Deep Learning for Natural Language Processing

Cornelia Caragea

Computer Science

University of Illinois at Chicago

Credits for slides: Chris Manning, Richard Socher

Why Deep Learning for NLP?

Today

Lecture Structure:

- What is NLP?
- Why deep learning for NLP?

What is Natural Language Processing?

Natural language processing (NLP) is a field at the intersection of computer science, artificial intelligence, and linguistics.

- It is an important technology due to the large and growing amount of online text that needs to be understood in order to get the enormous value out of it.

Recent Collections[1]

- Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased)
- Common Crawl (42B tokens, 1.9M vocab, uncased)
- Common Crawl (840B tokens, 2.2M vocab, cased)
- Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased)

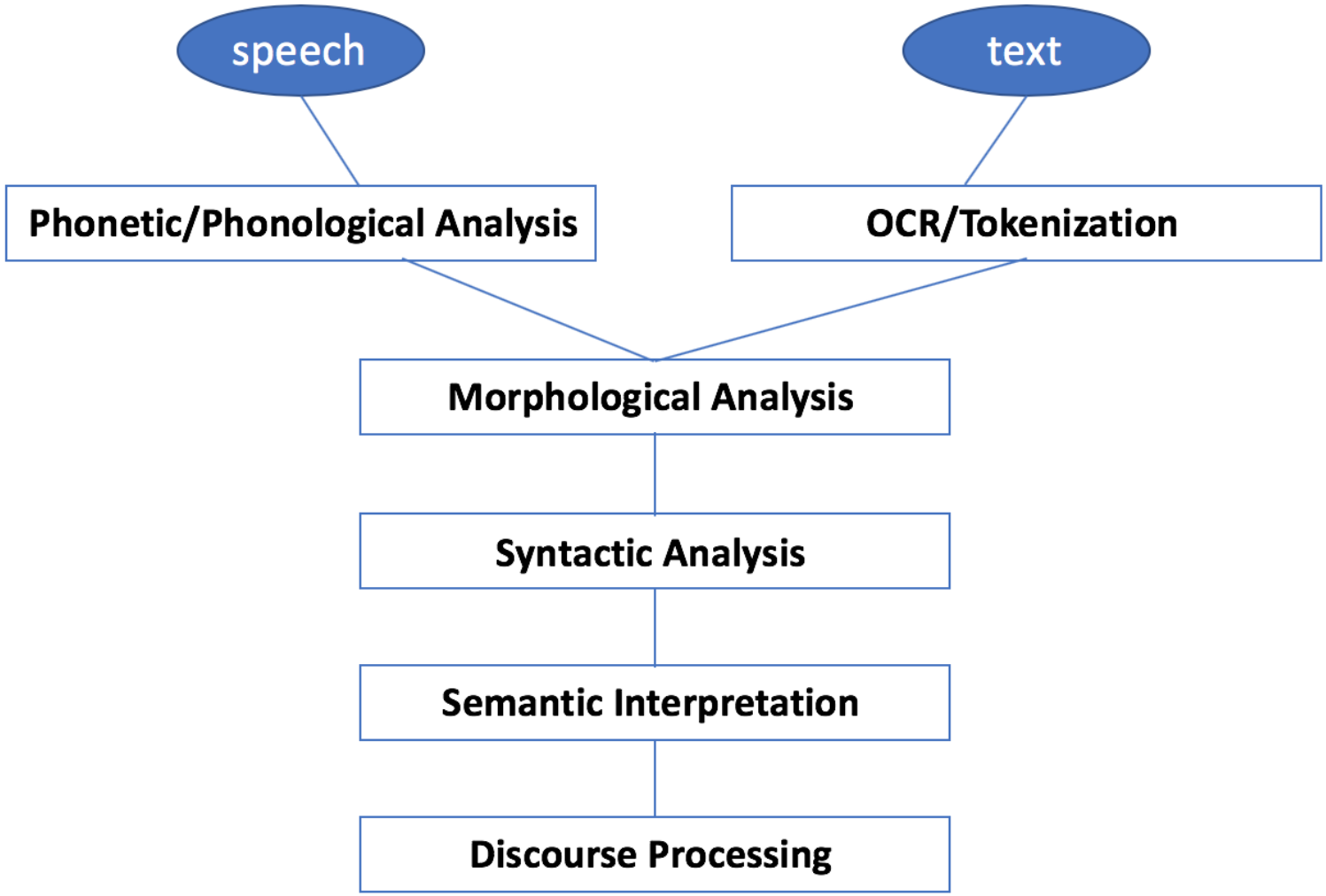
[1] Source: https://nlp.stanford.edu/projects/glove/

The Goal of Natural Language Processing

- For computers to process or "understand" natural language in order to perform tasks that are useful, e.g.,
  - Sentiment Analysis
  - Question Answering
  - Discourse Analysis

- Challenges:
  - Computers do not understand language the same way we understand language.
  - Fully **understanding and representing** the **meaning** of language (or even defining it) is an illusive goal.

