KU LEUVEN

Transferring Visual Knowledge into Semantic Roles

HCI-LIIR

Guillem Collell
Ted Zhang
Quynh Ngoc Do
Marie-Francine Moens

MUSTER project

28/04/2016

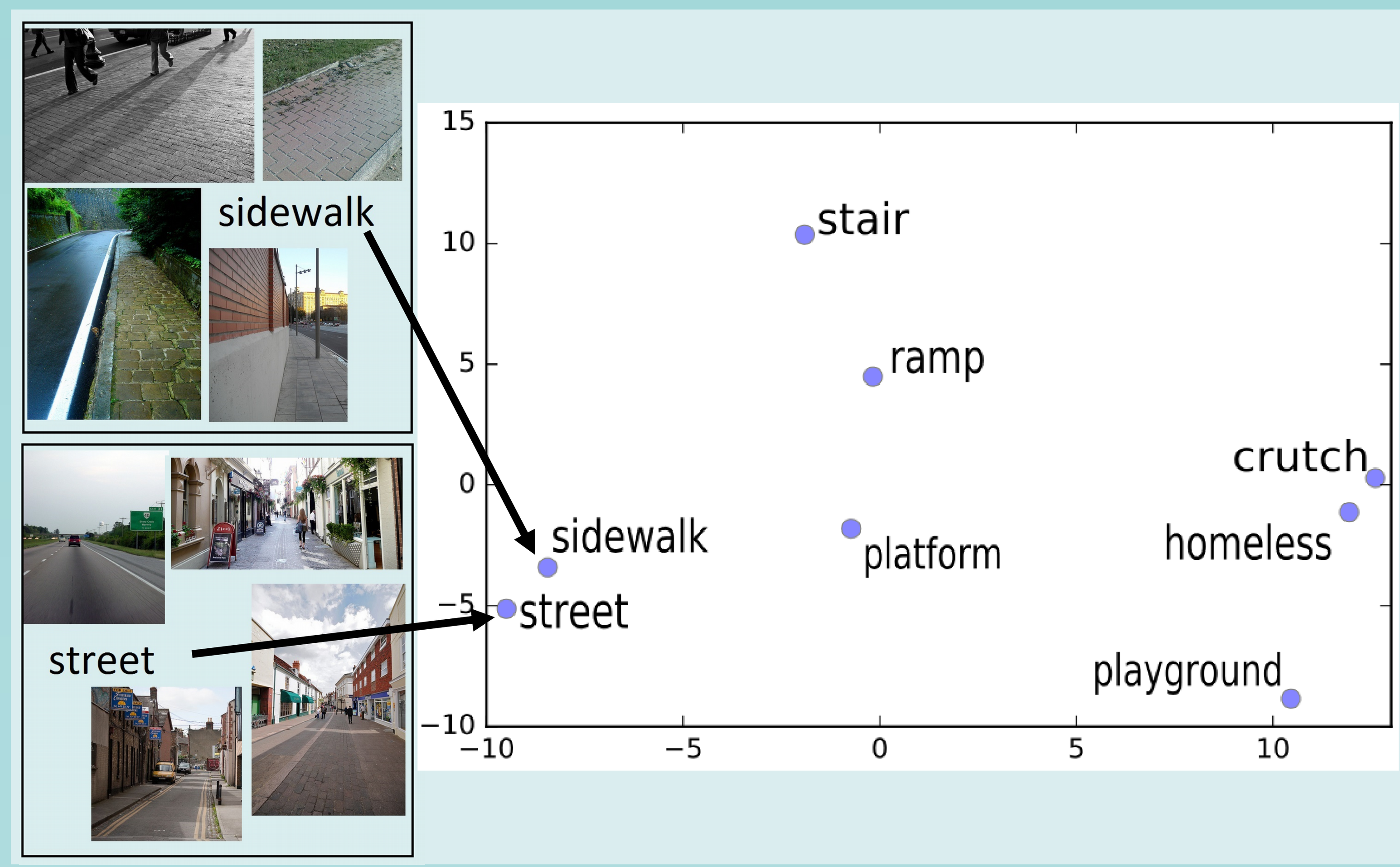
Semantic Role Labeling (SRL)

- **Definition:** SRL is the task of recognizing “who”, “does what”, “to whom”, “where”, “when” and “how” in a given sentence.
- **Main roles:**

