• **Instructor:** Dr. Mubarak Shah  
• **Email:** shah@crcv.ucf.edu  
• **Office:** HEC 245  
• **Phone:** 4078235077  
• **Time:** Tuesdays and Thursdays 3:00 to 4:15PM  
• **Location:** HEC 117  
• **Office Hours:** Tuesday 4:15 to 5:00PM; Thursdays 2:00 to 3:00PM and by appointment  
• **Extra Discussion Session:** Wednesdays 3:00 to 4:00, HEC 356  
• **Pre-requisite:** CAP5415  
• **Course webpage:** [http://crcv.ucf.edu/courses/CAP6412/Spring2019/](http://crcv.ucf.edu/courses/CAP6412/Spring2019/)
To expose graduate students to the cutting-edge research.

In each class we will discuss one recent research paper related to active areas of current research in particular employing Deep Learning.
STUDENT LEARNING OUTCOMES

Read and understand a research paper.

Write a comprehensive review of the paper.

To identify strong and weak points of the paper.

To come up with own ideas to solve the same problem, which may lead to their first research paper.

To implement known method or work on and successfully complete individual project.
GRADING POLICY

Reports 20%
Presentation 10%
Attendance and Discussion 20%
Projects/Programs 50%
(Keras, PyTorch, TensorFlow,..)

Late Policy
- 0 for late reports
- Projects/Programs
  - 20% off per day
  - up to 4 days
Each student will get an account

There will be presentation about how to use the system:

- Wed January 16, 3:00PM, HEC 101-B
PAPERS/ REPORTS

We will discuss one paper in each class

You can select the paper you want to present from the list on course webpage or
Suggest other any other paper you want to present with my approval

All students will read the assigned paper before the class and write a report

Reports will be due just before the class meeting through Web Courses

One student will be responsible to present the paper

Each presentation will be of maximum 30 minutes

Papers: http://crcv.ucf.edu/courses/CAP6412/Spring2019/

Schedule Table
REPORTS (ONE PAGE)

Summary
Good points
Weak points
Questions
Ideas
PAPER PRESENTATION REVIEW/REHEARSAL SCHEDULE

Review of power point presentation:

For Thursday presentation
- Review Monday 2:30PM/HEC 245
- Rehearsal Tuesday during office hours 4:15 PM/HEC 245

For Tuesday presentation
- Review Thursday a week before during office hours: 2:00 to 3:00PM/HEC 245
- Rehearsal Monday 2:00PM/HEC 245
STATEMENT OF ACADEMIC INTEGRITY

The UCF Golden Rule (http://goldenrule.sdes.ucf.edu/) will be observed in the class. Plagiarism and cheating of any kind on an examination, quiz, or assignment will result at least in an "F" for that assignment (and may, depending on the severity of the case, lead to an "F" for the entire course) and may be subject to appropriate referral to the Office of Student Conduct for further action. I will assume for this course that you will adhere to the academic creed of this University and will maintain the highest standards of academic integrity. In other words, don't cheat by giving answers to others or taking them from anyone else. I will also adhere to the highest standards of academic integrity, so please do not ask me to change (or expect me to change) your grade illegitimately or to bend or break rules for one person that will not apply to everyone.
There is no text book for this class. We will discuss recent research papers.

Recommended supplemental textbook:
UCF RESOURCES

Tutorial on Keras: Kishan Athrey


CAP6412 Spring 2018
RECOMMENDED ONLINE COURSES AND TUTORIALS

https://www.youtube.com/watch?v=CS4cs9xVecg&list=PLkDaE6sCZn6Ec-XTbcX1uRg2_u4xOEky0  Andrew Ng
http://cs231n.stanford.edu/CS231n: Convolutional Neural Networks for Visual Recognition
http://web.stanford.edu/class/cs224n/CS224n: Natural Language Processing with Deep Learning
http://rll.berkeley.edu/deeprlcourse/CS 294: Deep Reinforcement Learning
http://distill.pub/ Very nice explanations of some DL concepts
https://class.coursera.org/ml003/lecture/preview
https://github.com/adeshpande3?tab=repositories
HOW TO READ A RESEARCH PAPER?

