Information Extraction

Lecture 11 – Event Extraction and Multimodal Extraction

CIS, LMU München

Winter Semester 2017-2018

Prof. Dr. Alexander Fraser, CIS

Klausur

- 14.02 um 16:15

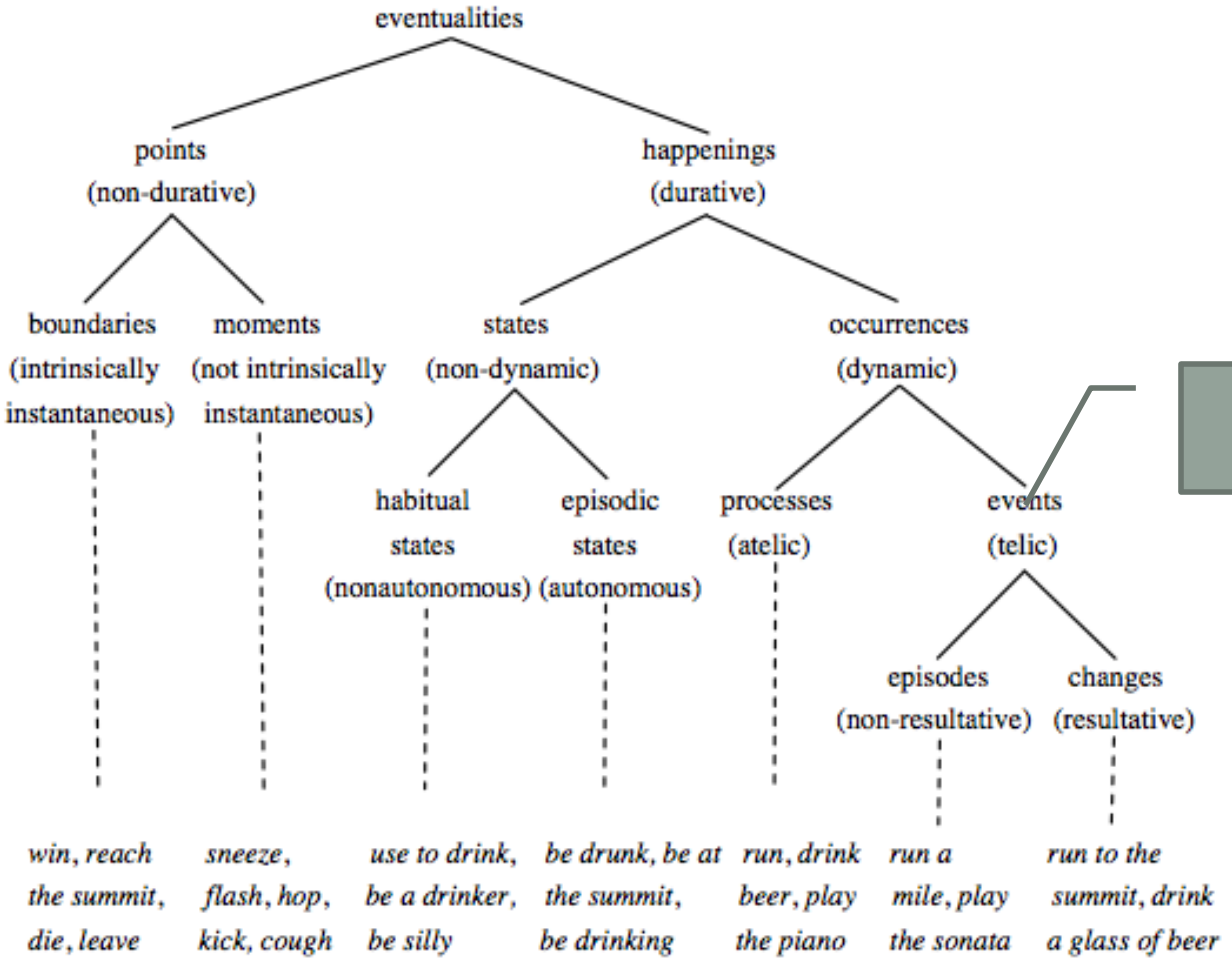
- Anmelden in LSF!
  - Einmal Seminar
  - Einmal Vorlesung

Event Extraction

- We'll now discuss event extraction, as defined in state-of-the-art statistical systems
  - This is an extension of the ideas in relation extraction (as discussed by Matthias) to events
  - Event extraction also offers a good opportunity to think about cross-sentence and cross-document extraction
  - The lecture on Ontologies and Open IE will be next week
- Later in this lecture we'll briefly discuss multimodal extraction (speech, images, etc)
  - Just to give a basic idea about what is possible

General Event Definition

- An Event is a specific occurrence involving participants.
- An Event is something that happens.
- An Event can frequently be described as a change of state.



Most of current NLP work focused on this

Chart from (Dölling, 2011)

Slide from Heng Ji

Event Mention Extraction: Task

- An event is specific occurrence that implies a change of states
- **event trigger**: the main word which most clearly expresses an event occurrence
- **event arguments**: the mentions that are involved in an event (participants)
- **event mention**: a phrase or sentence within which an event is described, including trigger and arguments
- Automatic Content Extraction defined 8 types of events, with 33 subtypes

Argument, role=victim *trigger*

ACE event type/subtype	*Event Mention Example*
Life/Die	**Kurt Schork died** in **Sierra Leone yesterday**
^	
Transaction/Transfer	**GM sold the company** in **Nov 1998** to **LLC**
Movement/Transport	**Homeless people** have been **moved** to **schools**
Business/Start-Org	**Schweitzer founded a hospital** in **1913**
Conflict/Attack	the **attack** on **Gaza** killed **13**
Contact/Meet	**Arafat's cabinet met** for **4 hours**
Personnel/Start-Position	**She** later **recruited the nursing student**
Justice/Arrest	**Faison** was wrongly **arrested** on suspicion of murder

Slide from Heng Ji

Supervised Event Mention Extraction: Methods

- Staged classifiers
  - Trigger Classifier
    - to distinguish event instances from non-events, to classify event instances by type
  - Argument Classifier
    - to distinguish arguments from non-arguments
  - Role Classifier
    - to classify arguments by argument role
  - Reportable-Event Classifier
    - to determine whether there is a reportable event instance
- Can choose any supervised learning methods such as MaxEnt and SVMs

(Ji and Grishman, 2008)

Slide from Heng Ji

Typical Event Mention Extraction Features

- Trigger Labeling
  - Lexical
    - Tokens and POS tags of candidate trigger and context words
  - Dictionaries
    - Trigger list, synonym gazetteers
  - Syntactic
    - the depth of the trigger in the parse tree
    - the path from the node of the trigger to the root in the parse tree
    - the phrase structure expanded by the parent node of the trigger
    - the phrase type of the trigger
  - Entity
    - the entity type of the syntactically nearest entity to the trigger in the parse tree
    - the entity type of the physically nearest entity to the trigger in the sentence

- Argument Labeling
  - Event type and trigger
    - Trigger tokens
    - Event type and subtype
  - Entity
    - Entity type and subtype
    - Head word of the entity mention
  - Context
    - Context words of the argument candidate
  - Syntactic
    - the phrase structure expanding the parent of the trigger
    - the relative position of the entity regarding to the trigger (before or after)
    - the minimal path from the entity to the trigger
    - the shortest length from the entity to the trigger in the parse tree

(Chen and Ji, 2009)

Slide from Heng Ji

Why Trigger Labeling is so Hard?

- A suicide bomber detonated explosives at the entrance to a crowded
- medical teams carting away dozens of wounded victims
- dozens of Israeli tanks advanced into the northern Gaza Strip
- Many nouns such as “death”, “deaths”, “blast”, “injuries” are missing

Slide from Heng Ji

Why Argument Labeling is so Hard?

