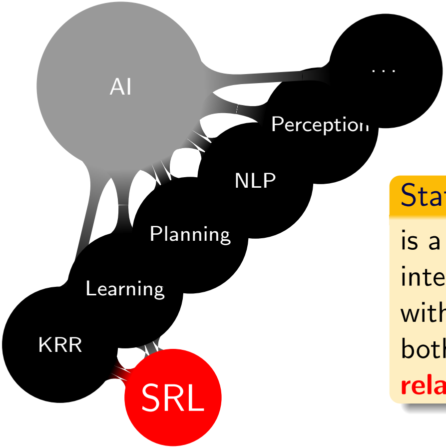
Learning and Reasoning in Logic Tensor Networks: Theory and Application to Semantic Image Interpretation

Luciano Serafini[1], Ivan Donadello[1,2], Artur d'Avila Garces[3]

[1]Fondazione Bruno Kessler, Italy
[2]University of Trento, Italy
[3]City University London, UK

April 5, 2017

The SRL Mindmap



Statistical Relational Learning

is a subdiscipline of artificial intelligence that is concerned with domain models that exhibit both **uncertainty** and **complex relational structure**.

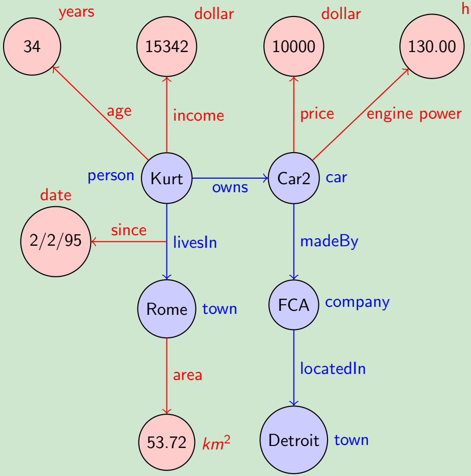
Hybrid domains

We are interested in Statistical Relational Learning over **hybrid domains**, i.e., domains that are characterized by the presence of

- structured data (categorical/semantic);
- continuous data (continuous features);

Hybrid domains

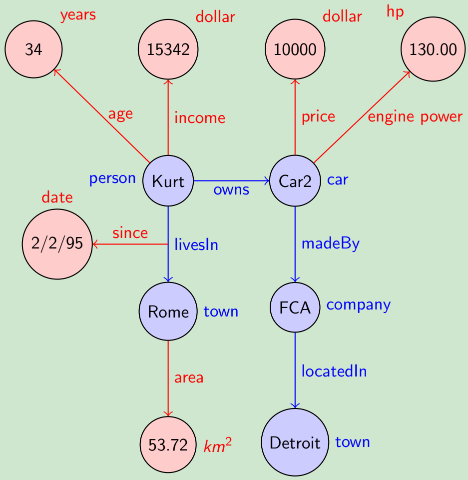
Example (SRL domain)



Tasks in Statistical Relational Learning

- Object Classification:
  Predicting the type of an object based on its relations and attributes;
- Reletion detection:
  Predicting if two objects are connected by a relation, based on types and attributes of the participating objects;
- Regression: predicting the (distribution of) values of the attributies of an object, (a pair of related objects) based on the types and relations of the object(s) involved.

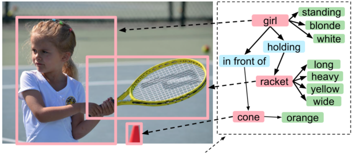
Example (SRL domain)



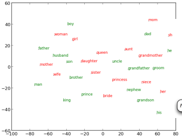
Real-world uncertain, structured and hybrid domains

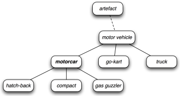
Robotics: a robot's location is a continuous values while the the types of the objects it encounters can be described by discrete set of classes

Semantic Image Interpretation: The visual features of a bounding box of a picture are continuous values, while the types of objects contained in a bounding box and the relations between them are taken from a discrete set