You have to read the paper several times to understand it.
- When you read the paper first time, if you do not understand something do not get stuck, keep reading assuming you will figure out that later.
- When you read it the second time, you will understand much more, and the third time even more ...
HOW TO READ A RESEARCH PAPER?

Try first to get a general idea of the paper
- What problem is being solved?
- What are the main steps?
- How can I implement the method?, even though I do not understand why each step is performed the way it is performed?

Try to relate the method to other methods you know, and conceptually find similarities and differences.
HOW TO READ A RESEARCH PAPER?

In the first reading it may be a good idea to skip the related work.

Do not use dictionary to just look up the meaning of technical terms like

Try to understand each concept in isolation, and then integrate them to understand the whole paper.
DEEP LEARNING HAS BEEN DISRUPTIVE

Very Different Paradigm

Real Learning

Excellent results

Rapid Progress

Computer Vision is impacting other areas
DEEP LEARNING HAS BEEN DISRUPTIVE

Dramatic Increase in Number of
  Publications
  Attendance in Conferences
  Datasets
  Startups
  Academics moving to Industry
  Software platforms/libraries
    Café, Keras, Tensor Flow, Chainer, ..
  GPUs
  Tutorials, videos, online courses
COMPUTER VISION CONFERENCES AND JOURNALS

- **Conferences**
  - International Conference on Computer Vision (ICCV)
  - Computer Vision and Pattern Recognition (CVPR)
  - European Conference on Computer Vision (ECCV)

- **Journals**
  - IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)
  - International Journal of Computer Vision (IJCV)
  - Computer Vision and Image Understanding (CVIU)
MACHINE LEARNING CONFERENCES

- Neural Information Processing Systems (NIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
NIPS 2018 TICKETS SELL OUT IN LESS THAN 12 MINUTES

Faster than this year's Burning Man, but still not as fast as a Beyonce concert! Maybe next year.
CVPR Attendance
CVPR Papers
Deep learning has been disruptive

- Computer Vision is impacting other areas
  - Natural Language Understanding
  - Robotics
  - Speech Recognition
  - Audio/Sketches
  - Alpha Go
DEEP LEARNING HAS BEEN DISRUPTIVE

- Learning
  - Supervised
  - Semi-Supervised
  - Weakly-Supervised
  - Unsupervised
  - Self Supervised
  - Reinforcement
OUR RESEARCH
SOME OF OUR LAST YEAR’S PAPERS


Human Semantic Parsing for Person Re-identification

Mahdi M. Kalayeh, Emrah Basaran, Muhittin G"okmen, Mustafa E. Kamasak Mubarak Shah
CVPR 2018

http://crcv.ucf.edu/people/phd_students/mahdi/papers/CVPR18.pdf
Person Re-Identification

- Retrieving the images of a person from a large gallery
- Query and gallery images are captured by different cameras
- A cross-camera data association problem
Person Re-Identification
Challenges

- Illumination conditions
- Observable human body parts
- Perceived posture of the person
- Background clutter
- Occlusion
SPReID: Human Semantic Parsing for Person Re-identification
SPReID - Global Representation
SPReID - Human Semantic Parsing
SPReID – Foreground & Region Representations

Person Re-identification Backbone
Human Semantic Parsing

Foreground

Body Parts
SPReID Representation

- Concatenation of
  - Global representation
  - Foreground representation
  - Body Part representation
Experimental Results

- Datasets
  - Person Re-Identification
    - CUHK03
    - Market-1501
    - DukeMTMC-reID
  - Semantic parsing
    - Look into Person (LIP)
Human semantic parsing
DukeMTMC-reID
### Results – Duke

