Introduction

Arthur J. Redfern
axr180074@utdallas.edu
Aug 20, 2018

---

0 Outline

Previous
1. None

Current
1. Welcome
2. Class
3. Logistics
4. Expectations

Next
1. Linear algebra

1 Welcome

1.1 Hello

Welcome to the special topics in computer science convolutional neural networks course

Brief introduction
- The goal of this lecture is to create a framework for understanding the design of the whole class and how the semester will progress

So this introduction lecture will be a little different than the normal lectures
- More words
- More stories
- Less math
- Then on Wed we'll start the more traditional lecture format

1.2 Me

Grew up a little south of Richmond, Virginia

BS in EE from University of Virginia

PhD in ECE from Georgia Tech

Moved to Dallas to work at Texas Instruments
- Physical layer communication system design
- Signal processing for analog systems
- Machine learning

Currently I manage a machine learning lab in the TI Embedded Processors organization
- Work on algorithms, software and hardware for different applications
- This class will cover much of the same (that's not an accident)

Live in Plano, Texas

2 Class

Official objectives (from the syllabus)
- Ability to understand, design and train convolutional neural networks
- Ability to create software for mapping convolutional neural network designs to hardware
- Ability to specify hardware for convolutional neural network optimized data movement and compute
- Ability to evaluate convolutional neural network performance
- Ability to apply convolutional neural networks to applications including vision, speech, language and games

2.1 Theory And Application

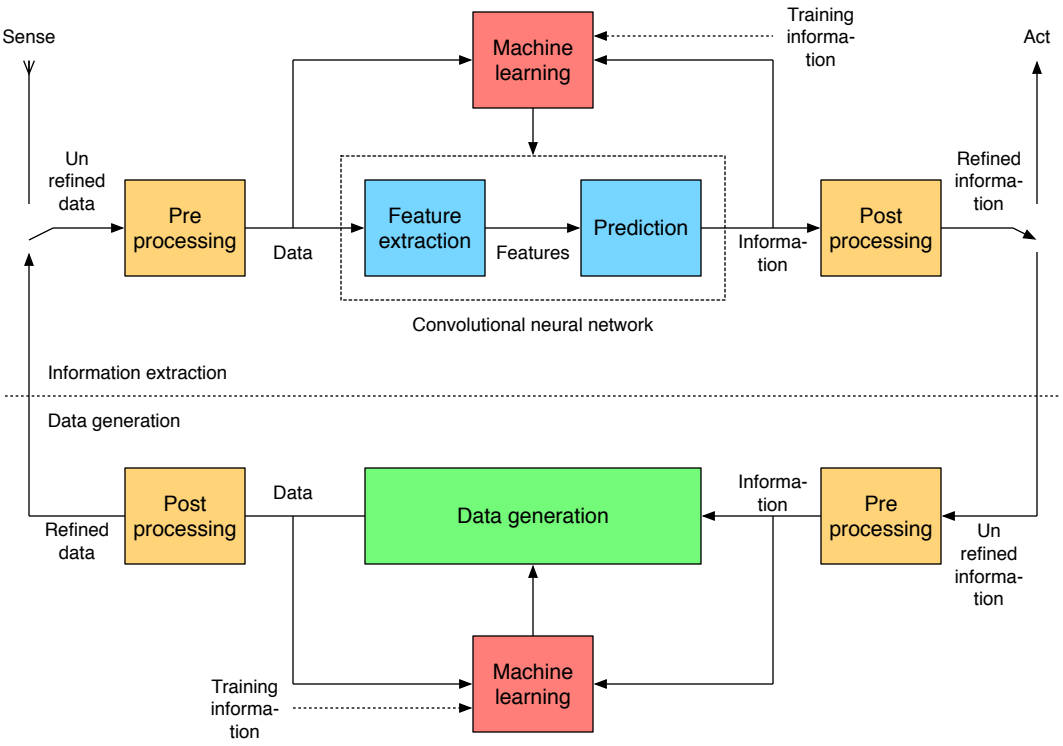

Figure: Information extraction and data generation

2.1.1 Information Extraction

Information extraction

- Mapping of data to information
- Typical focus of a machine learning class (this class is no exception)
- Examples: images to labels, sounds to words, …

Generic framework for information extraction

- Flow: pre processing, feature extraction, prediction, post processing
  - Transformation from un refined data space to data space to feature space to information space to refined information space
- Pre processing
  - Make feature extraction easier
  - Data cleaning, dimensionality reduction, …
  - Frequently uses application specific side information
- Feature extraction
  - Make prediction easier
  - Hand engineered or learned
- Prediction
  - Classification (discrete)

    - Regression (continuous)
- Post processing
    - Clean up predictions
    - Frequently uses application specific side information

Definitions (not Webster quality)

