Force Banner for the recognition of spatial relations

Robin Deléarde\textsuperscript{1,2}, Camille Kurtz\textsuperscript{1}, Philippe Dejean\textsuperscript{2}, Laurent Wendling\textsuperscript{1}

\textsuperscript{1}: LIPADE/SIP – Université de Paris, \textsuperscript{2}: Magellium

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

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

⇒ Interest of considering spatial relations between objects.

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

- 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:

  $F_{BA}^{AB} : [0, 2\pi ] \times [r_s, r_e] \rightarrow \mathbb{R}_+$

  $$ (\theta, r) \mapsto F_{r}^{AB}(\theta) $$

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

<table>
<thead>
<tr>
<th>Datasets</th>
<th>\textit{dF-banner}</th>
<th>\textit{bbox image}</th>
<th>\textit{dFB + image}</th>
<th>\textit{bbox coords}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train &amp; Test on N1</td>
<td>92.66% 0.94%</td>
<td>88.39% 0.50%</td>
<td>92.13% 0.25%</td>
<td>90.73% 1.71%</td>
</tr>
<tr>
<td>Train &amp; Test on N1+N2</td>
<td>92.70% 0.62%</td>
<td>87.53% 0.46%</td>
<td>92.90% 0.78%</td>
<td>90.13% 0.34%</td>
</tr>
<tr>
<td>Train &amp; Test on N1+N2+N3</td>
<td>91.47% 1.36%</td>
<td>87.30% 0.55%</td>
<td>92.53% 1.62%</td>
<td>88.96% 0.89%</td>
</tr>
<tr>
<td>Train on N1 &amp; Test on N3</td>
<td>76.03% 3.76%</td>
<td>73.86% 1.15%</td>
<td>78.13% 2.65%</td>
<td>72.75% 3.46%</td>
</tr>
<tr>
<td>Train on N1+N2 &amp; Test on N3</td>
<td>75.54% 2.70%</td>
<td>72.82% 1.54%</td>
<td>78.55% 2.79%</td>
<td>73.17% 0.96%</td>
</tr>
<tr>
<td>Train on N1+N2+N3 &amp; Test on GIS</td>
<td>\textbf{91.81%} 0.79%</td>
<td>61.75% 3.65%</td>
<td>86.02% 2.43%</td>
<td>86.67% 3.17%</td>
</tr>
</tbody>
</table>

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
Force Banner for the recognition of spatial relations

Robin Deléarde\textsuperscript{1,2}, Camille Kurtz\textsuperscript{1}, Philippe Dejean\textsuperscript{2}, Laurent Wendling\textsuperscript{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)