COURS deep Image

http://www-poleia.lip6.fr/~cord/teaching-multimedia/

Matthieu Cord
Sorbonne University
Computer Science - LIP6
Course Outline

2. Introduction to Neural Nets (NNs)
3. Introduction to Machine Learning: Risk, Classification, Datasets, benchmarks and evaluation, Linear classification (SVM)
4. Convolutional NNs
5. Large deep convnets
6. FCN, localization, visualization
7. Generative models with GANs
8. Generative models with conditional GANs
9. Segmentation
10. Challenges
COMPUTER VISION: WHERE ARE WE NOW?

Source (next slides): Cornell CV course
Deployed: depth cameras

https://realsense.intel.com/stereo/

Microsoft Kinect
Deployed: shape capture

*The Matrix* movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz
Deployed: Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Source: S. Seitz

Digit recognition, AT&T labs
http://www.research.att.com/~yann/

Automatic check processing
Deployed: Face detection

- Cameras now detect faces
  - Canon, Sony, Fuji, ...
Significant progress: Face Recognition
Significant progress: Recognizing objects

Mask R-CNN. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. ICCV 2017
Recognition-based product search

GrokStyle Visual Search Demo
Visual Search Solutions
for the Retail Industry
Challenges: Other imaging domains

**Fig.1: Glioma sub-regions.** Shown are image patches with the tumor sub-regions that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the BraTS IEEE TMI paper.)
Challenges: Integrating Vision and Action

Saurabh Gupta, James Davidson, Sergey Levine, Rahul Sukthankar, Jitendra Malik
CVPR 2017
Challenge: Visual Reasoning
VQA task: Why is this funny?

The picture above is funny.

Andrej Karpathy
Course Outline

2. Classification: Datasets, benchmarks and evaluation, Linear classification (SVM)
3. Use-case for BoVW
4. Deep (1) the basics
5. Deep (2) convolutional NNs
6. Deep (3) Large deep convnets
7. Deep (4) classification with localization
10. Deep (7): Generative models / GAN
11. Deep (8): Generative models with conditional GANs
12. Bayesian Deep Nets
Local feature detection and description

Points/Regions of Interest detection

One example: Corner detection (Harris corner detector)
Corner detection

- Main idea: Translating window should cause large differences in patch appearance
Corner Detection: Basic Idea

Recognize the type of point (flat, edge, corner) by looking through a small window W

- **"flat"** region: no change in all directions
- **"edge"**: no change along the edge direction
- **"corner"**: significant change in all directions

Corner detection op == Shifting a window in *any direction*, keep the ones that give *a large change* in intensity
Consider shifting the window $W$ by $(u, v)$

- how do the pixels in $W$ change?
- compare each pixel before and after by summing up the squared differences (SSD)
- this defines an SSD “error” $E(u,v)$:

$$E(u, v) = \sum_{(x,y)\in W} [I(x + u, y + v) - I(x, y)]^2$$

- We want $E(u,v)$ to be as high as possible for all $u$, $v$!
Taylor Series expansion of $I$:

$$I(x + u, y + v) = I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$

If the motion $(u,v)$ is small, then first order approximation is good

$$I(x + u, y + v) \approx I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

$$\approx I(x, y) + [I_x I_y] \begin{bmatrix} u \\ v \end{bmatrix}$$

shorthand: $I_x = \frac{\partial I}{\partial x}$

Plugging this into the formula on the previous slide...
Corner detection: the math

\[ E(u, v) = \sum_{(x,y) \in W} [I(x + u, y + v) - I(x, y)]^2 \]

\[ \approx \sum_{(x,y) \in W} [I(x, y) + I_x u + I_y v - I(x, y)]^2 \]

\[ E(u, v) \approx \sum_{(x,y) \in W} [I_x u + I_y v]^2 \]

\[ \approx A u^2 + 2B uv + C v^2 \]

\[ A = \sum_{(x,y) \in W} I_x^2 \quad B = \sum_{(x,y) \in W} I_x I_y \quad C = \sum_{(x,y) \in W} I_y^2 \]

\( E(u,v) \) is locally approximated as a quadratic error function
Interpreting the second moment matrix

\[ M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix} \]

\[ E(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \]

Recall that we want \( E(u,v) \) to be as large as possible for all \( u,v \)

What does this mean in terms of \( M \)?
\[ E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \]

\[
A = \sum_{(x,y) \in W} I_x^2 \\
B = \sum_{(x,y) \in W} I_x I_y \\
C = \sum_{(x,y) \in W} I_y^2 
\]

\[
M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\
M \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
E(u, v) = 0 \quad \forall u, v
\]

Flat patch:
\[
I_x = 0 \\
I_y = 0
\]
\[ E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \]

\[ A = \sum_{(x,y)\in W} I_x^2 \]

\[ B = \sum_{(x,y)\in W} I_x I_y \]

\[ C = \sum_{(x,y)\in W} I_y^2 \]

