# VISION FOR MULTIPLE OR MOVING CAMERAS

## 1. SYLLABUS INFORMATION

### 1.1. Course title

Deep Learning for Computer Vision 2

### 1.2. University

Universidad Autónoma de Madrid

### 1.3. Semester

2nd semester

## 2. COURSE DETAILS

### 2.1. Course nature

Compulsory

### 2.2. ECTS Credit allotment

6

### 2.3. Recommendations

This course requires previous knowledge in several aspects: fundamentals of deep learning, image processing at an introductory level, and programming and image handling with Python and Pytorch.
This course is given in parallel with the course, Deep Learning for Computer Vision1.Both courses are part of the same study area, and they present a high amountof interdependencies between them. It is highly recommended that the student attend both courses simultaneously.

### 2.4. Faculty data

Escuela Politécnica Superior

## 3. COMPETENCES AND LEARNING OUTCOMES

### 3.1. Course objectives

The aim of this course, together with its parallel subject Deep Learning for Computer Vision1, is to help the student understand and apply the theoretical and practical fundamentals behind deep learning techniques applied to image analysis and synthesis.

Deep learning technology represents the current state-of-the-art for image processing and computer vision applications, and is one of the hottest research and industrial topic nowadays, which is in constant development. This course intends to provide the student with the necessary tools to be ready to apply current computer vision technology in the research and industrial fields, and interiorize the basic concepts that will allow for the monitorization of the future development of the technology.
Specifically, this course covers: i) the fundamentals and development of core image deep learning architectures, which are prevalent in almost any computer vision pipeline, ii)and basic mechanisms for generative architectures, iii) the understanding and application of visual models learning from natural

language supervision and the associated image, text and coordinated representation learning techniques, and iv)an introduction to the social and environmental challenges associated deep learning computer vision; including a discussion on the implications of regulations such as the EU AI Act, addressing sustainability, interpretability, and reliability challenges of these models.

### 3.2. Course contents

**UNIT I: Deep learning architectures for image analysis and generation**

Introduction to image classification

- ImageNet and the blossom of deep learning

Basic image core architectures: Convolution Neural Network (CNNs)

- Plain CNN architectures: AlexNet and VGG
- Inception Networks
- Residual Networks

Advanced core image architectures:

- Recurrent Neural Networks and attention mechanisms
- Vision Transformers

Basic generative architectures:

- Variational AutoEncoders
- Generative Adversarial Networks

**UNIT II: Visual models learning from natural language supervision**

Visual models learning from natural language supervision

- Architecture and components
- Image, text and joint representation learning techniques
- Prompting strategies for zero-shot image classification

**UNIT III: Social and environmental challenges of image deep learning models**

Current socioenvironmental challenges in deep learning computer vision:

- Overview of the EU Artificial Intelligence Act and discussion of its potential implications
- Introduction to the environmental and sustainability risks of deep learning computer vision models
- Introduction to techniques for the explainability and interpretability of deep learning computer vision models
- Introduction to reliability and robustness challenges of deep learning computer vision models

### 3.3. Course bibliography

1. Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, MIT Press,2016,http://www.deeplearningbook.org

2. Dive into deep learning:Interactive deep learning book with code, math, and discussions,https://d2l.ai/index.html

3. A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012

4. Y. Lecun, et al.,"Gradient-based learning applied to document recognition,"inProc. of the IEEE,Nov. 1998

5. Simonyam et al.Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR 2015

6. Szegedy et al.Going deeper with convolutions. CVPR 2015

7. He et al.Deep Residual Learning for Image Recognition. CVPR 2016

8. Canziani et al. An Analysis of Deep Neural Network Models for Practical Applications. arXiv 2016 IEEE Access, Benchmark Analysis of Representative Deep Neural Network Architectures

9. Dosovitskiy, Alexey, etal. "An image is worth 16x16 words: Transformers for image recognition at scale."arXiv preprint arXiv:2010.11929(2020)

10. Guo, X., Liu, X., Zhu, E., Yin, J. (2017).Deep Clustering with Convolutional Autoencoders.In: Liu,D.,

11. Xie, S., Li, Y., Zhao, D., El-Alfy, ES.(eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science(), vol 10635. Springer

12. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.,and

13. Bengio, Y. Generative adversarial networks. In NIPS'2014.

Tutorial on Multimodal Machine Learning @ ICML 2023.MML Tutorial (cmu-multicomp-lab.github.io).2023

14. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish

15. Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pp. 8748–8763. PMLR,18–24 Jul 2021.