NLP Levels


speech
text
Phonetic/Phonological Analysis
OCR/Tokenization
Morphological Analysis
Syntactic Analysis
Semantic Interpretation
Discourse Processing

Examples of Practical NLP Applications

- Applications range from simple to complex

- Spell checking
- Keyword search, Search engines (e.g, Google)
- Part-of-Speech Tagging
- Information Extraction
- Named Entity Recognition
- Keyphrase/Relation Extraction
- Lexical Semantics
- Sentiment Analysis
- Emotion Detection
- Text Mining
- Summarization and Generation
- Neural Machine Translation
- Semantic Parsing
- Content Analysis
- ... from Noisy User-Generated Data
- Social Media and Computational Social Science
- Speech Recognition
- Phonology, Morphology and Word Segmentation
- Dialog and Interactive Systems
- Question Answering
- ... and many more

NLP in Industry

- Search (written and spoken)
- Online advertisement
- Automated translation
- Sentiment analysis for marketing or finance/trading
- Speech recognition
- Automating customer support

3/18/11 at 4:00 PM | 17 Comments

Mentions of the Name 'Anne Hathaway' May Drive Berkshire Hathaway Stock

By Patrick Huguenin

The Huffington Post recently pointed

Why is NLP Hard?

- Complexity in representing, learning and using linguistic/situational/world/visual knowledge.

- Example: Jane hit June and then **she** [fell/ran].

- Ambiguity: "I made her duck."

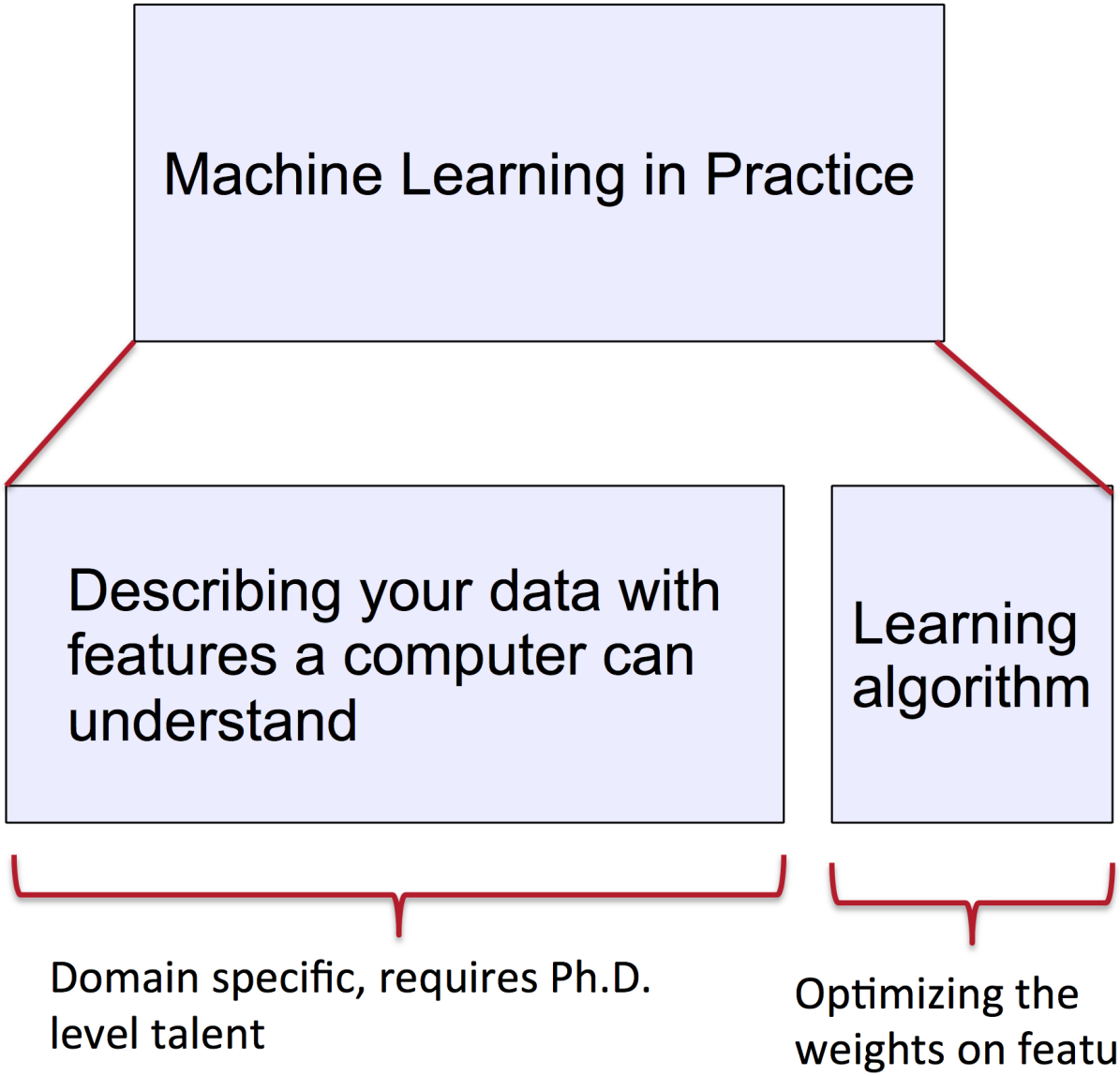
What is Deep Learning (DL)?

