Visual Sense

Tagging visual data with semantic descriptions

Imperial College London
University of Surrey
Institut de Robòtica i Informàtica Industrial
Ecole Central de Lyon
University of Sheffield

Start: Jan 2013

Speaker: Krystian Mikolajczyk

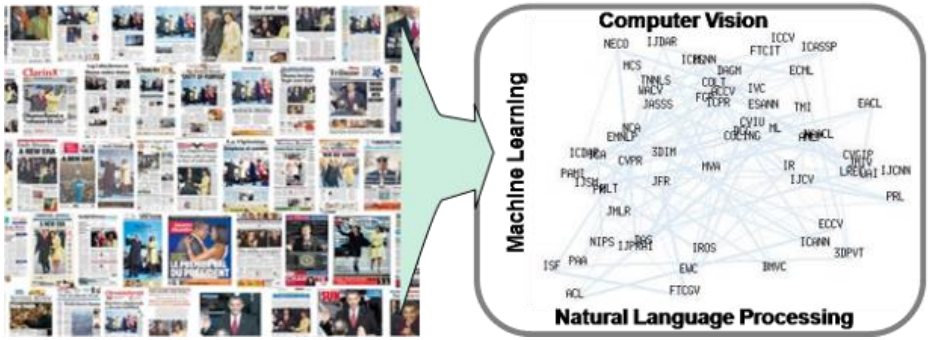
Imperial College London
UNIVERSITY OF SURREY
Institut de Robòtica i Informàtica Industrial
ÉCOLE CENTRALE LYON
The University Of Sheffield.

Tagging visual data with semantic descriptions - objectives

- Extract a semantic representation of visual content

- Generate rich description of images

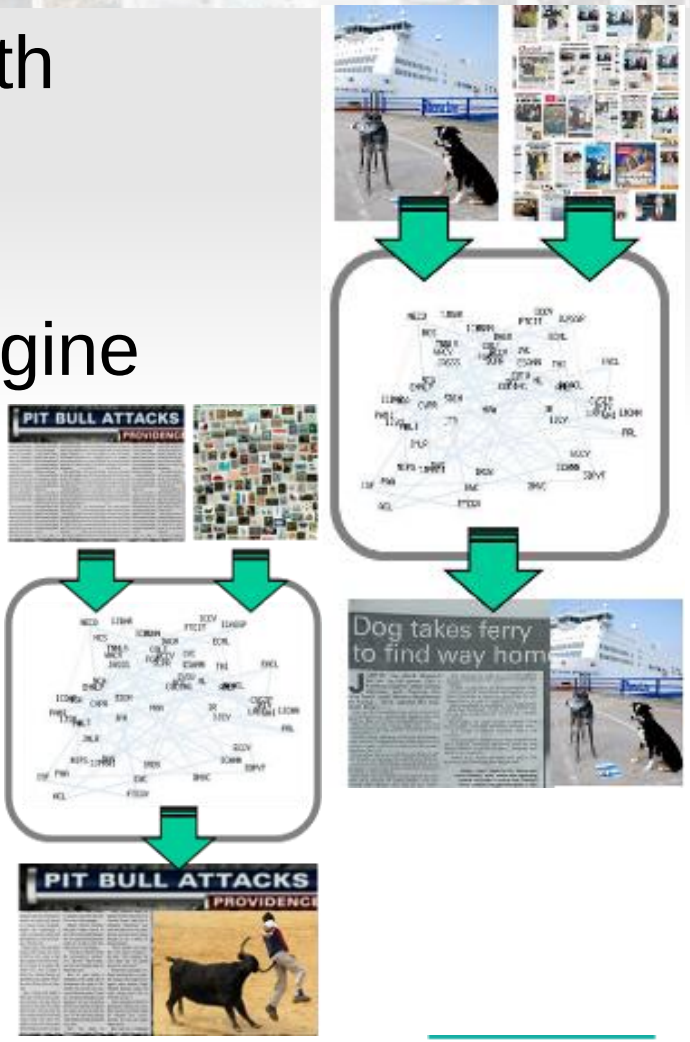
- Exploit available multi-modal data to discover mappings between visual and textual content.