- Intelligence is the ability to acquire and apply knowledge
    - Artificial intelligence is intelligence exhibited by algorithms
- Learning is the acquisition of knowledge from experience
    - Machine learning is learning from data (experience) applied to an algorithm such that it exhibits artificial intelligence
    - Deep learning is machine learning applied to a deep structure

How CNNs fit in

- CNNs are deep structures trained using deep learning to exhibit artificial intelligence
- Perform both feature extraction and prediction
- This semester we'll look at design and training CNNs
    - Tail body head approach to design
    - Supervised learning using back propagation and a variant of stochastic gradient descent for training
- Neural networks are universal approximators
    - Can work on all sorts of problems

Why now

- Data
- Compute
- Better network designs
- Better training algorithms
- Snowball of success

Why this is important in practice

- Instrumentation of everything (objects, people, spaces)
- Analysis of data, decision optimization and heuristic replacement
- Successful applications
    - Vision, speech, language, games, …

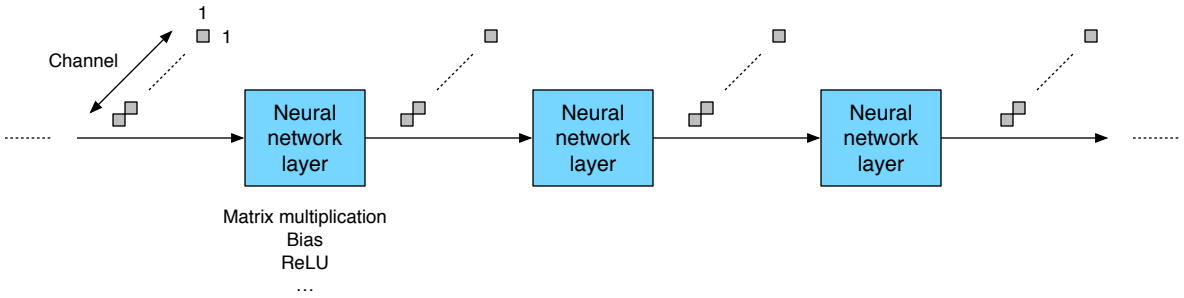

Figure: Generic neural network; layers take input vectors and produce output vectors of size channel x 1 x 1

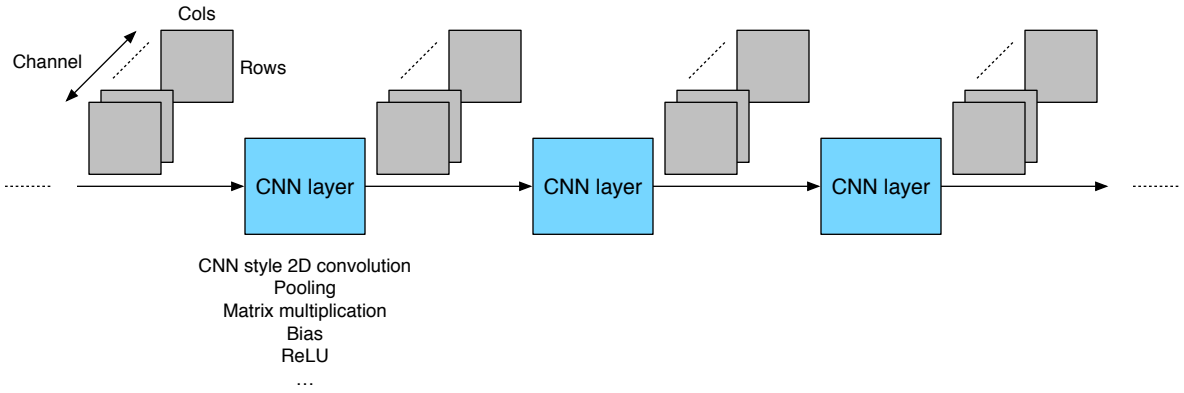

Figure: Generic convolutional neural network; layers take input feature maps and produce output feature maps of size channel x rows x cols

2.1.2 Data Generation

There's another problem that's the mirror of information extraction: data generation

Data generation

- Mapping of information to data
- Not the typical focus of a machine learning class (this class is no exception)
    - But we'll talk about it a little
- Examples: labels to images, words to sounds, …

Data generation problems are the complement of information extraction problems

- Can be used as an alternative input to information extraction
    - Synthetic vs natural options
    - Synthetic is helpful when natural is difficult
- Can also use CNNs as a component of data generation

2.2 Background And Implementation

So far

- What we've talked about will be discussed in the (2) theory and (4) application parts of the course
- 2 other key parts of the course are (1) background and (3) implementation

How background maps onto the information extraction and data generation figure

- Linear algebra
    - CNN style 2D convolution
    - Fully connected layers
    - Computation strategies later used in hardware
    - …
- Calculus
    - Back propagation for training
    - Limits and strategy for universal approximation
    - …
- Probability
    - Training extracts knowledge (information) from the training set, testing extracts information from the input given past knowledge (information) extracted from training
    - Initializing coefficients, batch norm, comparing output pdf to target pdf
    - Pdf of the output of a FFT or matrix multiplication
    - Compression
    - …