- Deep learning is a subfield of machine learning
- Most machine learning methods work well because of human-designed representations and input features
  - For example, for finding named entities like locations and organization names, the following features have been designed:

Feature	**NER**
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

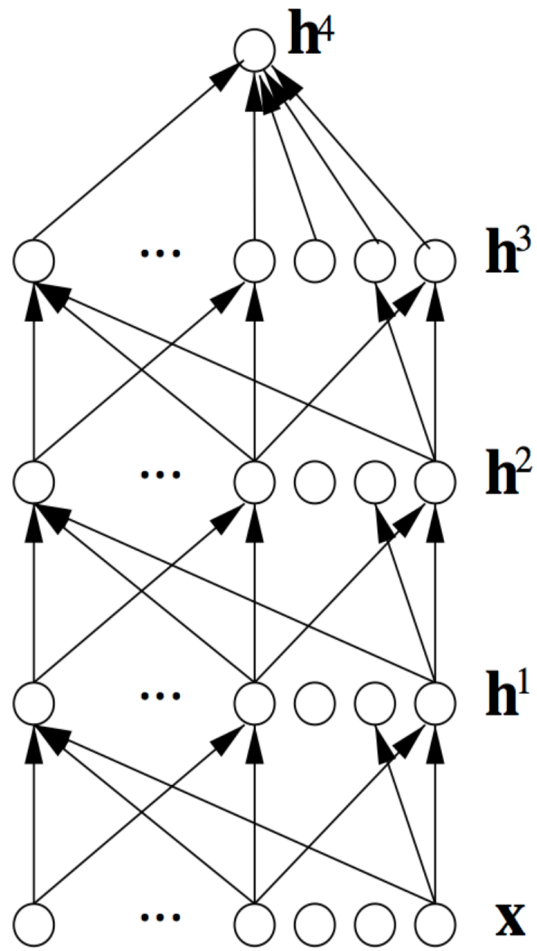
- Machine learning becomes just optimizing weights to best make a final prediction.

Machine Learning vs. Deep Learning



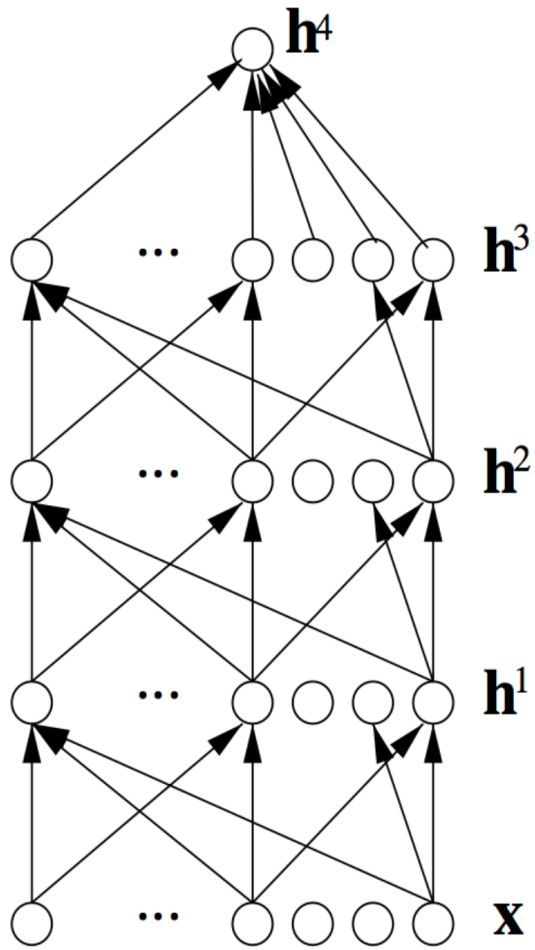
What is Deep Learning (DL)?

- **Representation learning** attempts to automatically learn good features or representations
- **Deep learning** algorithms attempt to learn (multiple levels of) representation and an output
- From "raw" inputs **x** (e.g. words)



What is Deep Learning (DL)?

- **Representation learning** attempts to automatically learn good features or representations
- **Deep learning** algorithms attempt to learn (multiple levels of) representation and an output
- From "raw" inputs **x** (e.g. words)



  - We will focus on different kinds of **neural networks**
  - The dominant model family inside deep learning

Reasons for Exploring Deep Learning

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- **Learned Features** are easy to adapt, fast to learn
- Deep learning provides a very flexible, (almost?) universal, learnable framework for **representing** world, visual and linguistic information.
- Deep learning can learn **unsupervised** (from raw text) and **supervised** (with specific labels like positive/negative).