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (%)</th>
<th>rank-1</th>
<th>rank-5</th>
<th>rank-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basel. + LSRO</td>
<td>47.1</td>
<td>67.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Basel. + OIM</td>
<td>-</td>
<td>68.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zheng et. Al.</td>
<td>49.3</td>
<td>68.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ACRN</td>
<td>52.0</td>
<td>72.6</td>
<td>84.8</td>
<td>88.9</td>
</tr>
<tr>
<td>SVDNet</td>
<td>56.8</td>
<td>76.7</td>
<td>86.4</td>
<td>89.9</td>
</tr>
<tr>
<td>Chen et. Al.</td>
<td>60.6</td>
<td>79.2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**State-of-the-art**

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (%)</th>
<th>rank-1</th>
<th>rank-5</th>
<th>rank-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inception-V3\textsuperscript{ft}</td>
<td>63.27</td>
<td>80.48</td>
<td>88.78</td>
<td>91.65</td>
</tr>
<tr>
<td>ResNet-152\textsuperscript{ft}</td>
<td>67.02</td>
<td>83.26</td>
<td>90.93</td>
<td>92.95</td>
</tr>
<tr>
<td>combined</td>
<td>72.0</td>
<td>85.37</td>
<td>92.15</td>
<td>94.21</td>
</tr>
</tbody>
</table>

**Ours - without semantic parsing**

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (%)</th>
<th>rank-1</th>
<th>rank-5</th>
<th>rank-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPReID\textsuperscript{ft}</td>
<td>70.97</td>
<td>84.43</td>
<td>91.88</td>
<td>93.72</td>
</tr>
<tr>
<td>+ re-ranking</td>
<td>83.16</td>
<td>87.21</td>
<td>92.37</td>
<td>93.9</td>
</tr>
</tbody>
</table>

**Ours - with semantic parsing**

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (%)</th>
<th>rank-1</th>
<th>rank-5</th>
<th>rank-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-152\textsuperscript{ft} + SPReID\textsuperscript{ft}</td>
<td>73.34</td>
<td>85.95</td>
<td>92.95</td>
<td>94.52</td>
</tr>
<tr>
<td>+ re-ranking</td>
<td>84.99</td>
<td>88.96</td>
<td>93.27</td>
<td>94.75</td>
</tr>
</tbody>
</table>
Human Semantic Parsing for Person Re-identification

Mahdi M. Kalayeh, Emrah Basaran, Muhittin Gökmen, Mustafa E. Kamasak, Mubarak Shah
CVPR 2018
http://crcv.ucf.edu/people/phd_students/mahdi/papers/CVPR18.pdf
Contents

- Target Detection in WAMI
- Facial Attributes Detection
- Human Re-Identification
- Semantic Segmentation
- Human Action Localization
- Video Fill In The Blank
- Reading The Mind
- Anomaly Detection
Fully Convolutional Deep Neural Networks for Persistent Multi-Frame Multi-Object Detection in Wide Area Aerial Videos

Rodney LaLonde, Dong Zhang and Mubarak Shah
CVPR-2018
The Goal

Multiple Vehicle Detection

Red dots are the ground truth \((x,y)\) annotations.

(WPAFB 2009)
WPAF 2009 Dataset

- Covering an area of over **19.5 sq. km.**
- Frame rate of roughly **1.25 Hz.**
- Over **315 million** pixels per frame, with each pixel corresponding to roughly **1/4 meter.**
- Vehicle makes up only approximately 9 X18 pixels
  - **2.4 million** vehicles in 1,025 frames of video,
  - Over **2,000** in every frame.