- Two 13-year-old children were among those killed in the Haifa bus bombing, Israeli public radio said, adding that most of the victims were youngsters
- Fifteen people were killed and more than 30 wounded Wednesday as a suicide bomber blew himself up on a student bus in the northern town of Haifa
- Two 13-year-old children were among those killed in the Haifa bus bombing

Slide from Heng Ji

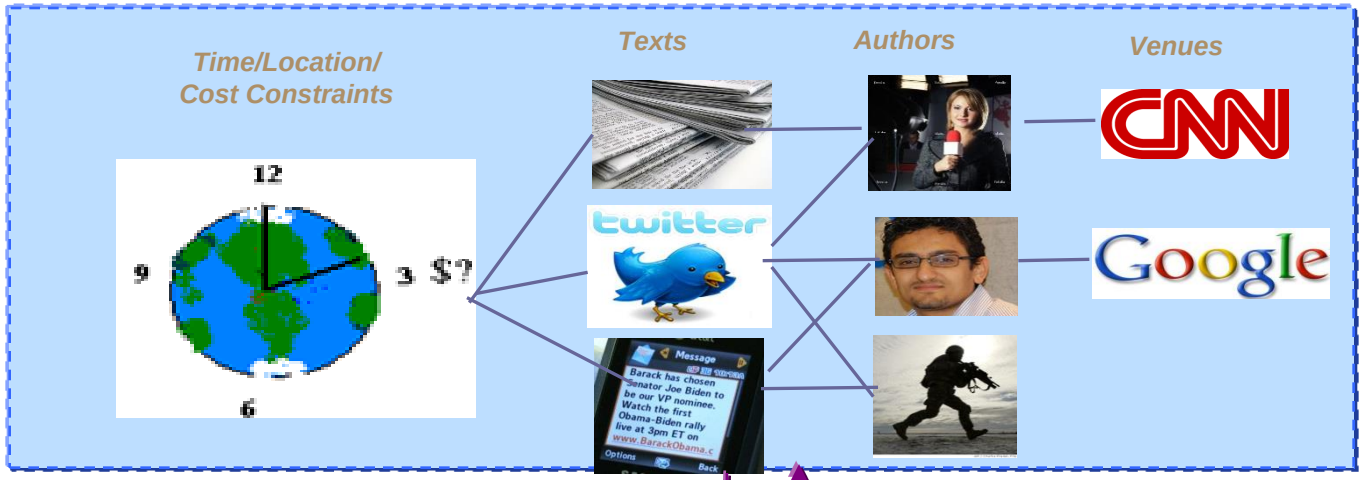
State-of-the-art and Remaining Challenges

- State-of-the-art Performance (F-score)
  - English: Trigger 70%, Argument 45%
  - Chinese: Trigger 68%, Argument 52%
  - Single human annotator: Trigger 72%, Argument 62%
- Remaining Challenges
  - Trigger Identification
    - Generic verbs
    - Support verbs such as “take” and “get” which can only represent an event mention together with other verbs or nouns
    - Nouns and adjectives based triggers
  - Trigger Classification
    - “named” represents a “Personnel_Nominate” or “Personnel_Start-Position”?
    - “hacked to death” represents a “Life_Die” or “Conflict_Attack”?
  - Argument Identification
    - Capture long contexts
  - Argument Classification
    - Capture long contexts
    - Temporal roles

(Ji, 2009; Li et al., 2011)