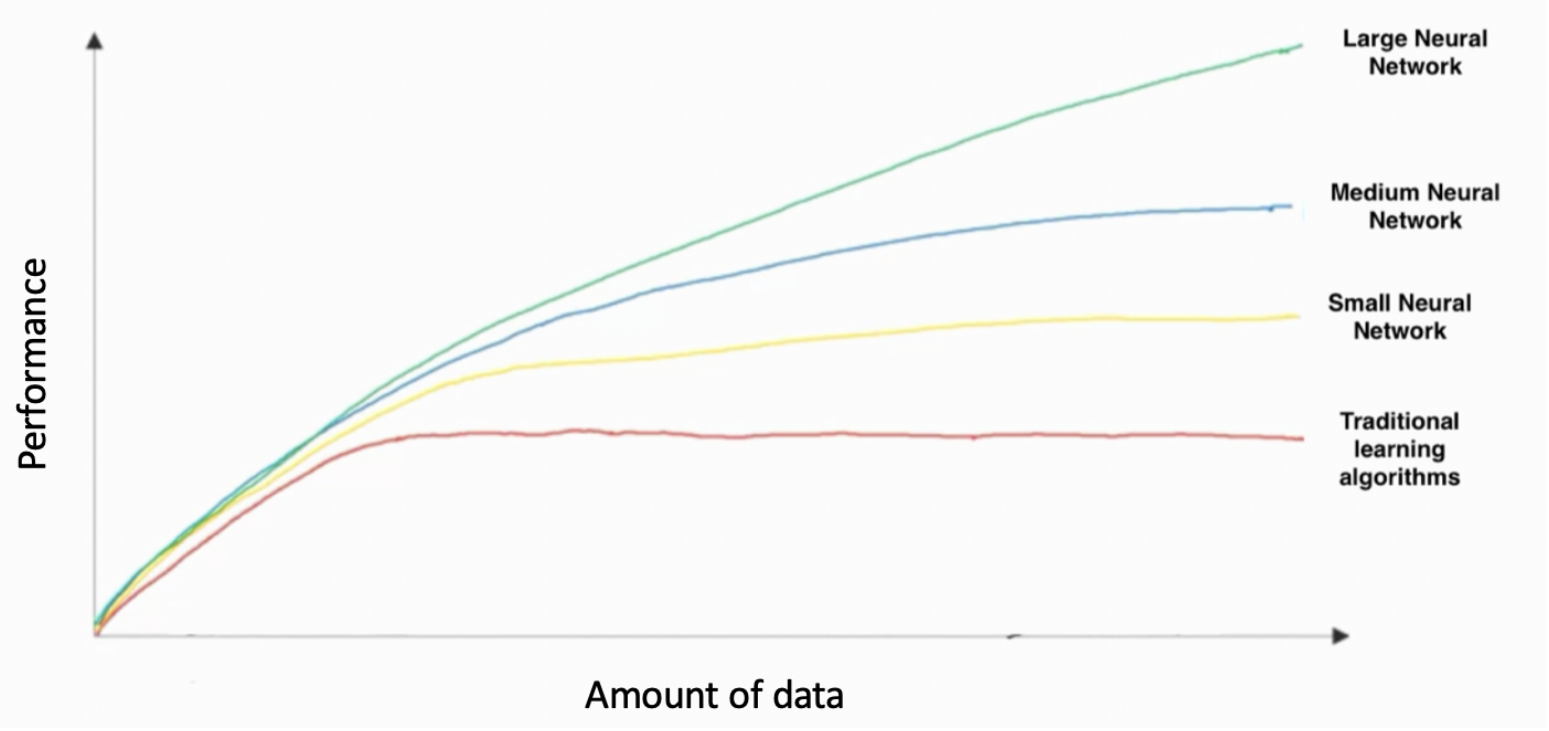
Reasons for Exploring Deep Learning

- In 2006 **deep learning techniques** started to outperform other machine learning techniques. Why now?

- DL techniques benefit more from a lot of data
- Faster machines and multicore CPU/GPU help DL
- New models, algorithms, ideas

$\rightarrow$ **Improved performance** (first in speech and vision, then NLP).

More Data Improves Performance



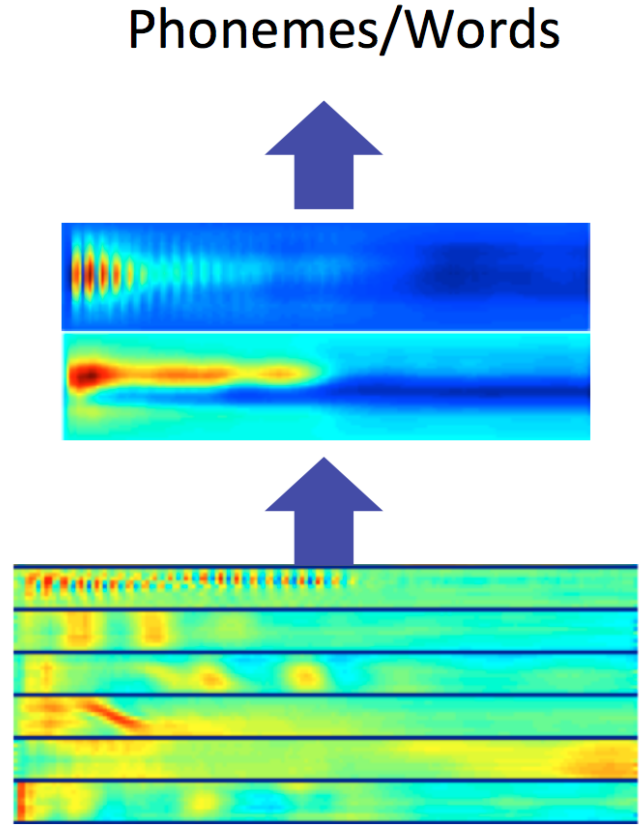
Credit: Andrew Ng

Deep Learning for Speech

- The first breakthrough results of "deep learning" on large datasets happened in speech recognition

- Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition Dahl et al. (2010)