**Eight AOIs**
- AOI-1 to AOI-4: 2,278 X2,278 size
- AOI 34 is 4,260 X2,604 . AOI 40 is 3,265X2,542 . AOI 41 is 3,207X 2,892
Video Frame Patch Creation
Ground Truth
Heat-maps

- Gaussian heat-maps have their colors inverted and $\sigma$ slightly increased for visualization purposes.
Deep Purely Convolutional Neural Network

Effective Receptive Field at Each Layer of the CNN
Supervised Learning

- 2D Convolutional Layer
- ReLU Layer
- Max Pooling Layer
- Dropout Layer (50%)
- Euclidean Loss Layer
- Solver Type
  - Adam

- Deep Learning Framework
  - Caffe
  - MATLAB interface
- GPU
  - 4 NVIDIA Titan X GPUs
- Training
  - Full training on a single GPU: 2 days
Experiments
Results AOI 34 Gaussian Heat-map
Results AOI 41
Gaussian Heat-map

Object Detection Results: AOI 41

precision vs recall graph
Results AOI 40
Gaussian Heat-map
Results AOI 41
Single Frame
## Comparison of $F_1$ Scores on Eight Crop and Aligned Sections of the WPAFB 2009 Dataset

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sommer et al. [23]</td>
</tr>
<tr>
<td>Shi [22]</td>
</tr>
<tr>
<td>Liang et al. [13]</td>
</tr>
<tr>
<td>Kent et al. [12]</td>
</tr>
<tr>
<td>Aeschliman et al. [2]</td>
</tr>
<tr>
<td>Pollard &amp; Antone (3-frame + N) [16]</td>
</tr>
<tr>
<td>Saleemi &amp; Shah [21]</td>
</tr>
<tr>
<td>Xiao et al. [25]</td>
</tr>
<tr>
<td>Keck et al. [11]</td>
</tr>
<tr>
<td>Reilly et al. [18]</td>
</tr>
<tr>
<td>Pollard &amp; Antone (IGMM) [16]</td>
</tr>
<tr>
<td>Teutsch &amp; Grinberg [24]</td>
</tr>
<tr>
<td>Prokaj &amp; Medioni [17]</td>
</tr>
</tbody>
</table>

**Proposed Multi-Frame**
Fully Convolutional Deep Neural Networks for Persistent Multi-Frame Multi-Object Detection in Wide Area Aerial Videos

Rodney LaLonde, Dong Zhang and Mubarak Shah
CVPR-2018
Real-World Anomaly Detection in Surveillance videos

Waqas Sultani, Chen Chen, Mubarak Shah

Computer Vision and Pattern Recognition (CVPR), 2018
Motivation

- Over 30 Millions cameras in US
- Over 4 Billions hours of videos per week
- Manual supervision is impossible
- Automatic Analysis is highly needed
Arson
Arrest
Abuse
Burglary
Assault
Stealing

Note: we fast play or trim some videos due to their long durations.
Anomaly Detection

- Signal an activity that deviates normal patterns.
- Anomalous events:
  - traffic accidents, crimes or illegal activities, etc
- Anomalous events rarely occur.
Our Approach

- Learn anomalies by exploiting both normal and anomalous videos.
- Avoid annotating the anomalous clips in training videos.
- Learn anomaly through the deep multiple instance ranking:
  - By leveraging weakly labeled training videos:
  - A video is normal or contains anomaly somewhere, but we do not know where.
Weakly labeled Crime Detection Framework

Ranking

\[ f(\mathcal{V}_a) > f(\mathcal{V}_n), \]

MIL Ranking

\[ \max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) > \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i), \]

Loss Function

\[ l(\mathcal{B}_a, \mathcal{B}_n) = \max(0, 1 - \max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) + \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i)) \]
Weakly labeled Crime Detection Framework
Qualitative Results
Ground Truth

Road Accident
Ground Truth

Anomaly Score

Frames

Shooting

Frames No = 001
Normal video
Quantitative Results
Comparison with stat-of-art anomaly detection methods
Comparison with state-of-art anomaly detection methods
Comparison with state-of-art anomaly detection methods

Comparison with state-of-art anomaly detection methods


Comparison with state-of-art anomaly detection methods


Real-World Anomaly Detection in Surveillance videos

Waqas Sultani, Chen Chen, Mubarak Shah

Computer Vision and Pattern Recognition (CVPR), 2018
Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds

Haroon Idrees, Muhammad Tayyab, Kishan Athrey, Dong Zhang, Somaya Al-Maadeed, Nasir Rajpoot and Mubarak Shah