Natural Language Processing: The distributional semantics provide a vectorial (numerical) representation of the meaning of words, while WordNet associates to each word a set of synsets and a set of relations with other words which are finite and discrete



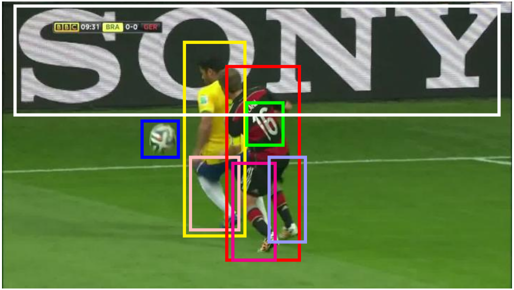


Semantic Image interpretation

semantic Image Interpretation (SII)

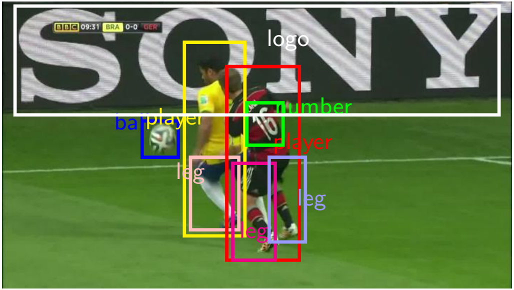
semantic Image Interpretation (SII)

- detect the main objects shown in the picture;



semantic Image Interpretation (SII)

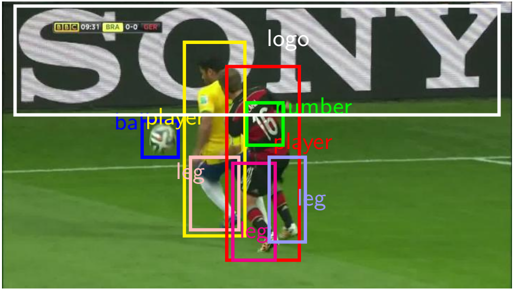
- detect the main objects shown in the picture;
- assign to each object an object type;

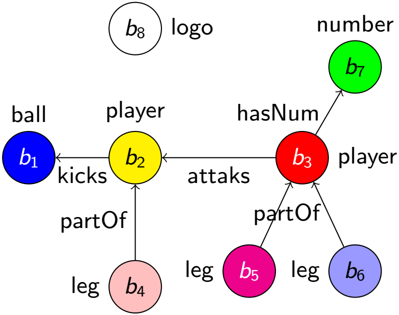


Semantic Image interpretation

semantic Image Interpretation (SII)

- detect the main objects shown in the picture;
- assign to each object an object type;
- determine the relations between the objects as shown in the picture
- represent the outcome of the detection in a semantic structure.





Language - to specify knowledge about models

Two sorted first order language: (abstract sort and numeric sort)

- Abstract constant symbols ($b_1, b_2, \dots, b_8$);
- Abstract relation symbols (player(x), ball(x), partOf(x,y),hasNum(x,y));
- Numeric function symbols (xBL(x),yBL(x),width(x),height(h) area(x),color(x),contRatio(x,y);

COLOR CODE:

- denotes objects and relations of the domain structure;
- denotes attributes and relations between attributes of the numeric part of the domain.

Domain description and queries

Example (Domain descritpion:)

knowledge about object detection:
$xBL(b_1) = 23$, $yBL(b_1) = 73$,
$width(b_1) = 20$, $height(b_1) = 21$
$xBL(b_2) = 45$, $yBL(b_1) = 70$,
$width(b_1) = 40$, $height(b_1) = 104 \ldots$
$contRatio(b_2, b_4) = 1.0$,
$contRatio(b_2, b_5) = 0.4, \ldots$