Vertical edge: \( I_y = 0 \)

\[ M = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \]

\[ M \begin{bmatrix} 0 \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]

\[ E(0, v) = 0 \quad \forall v \]
\[ E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \]

\[ A = \sum_{(x,y) \in W} I_x^2 \]
\[ B = \sum_{(x,y) \in W} I_x I_y \]
\[ C = \sum_{(x,y) \in W} I_y^2 \]

Horizontal edge: \( I_x = 0 \)

\[ M = \begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix} \]

\[ M \begin{bmatrix} u \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]

\[ E(u, 0) = 0 \quad \forall u \]
What about edges in arbitrary orientation?
\[ E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \]

\[ M \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow E(u, v) = 0 \]

\[ M \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow E(u, v) = 0 \]

Solutions to \( Mx = 0 \) are directions for which \( E \) is 0: window can slide in this direction without changing appearance.
\[ E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \]

Solutions to \( Mx = 0 \) are directions for which \( E \) is 0: window can slide in this direction without changing appearance.

For corners, we want no such directions to exist.
Eigenvalues and eigenvectors of $M$

- $Mx = 0 \Rightarrow Mx = \lambda x$: $x$ is an eigenvector of $M$ with eigenvalue 0.
- $M$ is $2 \times 2$, so it has 2 eigenvalues ($\lambda_{\text{max}}, \lambda_{\text{min}}$) with eigenvectors $(x_{\text{max}}, x_{\text{min}})$.
- $E(x_{\text{max}}) = x_{\text{max}}^T M x_{\text{max}} = \lambda_{\text{max}} \|x_{\text{max}}\|^2 = \lambda_{\text{max}}$ (eigenvectors have unit norm).
- $E(x_{\text{min}}) = x_{\text{min}}^T M x_{\text{min}} = \lambda_{\text{min}} \|x_{\text{min}}\|^2 = \lambda_{\text{min}}$
Corner detection: the math

How are $\lambda_{\text{max}}$, $x_{\text{max}}$, $\lambda_{\text{min}}$, and $x_{\text{min}}$ relevant for feature detection?

- Need a feature scoring function

Want $E(u,v)$ to be large for small shifts in all directions

- the minimum of $E(u,v)$ should be large, over all unit vectors $[u \ v]$
- this minimum is given by the smaller eigenvalue ($\lambda_{\text{min}}$) of $M$
Corner detection summary

Here’s what you do

- Compute the gradient at each point in the image
- Create the $M$ matrix from the entries in the gradient
- Compute the eigenvalues
- Find points with large response ($\lambda_{\text{min}} > \text{threshold}$)
- Choose those points where $\lambda_{\text{min}}$ is a local maximum as features
Corner detection summary

Here’s what you do

• Compute the gradient at each point in the image
• Create the $H$ matrix from the entries in the gradient
• Compute the eigenvalues
• Find points with large response ($\lambda_{\text{min}} > \text{threshold}$)
• Choose those points where $\lambda_{\text{min}}$ is a local maximum as features
The Harris operator

$\lambda_{\text{min}}$ is a variant of the “Harris operator” $R$ for feature detection:

$$R = \det(M) - \alpha \text{trace}(M)^2 = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2$$

- The trace is the sum of the diagonals, i.e., $\text{trace}(H) = h_{11} + h_{22}$
- Very similar to $\lambda_{\text{min}}$ but less expensive (no square root)
- Called the “Harris Corner Detector” or “Harris Operator”
The Harris operator

Harris operator

$\lambda_{\text{min}}$
Harris detector example
f value (red high, blue low)
Threshold (f > value)
Find local maxima of $f$
Harris features (in red)
Local feature detection

Looking for repeatability
Local feature description

Local description (always looking for invariance)

SIFT descriptors/features
Extension: Daisy

Circular gradient binning

SIFT

1 Ring 6 Segments 1 Ring 8 Segments
2 Rings 6 Segments 2 Rings 8 Segments

Picking the best DAISY, S. Winder, G. Hua, M. Brown, CVPR 09
Feature descriptors

• Expected properties?
  – Similar patches => close descriptors
  – Invariance (robustness) to geom. transformation: rotation, scale, view point, luminance, semantics? …
BoF: (First) Image representation

Sparse, at interest points

Dense, uniformly

Randomly

Multiple interest operators

Feature extraction

A bag of features BoF

© F-F. Li, E. Nowak, J. Sivic
BoF -- Image representation

- Image similarity based on matching of local features + voting
Applications to Image Retrieval

Target (if in)
Most similar to Q
+ infos: The Wedding at Cana -- Véronèse
Image Retrieval

• Context: Instance search (second example)
Advanced Understanding

Two pizzas sitting on top of a stove top oven
Image Understanding

• Focus of this course: recognition, classification and understanding

• Fundamental Pbs:
  – Image representation,
  – Data similarity,
  – Decision function

• Examples of applications:
  – face recognition
  – human action recognition in movies
  – Search Engines
  – Pattern recognition in remote sensing,
  – Medical imagery