Force Banner for the recognition of spatial relations

Robin Deléarde[1,2], Camille Kurtz[1], Philippe Dejean[2], Laurent Wendling[1]

1: LIPADE/SIP – Université de Paris, 2: Magellium

{firstname.lastname}@{u-paris,magellium}.fr

with the support of the French Defence Innovation Agency (AID)

Université de Paris

magellium
artal group

AGENCE INNOVATION DÉFENSE

Force Banner for the recognition of spatial relations

Introduction

- We consider the general task of the description of a scene in a 2D image.
- Common approaches: "bag-of-objects/features", CNNs, deformable parts models...
- But common features (shape, texture, points, "deep features"...) are not sufficient to describe complex objects/scenes and to model spatial configurations.

$\Rightarrow$ Interest of considering spatial relations between objects.

Segmented scene from the *CityScapes* dataset.

Aerial scene from the ICG drone dataset.

Scene from the *SpatialSense* dataset, with annotation *"street under bus"*.

Two ways to characterize spatial relations:

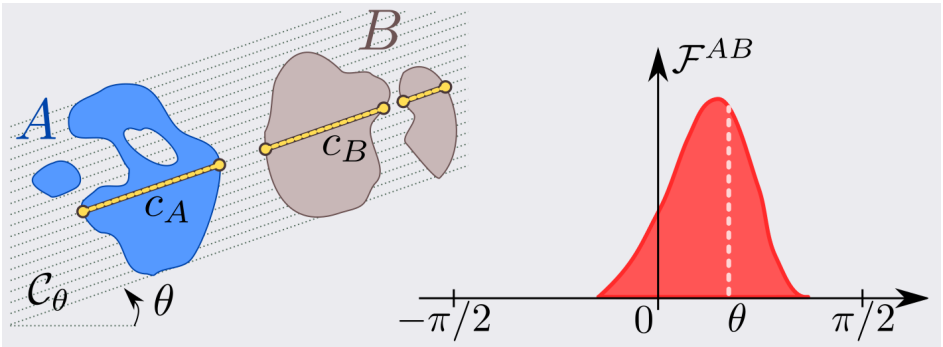
- evaluation of spatial relations in natural language (ex: to the left of, above, in...)
- **relative position descriptors** which aim at giving a complete description of the configuration, and which can be translated into spatial relations in natural language.

Force Banner for the recognition of spatial relations

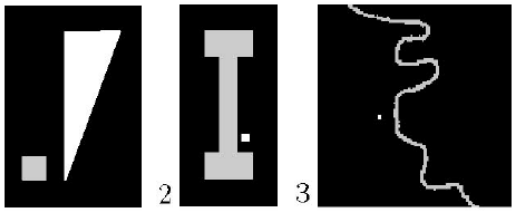
From Force Histogram to Force Banner

- The force histogram is a relative position descriptor, robust to similitudes (pose variation).

  For a given direction θ and a given force r, it integrates $\varphi_r(d) = \frac{1}{d^r}$ the attraction force between two points.



Force histogram computation (from [Matsakis et Wendling, 1999]).



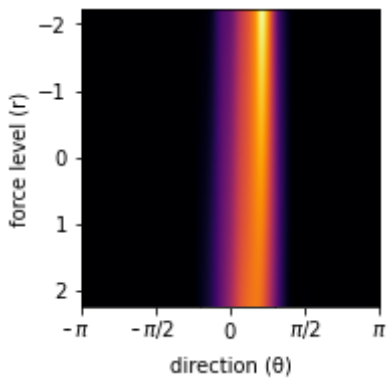
	1		2		3	
	M0	M2	M0	M2	M0	M2
Right	41	**64**	38	**58**	03	02
Left	00	00	05	07	51	**75**
Above	**60**	40	56	48	37	21
Below	05	05	**67**	33	**59**	28

F0 and F2 opinions in ambiguous configurations.

- It was proposed for a single force, providing variable opinions [Matsakis et Wendling, 1999].
- We suggest to extend it to a range of force levels, which makes a continuous 2D descriptor:

$$FB^{AB} : [0, 2\pi[\times [r_s, r_e] \to \mathbb{R}_+$$
$$(\theta, r) \mapsto F_r^{AB}(\theta)$$