Domain description and queries

Example (Domain descritpion:)

knowledge about object detection:
$xBL(b_1) = 23$, $yBL(b_1) = 73$,
$width(b_1) = 20$, $height(b_1) = 21$
$xBL(b_2) = 45$, $yBL(b_1) = 70$,
$width(b_1) = 40$, $height(b_1) = 104 \ldots$
$contRatio(b_2, b_4) = 1.0$,
$contRatio(b_2, b_5) = 0.4, \ldots$

---

partial knowledge about object types and relations
$ball(b_1)$, $player(b_2)$, $player(b_3)$,
$leg(b_4)$, $leg(b_5)$, $partOf(b_3, b_2)$,
$kicks(b_2, b_1)$, $hasNum(b_3, b_7), \ldots$

Domain description and queries

Example (Domain descritpion:)

knowledge about object detection:

$xBL(b_1) = 23$, $yBL(b_1) = 73$,
$width(b_1) = 20$, $height(b_1) = 21$
$xBL(b_2) = 45$, $yBL(b_1) = 70$,
$width(b_1) = 40$, $height(b_1) = 104 \ldots$
$contRatio(b_2, b_4) = 1.0$,
$contRatio(b_2, b_5) = 0.4, \ldots$

---

partial knowledge about object types and relations

$ball(b_1)$, $player(b_2)$, $player(b_3)$,
$leg(b_4)$, $leg(b_5)$, $partOf(b_3, b_2)$,
$kicks(b_2, b_1)$, $hasNum(b_3, b_7), \ldots$

---

ontological axioms

$\forall xy.partOf(x, y) \wedge leg(x) \rightarrow player(y)$,
$\forall xy, kick(x, y) \rightarrow player(x) \wedge ball(y)$,
$\forall xy partOf(x, y) \rightarrow contRatio(x, y) > .9$
$\forall x player(x) \rightarrow \neg ball(x)$,

Domain description and queries

Example (Domain descritpion:)

knowledge about object detection:

$xBL(b_1) = 23$, $yBL(b_1) = 73$,
$width(b_1) = 20$, $height(b_1) = 21$
$xBL(b_2) = 45$, $yBL(b_1) = 70$,
$width(b_1) = 40$, $height(b_1) = 104 \ldots$
$contRatio(b_2, b_4) = 1.0$,
$contRatio(b_2, b_5) = 0.4, \ldots$

partial knowledge about object types and relations

$ball(b_1)$, $player(b_2)$, $player(b_3)$,
$leg(b_4)$, $leg(b_5)$, $partOf(b_3, b_2)$,
$kicks(b_2, b_1)$, $hasNum(b_3, b_7), \ldots$

ontological axioms

$\forall xy.partOf(x, y) \wedge leg(x) \rightarrow player(y)$,
$\forall xy, kick(x, y) \rightarrow player(x) \wedge ball(y)$,
$\forall xy partOf(x, y) \rightarrow contRatio(x, y) > .9$
$\forall x player(x) \rightarrow \neg ball(x)$,

Example (Queries)

Query about missing knowledge about object tipes and relations

$$player(b_{10}) \left| \begin{array}{l} xBL(b_{10}) = 83, \\ yBL(b_{10} = 42, \\ width(b_{10} = 30 \\ \ldots \end{array} \right.$$

$$partOf(b_{10}, b_{11}) \left| \begin{array}{l} xBL(b_{10}) = 83, \\ yBL(b_{10} = 42, \\ width(b_{10} = 30 \\ \ldots \\ xBL(b_{11}) = 83, \\ yBL(b_{11} = 42, \\ width(b_{11} = 30 \\ \ldots \\ contRatio(b_{10}, b_{11}) = 0.6 \\ contRatio(b_{11}, b_{10}) = 0.9 \\ \ldots \end{array} \right.$$