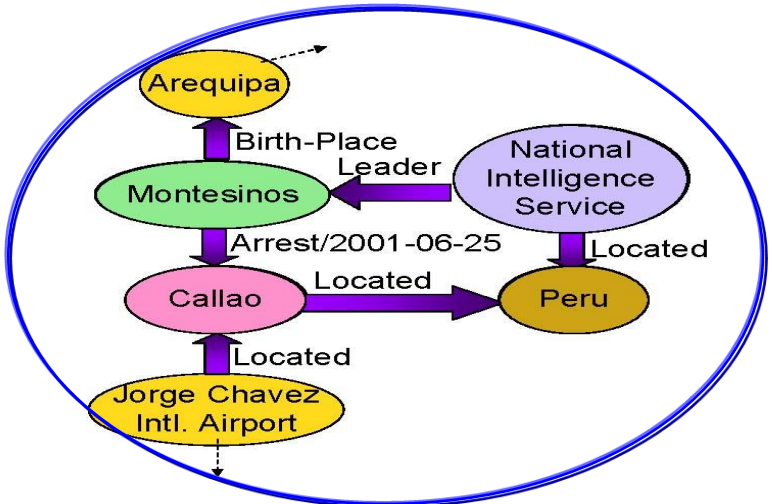
Slide from Heng Ji

IE in Rich Contexts



IE

Information Networks



Human Collaborative Learning

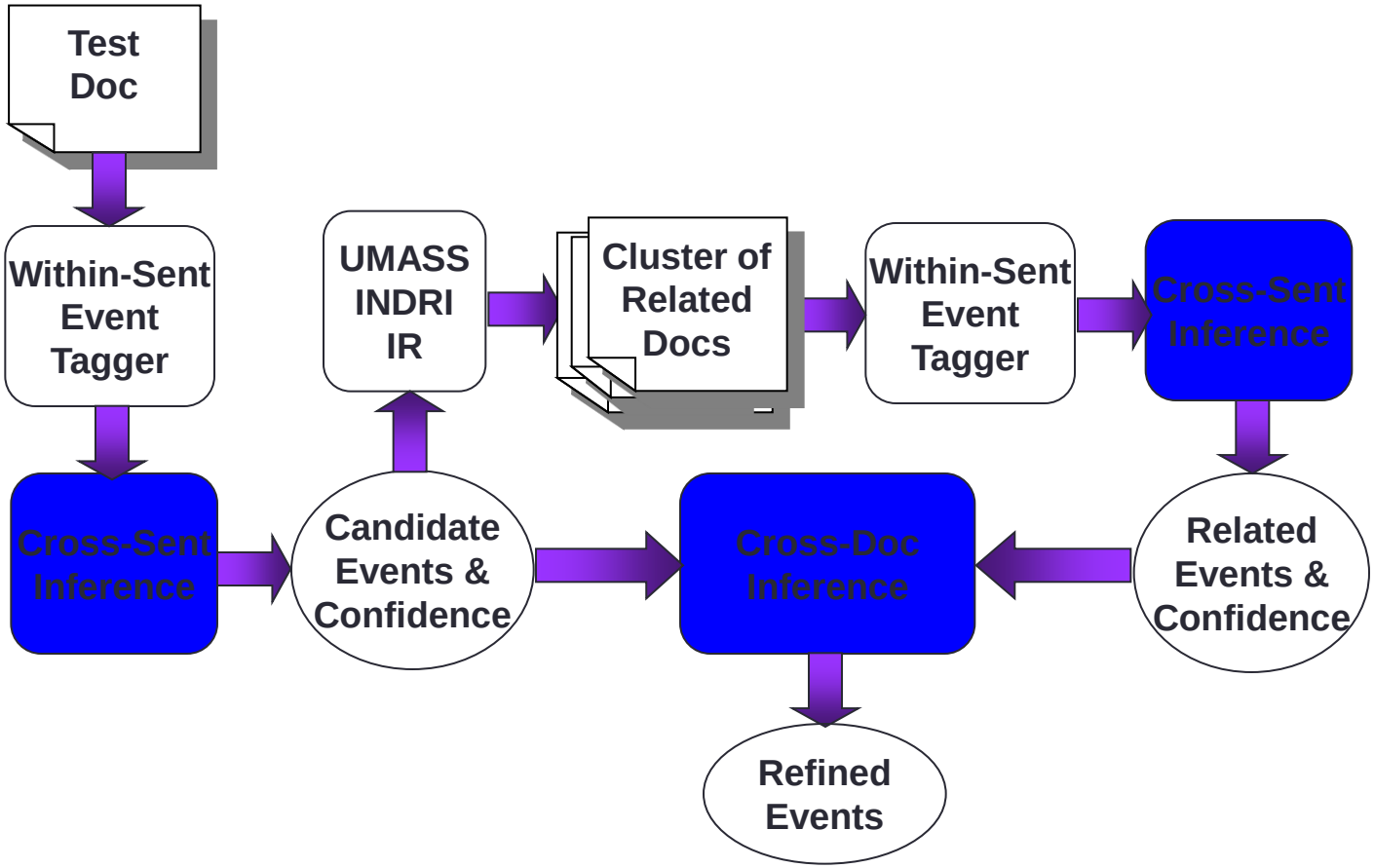
Slide from Heng Ji

Capture Information Redundancy

- When the data grows beyond some certain size, IE task is naturally embedded in rich contexts; the extracted facts become inter-dependent
- Leverage Information Redundancy from:
  - Large Scale Data (Chen and Ji, 2011)
  - Background Knowledge (Chan and Roth, 2010; Rahman and Ng, 2011)
  - Inter-connected facts (Li and Ji, 2011; Li et al., 2011; e.g. Roth and Yih, 2004; Gupta and Ji, 2009; Liao and Grishman, 2010; Hong et al., 2011)
  - Diverse Documents (Downey et al., 2005; Yangarber, 2006; Patwardhan and Riloff, 2009; Mann, 2007; Ji and Grishman, 2008)
  - Diverse Systems (Tamang and Ji, 2011)
  - Diverse Languages (Snover et al., 2011)
  - Diverse Data Modalities (text, image, speech, video…)
- But how? Such knowledge might be overwhelming…

Slide from Heng Ji

Cross-Sent/Cross-Doc Event Inference Architecture



Slide from Heng Ji

Baseline Within-Sentence Event Extraction

1. Pattern matching
   - Build a pattern from each ACE training example of an event
     - British and US forces reported gains in the advance on Baghdad
       → PER report gain in advance on LOC

2. MaxEnt models
   ① Trigger Classifier
   - to distinguish event instances from non-events, to classify event instances by type

   ② Argument Classifier
   - to distinguish arguments from non-arguments

   ③ Role Classifier
   - to classify arguments by argument role

   ④ Reportable-Event Classifier
   - to determine whether there is a reportable event instance

Slide from Heng Ji

Global Confidence Estimation

- Within-Sentence IE system produces local confidence
- IR engine returns a cluster of related docs for each test doc
- Document-wide and Cluster-wide Confidence
  - Frequency weighted by local confidence
  - *XDoc-Trigger-Freq(trigger, etype):* The weighted frequency of string *trigger* appearing as the trigger of an event of type *etype* across all related documents
  - *XDoc-Arg-Freq(arg, etype):* The weighted frequency of *arg* appearing as an argument of an event of type *etype* across all related documents
  - *XDoc-Role-Freq(arg, etype, role):* The weighted frequency of *arg* appearing as an argument of an event of type *etype* with role *role* across all related documents
  - *Margin* between the most frequent value and the second most frequent value, applied to resolve classification ambiguities
  - ……

Slide from Heng Ji

Cross-Sent/Cross-Doc Event Inference Procedure

- Remove triggers and argument annotations with local or cross-doc confidence lower than thresholds
  - *Local-Remove*: Remove annotations with low local confidence
  - *XDoc-Remove*: Remove annotations with low cross-doc confidence

- Adjust trigger and argument identification and classification to achieve document-wide and cluster-wide consistency
  - *XSent-Iden/XDoc-Iden*: If the highest frequency is larger than a threshold, propagate the most frequent type to all unlabeled candidates with the same strings
  - *XSent-Class/XDoc-Class*: If the margin value is higher than a threshold, propagate the most frequent type and role to replace low-confidence annotations

Slide from Heng Ji

Experiments: Data and Setting

- Within-Sentence baseline IE trained from 500 English ACE05 texts (from March – May of 2003)
- Use 10 ACE05 newswire texts as development set to optimize the global confidence thresholds and apply them for blind test
- Blind test on 40 ACE05 texts, for each test text, retrieved 25 related texts from TDT5 corpus (278,108 texts, from April-Sept. of 2003)

Slide from Heng Ji

Experiments: Trigger Labeling

Performance / System/Human	Precision	Recall	F-Measure
Within-Sent IE (Baseline)	67.6	53.5	59.7
After Cross-Sent Inference	64.3	59.4	61.8
After Cross-Doc Inference	60.2	76.4	67.3
Human Annotator 1	59.2	59.4	59.3
Human Annotator 2	69.2	75.0	72.0
Inter-Adjudicator Agreement	83.2	74.8	78.8

Slide from Heng Ji

Experiments: Argument Labeling

Performance / System/Human	Argument Identification			Argument Classification Accuracy	Argument Identification +Classification		
	P	**R**	**F**		**P**	**R**	**F**
Within-Sent IE	47.8	38.3	42.5	86.0	41.2	32.9	36.3
After Cross-Sent Inference	54.6	38.5	45.1	90.2	49.2	34.7	40.7
After Cross-Doc Inference	55.7	39.5	46.2	92.1	51.3	36.4	42.6
Human Annotator 1	60.0	69.4	64.4	85.8	51.6	59.5	55.3
Human Annotator 2	62.7	85.4	72.3	86.3	54.1	73.7	62.4
Inter-Adjudicator Agreement	72.2	71.4	71.8	91.8	66.3	65.6	65.9

Slide from Heng Ji

Event Extraction: Summary

- Event extraction is an interesting topic which has recently started to undergo significant changes
  - In these slides we talked about cross-document reference
  - One can go further and include the web and/or ontologies (next lecture)
- It is a very difficult problem but clearly necessary if we want to reason about changes of state, rather than facts that hold over long periods of time
- Now let's briefly talk about Multimodal IE

