ECE59500CV: Deep Learning for Computer Vision
Fall 2021

Course Information
Course number and title: ECE59500CV (013) Deep Learning for Computer Vision
CRN: 24471
Meeting days(s) and times(s): MWF 3:30pm–4:20pm WALC 3154
Instructional Modality: Face-to-Face
Course credit hours: 3.000
Prerequisites: ECE26400, ECE36900, ECE30200, ECE60000 for graduate students, and MA26200 or MA26500, or permission of instructor
Course web page: https://engineering.purdue.edu/ece595cv
Course Brightspace page: https://purdue.brightspace.com/d2l/home/336301

Instructor(s) Contact Information
Name of the instructor: Jeffrey Mark Siskind
Office Location: EE313e
Office Phone Number: 765/496-3197
Purdue Email Address: qobi@purdue.edu
Student Consultation hours, times, and location: T 5:00pm–6:00pm EE313e

Course Description
An introduction to modern computer vision using methods from machine learning and deep learning. Covers segmentation, object classification and localization, activity classification and localization, semantic segmentation, depth reconstruction, 3D reconstruction, generative adversarial networks, image and video captioning, and image and video retrieval. The course will cover fundamental topics as well as recent advances from the literature.

Learning Resources, Technology & Texts
There is no required text for this course. We will cover a recent paper from the literature during each lecture. The papers will be posted at http://engineering.purdue.edu/ece595cv before each lecture. Students should read each paper before the associated lecture and come to class prepared to discuss and ask questions.

Learning Outcomes
Outcome i) Ability to design and implement an object classifier and localizer. [1, 2, 4, 6, 7]
Outcome ii) Ability to design and implement an activity classifier and localizer. [1, 2, 4, 6, 7]
Outcome iii) Ability to design and implement an image or video captioning system. [1, 2, 4, 6, 7]
Outcome iv) Ability to present research results in Computer Vision to peers. [3, 5]

Assignments
There will be no problem sets and no exams. The only assignment is a term paper/project/presentation. Beyond this, there will be a reading assignment for each lecture. A recent paper will be posted on the course web page for each lecture. Students are expected to read the paper prior to lecture and come to class prepared to discuss and ask questions.
The term paper/project/presentation is organized like a conference. Students will be required to select and read three recent conference or journal papers in the field of computer vision. Nominally, the papers should have been published within the last three years in one of the following venues: CVPR, ICCV, ECCV, TPAMI, IJCV, NIPS, ICML, ICLR, or JMLR.
You are welcome to read more than three. Your selection of papers must be approved by me. I may be willing to accept papers from venues other than the above and/or older than three years old. But all paper selections, whether or not they meet the above criteria must be approved by me. You must submit your paper selections by 5:00pm Friday 3 September 2021 with the following information:
(a) a BibTeX entry for each of the papers that you would like to read. For a journal paper this should contain (at least) the paper title, authors, journal, volume, year, and pages. For a conference paper this should contain (at least) the

paper title, authors, conference, year, and pages.

(b) a URL containing a pdf file of each paper, included in the BibTeX entry

Paper selections will be submitted through Brightspace. Your submission should be a single BibTeX file with the extension `.bib`. The filename does not matter.

Please note that you should not send me the paper itself. Not even the abstract.

If you intend to select papers that are older than three years old or from a venue other than listed above, I suggest that you discuss this with me prior to Friday 3 September 2021.

You will be required to read all three papers and implement and evaluate the ideas from (at least) one of the papers. Thus (at least) one of the papers that you select should contain material that is suitable for implementation. The implementation must be nontrivial. A good guideline is that the implementation should be at least four pages of code. This is not a strict guideline. Ultimately, I will determine whether or not the implementation meets the non-triviality requirement. The implementation can be in any programming language that you choose, though obviously lower level languages may require much more code to implement the same functionality that could be achieved with less code in a higher level language. You must also conduct a substantive evaluation of your implementation to determine how well it solves the intended problem. Ideally, you should replicate the experiments presented in the paper but I will not require this.