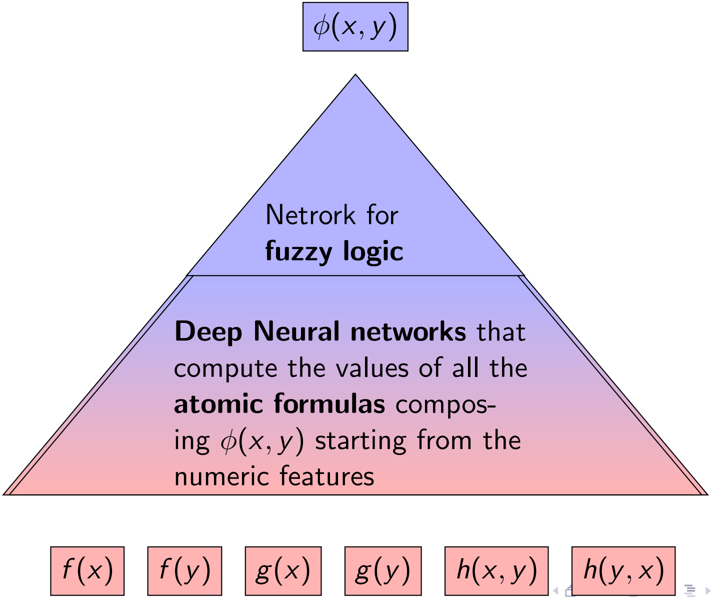
Logic Tensor Network basic idea

$\phi(x, y)$

Logic Tensor Network that computes the truth value of the formula $\phi(x, y)$ on the basis of the numeric features of x, y and the pair $\langle x, y \rangle$

$f(x)$ $f(y)$ $g(x)$ $g(y)$ $h(x, y)$ $h(y, x)$

Logic Tensor Network basic idea



LTN for predicates

n unary numeric function $f_1(x), \ldots, f_n(x)$ and m binary numeric function $g_1(x, y), \ldots, g_m(x, y)$

LTN for unary predicate/type $P(x)$

$$\boxed{LTN_P(\mathbf{v}) = \sigma\left(u_P^\intercal \tanh\left(\mathbf{v}^\intercal W_P^{[1:k]}\mathbf{v} + V_P\mathbf{v} + b_P\right)\right)}$$

$w_P \in \mathbb{R}^{k\times n\times n}$, $V_P \in \mathbb{R}^{k\times n}$, $b_P \in \mathbb{R}^k$, and $u_P \in \mathbb{R}^k$ are **parameters**.

LTN for binary relation $R(x, y)$

$$\boxed{LTN_P(\mathbf{v}) = \sigma\left(u_P^\intercal \tanh\left(\mathbf{v}^\intercal W_P^{[1:k]}\mathbf{v} + V_P\mathbf{v} + b_P\right)\right)}$$

$w_P \in \mathbb{R}^{k\times h\times h}$, $V_P \in \mathbb{R}^{k\times h}$, $b_P \in \mathbb{R}^k$, and $u_P \in \mathbb{R}^k$ are **parameters**, and $h = 2(n + m)$ = the total number of numeric features that can be obtained applying f_i and g_i to x and y.

Fuzzy semantics for propositional connectives

- In fuzzy semantics atoms are assigned with some truth value in real interval [0,1]
- connectives have functional semantics. e.g., a binary connective $\circ$ must be interpreted in a function $f_\circ : [0,1]^2 \rightarrow [0,1]$.
- Truth values are ordeblue, i.e., if $x > y$, then x is a stronger truth than y
- Generalization of classical propositional logic:

  0 corresponds to FALSE and
  1 corresponds to TRUE

Fuzzy semantics for connectives and quantifiers

Lukasiewicz T-norm, T-conorm, residual, and precomplement

T-norm	$a \wedge b$	$=$	$\max(0, a + b - 1)$
T-conorm	$a \vee b$	$=$	$\min(1, a + b)$
residual	$a \rightarrow$	$=$	$\begin{cases} \text{if } a > b & 1 - a + b \\ \text{if } a \leq b & 1 \end{cases}$
precomplement	$\neg a$	$=$	$1 - a$
aggregation	$\forall x.a(x)$	$=$	$\lim_{n\rightarrow\infty} \left(\frac{1}{n}\sum_{i=1}^{n}(a(i)^{-1}\right)^{-1}$

Alternatively, use Gödel or Product T-norm, and geometric or aritmetic mean as aggregator.