Multimodal Extraction

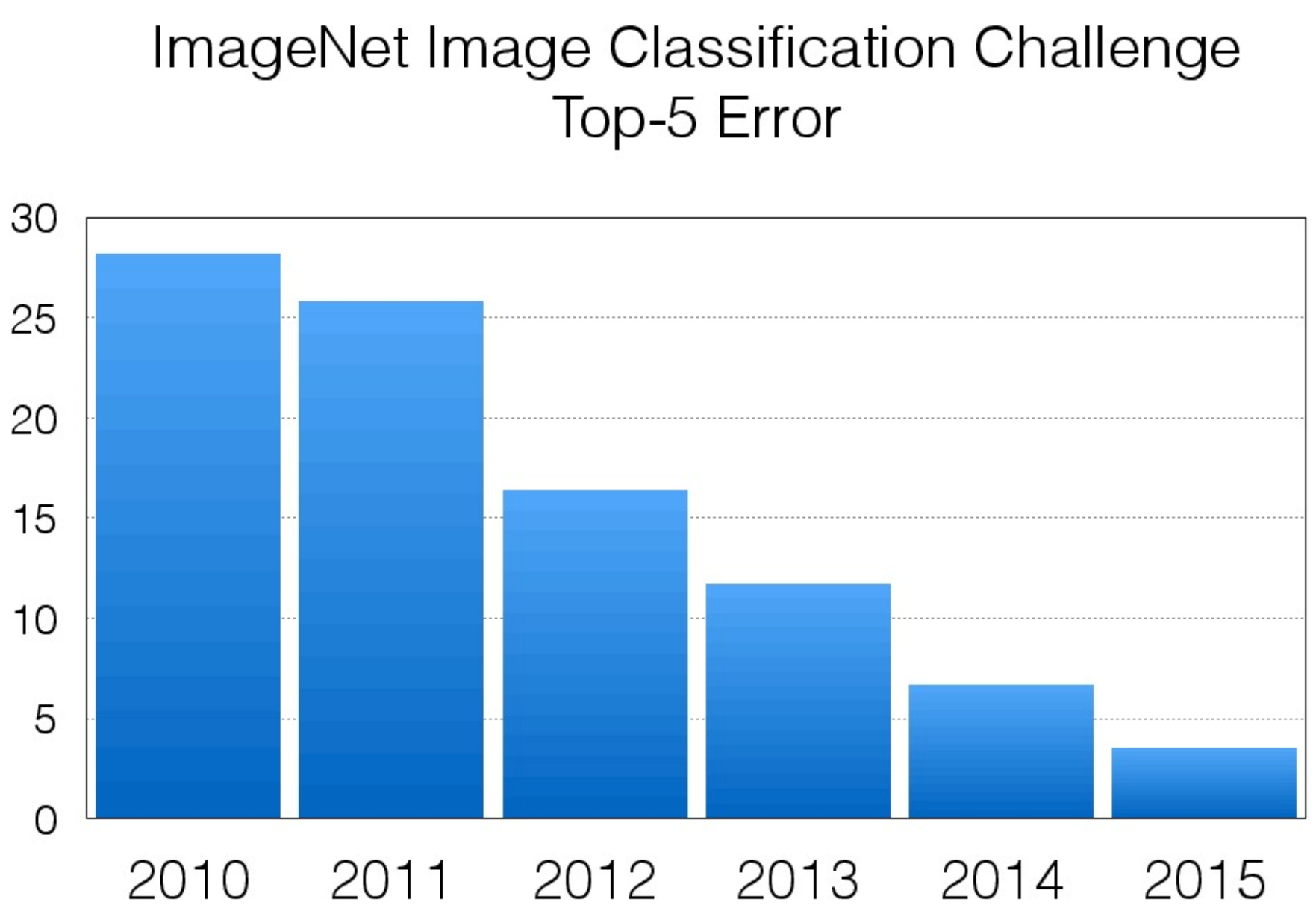
- The purpose of these slides is to give a basic idea about what can be done in a multimodal setting
- Details of how the systems work in detail is out of scope here (i.e., don't worry about this)

Extraction from Speech

- Extraction from speech is typically addressed by adapting text-based NLP tools to ASR (Automatic Speech Recognition) output
  - Neural systems are typically used for ASR
- Some significant challenges using ASR output as input to NLP
  - ASR errors (in recognizing speech)
  - No or little punctuation in ASR output
  - Disfluencies (e.g., when people, are, um, sp..., speaking)
- Some new work tries to train end-to-end systems to do tasks like ASR and NER at the same time
  - Make sense, because many names are likely to be out-of-vocabulary items to the ASR system
  - Allows use of specialized ASR sub-model

Extraction from Images

- Approaches for image classification and related problems have been dramatically changed by deep learning
- Current explosion of new work and dramatically different problems being addressed
- First let's look at accuracies on the ImageNet task (next slide)
- Then let's take a brief look at image captioning, as a prototypical text/image task

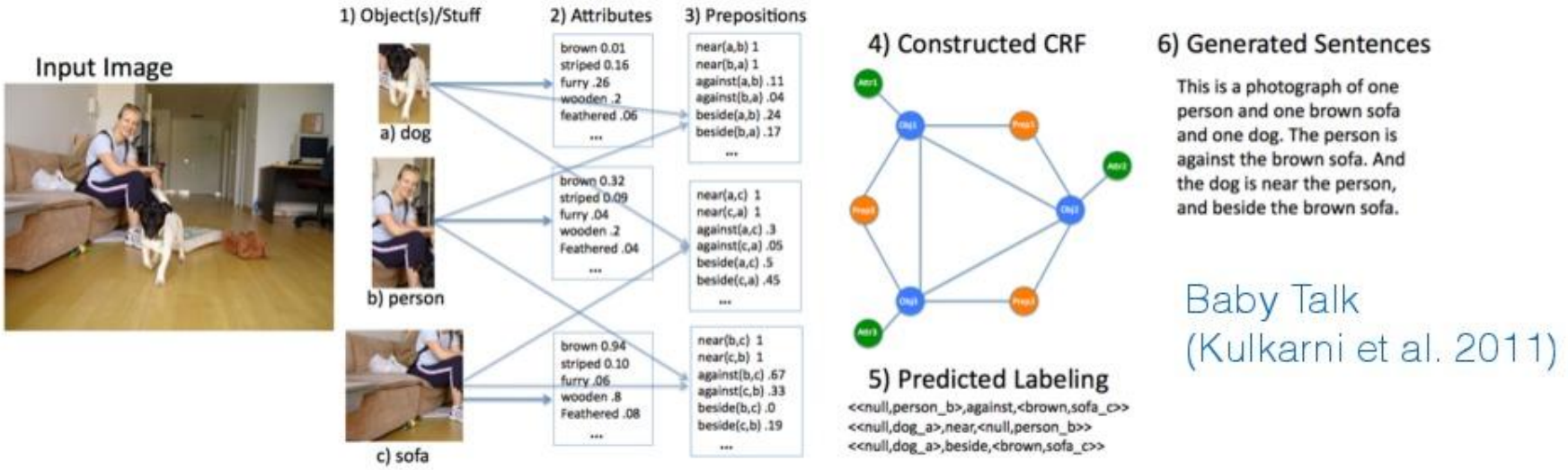


Slide from Andrej Karpathy, results from ILSVRC, Russakovsky et al., 2015

From image classification to image captioning

- Image classification has gotten much better
- The basic approach is the same as training a linear model like perceptron
  - Check if we get the right answer
  - If yes, do nothing
  - If no, update the parameters to make the right answer more likely
- But how can we generate captions?

Generating Descriptions: Prior work



example template:

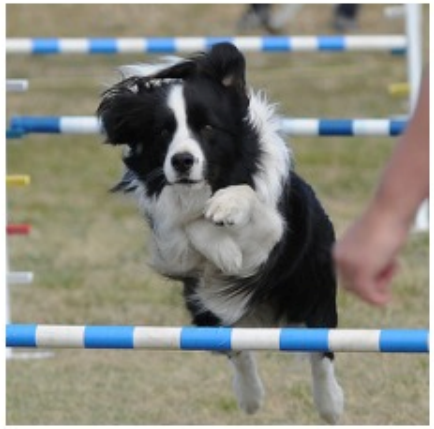
(noun) _____ in (noun) _____ is (verb) _____ in (noun) _____.

[Barnard '03] [Duygulu '02] [Frome '13] [Gupta & Mannem '12] [Elliott & Keller '13] [Yatskar '14] [Kiros '14] [Yao '10] [Yang '11] [Barbu '12] [Mitchell '12]

Slide from Andrej Karpathy

Core Challenge

how can we predict sentences?



???

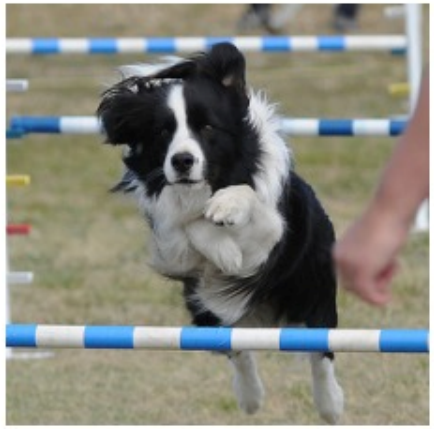
"A dog jumping over a hurdle"

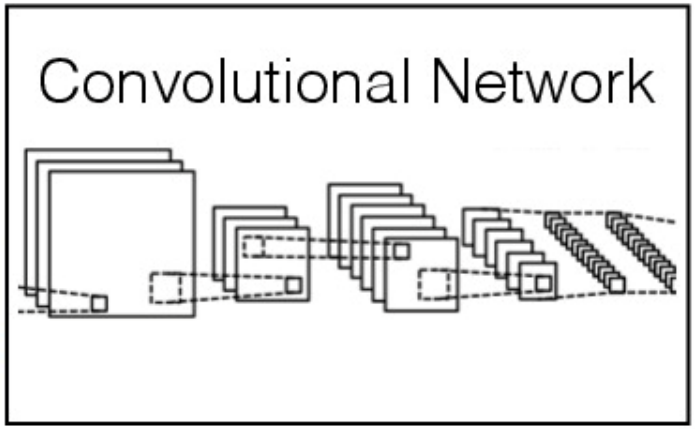
differentiable function

Slide from Andrej Karpathy

Core Challenge

how can we predict sentences?





