CSE 659A: Advances in Computer Vision

Spring 2019: T-R: 2:30-4pm @ Cupples II/230

Instructor: Ayan Chakrabarti (ayan@wustl.edu).

http://www.cse.wustl.edu/~ayan/courses/cse659a/

Jan 15, 2019
INTRODUCTION

No. of Papers Accepted at CVPR (CVPR 2018 Opening Slides)
No. of Papers Submitted & Accepted at CVPR (CVPR 2018 Opening Slides)

(Many rejected papers still have good ideas, and just end up on arXiv)
Computer Vision is a rapidly changing field!

- Every conference (CVPR, ICCV/ECCV, NIPS, ...) leads to improvements in performance on many tasks.
- At any one time, researchers are exploring different strategies in parallel for each task.
- At the same time, people are coming up with new tasks as old ones get "solved".

This is both exciting, and overwhelming ...

What it means for you:

- Need to do have an idea of the different advanced techniques that people are trying to solve problems.
- Have a broad sense of how state-of-the-art methods for different tasks work.
- Have the ability to read and understand papers that are, and continue, appearing in conferences.
## PRE-REQUISITE

CSE 559A (or some other foundational vision course)

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 28</td>
<td>Introduction. Cameras &amp; Image Formation I.</td>
</tr>
<tr>
<td>Aug 30</td>
<td>Image Formation II. Per-pixel Image Operations. Convolutions I.</td>
</tr>
<tr>
<td>Sep 4</td>
<td>Convolutions II. Edge &amp; Line detection.</td>
</tr>
<tr>
<td>Sep 6</td>
<td>Other spatial Image operations. Fourier Transforms &amp; Convolution Theorem.</td>
</tr>
<tr>
<td>Sep 11</td>
<td>Scale &amp; Multi-scale Representations. Efficient Convolutions.</td>
</tr>
<tr>
<td>Sep 13</td>
<td>Image Restoration &amp; Optimization.</td>
</tr>
<tr>
<td>Sep 18</td>
<td>Restoration &amp; Optimization II. Color, spectrum, and Illumination.</td>
</tr>
<tr>
<td>Sep 20</td>
<td>Material &amp; Shading. Lambertian, specular, BRDFs.</td>
</tr>
<tr>
<td>Sep 25</td>
<td>Light Sources. Photometric Stereo.</td>
</tr>
<tr>
<td>Sep 27</td>
<td>Photometric Stereo Roundup. Camera Projection &amp; Geometry I.</td>
</tr>
<tr>
<td>Oct 2</td>
<td>Camera Projection &amp; Calibration. Multi-view Geometry</td>
</tr>
<tr>
<td>Oct 4</td>
<td>Neural Networks for Physical Tasks I.</td>
</tr>
<tr>
<td>Oct 11</td>
<td>Stereo Globalization II.</td>
</tr>
<tr>
<td>Oct 16</td>
<td>Fall Break</td>
</tr>
<tr>
<td>Oct 18</td>
<td>Optical Flow.</td>
</tr>
<tr>
<td>Oct 23</td>
<td>Segmentation &amp; Grouping I.</td>
</tr>
<tr>
<td>Oct 25</td>
<td>Segmentation &amp; Grouping II. Intro to ML.</td>
</tr>
<tr>
<td>Nov 1</td>
<td>Intro to ML II. Learning with SGD.</td>
</tr>
<tr>
<td>Nov 8</td>
<td>Back prop II.</td>
</tr>
<tr>
<td>Nov 13</td>
<td>Semantic Vision Tasks.</td>
</tr>
<tr>
<td>Nov 20</td>
<td>Training in Practice: Regularization, optimization.</td>
</tr>
<tr>
<td>Nov 22</td>
<td>Thanksgiving</td>
</tr>
<tr>
<td>Nov 27</td>
<td>Neural Networks for Physical Tasks II.</td>
</tr>
<tr>
<td>Dec 4</td>
<td>Generative Adversarial Networks.</td>
</tr>
</tbody>
</table>

[http://www.cse.wustl.edu/~ayan/courses/cse559a/](http://www.cse.wustl.edu/~ayan/courses/cse559a/)
PRE-REQUISITE

If you haven't taken CSE 559A (with me)

• Go to http://www.cse.wustl.edu/~ayan/courses/cse559a/
• Read the list of topics, and make sure you are familiar with all / most.