Constructive semantics for Existential quantifier

- LTN interprets existential quantifiers constructively via Skolemization.
- Every formula $\forall x_1, \ldots, x_n \exists y \phi(x_1, \ldots, x_n, y)$ is rewritten as $\forall x_1, \ldots, x_m \phi(x_1, \ldots, x_n, f(x_1, \ldots, x_m))$,
- by introducing a new *m*-ary function symbol f,

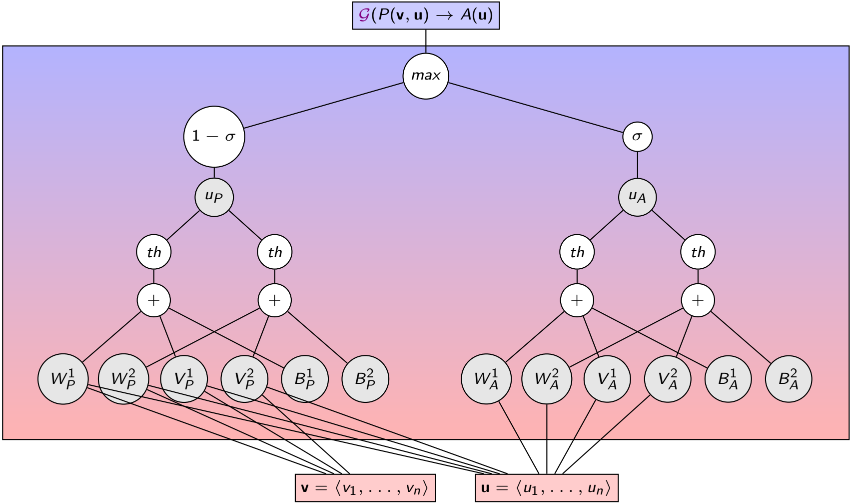
Example

$$\forall x.(cat(x) \rightarrow \exists y.partof(y, x) \wedge tail(y))$$

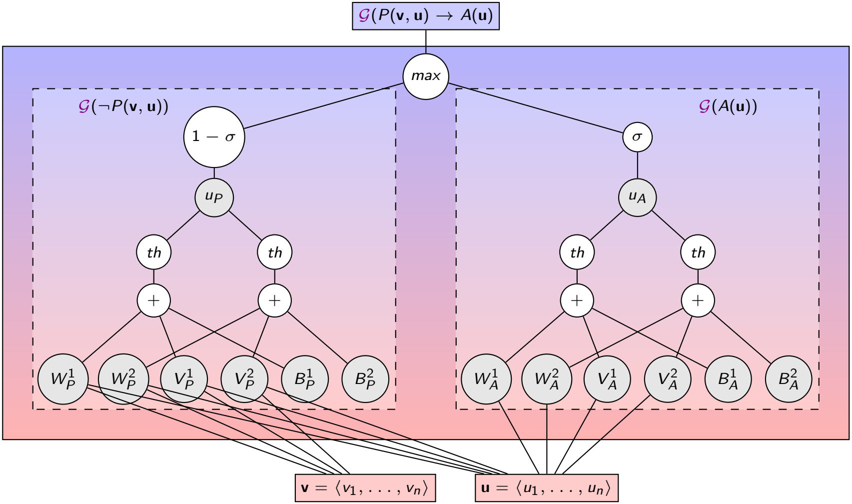
is transformed in

$$\forall x(cat(x) \rightarrow partOf(tailOf(x), x) \wedge tail(tailOf(x)))$$

Grounding = relation between logical symbols and data



Grounding = relation between logical symbols and data



Parameter learning = best satisfiability

Given a FOL theory $\mathtt{K}$ the **best satisfiability problem** as the problem of finding the set of parameters Θ of the LTN, then the problems become $\mathcal{G}^* = LTN(K, \Theta^*)$

$$\Theta^* = \operatorname*{argmax}_{\Theta} \left(\min_{\mathtt{K} \models \phi} LTN(K, \Theta)(\phi) \right)$$