How implementation maps onto the information extraction and data generation figures

- Software
    - High level application specification
    - Low level software runtime
    - Bridge between network specification and hardware implementation
- Hardware
    - Memory, data movement and compute to run the software that runs the network
- We'll consider the co design of software and hardware using a low level graph framework to unite them both

2.3 This Class In Context

Not comprehensive of all convolutional neural network information

- What is presented in lecture is less than what is in the references

- What is in the references is less than all information on the topic
- A role of a professor is a guide through information and I'll attempt to provide that in this class
- But realize that there's a lot more useful information out there that could later be critical to you depending on the specifics of your interests
- So an unofficial goal of this class (perhaps more important than any of the official goals) is to help you learn how to learn in this field

3 Logistics

3.1 Grades

4 parts of the semester

- Background
- Theory
- Implementation
- Application

4 components of the grade

- 25% background and theory test
- 25% implementation tool
- 25% project
- 25% homework

Note

- No final

3.2 Background And Theory Test

Covers

- Background: linear algebra, calculus and probability
- Theory: machine learning and convolutional neural networks

Format

- In class closed book
- Pencil and paper only

Goals

- Make sure that key concepts are well understood from each chapter

3.3 Implementation Tool

Tools are critical to making practical progress in CNNs

Examples of tools

- High level network design
  - Creating networks from building blocks
  - Visualization
- High level network training
  - Monitoring and optimization
  - Transformation
  - Quantization
- Low level graph software runtime
  - High level to low level mapping
  - Low level compilation
  - Performance prediction

Format

- Designed to complement the implementation part
- Basic idea is to create a tool and also give a quick ~ 1 min demo in class
- Details closer to time

3.4 Project

Can be on anything related to the class

- Requirements
  - I approve
  - Can be in theory, implementation or application
- Sources
  - Ideas I provide (I'll give many)
  - Ideas related to your thesis work
  - Ideas related to your hobbies
  - Ideas out of nowhere you find interesting

Format

- Work in groups of 1 – 3 (average size 2 for ~ 30 projects)
- 4 min presentation / demonstration ((4 + 1) x 30 = 150 min so we'll use ~ 2.5 classes at the end)

Expectations

- Make it meaningful
- I want to be impressed

3.5 Homework

Basic strategy is to assign on Wed, due the following Wed

- Most weeks

Goal is to make it useful for understanding but not tedious

- Some overlap of key ideas in class that need quiet time and thought
- Some new ideas not covered in class
- Some fun (reading, videos / movies)

3.6 Practice And Review Lectures

I want people to get setup as early as possible with a high level framework

- Pick 1 (I don't care): PyTorch, TensorFlow, Caffe, …

More uses of these lectures

- Review previous material
- Brainstorm ideas for projects etc.

3.7 Class Web Site

Follow this page, it will be updated on a regular basis

GitHub https://github.com/arthurredfern/UT-Dallas-CS-6301-CNNs

- Syllabus Course syllabus, will update plan as necessary
- Lectures Post after class
- References Book (early draft form, also see references), links to others
- Homework (Usually) post on Wed due next Wed
- Tests Will eventually contain theory test
- Tools Will eventually contain implementation tool information
- Projects Will eventually contain project information
- Code Update as necessary

4 Expectations

4.1 Of Me

My best every class

My opinions
- It’s a special topics class

I speak to adults like adults
- It's a grad class
- I want to be precise
- But I don't want to make things unnecessarily complicated

A logically laid out plan for both the whole course and individual lectures

A willingness to modify the plan as needed
- It's a new class

I don't have a perfect picture with respect to how long different topics will take
- If I go short in a lecture we'll figure out a way to make use of the extra time
    - I'll answer questions
    - We'll do some practical items
    - We'll discuss project and paper ideas
- If I run out of time in a lecture we'll make it up in the next lecture or have reading for homework
- Structured the implementation and application sections to give a buffer we can shrink or expand

I don't have a perfect knowledge of what you do and don't know
- The course covers a lot
- It's unlikely that you have a perfect background in everything
- That's ok
- Part of the purpose of the course will be to fill in those gaps
- I'll help via the structuring of the material

4.2 Of Students

Honesty
- In your work
- In your interactions with other students
- In your interactions with me

Hard work
- Nothing meaningful in life is easy
- This won't be an exception

Correct me if I'm wrong

- Politely

Friendly environment

- Great to shine as an individual through individual accomplishments
- Great to shine by helping others shine
    - Characteristic of a leader

Be engaged

- Ask questions freely

I would like to learn everyone's name

- Help me out and say it when you ask a question