You will be required to write a paper in LaTeX meeting all of the submission format requirements of CVPR 2022 including the typesetting conventions and page length. Approximately half of this paper should be a substantive critique of the three papers that you have read. And approximately half of this paper should be a description of your implementation and evaluation of the material from one of the papers. The term paper must be submitted by 5:00pm Friday 12 November 2021.

Term papers will be submitted through Brightspace Your submission should be a single PDF file with the extension `.pdf`. The filename does not matter.

The term papers will be reviewed by other students in the class. Like all conferences, this process will be double blind: reviewers will not know the identity of authors and vice versa. To support this, like all conferences, you should NOT put your name on the term paper submission. In place of your name, you should put your PUID. Also like conferences, reviews will be confidential. The only person who will be privy to the reviews will be the reviewer, the instructor, and the author.

I will assign each term paper to five students to read. Each student will be required to read five student term papers (other than their own) and prepare conference-style reviews, primarily indicating clarity and the quality of the implementation effort. Peer reviews will be due by 5:00pm Friday 3 December 2021. The exact format for the review will be announced later in the semester.

Reviews will be submitted through Brightspace using a mechanism that will be announced later in the semester. Your reviews should be formatted as text files with the extension `.txt`. Your reviews should be enclosed in a single ZIP file with the extension `.zip`. The filename of the ZIP does not matter. But you should take care to use exactly the names of the text files as will be described later in the semester

You will also be required to make a 12 minute conference-like presentation in class during the last six weeks of classes. The exact schedule will be determined later in the semester. But to accommodate all students in class, presentations will start on Friday 12 November 2021. So you should be prepared to give your presentation at any time after that. This is necessarily before the due date for paper submissions and the due date for reviews. This presentation should cover both a summary of the three papers that you have read, a summary of your critique of those papers, and a description of your implementation and evaluation. The presentation must be given in class presented either by a laptop connected to the wall screens or on the computer at the podium. You can use whatever tools you wish to prepare your presentation (i.e. LaTeX/beamer or PowerPoint). On the day that you are scheduled to give you presentation, make sure ahead of class that you presentation setup works and that you are prepared to give your presentation in the allotted time slot.

Students will be required to attend all student conference-like presentations. Attendance will be taken in the student conference-like presentations.

Grading Scale

On the first day of class, Mr Silver, my eight grade science teacher, said: "This is an honors class. You all are smart and deserve a good grade. A 90 is a good grade. None of you are perfect yet. So nobody deserves a 100. Thus I will give everyone a 90. Now that we dispensed with grades, we can get down to learning."

I will follow the same policy in this class this semester. The grading policy is simple. This is a graduate class. You all are smart and deserve a good grade. An A is a good grade. None of you are perfect yet. So nobody deserves an A+. Thus I will give everyone who completes all of the assignments by the deadlines, does not violate the anti-plagiarism policy, and attends all of the student presentations an A. Everyone else gets an F. Now that we dispensed with grades, we can get down to learning.
Grades are cheap. An education is expensive. I can manufacture As. When the marginal cost of production goes to zero, the value of the product goes to zero. The As that I give in this course are worthless. I strive to make the education that I give in this course worth the price of tuition. Please don't ask me for a letter of recommendation, to hire you as an RA, or to take you on as a graduate student on the basis of your grade in this course, or even on the basis of the term paper/project/presentation you prepared for this course. You need to do something else that distinguishes you for me to do those things.