16. Zhou, Kaiyang, et al. "Conditional prompt learning for vision-language models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

17. Dhar, Payal. "The carbon impact of artificial intelligence. "Nat. Mach. Intell.2.8pp.423-425.2020 'Artificial Intelligence Act, Corrigendum, 19 April 2024'. Interinstitutional File: 2021/0106(COD)(https://artificialintelligenceact.eu/ai-act-explorer/).2024.

18. CVPR'21 Online Tutorial on Interpretable Machine Learning in Computer Vision (interpretablevision.github.io).2021

19. M. Escudero-Viñolo, J. Bescos, A. López-Cifuentes, and A. Gajic.Explainable Deep Learning AI, chapter 5: Characterizing a scene recognition model by identifying the effect of input features via semantic-wise attribution. Academic Press. Ed. J. Benois-Pineau, R. Bourqui, D. Petkovic, G.Quenot. ISBN 9780323960984. Feb. 2023.

20. K. Sirotkin, P. Carballeira, M. Escudero-Viñolo: "A Study on the Distribution of Social Biases in Self-Supervised Learning Visual Models", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022

21. Gandelsman, Yossi, Alexei A. Efros, and Jacob Steinhardt. "Interpreting CLIP's Image Representation via Text-Based Decomposition. "The Twelfth International Conference on Learning Representations.ICLR2024.

22. Walmer, Matthew, et al. "Teaching matters: Investigating the role of supervision in vision transformers. "Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

23. GitHub-huytransformer/Awesome-Out-Of-Distribution-Detection.

24. Yang, Jingkang, et al. "Generalized out-of-distribution detection: A survey. "International Journal of Computer Vision:pp.1-28.2024.

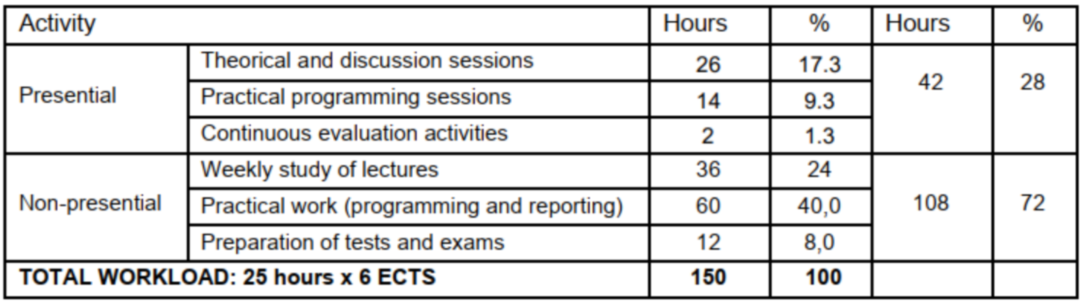
## 4. TEACHING-AND-LEARNING METHODOLOGIES AND STUDENT WORKLOAD

### 4.1. Contact hours

| Modality | Hours | % |
|---|---|---|
| Presential | 42 | 28 |
| Non-presential | 108 | 72 |

### 4.2. List of training activities

The course involves lectures, lab assignments (including implementation and documentation), and evaluation activities, according to the following distribution.