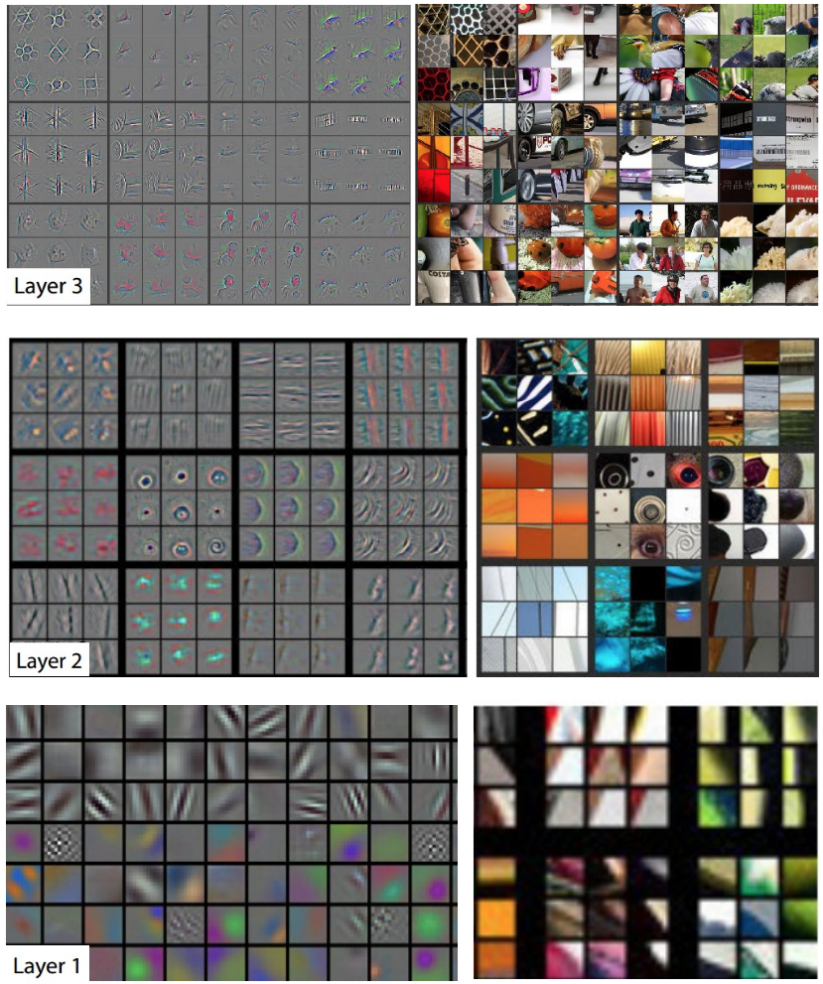
Acoustic model	Recog \ WER	RT03S FSH	Hub5 SWB
Traditional features	1-pass –adapt	**27.4**	**23.6**
Deep Learning	1-pass –adapt	**18.5** (–33%)	**16.1** (–32%)



