- Bloch I. (2005). Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, vol. 23, n° 2, p. 89 - 110.
- Chein M., Mugnier M.-L. (2008). *Graph-based knowledge representation: computational foundations of conceptual graphs*. Springer Science & Business Media.
- Colucci S., Di Noia T., Di Sciascio E., Donini F. M., Mongiello M. (2004). A uniform tableaux-based approach to concept abduction and contraction in \mathcal{ALN} . In *17th International Workshop on Description Logics (DL)*, vol. 104, p. 158–167.
- Coradeschi S., Saffiotti A. (2000). Anchoring symbols to sensor data: preliminary report. In *AAAI conference on Artificial Intelligence*, p. 129–135.
- Dubois D., Prade H. (1980). *Fuzzy sets and systems: theory and applications*. Academic Press.
- Eiter T., Gottlob G. (1995). The complexity of logic-based abduction. *Journal of the ACM (JACM)*, vol. 42, n° 1, p. 3–42.
- Elsenbroich C., Kutz O., Sattler U. (2006). A case for abductive reasoning over ontologies. In *OWL: Experiences and Directions*, vol. 216, p. 10–20.
- Fouquier G., Atif J., Bloch I. (2012). Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. *Computer Vision and Image Understanding*, vol. 116, n° 1, p. 146–165.
- Harnad S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, vol. 42, n° 1-3, p. 335–346.
- Hobbs J. R. (2004). Abduction in natural language understanding. *Handbook of pragmatics*, p. 724–741.
- Horrocks I., Sattler U. (1999). A description logic with transitive and inverse roles and role hierarchies. *Journal of Logic and Computation*, vol. 9, n° 3, p. 385–410.
- Hudelot C., Atif J., Bloch I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, vol. 159, n° 15, p. 1929–1951.
- Klarman S., Endriss U., Schlobach S. (2011). ABox abduction in the description logic \mathcal{ALC} . *Journal of Automated Reasoning*, vol. 46, n° 1, p. 43–80.
- Krizhevsky A., Sutskever I., Hinton G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, p. 1097–1105.
- Nempont O., Atif J., Bloch I. (2013). A constraint propagation approach to structural model based image segmentation and recognition. *Information Sciences*, vol. 246, p. 1–27.
- Neumann B., Möller R. (2008). On scene interpretation with description logics. *Image and Vision Computing*, vol. 26, n° 1, p. 82–101.

- Neuranat*. (2006). <http://rprcsgi.rprc.washington.edu/neronames/>.
- Peirce C. (1958). *Collected papers of Charles Sanders Peirce, vol. I–VI edited by C. Hartshorne and P. Weiss, 1931–1935, vol. VII–VIII edited by A. W. Burks*. Harvard University Press, Cambridge MA.
- Peraldi S. E., Kaya A., Melzer S., Möller R., Wessel M. (2007). Multimedia interpretation as abduction. In *20th International Workshop on Description Logics*, p. 323–330.
- Reiter R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, vol. 32, n° 1, p. 57–95.
- Rosse C., Mejino Jr J. L. *et al.* (2003). A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of biomedical informatics*, vol. 36, n° 6, p. 478–500.
- Russell B. C., Torralba A., Murphy K. P., Freeman W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, vol. 77, n° 1-3, p. 157–173.
- Shanahan M. (2005). Perception as abduction: Turning sensor data into meaningful representation. *Cognitive Science*, vol. 29, n° 1, p. 103–134.
- Smeulders A. W., Worring M., Santini S., Gupta A., Jain R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n° 12, p. 1349–1380.
- Soomro K., Zamir A. R., Shah M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.
- Town C. (2006). Ontological inference for image and video analysis. *Machine Vision and Applications*, vol. 17, n° 2, p. 94–115.
- Vinyals O., Toshev A., Bengio S., Erhan D. (2015). Show and tell: A neural image caption generator. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 3156–3164.
- Waxman S. G. (2000). *Correlative Neuroanatomy* (24^e éd.). New York, McGraw-Hill.
- Zhu S.-C., Mumford D. (2006). A stochastic grammar of images. *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, n° 4, p. 259–362.