Attendance Policy
Nobody ever listens to Grateful Dead studio albums. The Dead exist solely for their live concerts. People went to Dead concerts for two reasons. To hear Jerry Garcia jam, and to get stoned.
I don't condone getting stoned at Purdue or in Indiana. But I do strive to jam like Jerry Garcia. If you want studio albums, you can watch online videos of AI courses taught at Stanford by well-known researchers like Peter Norvig, Sebastian Thrun, Andrew Ng, Daphne Koller, and Feifei Li. Hundreds of thousands, perhaps millions, of people have taken these MOOCs. But with my course, you get to see me jam.
My course is like a concert. You already bought the ticket. I perform. You get out of the class what you put into it. It is your choice whether you come to class (prepared) and do the coursework. If you do, you get an opportunity to learn. If you don't, you wasted your money (or someone else's money).
If you bought a ticket to a Taylor Swift concert, but need to miss the concert, you wouldn't run up on stage and say "Ms. Swift, I'm sorry but I need to miss your concert. But I really like your new songs. So can I come by your home and have you give me a private rendition of the songs I missed at your concert?" So please don't ask me for permission to miss class. And please don't ask me to repeat material from a missed class.

Academic Guidance in the Event a Student is Quarantined/Isolated
I will make an attempt to record lectures and post on the course web page. Recorded lectures from the Fall 2020 offering are posted on the course web page.

Course Schedule
selections due 5pm ET Friday 3 September 2021
paper due 5pm ET Friday 12 November 2021
reviews due 5pm ET Friday 3 December 2021

Classroom Guidance Regarding Protect Purdue
Any student who has substantial reason to believe that another person is threatening the safety of others by not complying with Protect Purdue protocols is encouraged to report the behavior to and discuss the next steps with their instructor. Students also have the option of reporting the behavior to the Office of the Student Rights and Responsibilities. See also Purdue University Bill of Student Rights and the Violent Behavior Policy under University Resources in Brightspace.

Academic Integrity
For the course, the only requirement is the term paper/project/presentation. There is a simple requirement: you cannot plagiarize. You must follow standard academic practice and cite the source of all material that you (re)use. If you violate this anti-plagiarism policy you will be prosecuted by the standard University mechanisms for such violations.

Nondiscrimination Statement
Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members

strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. A hyperlink to Purdue's full Nondiscrimination Policy Statement is included in our course Brightspace under University Policies.

Accessibility
Purdue University is committed to making learning experiences accessible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247.

Mental Health/Wellness Statement
If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack. Sign in and find information and tools at your fingertips, available to you at any time.
If you need support and information about options and resources, please contact or see the Office of the Dean of Students. Call 765-494-1747. Hours of operation are M–F, 8 am–5 pm.
If you find yourself struggling to find a healthy balance between academics, social life, stress, etc. sign up for free one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell. Student coaches can help you navigate through barriers and challenges toward your goals throughout the semester. Sign up is completely free and can be done on BoilerConnect. If you have any questions, please contact Purdue Wellness at evans240@purdue.edu.
If you're struggling and need mental health services: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of mental health support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office on the second floor of the Purdue University Student Health Center (PUSH) during business hours.

Basic Needs Security
Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. There is no appointment needed and Student Support Services is available to serve students 8 a.m.–5 p.m. Monday through Friday. Considering the significant disruptions caused by the current global crisis as it related to COVID-19, students may submit requests for emergency assistance from the Critical Needs Fund

Emergency Preparation
In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor's control. Relevant changes to this course will be posted onto the course website or can be obtained by contacting the instructors or TAs via email or phone. You are expected to read your @purdue.edu email on a frequent basis.

Instructor Email Policies
Please do not send any course-related email to my personal email account. Please send all course-related email to one of the following two mailing lists:

- ece59500cv-staff-list@ecn.purdue.edu
- ece59500cv-students-list@ecn.purdue.edu

The former just goes to me, and any TAs if we have any in the future. The latter goes to me, any TAs if we have any in the future, and to all students enrolled. All students are automatically added to the latter under their Purdue Career account. Please send all course-related email from your Purdue Career account.