Speech Recognition on Smart Phones or Alexa

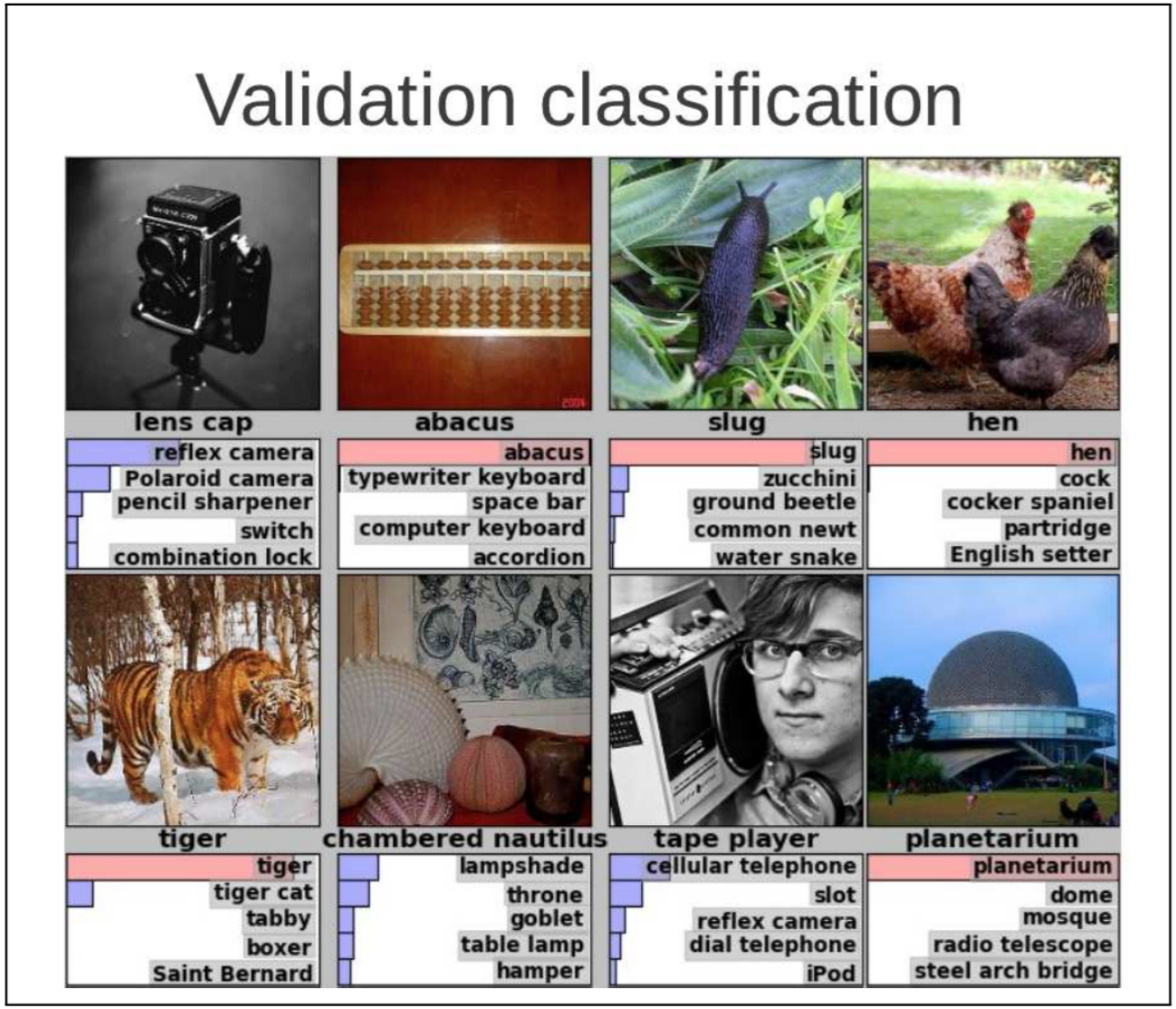
Deep Learning for Computer Vision

- Most deep learning groups have focused (until recently) on computer vision
- Break through paper: ImageNet Classification with Deep Convolutional Neural Networks by Krizhevsky et al. (2012)



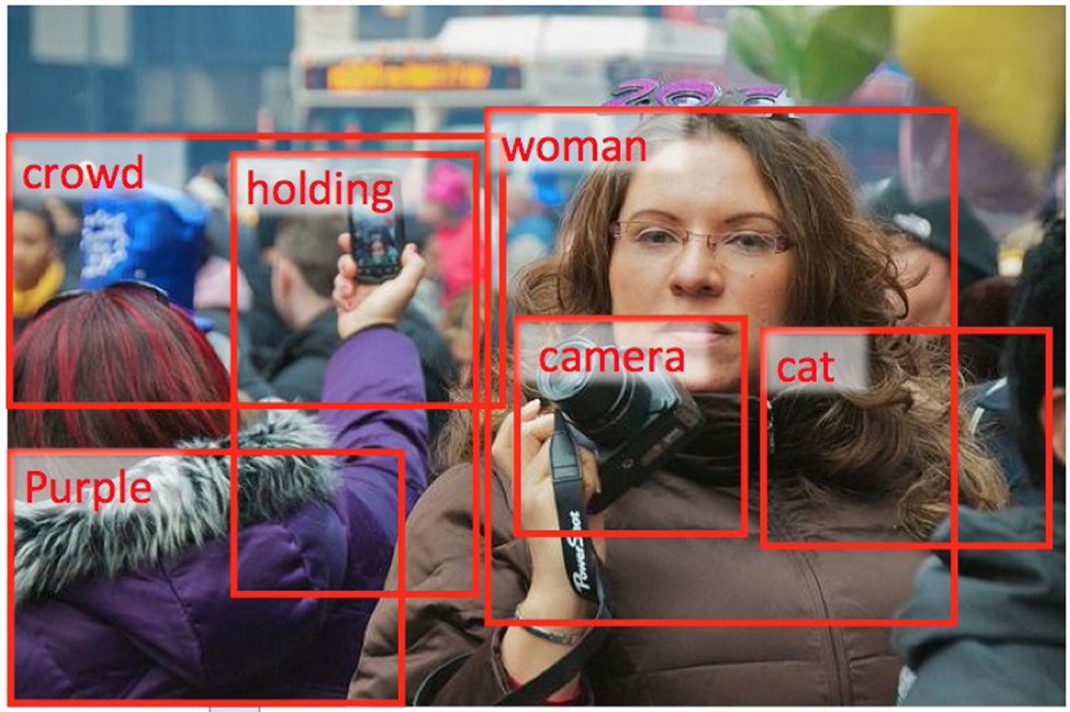
Zeiler and Fergus (2013)