Learning from model description and answering queries

K

$xBL(b_1) = 23,\ yBL(b_1) = 73,$
$width(b_1) = 20,\ height(b_1) = 21$
$xBL(b_2) = 45,\ yBL(b_1) = 70,$
$width(b_1) = 40,\ height(b_1) = 104\ \ldots$
$contRatio(b_2, b_4) = 1.0,\ contRatio(b_2, b_5) = 0.4,\ \ldots$
$ball(b_1),\ player(b_2),\ player(b_3),$
$leg(b_4),\ leg(b_5),\ partOf(b_3, b_2),$
$kicks(b_2, b_1),\ hasNum(b_3, b_7), \ldots$
$\forall xy.partOf(x, y) \wedge leg(x) \rightarrow player(y),$
$\forall xy, kick(x, y) \rightarrow player(x) \wedge ball(y),$
$\forall xypartOf(x, y) \rightarrow contRatio(x, y) > .9$
$\forall xplayer(x) \rightarrow \neg ball(x),$

Learning from model description and answering queries

$$\boxed{\Theta^* = \text{argmax}_\Theta \left(\mathsf{min}_{\mathtt{K}\models\phi}\, \mathit{LTN}(K, \Theta)(\phi)\right)}$$

K

$xBL(b_1) = 23,\ yBL(b_1) = 73,$
$width(b_1) = 20,\ height(b_1) = 21$
$xBL(b_2) = 45,\ yBL(b_1) = 70,$
$width(b_1) = 40,\ height(b_1) = 104\ \ldots$
$contRatio(b_2, b_4) = 1.0,\ contRatio(b_2, b_5) = 0.4,\ \ldots$
$ball(b_1),\ player(b_2),\ player(b_3),$
$leg(b_4),\ leg(b_5),\ partOf(b_3, b_2),$
$kicks(b_2, b_1),\ hasNum(b_3, b_7), \ldots$
$\forall xy.partOf(x, y) \wedge leg(x) \rightarrow player(y),$
$\forall xy, kick(x, y) \rightarrow player(x) \wedge ball(y),$
$\forall xypartOf(x, y) \rightarrow contRatio(x, y) > .9$
$\forall xplayer(x) \rightarrow \neg ball(x),$

Learning from model description and answering queries

$$\boxed{\Theta^* = \text{argmax}_\Theta \left(\min_{\mathtt{K}\models\phi} LTN(K,\Theta)(\phi)\right)}$$

K

$xBL(b_1) = 23,\ yBL(b_1) = 73,$
$width(b_1) = 20,\ height(b_1) = 21$
$xBL(b_2) = 45,\ yBL(b_1) = 70,$
$width(b_1) = 40,\ height(b_1) = 104\ \ldots$
$contRatio(b_2, b_4) = 1.0,\ contRatio(b_2, b_5) = 0.4,\ \ldots$
$ball(b_1),\ player(b_2),\ player(b_3),$
$leg(b_4),\ leg(b_5),\ partOf(b_3, b_2),$
$kicks(b_2, b_1),\ hasNum(b_3, b_7), \ldots$
$\forall xy.partOf(x, y) \wedge leg(x) \rightarrow player(y),$
$\forall xy, kick(x, y) \rightarrow player(x) \wedge ball(y),$
$\forall xypartOf(x, y) \rightarrow contRatio(x, y) > .9$
$\forall xplayer(x) \rightarrow \neg ball(x),$