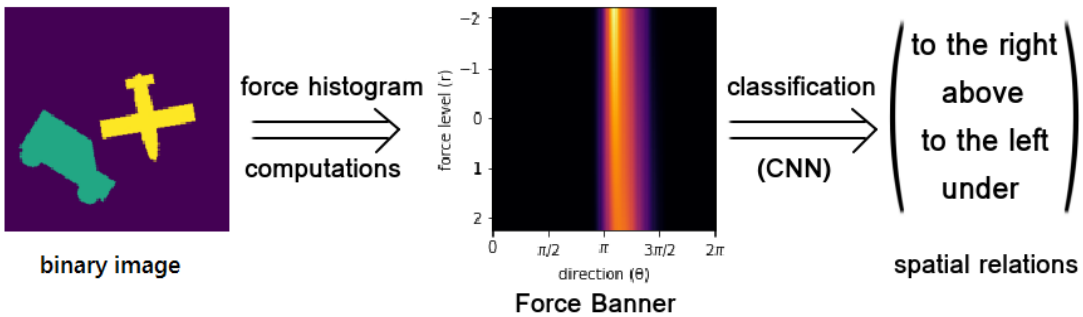
This provides a more complete description, which can be used as input of a CNN.



Force Banner for the recognition of spatial relations

Translation to spatial relations

Such relative position descriptors can be used for the recognition of spatial relations.
We propose to use the Force Banner in a learning scheme with a CNN:



Experimental settings

Task: classification of simple spatial relations (4 classes: right, left, above, under)

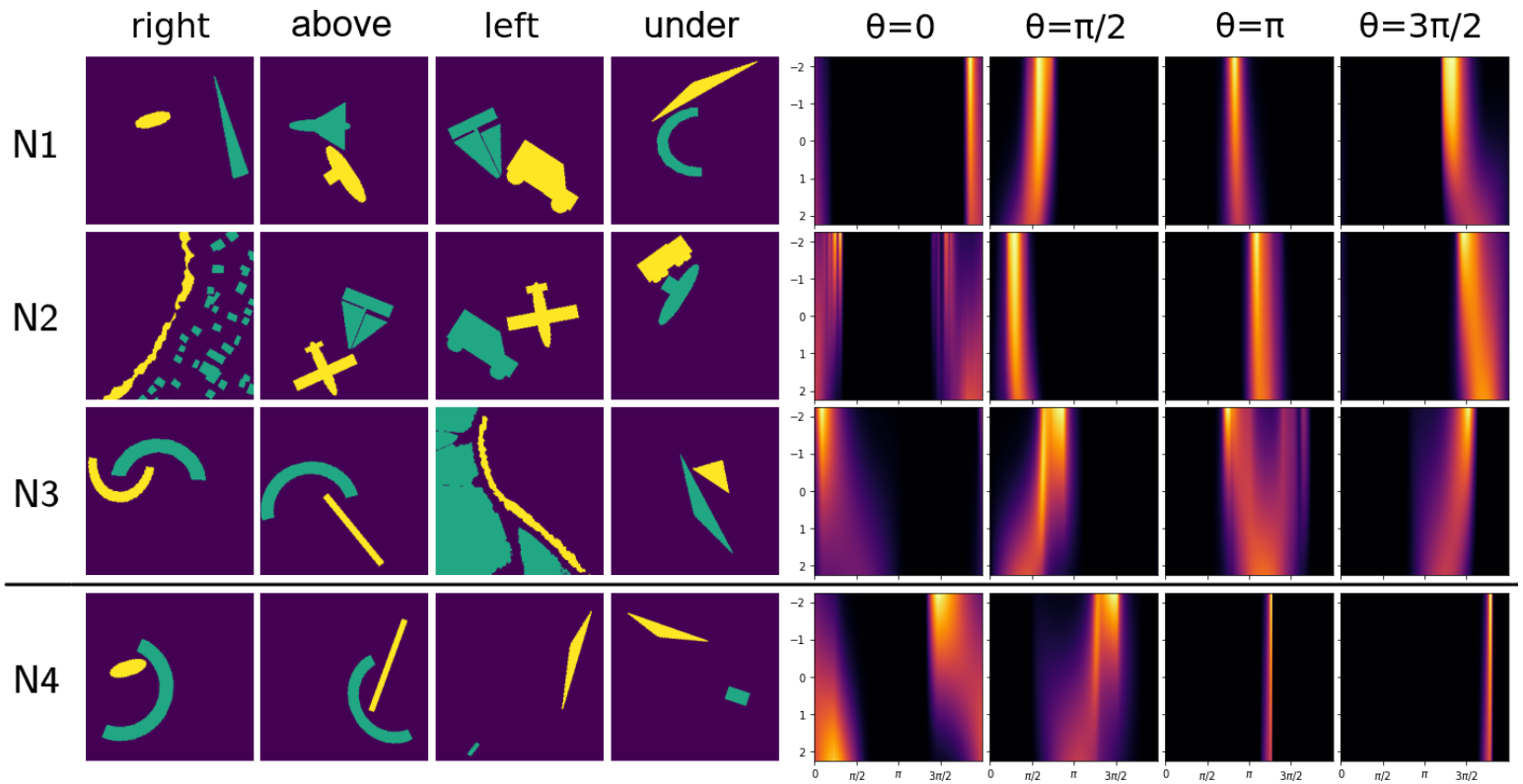
- data: 2 synthetic datasets (2 280 images) + test on patches from a remote sensing image
- model: SqueezeNet pre-trained on ImageNet and fine-tuned / trained from scratch
- baselines: same CNN trained on the raw images / MLP on the bounding box coordinates

Samples of images from our 3 experimental datasets.