???

"A dog jumping over a hurdle"

differentiable function

Slide from Andrej Karpathy

Core Challenge

how can we predict sentences?

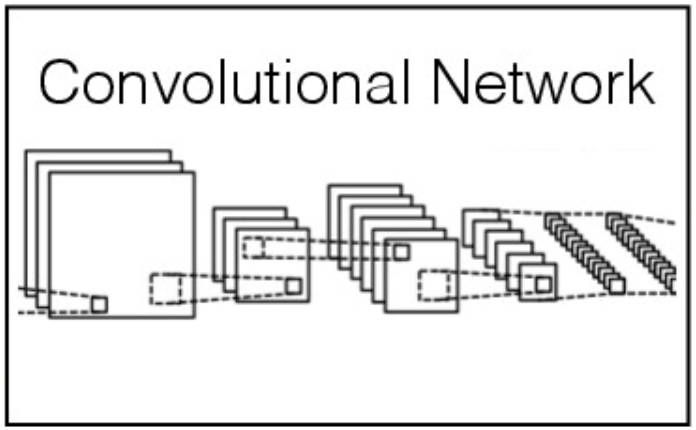


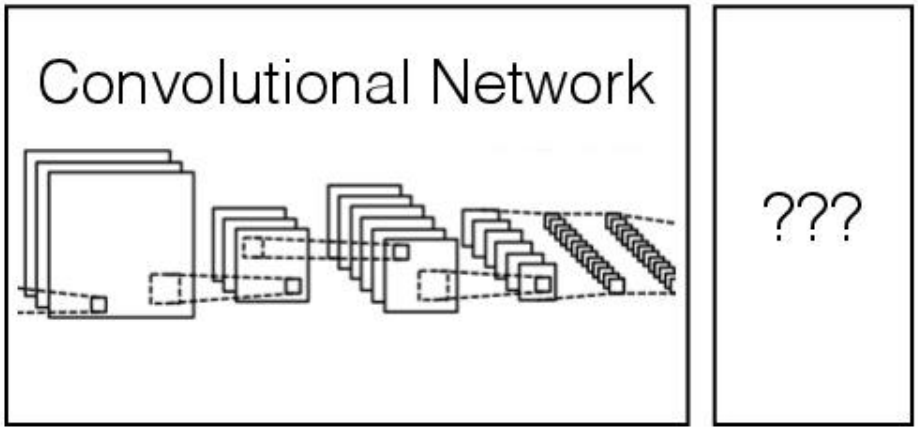
image classification

dog

differentiable function

Slide from Andrej Karpathy

Core Challenge



a

dog

jumping

over

a

hurdle

<end>

sentences have variable number of words
=> output not fixed size!

Slide from Andrej Karpathy

Language Model

words

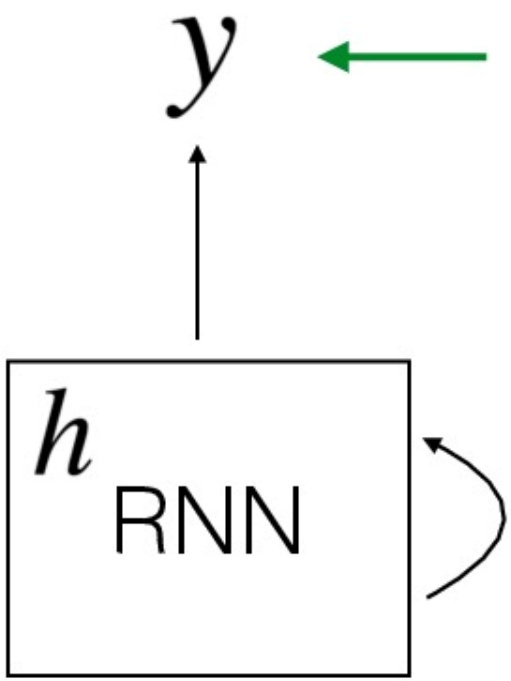
$$P(x_1, x_2, \ldots, x_n)$$

$$= \prod_{i=1}^{n} P(x_i | x_1, \ldots, x_{i-1})$$

P(word | previous words)

Slide from Andrej Karpathy

Recurrent Neural Network Language Model



predict next word distribution

feed in words one at a time
e.g. “one-hot” encoding

Slide from Andrej Karpathy

Image Classification

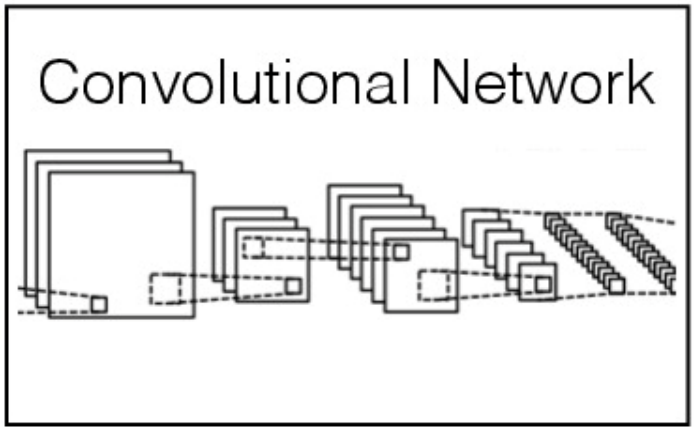
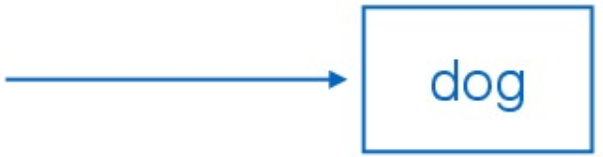
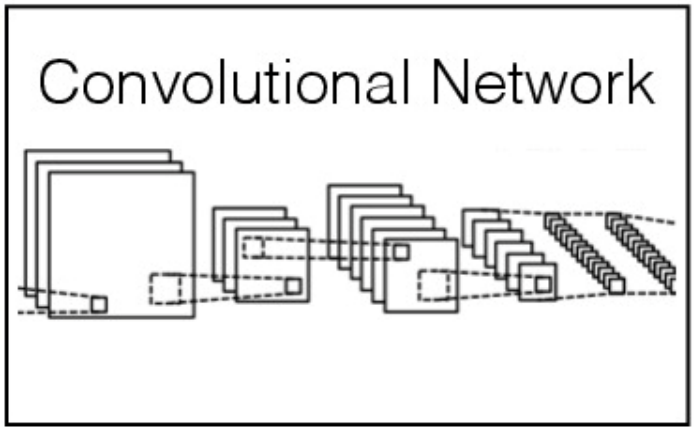