Missed or Late Work
The only assignments this semester are related to the term paper/project/presentation. These assignments attempt to model the actual experience of submitting to, reviewing for, presenting at, and attending a research conference. Research conferences have submission and reviewing deadlines, and papers are presented according to a fixed schedule. In order to teach what it is like to participate in a research conference, we have strict deadlines for all components of the term paper/project/presentation. There will be no exceptions. No late work will be accepted and no credit will be

given for missed or late assignments.

Incompletes
No incompletes will be given.

References

A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research (JMLR)*, 18(153):1–43, 2018.

J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the Kinetics dataset. In *Computer Vision and Pattern Recognition (CVPR)*, pages 6299–6308, 2017.

N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005.

D. F. Dementhon and L. S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision (IJCV)*, 15(1–2):123–141, 1995.

A. Diba, V. Sharma, and L. Van Gool. Deep temporal linear encoding networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2329–2338, 2017.

C. Feichtenhofer. X3D: Expanding architectures for efficient video recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 203–213, 2020.

C. Feichtenhofer, A. Pinz, and A. Zisserman. Convolutional two-stream network fusion for video action recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1933–1941, 2016.

C. Feichtenhofer, H. Fan, J. Malik, and K. He. Slowfast networks for video recognition. In *International Conference on Computer Vision (ICCV)*, pages 6202–6211, 2019.

P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(9):1627–1645, 2010.

R. Girshick. Fast R-CNN. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1440–1448, 2015.

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 580–587, 2014.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *International Conference on Computer Vision (ICCV)*, pages 2961–2969, 2017.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.

S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, pages 448–456, 2015.

A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1725–1732, 2014.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1097–1105, 2012.

Y. Li, B. Ji, X. Shi, J. Zhang, B. Kang, and L. Wang. TEA: Temporal excitation and aggregation for action recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 909–918, 2020.

J. Lin, C. Gan, and S. Han. TSM: Temporal shift module for efficient video understanding. In *International Conference on Computer Vision (ICCV)*, pages 7083–7093, 2019.

T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2117–2125, 2017.

W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In *European Conference on Computer Vision (ECCV)*, pages 21–37, 2016.

D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence (AIJ)*, 31(3):355–395, 1987.

B. A. Pearlmutter and J. M. Siskind. Reverse-mode AD in a functional framework: Lambda the ultimate backpropagator. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 30(2):1–36, 2008.

J. Redmon and A. Farhadi. YOLO9000: better, faster, stronger. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7263–7271, 2017.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 91–99, 2015.

K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 568–576, 2014.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015.

J. M. Siskind and B. A. Pearlmutter. Nesting forward-mode AD in a functional framework. *Higher-Order and Symbolic Computation (HOSC)*, 21(4):361–376, 2008.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception architecture for computer vision. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *Conference on Artificial Intelligence (AAAI)*, pages 4278–4284, 2017.

D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In *International Conference on Computer Vision (ICCV)*, pages 4489–4497, 2015.

L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4305–4314, 2015.

L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool. Temporal segment networks: Towards good practices for deep action recognition. In *European Conference on Computer Vision (ECCV)*, pages 20–36, 2016.

X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7794–7803, 2018.

C.-Y. Wu, M. Zaheer, H. Hu, R. Manmatha, A. J. Smola, and P. Krähenbühl. Compressed video action recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 6026–6035, 2018.

J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. Beyond short snippets: Deep networks for video classification. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4694–4702, 2015.

B. Zhou, A. Andonian, A. Oliva, and A. Torralba. Temporal relational reasoning in videos. In *European Conference on Computer Vision (ECCV)*, pages 803–818, 2018.

Y. Zhu, Z. Lan, S. Newsam, and A. Hauptmann. Hidden two-stream convolutional networks for action recognition. In *Asian Conference on Computer Vision (ACCV)*, pages 363–378, 2018.

M. Zolfaghari, K. Singh, and T. Brox. ECO: Efficient convolutional network for online video understanding. In *European Conference on Computer Vision (ECCV)*, pages 695–712, 2018.