What CSE 559A taught you: (hopefully)

• Basics of the math behind computer vision (image representations, geometry, photometry, ML, ...)
• Building blocks of image processing and vision algorithms (convolutions, optimization, ...)
• Linking up equations and concepts to implementations (guided by problem sets)
  □ Read about an algorithm, and figuring out how to implement it (efficiently).
THIS COURSE

Assuming you have the foundations covered:

- Will teach you about more advanced algorithms that build on these foundations.
- Significant fraction of the course will be talking about specific methods in published papers.
  - Including 'competing' alternate methods, since "the best approach" is not settled.
  - But as part of covering these specific examples, you'll learn about broader techniques and strategies.
- Will introduce you to emerging new applications.
Topic 1: Models for Spatial Relationships

- We saw simple regularizers and smoothness models in 559A (image restoration, stereo, ...)
- Describe more complex models for image intensities, and other continuous and discrete valued variables.
- Talk about optimization algorithms that incorporate these models for inference.
- Talk about how these interact with neural networks.
Topic 2: Depth and Motion Estimation

- More complex methods for stereo and optical flow
  - Complex in the kind of 'priors' they assume
  - Also look at un-rectified stereo
- CNN-based methods for stereo and flow
- Monocular depth and surface normal estimation
Topic 3: Classification and Recognition

- Begin by describing 'classical' methods for semantic reasoning
  - Interest point detectors and **descriptors**
  - Whole Scene Descriptors
  - Algorithms based on these
- Advanced CNN-based methods for detection (and how their architectures match up to traditional methods)
- Algorithms for segmentation (including interaction with spatial models)
- New tasks: image captioning, visual question answering, ...
Topic 4: Computational Photography I

"Advanced Image Processing"

- Texture synthesis, image retargeting (seam carving)
- Image harmonization (for image splicing), CG2Real
- Motion Magnification
- Image Inpainting and "Un-cropping"
- Image Editing with Smart Contours
Topic 5: Photometric Reasoning

- Deeper conceptual dive into what does shading and appearance tell us about shape?
- Modern algorithms for photometric stereo.
- Shape from shading, intrinsic image decomposition.
- Color Constancy with CNNs. Multiple illuminant estimation & separation.
Topic 6: Computational Photography II

*Wacky cameras*

- Dark flash photography
- Coded aperture and coded exposure photography. Light-field cameras.
- Others (flatcam, modulo cameras, ...)
- Learning camera designs and acquisition setups.
GRADING

(probably) less work than 559A, but more self-directed!

- Most of your grade (60%) will be based on two projects (30% each)
- You will also give a 15-20 minute presentation on your first project (15%)
  - Length will be based on no. of people.
- 25% of your grade will be based on completing 5 homework 'paper reviews'.

Again: No Exams!

But do read the late and collaboration policies on course website. (No late days!)
GRADING: HOMEWORK REVIEWS

• For each homework, I will assign a paper for you to read and review.
  ■ See website for (tentative) schedule of when papers assigned, and reviews due.
• Your review will be short: 1-2 pages (LaTeX template will be posted)

Should summarize three things:

• *Novelty/Significance* of the paper
  ■ Based on stuff we've been talking about in class, what's new and different about this paper?

• *Summary of Technical Content*
  ■ Generally brief. You should read and understand the paper to the level that you're confident you could implement it if you tried.
  ■ Summarize the content in your own words so that I know you've understood.
  ■ Think of effort being equal to writing a 'project proposal' if you were going to implement this paper as your project.