| Activity | | Hours | % | Hours | % |
|---|---|---|---|---|---|
| Presential | Theorical and discussion sessions | 26 | 17.3 | 42 | 28 |
| | Practical programming sessions | 14 | 9.3 | | |
| | Continuous evaluation activities | 2 | 1.3 | | |
| Non-presential | Weekly study of lectures | 36 | 24 | 108 | 72 |
| | Practical work (programming and reporting) | 60 | 40,0 | | |
| | Preparation of tests and exams | 12 | 8,0 | | |
| **TOTAL WORKLOAD: 25 hours x 6 ECTS** | | **150** | **100** | | |

## 5. EVALUATION PROCEDURES AND WEIGHT OF COMPONENTS IN THE FINAL GRADE

### 5.1. Regular assessment

The grading range is from 0.0 to 10.0. The maximum grade is 10.0 and each of the parts (Theory, TH and Practice, PR) will be also graded with the same grading range.

For the regular assessment, theory and practice will consider according to the following rule:

Final Mark =50%TH +50%PR

In order to pass the course, it is necessary to have a pass grade (equal or greater than 5.0) in the overall evaluation (Final Mark), as well as a pass grade (equal or greater than 5.0) in the two individual parts (TH and PR); otherwise, the applied rule will be:

Final Mark =min (TH, PR)

The individual grades for the TH and PR parts, if passed, are kept for the re-sitting exam.

**TH** is the grade obtained from the evaluation of the Theory lectures. The acquired knowledge will be evaluated through two exams; hence, the TH mark is obtained as follows:

TH = 50%TH1 + 50%TH2

where TH1 and TH2 are the marks of the two evaluation exams the first one corresponding to Unit I and the other one corresponding to Unit II. This rule will only be applied if THi >= 4.0; otherwise, the applied rule will be:

$TH = \min (TH1, TH2)$

In case TH<5.0, if THi>=5, the corresponding i evaluation is considered passed and there is no need to repeat the exam in future evaluation exams.

PRis the grade obtained from the lab assignments. There is a total of four lab assignments, each one is to be documented and scored. The PR mark is obtained as follows:

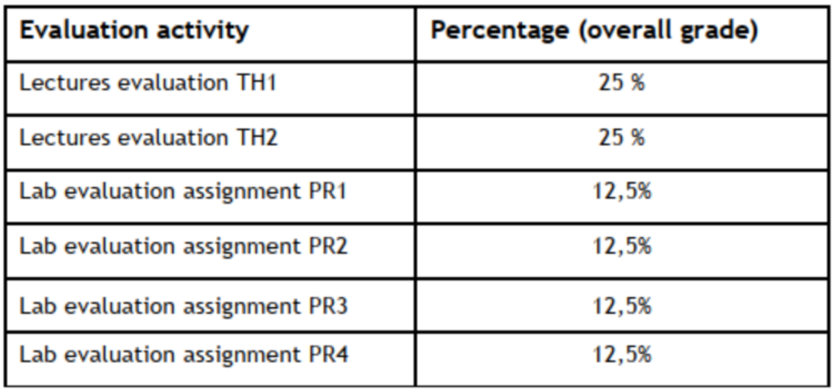
$PR = 25\%PR1 +25\%PR2 +25\%PR3 +25\%PR4,$

where PRi are the marks obtained for each of the lab assignments. This rule will only be applied if PRi>= 4.0; otherwise, the applied rule will be:
$PR = \min (PR1, PR2, PR3, PR4)$

In case PR <5.0, if PRi>=5, the corresponding lab is considered passed and there is no need to repeat the assignment in future evaluations.

Note: each lab assignment must be submitted within a corresponding deadline. The grade for late submissions is capped to a maximum of 7 points, i.e. PRi =min(7,PRi)

### 5.2. List of evaluation activities



| Evaluation activity | Percentage (overall grade) |
|---|---|
| Lectures evaluation TH1 | 25 % |
| Lectures evaluation TH2 | 25 % |
| Lab evaluation assignment PR1 | 12,5% |
| Lab evaluation assignment PR2 | 12,5% |
| Lab evaluation assignment PR3 | 12,5% |
| Lab evaluation assignment PR4 | 12,5% |