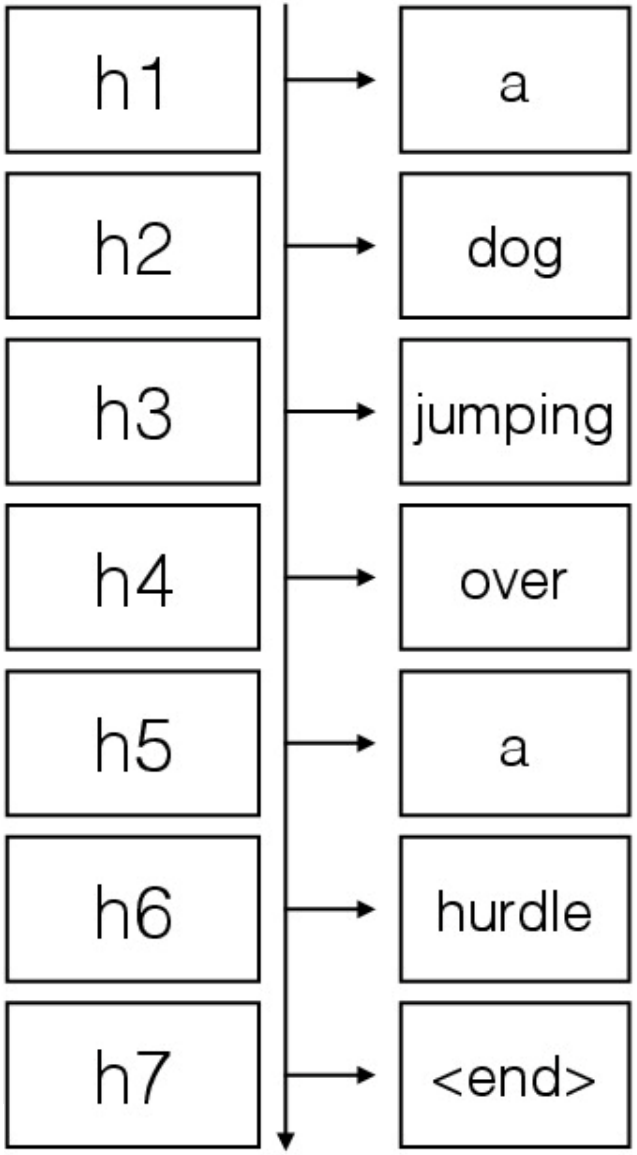
image classification



Slide from Andrej Karpathy

Image Captioning

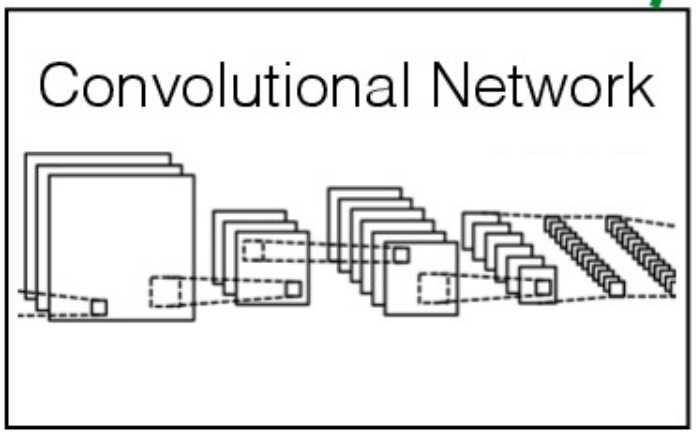


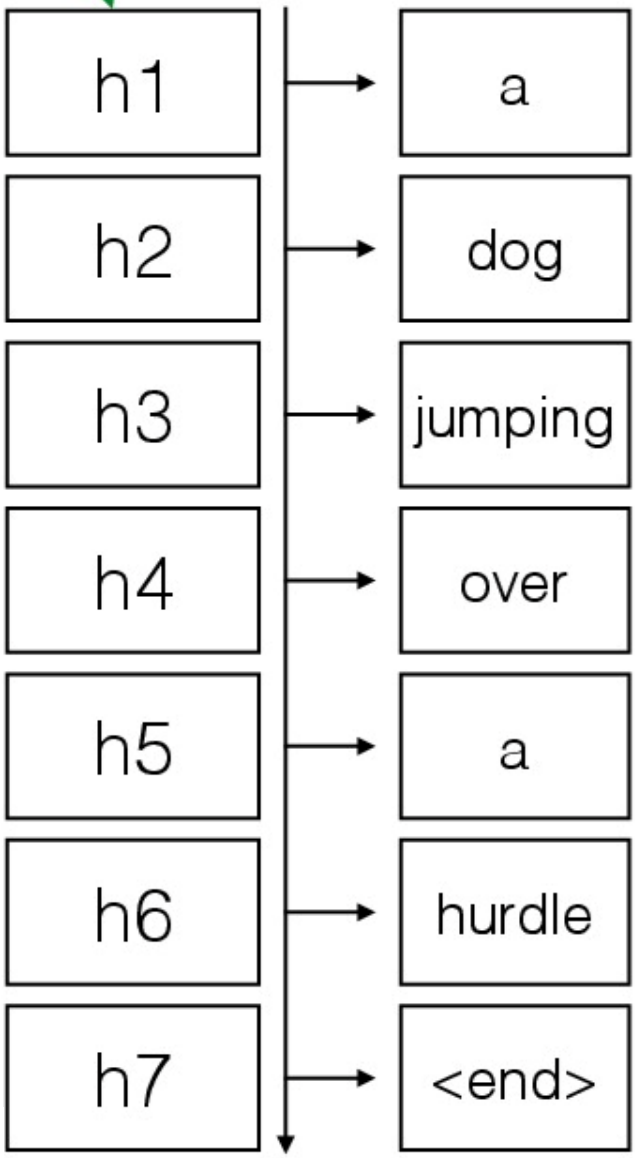


Q: how do we condition the generative process on the image information?

Slide from Andrej Karpathy

Image Captioning





$$h_0 = v$$

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$y_t = W_{hy} h_t$$

Slide from Andrej Karpathy

Image Sentence Datasets

1. A woman and her dog watch the cameraman in their living with wooden floors.
2. A woman sitting on the couch while a black faced dog runs across the floor.
3. A woman wearing a backpack sits on a couch while a small dog runs on the hardwood floor next to her.
4. A women sitting on a sofa while a small Jack Russell walks towards the camera.
5. White and black small dog walks toward the camera while woman sits on couch, desk and computer seen in the background as well as a pillow, teddy bear and moggie toy on the wood floor.

1. A man in a cowboy hat check approaches a small red sports car.
2. The back and left side of a red Ferrari and two men admiring it.
3. The sporty car is admired by passer by.
4. Two men next to a red sports car in a parking lot.
5. Two men stand beside a red sports car.

[1] **Pascal 1K**: 1,000 images (5 sentences per image)
[2] **Flickr8K**: 8,000 images
[3] **Flickr30K**: 30,000 images
[4] **MSCOCO**: 115,000 images

[1] Rashtchian et al., 2010
[2] Hodosh et al., 2013
[3] Young et al., 2014
[4] Lin et al., 2015

Slide from Andrej Karpathy

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

"little girl is eating piece of cake."

"baseball player is throwing ball in game."

"woman is holding bunch of bananas."

"black cat is sitting on top of suitcase."

Slide from Andrej Karpathy

Example Error

"a woman in a bikini is jumping over a hurdle."