• *Criticisms and Limitations*
  ■ Need to understand that papers, even those accepted to CVPR, aren't 'authoritative'. Read it critically.
  ■ Talk about where the paper falls short: e.g., certain cases not considered in experiments, novelty less than claimed, real-world utility less than claimed, ...
GRADING: PROJECTS

- Two projects: due in the middle and end of the semester.
- Expect you to read and implement 1-2 published papers in each.
- Restriction: Project 1 should be on a paper related to topics 1-3, and Project 2 on topics 3-5.
  - You can't have both projects based on topic 3 (recognition/classification)
- Paper can be one we discussed in class: but then bar is higher in terms of analysis.
- Like for 559A, a short proposal will be due where you tell me which papers you're working on.
- Analyze method. Implement yourself / modify or analyze existing implementation.
- Write a report (about 6 pages). Template and Rubric will be posted.
GRADING: PRESENTATION

- Prepare a 15-20 minute talk based on Project 1.
- Including 3-5 minutes for questions.
- Last month of semester, one day a week will be for these presentations.
- Schedule will be determined randomly.
- But irrespective of when you present, everyone's slides will need to be submitted at the same time (prior to the first day of presentations).
Go to the course website: http://www.cse.wustl.edu/~ayan/courses/cse659a/

Has links to piazza and canvas page for course.

My Office Hours: By appointment, make a private post on piazza. (No TAs unfortunately).

Slides will be posted on the course website after class.

Readings will also be posted:
  - Typically papers we are discussing, will discuss in class.
  - Slides will announce suggested papers to read before coming lectures.

Use Piazza to air questions and have more in-depth discussions about papers we're covering in class.
  - Take part in the discussion!

Any questions?
Consider the generic task of estimating an image like object $X$

- $X[n]$ is the value of $X$ at pixel location $n$

- Continuous valued: $X[n] \in \mathbb{R}^d$ like images, depth maps, flow fields, normal maps, ...

- Discrete valued: $X[n] \in \mathbb{L}$ for some label set $\mathbb{L}$
  - Per-pixel labels or segmentation maps, discrete valued disparities, etc.

Since inference problems in vision are ill-posed or under-determined, we want to impose a "prior" or regularizer on $X$

- Typically, this takes the form of constraining neighboring values of $X[n]$ to be related in some way (e.g., imposing smoothness).

- Different such models for spatial relationships: for continuous and discrete values.

- For each model, we must ensure that it faithfully encodes the expected structure of $X$, and allows efficient inference.
CONTINUOUS VALUED MAPS

\[ \hat{X} = \arg \min_X \phi(X) + R(X) \]

- \( \phi(X) \) is the data terms: derived from observations in some way.

E.g., image restoration:

\[ \phi(X) = \|X - Y\|^2, \quad \phi(X) = \|AX - Y\|^2 \]

- \( R(X) \) is based on our prior expectation for \( X \). And that's where we encode these spatial relationships.

Recap: Bayesian interpretation

\[ \Phi(X) = -\log p(Y|X), \quad R(X) = -\log p(X) \quad \Rightarrow \hat{X} = \arg \max_X p(X|Y) = \arg \max_X p(Y|X)p(X) \]
CONTINUOUS VALUED MAPS

- Simplest form of prior: gradients are small!
- Define $R(X)$ as sum of independent penalties on different gradients of $X$ (based on gradient filters $\{\nabla_f\}$)

$$R(X) = \sum_n \sum_f R_f \left((\nabla_f \ast X)(n)\right)$$

- For example: the summation could be over the different wavelet coefficients of $X$ ($\nabla_f$ correspond to wavelet filters).
- Each $R_f(\cdot)$ takes a scalar-valued input and imposes its own penalty.