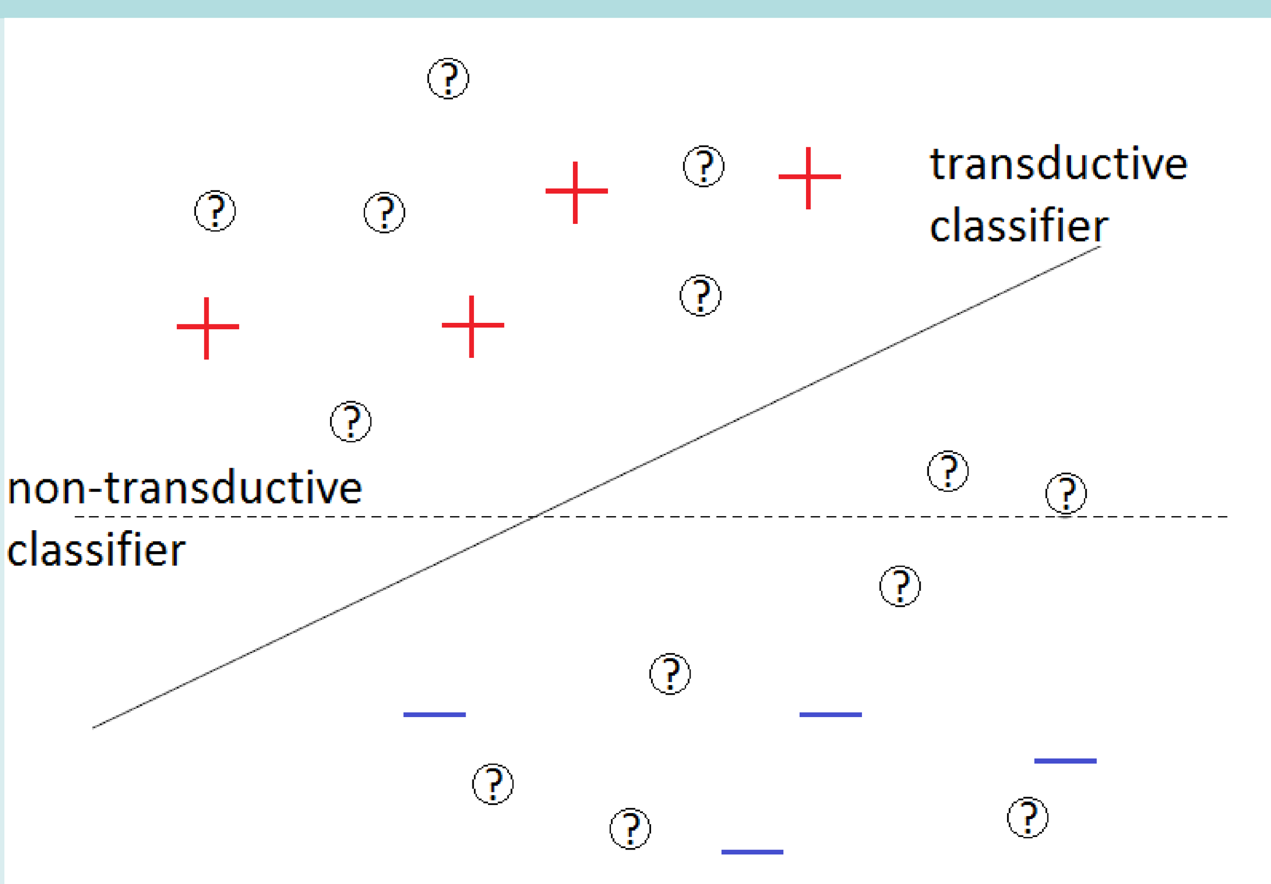
A0: Agent	**A1:** Patient
A2: Indirect object	**LOC:** Location
TMP: Temporal	**DIR:** Direction
MNR: Manner	

- **Example:** [Mary A0] gave [Peter A2] [a book A1] [at school LOC] [yesterday TMP].

Visual Information



Transductive Learning



•**Cluster assumption**: If points are in the same cluster they are likely to be of the same class (Chapelle et al. 2010).

Our Approach

What is new?

•Previous research focuses on using language to aid visual tasks. Here, we do the inverse.
•Transferring visual information has never been show to improve a complex language task such as SRL.

Approach:

•1. Obtain a visual representation for each concept

1. Search images in ImageNet (http://image-net.org/) of nouns from our (train and test) vocabulary.
2. Obtain **feature representation** for each image with a convolutional neural network (CNN).
3. Average feature representations of all the images of a given concept to obtain a single vector representation.
4. Reduce dimensionality from 4096 to 50 with a PCA.

•2. Generate new training examples with a recurrent neural network (RNN) language model. (Do et al., 2015)

- Replace arguments in the training sentences by visually similar concepts (nearest neighbors) from the test set.

An assassin in Colombia killed a federal judge on a ***street***
An assassin in Colombia killed a federal judge on a ***sidewalk***

- This helps the ***cluster assumption*** to hold: test examples of the same *visual cluster* as a particular training example will be “pushed” closer in the semantic *role clusters*.

•3. Train semantic role labeler

- **Pipeline:** predicate disambiguation → argument identification → argument classification.
- **Dataset:** CoNLL 2009 (in domain)

•4. Testing

- **Dataset:** Brown corpus (out of domain)

Results and Conclusions

- **Results***

	Precision	Recall	F1
DIR	-2.43	+1.89	+1.14
LOC	-0.92	+3.43	+1.27
MNR	-1.483242	-0.75757	-1.10948
TMP	0	0	0

Overall F1 (all the roles) does not decrease (+0.06%).

**Results indicate absolute gain/loss (in %) with respect to the baseline semantic role labeller that we employed.*

- **Conclusions**

•Our setting is specially useful for **spatial roles**. Visual information is not meant for every role (e.g., temporal).

•The effect of visual information and that of transductive learning are still moderated:

- 11% of our (training and test) vocabulary had an image representation available.
- 1% of our test examples *not present in the training* data had an image representation available.

References:

•Chapelle, Olivier, Bernhard Schoelkopf, and Alexander Zien (2010).Semi-Supervised Learning. MIT press.

•Do, Q., Bethard, S., & Moens, M. F. (2015). Domain adaptation in semantic role labeling a neural language model and linguistic resources.IEEE/ACMTrans. on Audio, Speech, and Language Processing, 23 (11), 1812-1823.