Computer Vision II (CSE 40536 / 60536)

Syllabus, Spring 2020 (ver. 1/15/2020)

Adam Czajka
University of Notre Dame

1 Goals and scope

The aim of Computer Vision is to give computers the ability to *understand* what they *see* in images and videos taken by one or more sensors (usually visible-light cameras). This course will be focused on advanced topics in computer vision, and specifically oriented around robotic vision. Completing *Computer Vision I* course, or prior skills and knowledge of basic computer vision topics and tools (evaluated on a case-by-case basis by the instructor prior to enrolling to the class), is required to attend this class.

This course is prepared jointly with our industrial partner (Amazon Robotics) and will encompass three main modules:

(a) interactive class meetings discussing advanced and current topics in computer vision, along with coding assignments, student presentations (one selected paper in the semester), practicals and quizzes;

(b) semester project focused on object detection and recognition in dense clutter;

(c) presentations of semester project solutions (mid-term and final).

Class meetings will be organized around selected, current problems presented at the leading computer vision meetings, such as CVPR, ECCV, ICCV and WACV. Rather than regular lectures, these will be highly interactive sessions, focused on understanding both the challenges presented in the papers, as well as strengths and weaknesses of the proposed solutions. The semester project will solve an example, real-world problem suggested by Amazon Robotics and related to robotic vision. Students will be given an opportunity to present their semester project solutions to the Amazon Robotics research team.

After completing this course students will understand current top-level literature on computer vision, know seminal solutions to selected advanced challenges in vision-related tasks, and understand how to apply complex methods and tools in real-world problem of object identification.

2 Prerequisites

- **Required:** completed *Computer Vision I* course, or demonstrating skills and knowledge equivalent to completing the *Computer Vision I* course (please contact Dr. Czajka at `aczajka@nd.edu` to discuss prerequisites, if you did not attend *Computer Vision I*);

- **Good to have:** advanced skills in Python programming, experience with OpenCV and deep learning libraries (such as Keras, Tensorflow, PyTorch); these will significantly help in smooth implementation of ideas for the semester project.

3 Course structure

- 2 introductory lectures on current advanced topics in computer vision, and discussing prerequisites required to start the semester project;
- 20 interactive class sessions, 75 minutes each (see Sec. 4 for details), including two sessions for setting up the acquisition and collecting image samples;
- 1 semester project, solving an industry-driven problem related to robotic vision (see Sec. 5 for details);
- 2 lectures offered by invited speakers;
- 3 class meetings devoted to semester project final presentations.

4 Interactive class sessions

Class meetings will not discuss basics in Computer Vision, as this is the aim of Computer Vision I (introductory) course. Instead, the class meetings will be organized around selected, current problems presented at leading computer vision conferences, including CVPR, ECCV, ICCV and WACV.

Students (and/or small teams, if desired) will be asked to read and prepare a critique review of one paper in the semester, and make a short presentation in class. The papers discussed will be selected in a way to help students in providing the best possible solutions to the semester project. The team's presentation will be complemented by the instructor's deeper dive into the topic and – when appropriate – practical coding exercises (started in class and to be completed at home). A discussion among all participants will conclude each class meeting. The discussed papers may be proposed both by students and by the instructor. All not-presenting students will be also asked to read the selected paper ahead of time, and actively participate in the discussion. A small quiz will be organized after each class to ensure all students were actively following the discussions.

Class topics to be discussed in class include:

- **Neural Networks for Vision Problems:** CNNs, Generative Adversarial Networks (GANs), Network Architecture Search, capsule networks, Recurrent Neural Networks (LSTM, GRU, NTM);
- **Vision-specific Neural Network Training:** data split, cross-validation, bias, overfitting and regularization, stopping criteria, optimization of layered structures, one-shot and zero-shot learning, triplet loss and its applications, hyperparameter optimization (grid search, TPE, adaptive TPE);
- **Segmentation and Retrieval:** deep learning-based image segmentation, image captioning and action recognition, large-scale instance retrieval;
- **Image Representation:** autoencoders, latent variables, compression;
- **Security:** neural network threats and defense mechanisms, physical and synthesized adversarial inputs, one-pixel attacks;
- **Tracking and Localization:** optical flow and video tracking, image stabilization, simultaneous localization and mapping (SLAM);
- **Visualization:** t-SNE, inceptionism, activation atlas;

- **Solving Vision Problems on Low-resource Platforms:** practical exercises with running deep learning-based object recognition on Raspberry Pi.

The semester will begin with two lectures discussing current, complex challenges in computer vision and the scope of the semester project. Additionally, two guest lectures are planned to be given by invited computer vision experts, either from academia or industry, including a researcher from Amazon Robotics.

5 Semester project

5.1 Background

Robotics is transforming the industrial and commercial landscape across a variety of domains allowing for faster and more efficient operations. While promising, many robotic systems are only functional in highly structured environments devoid of human interaction and robust exception handling. The goal of the semester project is to propose a solution to a robotic vision problem of object identification in a densely cluttered environment.

Amazon Robotics has historically sponsored a Picking Challenge to better understand (among other reasons) the practical limitations of machines, particularly robotic manipulators, in semi-unstructured operational environments that identify, grasp, and manipulate a variety of material goods. While the Picking Challenge has stopped (for now), Amazon continues to have interest in these research areas.

