230376 - DLV - Deep Learning for Vision

Coordinating unit: 230 - ETSETB - Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications
Academic year: 2018
Degree: MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 3
Teaching languages: English

Teaching staff
Coordinator: Xavier Giró i Nieto
Others: Xavier Giró i Nieto, Elisa Sayrol, Amaia Salvador, Kevin McGuinness and Eva Mohedano

Degree competences to which the subject contributes

Specific:
CE1. Ability to apply information theory methods, adaptive modulation and channel coding, as well as advanced techniques of digital signal processing to communication and audiovisual systems.

Transversal:
CT3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
CT4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Teaching methodology
- Lectures
- Application classes
- Group work
- Group work (distance)
- Short answer test (Test)

Learning objectives of the subject

The aim of this course is to train students in methods of deep learning for computer vision. Convolutional neural networks (convnets) will be presented and analyzed in detail to understand the potential of these state of the art tools in visual pattern recognition. Engineering tips and scalability issues will be addressed to solve tasks such as image classification, object detection or automatic textual captioning. Hands-on sessions will provide development skills so that attendees can solve a selected task in an open scientific benchmark and, if successful, submit their results.
230376 - DLV - Deep Learning for Vision

**Study load**

<table>
<thead>
<tr>
<th>Total learning time: 75h</th>
<th>Hours large group: 12h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours small group: 12h</td>
<td>16.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 51h</td>
<td>68.00%</td>
</tr>
</tbody>
</table>

**Content**

**1. Convolutional Neural Networks**

**Learning time:** 18h 30m
- Theory classes: 4h
- Laboratory classes: 3h
- Guided activities: 4h
- Self study: 7h 30m

**Description:**
- Architecture: Forward and recurrent networks.
- Backpropagation
- Layer Visualization.
- Memory and computational requirements.
- Best practices.
- Fine-tunning

**2. Applications**

**Learning time:** 44h
- Theory classes: 4h
- Practical classes: 7h
- Laboratory classes: 1h
- Guided activities: 15h
- Self study: 17h

**Description:**
- Image retrieval and classification
- Face and object detection/recognition.
- Semantic segmentation
- Saliency prediction
- Image captioning
- Multimodal fusion
# Planning of activities

| **Laboratory practical exercises** | **Hours**: 4h  
Laboratory classes: 4h |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td></td>
</tr>
</tbody>
</table>
- Training of a convnet for character recognition. (1 hour)  
- Visualization and ablation of convnet layers. (1 hour)  
- Fine-tuning a convnet for transfer learning. (1 hour)  
- Local image analysis. (1 hour) |
| **Extended answer test (Final examination)** | **Hours**: 1h  
Theory classes: 1h |
| **Final project presentations** | **Hours**: 3h  
Theory classes: 3h |
| **Description:**                  |  
Oral presentation of a solved Project (30 minutes) |

## Qualification system

Final examination: 30%  
Final project: 30%  
Laboratory assessments: 30%  
Communication skills: 10%

## Bibliography

**Basic:**  

**Others resources:**

- **Hyperlink**  
  Fei-Fei Li, Andrej Karpathy, "CS231n: Convolutional Neural Networks for Visual Recognition". Stanford University 2015  
  [http://cs231n.stanford.edu/](http://cs231n.stanford.edu/)

- **Audiovisual material**  
  Slides of the course and the bibliography referred within