Deep Learning for Computer Vision

Transitioning From Traditional Vision to Deep Learning

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Summarizing Topics So Far

Fundamental Operations

- **Convolution** is a unique operation
  - linear, shift-invariant
  - *Useful properties:* Commutative, Associative, Distributive (over addition)
- Forms basis of image operations and even modern-day neural networks working on images

Summarizing Topics So Far

Fundamental Operations

- **Convolution** is a unique operation
  - linear, shift-invariant
  - *Useful properties:* Commutative, Associative, Distributive (over addition)
- Forms basis of image operations and even modern-day neural networks working on images

Common Pipeline in Traditional Vision Tasks

- Extract corners or patches in images $\rightarrow$ Extract descriptors
- Use banks of filters, such as Steerable filters or Gabor filters
- Use descriptors for tasks such as retrieval, matching or classification

Traditional Vision: High-level Abstractions

Image-Level Understanding

- Going from low-level image understanding to aggregation of descriptors
- Banks of filters capture responses at different scales and orientations
- Histograms can be viewed as "*encoding*" and "*pooling*"
- Similarities to the human visual system

Traditional Vision: High-level Abstractions

Image-Level Understanding

- Going from low-level image understanding to aggregation of descriptors
- Banks of filters capture responses at different scales and orientations
- Histograms can be viewed as "*encoding*" and "*pooling*"
- Similarities to the human visual system

Local Features/Understanding

- Not all spatial regions important, depends on task (stereopsis, motion estimation, instance recognition compared to class recognition)
- Encoding makes features sparse
  - Many words in BoW have zero count
- Operators that detect local features can be viewed as "*convolution*" followed by some kind of "*competition*"

Traditional Vision: High-level Abstractions

Representing Images/Regions as Descriptors

- Learn descriptors/representations such that dot product is good enough for matching
- Some invariance to geometric transformations, designed or learned in certain cases

Traditional Vision: High-level Abstractions

Representing Images/Regions as Descriptors

- Learn descriptors/representations such that dot product is good enough for matching
- Some invariance to geometric transformations, designed or learned in certain cases

Moving on to Deep Learning...

Although not by design, Deep Learning seems to build on some of the above principles, but in a learnable manner...we will see soon