E.g., squared regularizers:

$$R_f(v) = \lambda v^2$$

We can also have L1 penalties (remember problem set 2 from 559)

$$R_f(v) = \lambda |v|$$

More generally,

$$R_f(v) = \lambda |v|^p, \quad p > 0$$
CONTINUOUS VALUED MAPS

\[ R(v) = \lambda |v|^p, \quad p > 0 \]

- What is the effect of different values of \( p \)?
- Two ways to analyze
  - Think of what probability distribution each corresponds to.
  - Think of the effect of minimizing \((v - v_0)^2 + R(v)\)

\[ R(v) = \lambda |v|^p \quad \Rightarrow \quad p(v) \propto \exp(-\lambda |v|^p) \]

- For \( p = 2 \), this is a zero-mean Gaussian distribution, with variance \( \sigma^2 = \frac{1}{2\lambda} \).
- For \( p = 1 \), this is a "Laplace" distribution. It also has zero mean, and variance \( \mathbb{E}v^2 = \frac{2}{\lambda^2} \).
- More generally for arbitrary \( p \), these correspond to Generalized Gaussian Distributions
- As you decrease the value of \( p \), the distribution gets "heavier tailed".
CONTINUOUS VALUED MAPS

- Plots of $\log p(v), p(v) \propto \exp(-\lambda |v|^p)$, for different values of $p$. Different $\lambda$ values were chosen for each $p$, so that these distributions have the same variance.

- We see that $p < 2$ has higher probability than a Gaussian ($p = 2$) at both 0 and high-magnitudes (tails)!

- The $p < 2$ distributions drop sharply from 0, but then flatten out.

- Turns out, heavier-tailed distributions are a closer match to actual distributions of gradients in natural images (and depth maps, etc.)

- This is because gradients are very small in most (smooth) regions except at discontinuities / edges. But at these edges, the gradients can be arbitrarily large.
CONTINUOUS VALUED MAPS

- Understanding the effect of heavy-tailed regularizers

Shrinkage Function

\[ s_{R, \alpha}(v_0) = \arg \min_v \alpha(v - v_0)^2 + R(v) \]

- Can be thought of as "denoising" with the prior corresponding to \( R \).
- Would correspond to an algorithm for actual denoising if you defined \( R(X) \) on the entire image in terms of sums of regularizers on wavelet coefficients.

\[ \hat{X} = \alpha|X - Y|^2 + \sum_i R_i ((WX)_i) \]

- \( W \) is a unitary matrix corresponding to the wavelet transform, and \((WX)_i\) is the \( i^{th} \) wavelet coefficient of \( X \).
- Then, you would take the wavelet transform to \( Y \), solve an independent minimization for each coefficient, and take the inverse Wavelet transform.

\[ C \leftarrow WY, \quad \hat{C}_i \leftarrow s_{R_i, \alpha}(C_i), \quad \hat{X} = W^{-1} \hat{C} \]
Shrinkage Function: $s_{R;\alpha}(v_0) = \arg \min_v \alpha (v - v_0)^2 + R(v)$ for different $p$

- Again, regularizers correspond to the same prior variance. Same value of $\alpha$ used.
- For $p = 2$ / Gaussian, we get a straight line with a lower slope. Because it shrinks by a constant factor.
- $p < 2$: The effect is non-linear. All values below some threshold are zero-ed out. But conversely, these better preserve higher-magnitude values which could correspond to true edges.
So, when talking about simple gradient regularizers, $p < 2$ or heavy-tailed regularizers are often preferable. Because they better encode true distributions of gradients, and "do what we want".

But denoising by shrinkage is (mostly) a toy example. How do you apply them in a general optimization scenario?

- When its not simple denoising?
- When we aren't considering coefficients of a Wavelet or some other 'unitary' transform?

\[
\hat{X} = \arg \min_X \|AX - Y\|^2 + \lambda \sum_n |(\nabla_x * X)[n]|^p + \lambda \sum_n |(\nabla_y * X)[n]|^p
\]

- We've seen the $p = 2$ case in 559A. This is a least-squares problem. Can be solved in the Fourier domain, or by conjugate gradient.

- $p = 1$ is a convex optimization problem, but not least-squares.

- $p < 1$ is not even convex!