ImageNet Vision Challenge



Mapping Images to Text

Image Caption Generator

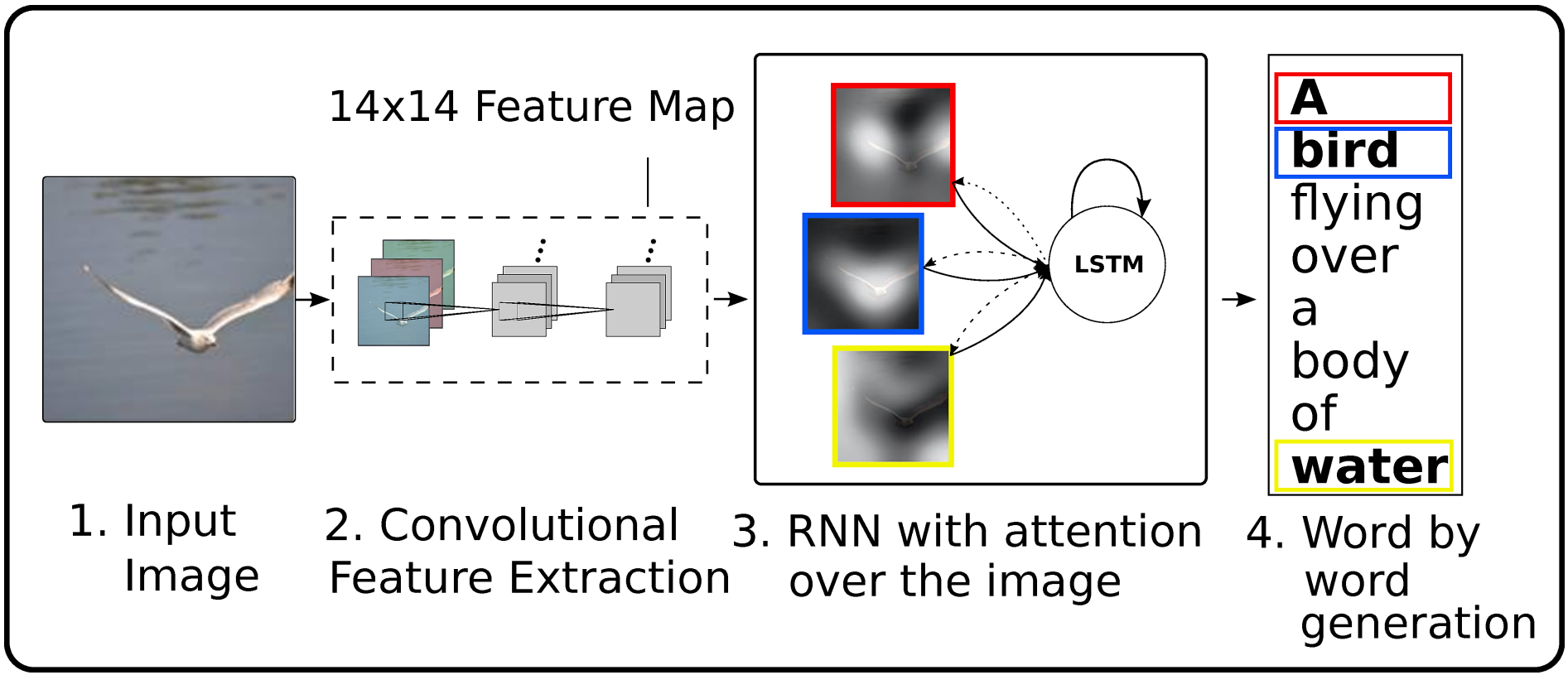


1. detect words → woman, crowd, cat, camera, holding, purple

2. generate sentences → A purple camera with a woman. A woman holding a camera in a crowd. ... A woman holding a cat.

3. re-rank sentences → #1 A woman holding a camera in a crowd.