Slide modified from Andrej Karpathy

Limitations

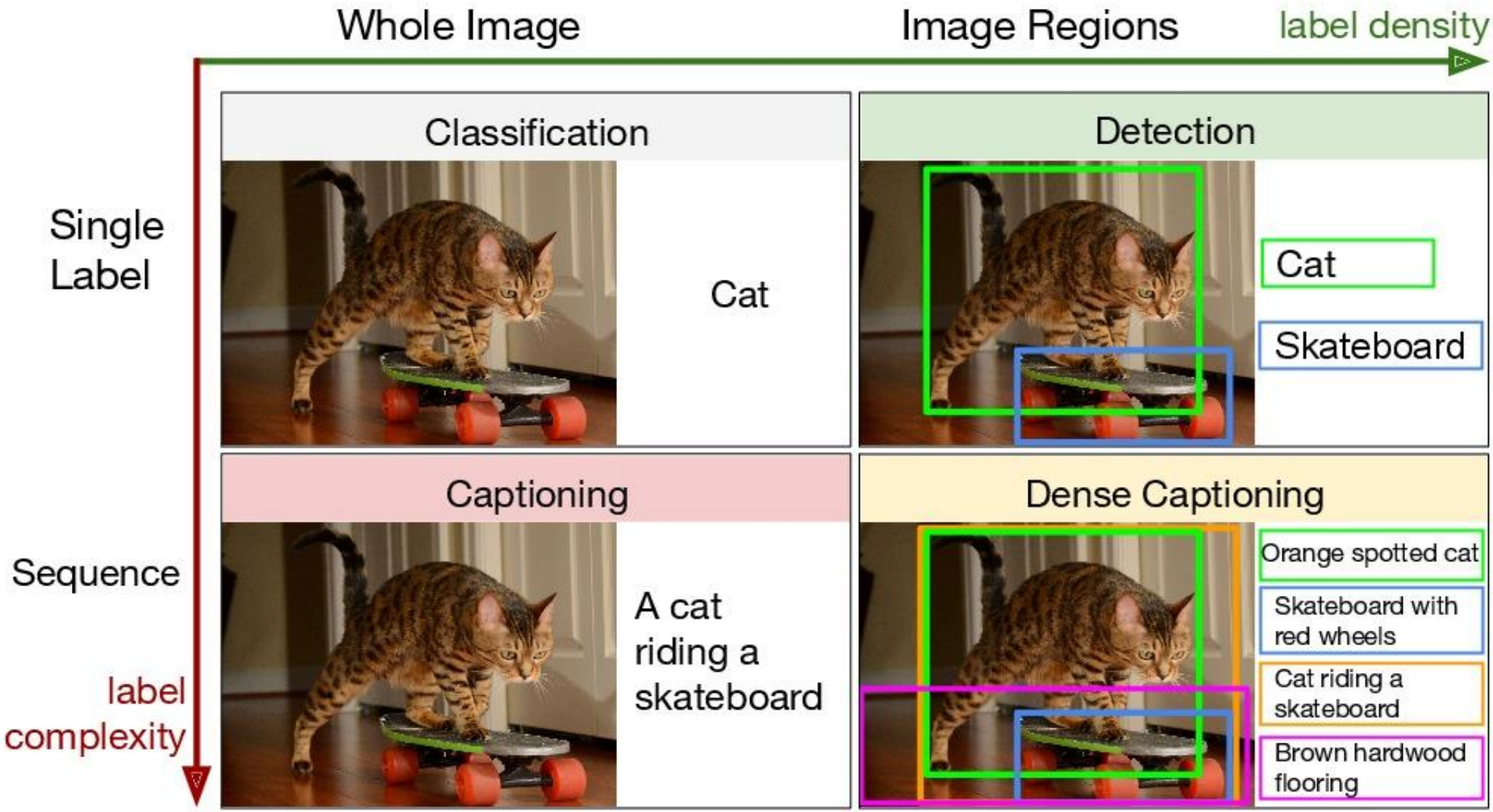
"A group of people in an office."

Slide from Andrej Karpathy

Can go even further...

- Deep learning enabled addressing image caption generation in a much more natural way
  - Also, cross-fertilization of ideas with machine translation (!)
  - Framework is actually very similar to neural machine translation
- Deep learning also enables solving new problems
  - For instance, there is now work on breaking images down into regions (next slide)

Dense Captioning



Slide from Andrej Karpathy

Putting it all together for IE

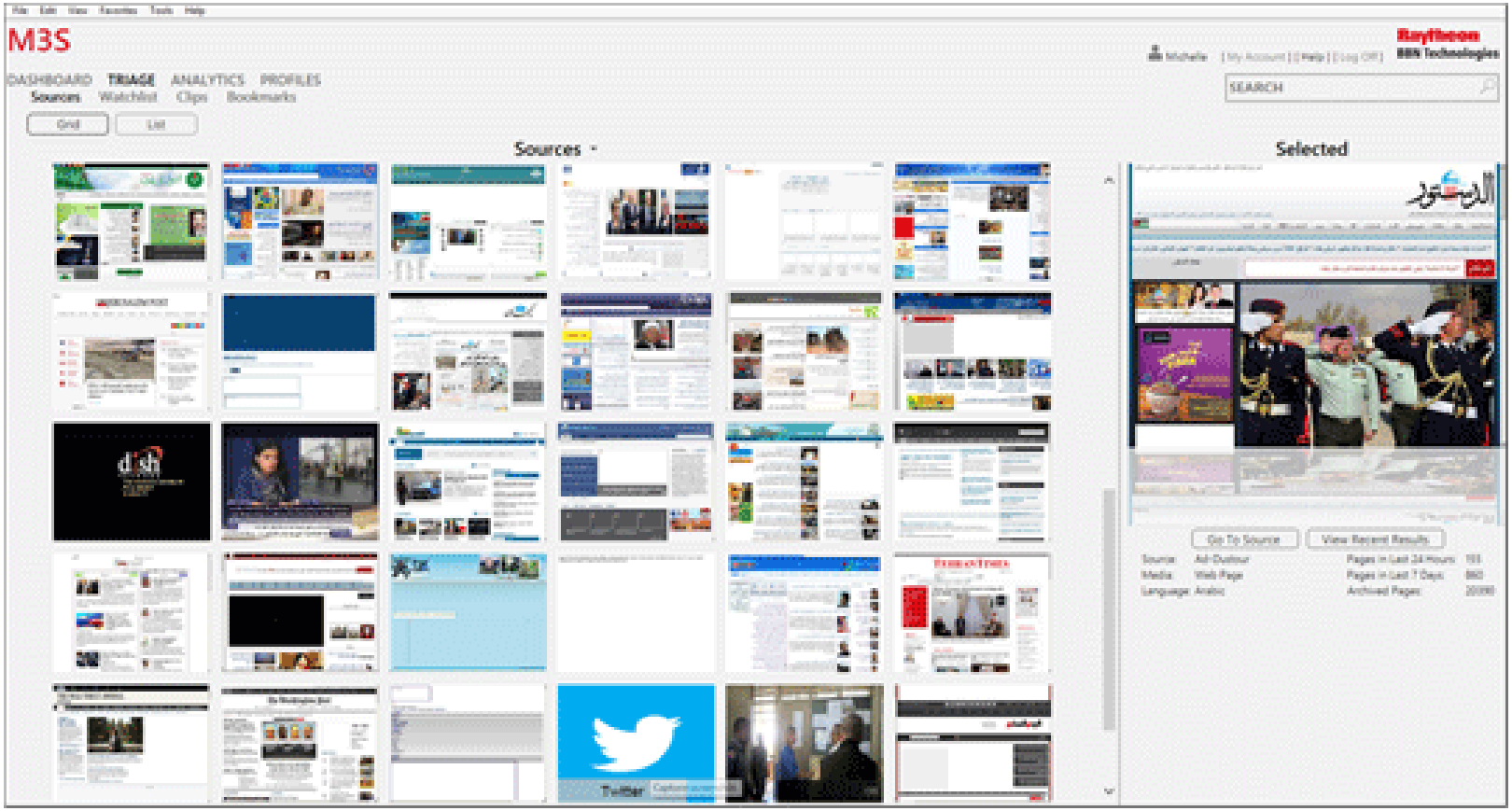
- Near term: gains in (static) image processing performance will continue, video processing and ASR will make big improvements
- IE: Here is an example of a state-of-the-art system for indexing multimodal news streams
  - Primarily working with speech and text though, only limited support for images and video (at least in the 2013 version I looked at)

BBN Multimedia Monitoring System (M3S)