Q

$$LTN_{K,\Theta^*}\left(player(b_{10}) \;\middle|\; \begin{array}{l} xBL(b_{10}) = 83, \\ yBL(b_{10} = 42, \\ width(b_{10}) = 30 \\ \ldots \end{array} \right)$$

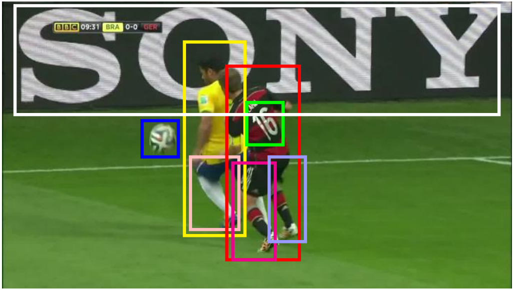
Semantic Image interpretation

semantic Image Interpretation (SII)

Semantic Image interpretation

semantic Image Interpretation (SII)

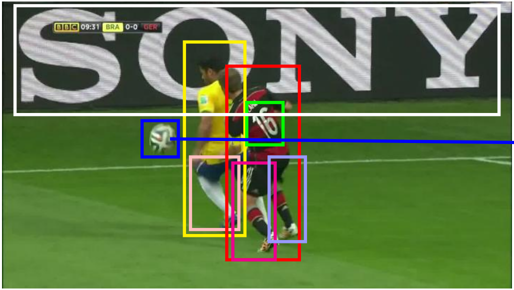
- object detection: Fast RCNN (state of the art object detector)



Semantic Image interpretation

semantic Image Interpretation (SII)

- object detection: Fast RCNN (state of the art object detector)
- Fast-RCNN returns candidate bounding boxes, associated with weights for each object class;

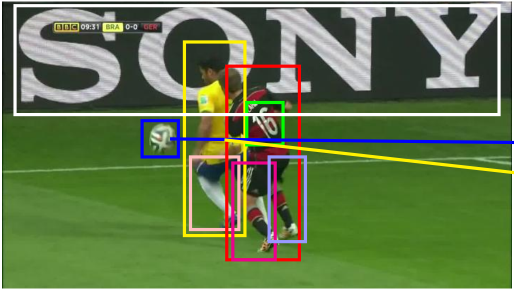


$$
\begin{aligned}
&xBL(b_1) = 12\\
&yBL(b_1) = 27\\
&width(b_1) = 30\\
&height(b_1) = 30\\
&rcnn_{ball}(b_1) = .8\\
&rcnn_{player}(b_1) = .3\\
&rcnn_{logo}(b_1) = .02\\
&\ldots
\end{aligned}
$$

Semantic Image interpretation

semantic Image Interpretation (SII)

- object detection: Fast RCNN (state of the art object detector)
- Fast-RCNN returns candidate bounding boxes, associated with weights for each object class;



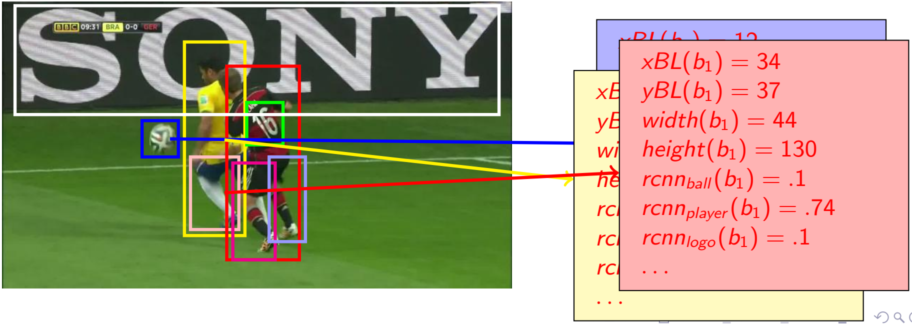
$xBL(b_1) = 12$
$yBL(b_1) = 27$

$xBL(b_1) = 14$
$yBL(b_1) = 17$
$width(b_1) = 40$
$height(b_1) = 100$
$rcnn_{ball}(b_1) = .1$
$rcnn_{player}(b_1) = .7$
$rcnn_{logo}(b_1) = .02$
$\ldots$