chist-era Imperial College London UNIVERSITY OF SURREY Institut de Robòtica i Informàtica Industrial ÉCOLE CENTRALE LYON The University Of Sheffield. EPSRC

Application scenarios

- Annotating visual documents with a rich semantic description
- More accurate image search engine
- Finding suitable illustrations for text documents



chist-era
Imperial College London
UNIVERSITY OF SURREY
Institut de Robòtica i Informàtica Industrial
ÉCOLE CENTRALE LYON
The University Of Sheffield.
EPSRC

Consortium

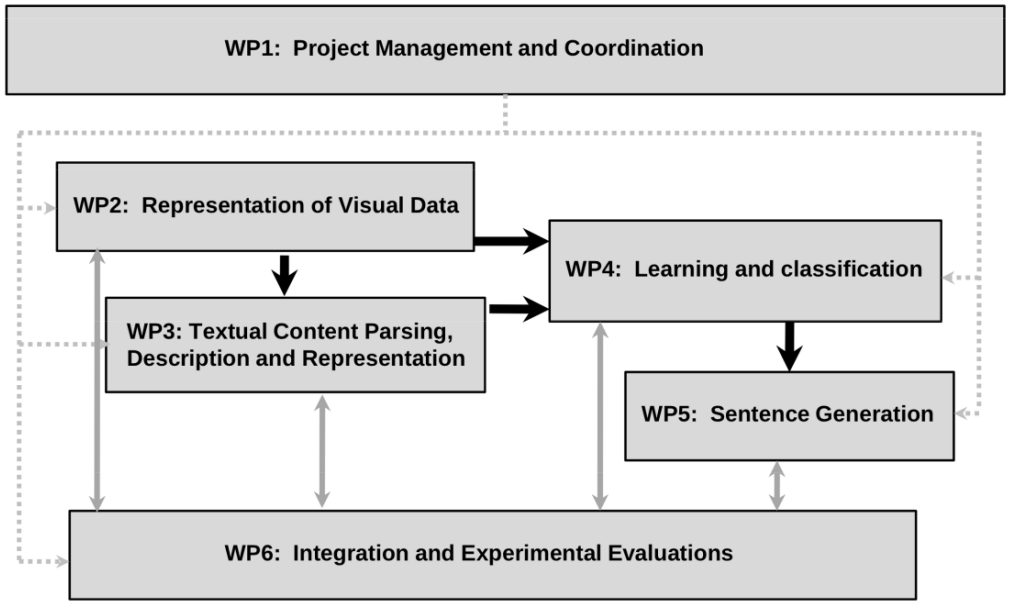
17 researchers actively involved in the project
6 academics, 6 researchers, 5 PhD students

Imperial College London, UK
Dr Krystian Mikolajczyk, project coordinator

Institut de Robotica i Informatica Industrial, Spain
Dr Francesc Moreno-Noguer

Ecole Centrale de Lyon, France
Dr Emmanuel Dellandréa

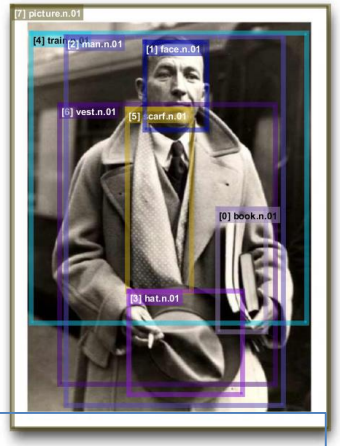
University of Sheffield, UK
Prof Robert Gaizauskas



chist-era Imperial College London UNIVERSITY OF SURREY Institut de Robòtica i Informàtica Industrial ÉCOLE CENTRALE LYON The University Of Sheffield. EPSRC

Change of landscape in V&L research

- 2012
  - Bridging the gap between vision output and input required for language generation
  - Little research activities in caption generation
  - The existing datasets not suitable for evaluations
- 2013
  - Neural network revolution in all related fields
  - End-to-end deep learning rather than multistage approaches
  - Insufficient datasets for V&L DL
- 2014
  - Google, Facebook, Microsoft became very active in V&L
  - End of 2014 Microsoft and Facebook release large datasets
  - Beginnings of joint language & vision modelling