Problem

- Estimate of number of people, per-pixel density and the location of each person’s head.
Applications

• Crowd management for safety and surveillance
  • Hajj stampedes
  • Beijing lantern festival stampede
• Political significance of rally or protest
• Volume of commuters
• Counting in the wild
Our Contributions

• Novel framework that simultaneously solves
  • Counting
  • Density map estimation and
  • Localization

• Largest annotated crowd data set so far, (in terms of number of annotations)
  • 1535 Images
  • 1.25 Million annotations
Challenges

- Perspective variation
- Occlusion
- Diverse scenes
- Low resolution
Composition Loss
Ideas

• Solving density estimation and localization as side tasks should help counting

• Starting with a coarse estimate of density and iteratively refining towards point localizations
Multiple Losses

- 56
- Count Loss
- Density Loss
- Localization Loss

Deep Neural Network
Gaussian Kernel

\[ D(x, f(\cdot)) = \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi f(\sigma_i)}} \exp\left( -\frac{(x - x_i)^2 + (y - y_i)^2}{2f(\sigma_i)^2} \right) \]

- Number of people in image
- Person Location
- Desired Kernel
- Variable kernel bandwidth
Multi-Level Density

- Image
Multi-Level Density

- Level 1
Multi-Level Density

- Level 2
Multi-Level Density

• Level 3
Multi-Level Density

• Level $\infty$
Network
Network

- Density prediction unit
- Training details
  - Adam solver
  - Step learning rate
  - 70 Epochs

<table>
<thead>
<tr>
<th>Layer</th>
<th>Output Size</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>512 × 28 × 28</td>
<td></td>
</tr>
<tr>
<td>Density Level 1</td>
<td>1 × 28 × 28</td>
<td>1 × 1 conv</td>
</tr>
<tr>
<td>Density Level 2</td>
<td>641 × 28 × 28</td>
<td>1 × 1 conv, 3 × 3 conv, × 4</td>
</tr>
<tr>
<td>Density Level ∞</td>
<td>771 × 28 × 28</td>
<td>1 × 1 conv, 3 × 3 conv, × 4</td>
</tr>
<tr>
<td></td>
<td>1 × 28 × 28</td>
<td>1 × 1 conv</td>
</tr>
</tbody>
</table>
Results
## Counting

<table>
<thead>
<tr>
<th>Method</th>
<th>C-MAE</th>
<th>C-NAE</th>
<th>C-MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idrees et al. [12]*</td>
<td>315</td>
<td>0.63</td>
<td>508</td>
</tr>
<tr>
<td>Resnet101 [30]*</td>
<td>190</td>
<td>0.50</td>
<td>277</td>
</tr>
<tr>
<td>Densenet201 [26]*</td>
<td>163</td>
<td>0.40</td>
<td>226</td>
</tr>
<tr>
<td>CMTL [29]</td>
<td>252</td>
<td>0.54</td>
<td>514</td>
</tr>
<tr>
<td>SwitchCNN [25]</td>
<td>228</td>
<td>0.44</td>
<td>445</td>
</tr>
<tr>
<td>MCNN [13]</td>
<td>277</td>
<td>0.55</td>
<td>426</td>
</tr>
<tr>
<td>Encoder-Decoder [28]</td>
<td>270</td>
<td>0.56</td>
<td>478</td>
</tr>
<tr>
<td><strong>Proposed</strong></td>
<td><strong>132</strong></td>
<td><strong>0.26</strong></td>
<td><strong>191</strong></td>
</tr>
</tbody>
</table>
Counting

• Best results

Ground Truth=236
Proposed=236

Ground Truth =1612
Proposed=1610
Counting

• Typical results

Ground Truth=556
Proposed=605

Ground Truth =2066
Proposed=2005
Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds

Haroon Idrees, Muhmmad Tayyab, Kishan Athrey, Dong Zhang, Somaya Al-Maadeed, Nasir Rajpoot and Mubarak Shah

Thank you