Semantic Image interpretation

semantic Image Interpretation (SII)

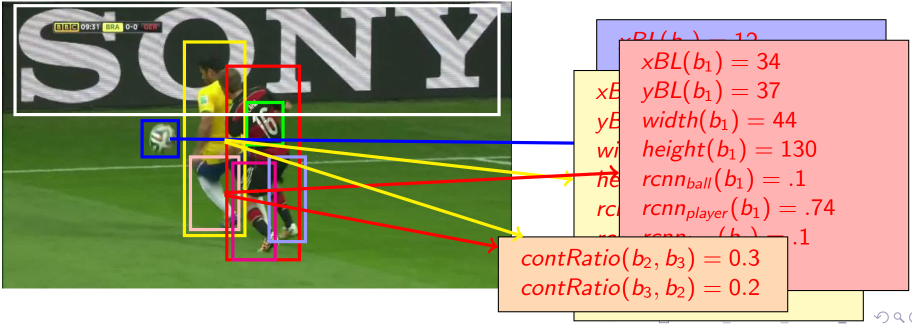
- object detection: Fast RCNN (state of the art object detector)
- Fast-RCNN returns candidate bounding boxes, associated with weights for each object class;



Semantic Image interpretation

semantic Image Interpretation (SII)

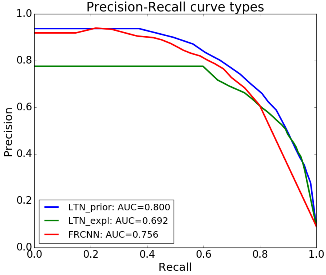
- object detection: Fast RCNN (state of the art object detector)
- Fast-RCNN returns candidate bounding boxes, associated with weights for each object class;
- For each pair of bounding boxe we compute additional binary feature that measure the mutual overlap between the two bounding boxes.

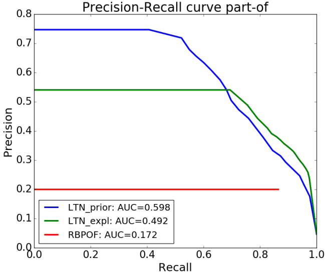


LTN evaluation on PascalPart dataset

- PascalPart contains **10103 pictures** annotated with a set of bounding boxes labelled with object types (60 classes among animals, vehicles, and indor objects)
- We train an LTN with the approx 2/3 pictures and test on 1/3. by including the following **background knowledge**
  - positive/negative examples for object classes (from training set) $weel(bb1),\ car(bb2),\ \neg horse(bb2), \neg person(bb4)$
  - positive/negative examples for relations (we focus on parthood relation). $partOf(bb1, bb2),\ \neg partOf(bb2, bb3), \dots,$
  - general axioms about parthood relation: $\forall x.car(x) \wedge partof(y, y) \rightarrow wheeel(y) \vee mirror(y) \vee door(y) \vee \dots,)$

LTN for SII results





- *LTN_{prior}* is an LTN trained with positive and negative examples + general axioms about partOo relation
- *LTN_{expl}* is an LTN trained only with positive and negative examples of types and partOf
- *FRCNN* is the baseline proposal classification for types given by Fast-RCNN
- *RBPOF* is the baseline for partOf based on the naive criteria

$$\textit{area containment} \geq \textit{threshold}$$

Conclusions

- we introduce **Logic Tensor Networks**, a general framework for SRL that integrates fuzzy logical reasoning and machine learning based on neural networks;
- We apply LTN to the challanging problem of **semantic image interpretation**;
- We experimentally show that the **usage of logic based background knowledge improves the performance** of automatic classification based only on numeric features.

Thanks

Thanks for your attention