- An example system for multimodal extraction is the BBN M3S system (version here from 2013)
- Features:
  - Automatic multi-lingual data collection and mirroring of user-identified Web sites, broadcast media, and social media (Twitter and Facebook)
  - Automatic extraction and translation of text
  - Search across multi-lingual sites, channels, and posts
  - Visualization tools and automatic topic detection for enhanced analysis
  - Collected media archived for later use
  - Browser-based user interface with personalized user dashboards
  - Story segmentation of broadcast media

(From BBN website)

BBN Multimedia Monitoring System (M3S)



(Example Graphic from BBN M3S website, downloaded 2017-01-07)

Discussion

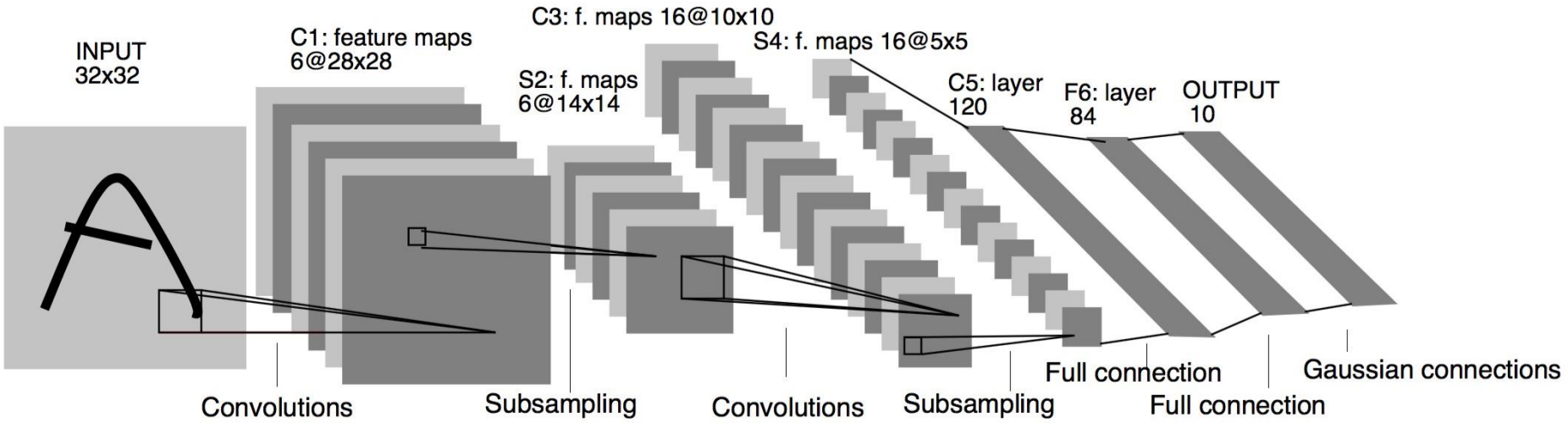
- Another prominent system: Europe Media Monitor
  - Check out their website (free access to a good amount of functionality, also free tablet and smartphone apps; and a special medical system)
- Overall: multimodal processing approaches are changing rapidly due to better modeling and new sub-tasks
- Deep learning approaches should enable IE systems to reason in a more deep way about video/audio streams
  - Much new academic work appearing here in many different venues
  - Exciting time for this research!

Slides

- The slides for event extraction are from Heng Ji. She is an IE researcher at RPI
- The slides on image captioning are from Andrej Karpathy (PhD student of Fei-Fei Li), now at OpenAI

- Thank you for your attention!

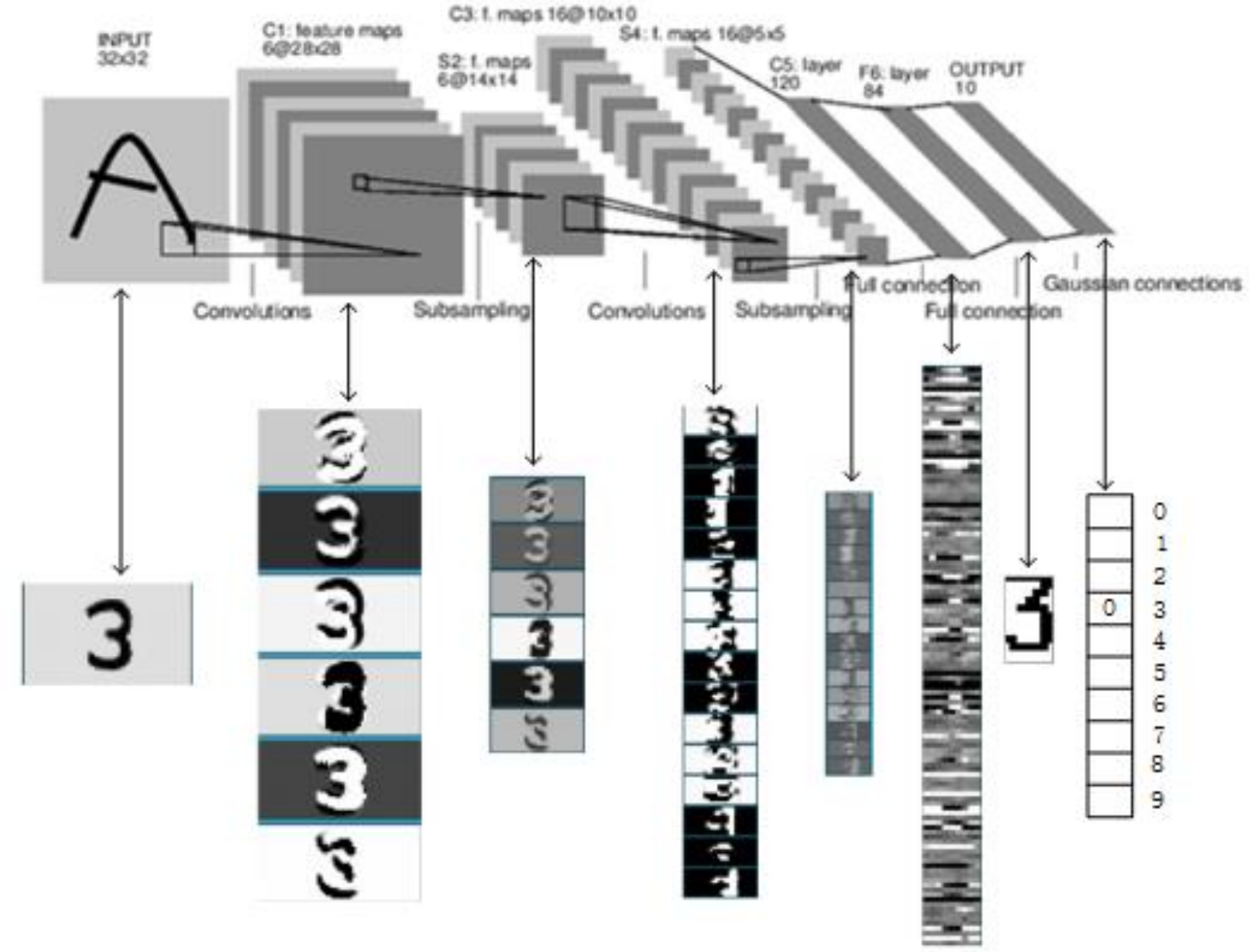
LeNet-5



- convolutional neural network use sequence of 3 layers: convolution, pooling, non-linearity –> This may be the key feature of Deep Learning for images since this paper!
- use convolution to extract spatial features
- subsample using spatial average of maps
- non-linearity in the form of tanh or sigmoids
- multi-layer neural network (MLP) as final classifier
- sparse connection matrix between layers to avoid large computational cost

(Graphic from Yann LeCun, Text from Culurciello et al.)

LeNet-5 recognizing "3"



(Graphic from Yann LeCun (and world4jason??))