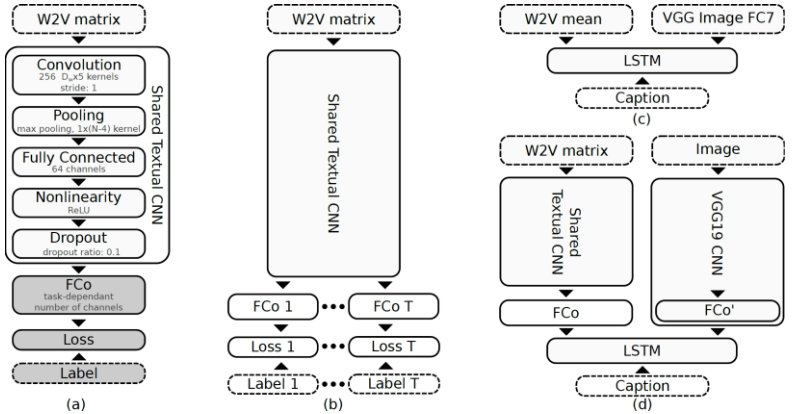
Text: Bigram [F: 0.77]
- [Man]^{2} around the [hat]^{3} along the [book]^{0} .

Visual: Bbox position [F: 0.41]
- [Picture]^{7} on [man]^{2} beside the [scarf]^{4} .

Visual: Bbox size [F: 0.49]
- [Picture]^{7} among [man]^{2} on the [train]^{4} .

Change of landscape in V&L research

- 2015
  - Computational resources for DNN processing of large scale data
  - Insufficient annotation for certain tasks
  - Performance measures for sentence generation
  - Building new benchmark datasets
  - Identifying visual language
  - LSTM for sentence generation
- ViSen challenges in 2016 and beyond
  - Content selection from scene understanding
  - Description of continuous visual data
  - Generalization problem
  - Insufficient hardware resources
  - Unsupervised learning



chist-era Imperial College London UNIVERSITY OF SURREY Institut de Robòtica i Informàtica Industrial ÉCOLE CENTRALE LYON The University Of Sheffield. EPSRC

ViSen activities

- Publications
  - 2 book chapters, 9 PhD thesis, 21 journals, 62 conference papers
  - 7 collaborative publications
- Sustainability/Valorisation
  - Annotation interfaces (crowdsourcing)
  - Datasets
    - Flickr8k, ImageCLEF2015 & 2016, TennisData, BreakingNews, Prepositions
  - Software release
    - segmentation code, image descriptors,
- Follow up projects exploiting Visen's outcomes
  - UK EPSRC Making Sense of Sound (€1.5m)
  - UK EPSRC Face and soft biometrics (€7m)
  - MINECO Robotics and natural communication (€160k)
  - FUI Robotics scene understanding (€1.2m)
  - More under review and preparation …

Dissemination

- Workshops
  - Language and Vision Workshop at CVPR 2015 Boston, co-organized with MIT.
  - The 5th Workshop on Vision and Language at ACL 2016 Berlin, co-organized with iV&L Net (European Network on Integrating Vision and Language)
- ImageCLEF Challenge 2015 and 2016
  - Introduced new tasks and new benchmarking dataset
  - Defined evaluation metrics to support the new task
  - 14 participating teams (China, France, Japan, Mexico, Romania,…)
- Training Schools
  - Deep Learning for Vision and Language, Spring School in Malta
- Invited Talks:
  - VL'15 at EMNLP2015,
  - CLEF2015

Summary

- Intensive collaboration between all partners
  - Exchange of expertise
  - Data processing/generation services
  - Large dissemination activities
    - ImageCLEF2015, 2016, 2 V&L workshops, V&L Spring School
- Main achievements
  - State-of-the-art performance in text2image and image2text
  - Resources shared with the community: benchmarks, code, interfaces,
  - New tasks in V&L
  - Publications
- Come to see our demos, real time text2image and caption generation