Broadly, Amazon Robotics seeks to visually identify (using only single monocular optical camera) a wide variety of known and unknown objects when densely packed inside plastic totes. These objects can present themselves in a variety of orientations while inside the tote and additionally have various levels of occlusions presented to the camera due to other objects resting on top. Students undertaking this class will form small teams and build algorithms that can identify as many object as possible in a given setup. Students will collect data which will be curated to form a benchmark gathering all samples, to be later used by all teams.

5.2 Project scope

The semester project aims at delivering an end-to-end algorithm, with a prototype software written in one of popular languages (*e.g.*, Python), to visually identify objects in dense clutter that considers a variety of occlusions for the object itself.

Teams will be presented a plastic tote, approximately 24” $\times$ 16” $\times$ 11” in dimension full of representative objects. Using a few different cameras (delivered by the instructor) each team will take pictures under varying lighting conditions, at different perspectives and at different distances from the tote. Once the tote is filled with objects, the objects cannot be manipulated or moved. However, the objects will be also imaged separately to form a reference database of items to be identified. Acquisitions from all teams will be mixed and used by all students in developing of their algorithms.

All teams will then develop methods, using the same benchmark, with a goal to visually identify as many objects within the tote as possible. The proposed solutions will be tested with a sequestered dataset collected by teaching assistants in a similar setup, but introducing (a) more variations of lighting/scale/perspective, and (b) unknown objects not present in the train set. The objects found in the test pictures should be segmented, identified and counted against a local database. “Unknown” objects (introduced in the test set, and not present in the train set) should not be counted. The solutions will be ranked by both accuracy and speed.

All teams will make 10-minute presentations of their solutions, research findings at the end of the semester. The audience will identify strengths and weaknesses of the proposed solutions, and compare them with own algorithms (What my colleagues did better than I did? What my colleagues could do better if they had more time? Are our solutions complementary in some sense, and could be combined to offer a better performance?). Additionally, all students will be given an opportunity to have their research reports submitted to the Amazon Robotics research team.

5.3 Research questions

Questions to be answered in this semester project include:

- What is the minimum amount of information necessary to robustly identify everyday objects within a tote?
- How much occlusion can be present on a single object yet still result in a positive high confidence match?
- What is the minimum amount of information necessary in the matching database and how small can the footprint of the database be?
- What is the robustness of the proposed methods to varying illumination, scale, rotation and selested properties of sensors?
- Is training necessary? If so, what is the least amount of training required to robustly match occluded objects?
- Is the approach scalable to large databases? If so, how can we test this?
- Could the approach extend to classes of objects or even novel objects?

5.4 Deliverables

"EoW X" means "End of week X of the semester." We have 15 weeks in total in the semester.

- **EoW 3:** Training database of images of a tote filled with objects, and a reference database of object images (10 points).
- **EoW 4:** Preliminary design and architectural description (5 points).
- **EoW 8:** Mid-semester report with findings to date. Refinement of the design, if needed (15 points).
- **EoW 13:** Final research report (20 points).
- **EoW 15:** End of semester final presentation and demo (10 points).

Note: Interested students may request having their reports submitted to the Amazon Robotics research team by the instructor.

6 Learning objectives

At the end of the course students should be able to:

- understand top-level and up-to-date literature in computer vision,
- select and apply current computer vision tools to solve an object identification problem in densely cluttered environment,
- prepare a constructive critique of computer vision solutions offered by other researchers and classmates,
- present in a concise way own computer vision solutions.

7 Instructors

Dr. Adam Czajka
Office: 180 Fitzpatrick Hall
Office hours: Wednesday 11:00 am – 12:00 pm
Email: `aczajka@nd.edu`

Lucas Parzianello
Teaching Assistant
Office: 212 Cushing Hall
Office hours: Monday 3:30 pm – 4:30 pm
Email: `lbarbosa@nd.edu`

8 Grading

- 54 points for the semester project (data collection: 5, preliminary design: 4, midterm report: 15, final report: 25, final presentation: 5)
- 18 for in-class practicals (9 practicals × 2)
- 18 points for in-class quizzes (9 quizzes × 2)
- 10 for paper presentation

Grade scale:

		Total		
A	4.000	[100	->	91)
A-	3.667	[91	->	83)
B+	3.333	[83	->	75)
B	3.000	[75	->	66)
B-	2.667	[66	->	58)
C+	2.333	[58	->	50)
C	2.000	[50	->	41)
C-	1.667	[41	->	25)
D	1.000	[25	->	25]
F	0.000	<25		

Late grading policy:

- -1 point / day of late delivery, but ...
- students are allowed to be late **once** during the semester (*i.e.*, there is no penalty for the first late submission).

9 Code of Honor

This class follows the binding Code of Honor at Notre Dame (`http://honorcode.nd.edu`). The graded work you do in this class must be your own. In the case where you collaborate with other students make sure to fairly attribute their contribution to your project.

10 Suggested reading

There are no required textbooks. Our class meetings will be based on open access versions of the research papers provided by the Computer Vision Foundation, ArXiv, as well as offered by IEEEXplore at no cost for Notre Dame students while accessed on campus. The following materials related to general computer vision topics are suggested:

- Open access vision papers offered by Computer Vision Foundation: `http://openaccess.thecvf.com/menu.py`
- R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011 (draft available on-line: `http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf`)
- J. Daugman, *Computer Vision – Lecture Notes*, available on-line: `https://www.cl.cam.ac.uk/~jgd1000/`
- CVonline: *Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision*: `http://homepages.inf.ed.ac.uk/rbf/CVonline`
- R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, 2004 (available on-line: `http://www.robots.ox.ac.uk/~vgg/hzbook`)
- Reinhard Klette, *Concise Computer Vision*, Springer, 2014