Force Banner for the recognition of spatial relations

Preliminary results

- Dataset generation with various simple shapes and random positions, orientations and scales
- Manual annotation: class + level of difficulty (from N1 to N4)
- Force Banner computation: force level from -2.12 to 2.12



Samples of input images and force banners for each class and each difficult level.

Classification results

Train & Test on synthetic images (SimpleShapes), using different subsets of difficulty levels + 1 test on the remote sensing image (GIS)

	dF-banner		*bbox image*		*dFB + image*		*bbox coords*	
Datasets	**OA**	**STD**	**OA**	**STD**	**OA**	**STD**	**OA**	**STD**
Train & Test on N1	**92.66%**	0.94%	88.39%	0.50%	92.13%	0.25%	90.73%	1.71%
Train & Test on N1+N2	92.70%	0.62%	87.53%	0.46%	**92.90%**	0.78%	90.13%	0.34%
Train & Test on N1+N2+N3	91.47%	1.36%	87.30%	0.55%	**92.53%**	1.62%	88.96%	0.89%
Train on N1 & Test on N3	76.03%	3.76%	73.86%	1.15%	**78.13%**	2.65%	72.75%	3.46%
Train on N1+N2 & Test on N3	75.54%	2.70%	72.82%	1.54%	**78.55%**	2.79%	73.17%	0.96%
Train on N1+N2+N3 & Test on GIS	**91.81%**	0.79%	61.75%	3.65%	86.02%	2.43%	86.67%	3.17%

Classification results (overall accuracy – OA and standard deviation – STD) on the test sets (on 3 runs).

Analysis:

- good performance of the Force Banner, over both baselines in all test cases
- particularly good on the GIS image $\Rightarrow$ allows to generalize to other kinds of configurations
- relatively low gain of the image in addition to the force banner $\Rightarrow$ complete descriptor

$\Rightarrow$ good descriptor for this task

Force Banner for the recognition of spatial relations

Conclusion

- complete representation of spatial configurations, robust to similitudes (pose variation)
- can be easily translated into spatial relations in natural language
- allows to recognize spatial relations with good generalization capacity
- can be associated to other features to provide a complete description of a scene

Perspectives

- more experiments to explore different parameters, classifiers, spatial relations, data...
- deeper analysis of the force levels that are used for given objects configurations
- integration of this descriptor into a larger model to describe a scene with several objects
- use this descriptor between an object and its background to make it a shape descriptor

Scene from the *SpatialSense* dataset, with the annotation *"street under bus"*.

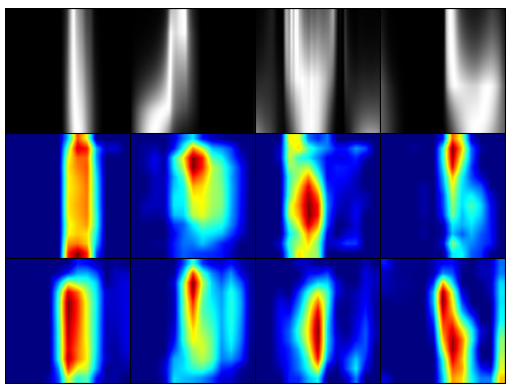


Illustration of GradCam on Force Banners.

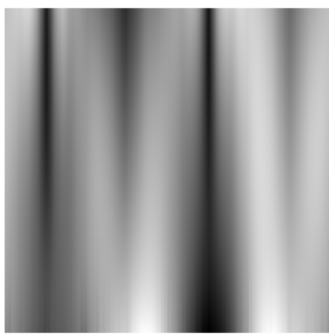


Image from the *Sharvit1* dataset, circular background and force banner.

Force Banner for the recognition of spatial relations

Robin Deléarde[1,2], Camille Kurtz[1], Philippe Dejean[2], Laurent Wendling[1]

1: LIPADE/SIP – Université de Paris, 2: Magellium

{firstname.lastname}@{u-paris,magellium}.fr

with the support of the French Defence Innovation Agency (AID)

Université de Paris

magellium
artal group

AGENCE INNOVATION DÉFENSE