Image Caption Generator

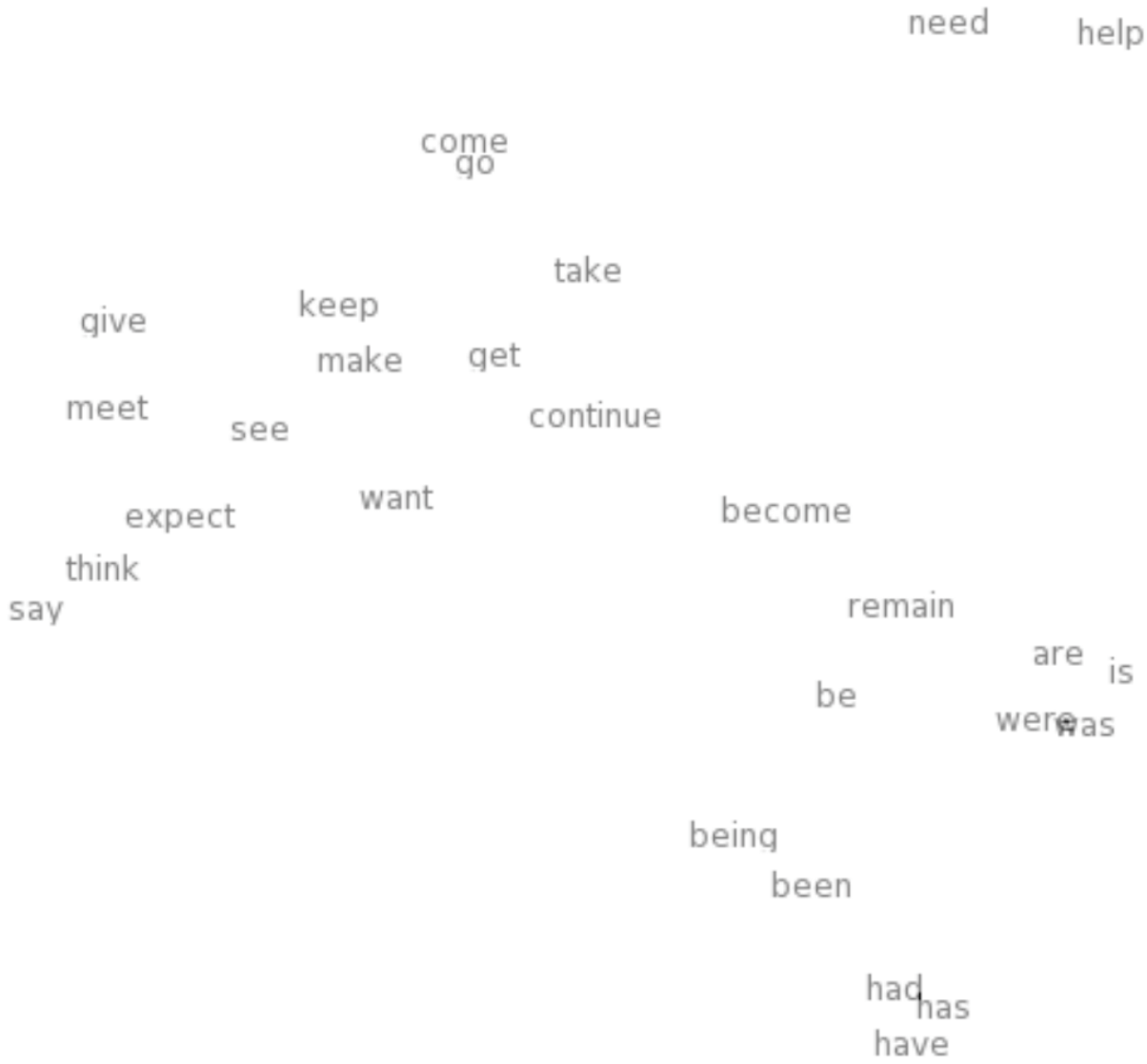


Deep Learning + NLP = Deep NLP

- Combine ideas and goals of NLP and use representation learning and deep learning methods to solve all kinds of different problems.

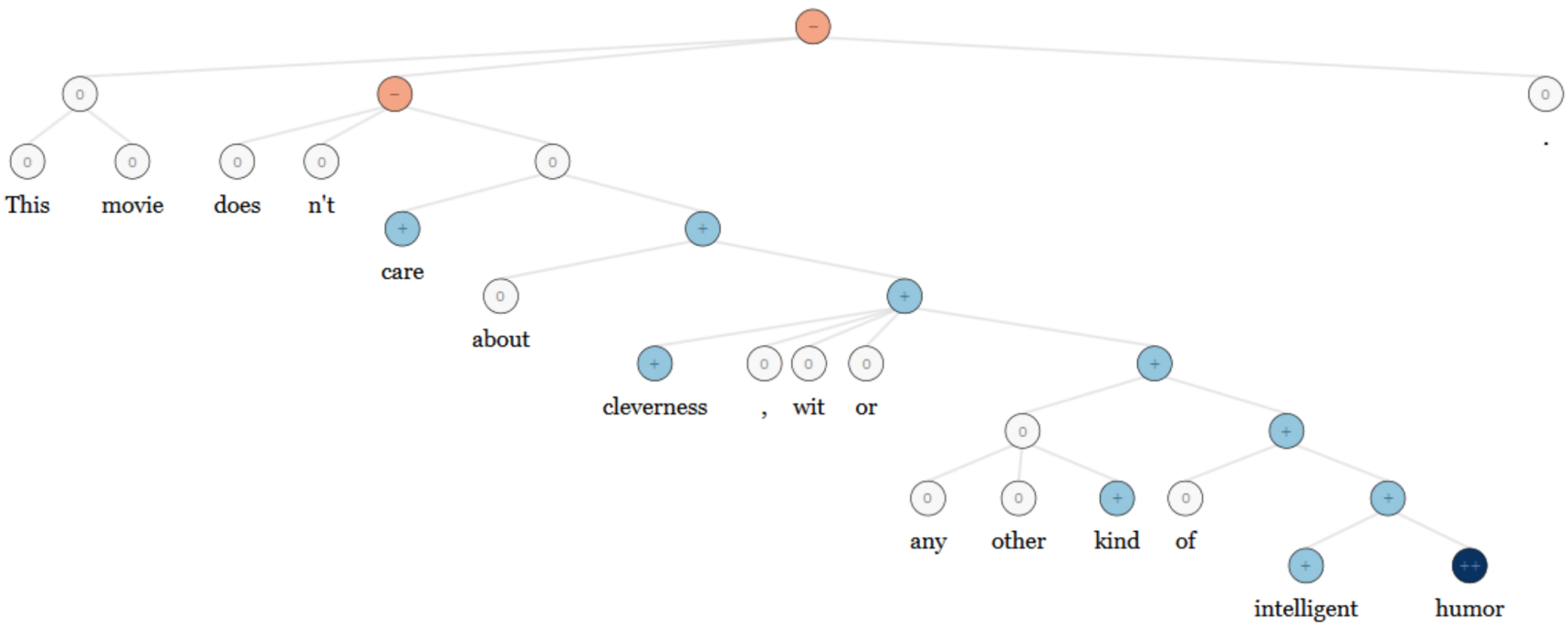
- Several big improvements in recent years across different NLP
  - **levels:** speech, morphology, syntax, semantics
  - **applications:** machine translation, sentiment analysis and question answering

Neural Word Vectors - Visualization



NLP Applications: Sentiment Analysis

- Traditional: Curated sentiment dictionaries combined with either bag-of-words representations (ignoring word order) or hand-designed negation features (ain't gonna capture everything)
- Same deep learning model that was used for morphology, syntax and logical semantics can be used → RecursiveNN



IBM Jeopardy Quiz Program vs. QA with RecursiveNNs

What Is This Course About?

- Understanding the basics of deep learning (DL)
- Using pre-trained or training your own DL models
- Applying DL models

$\cdots$ all to textual data!

Summary

We discussed:

- What is NLP?
- Why deep learning for NLP?