



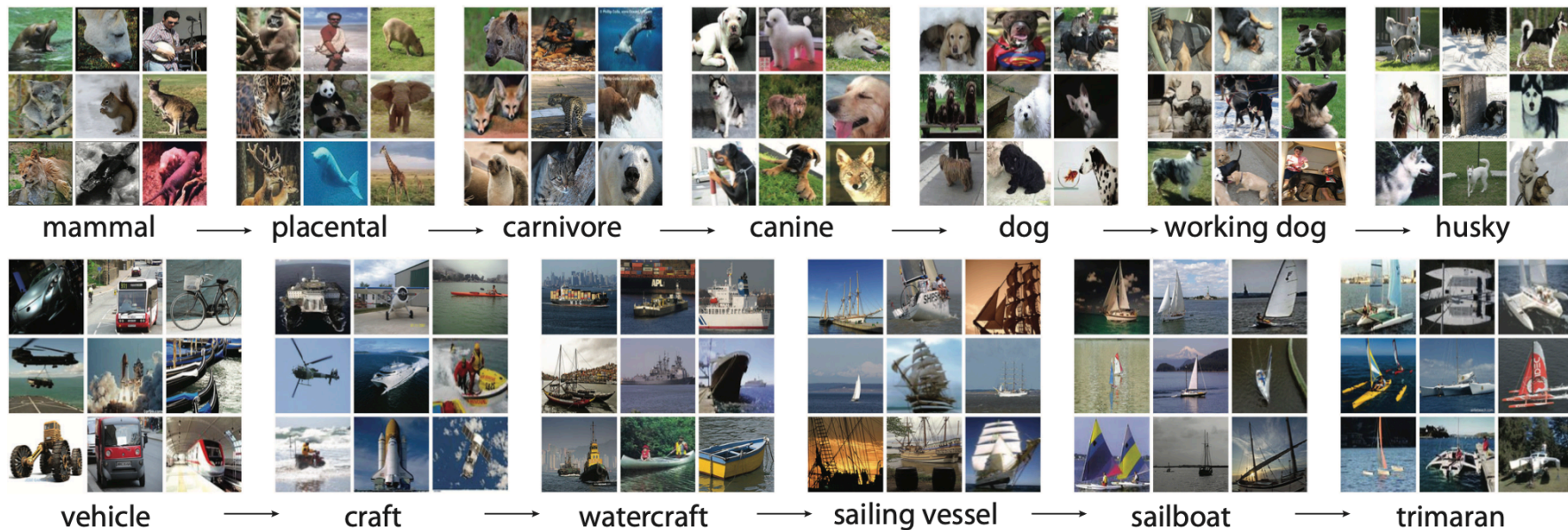
Natural Language Processing

Info 159/259

Lecture 25: Multimodal NLP (April 26, 2022)

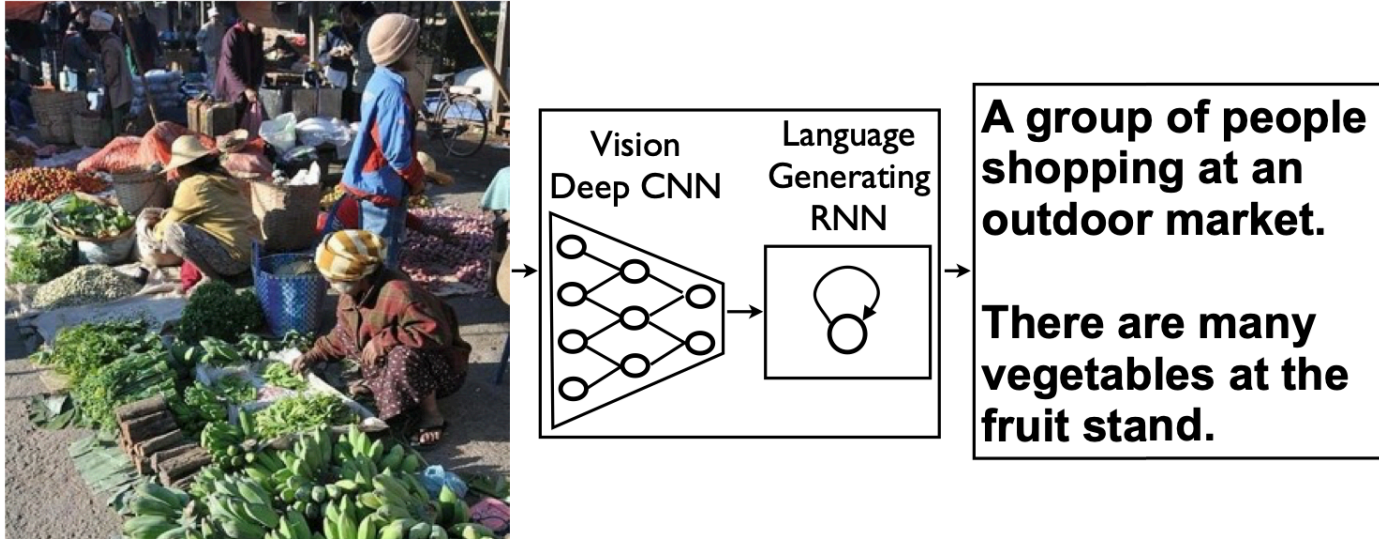
David Bamman, UC Berkeley

ImageNet



Deng et al. 2009, "ImageNet: A Large-Scale Hierarchical Image Database"

Image Captioning



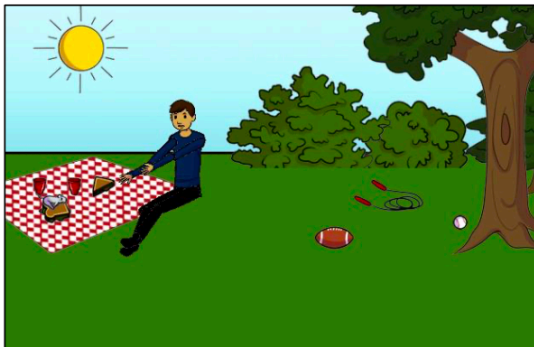
Visual QA



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?

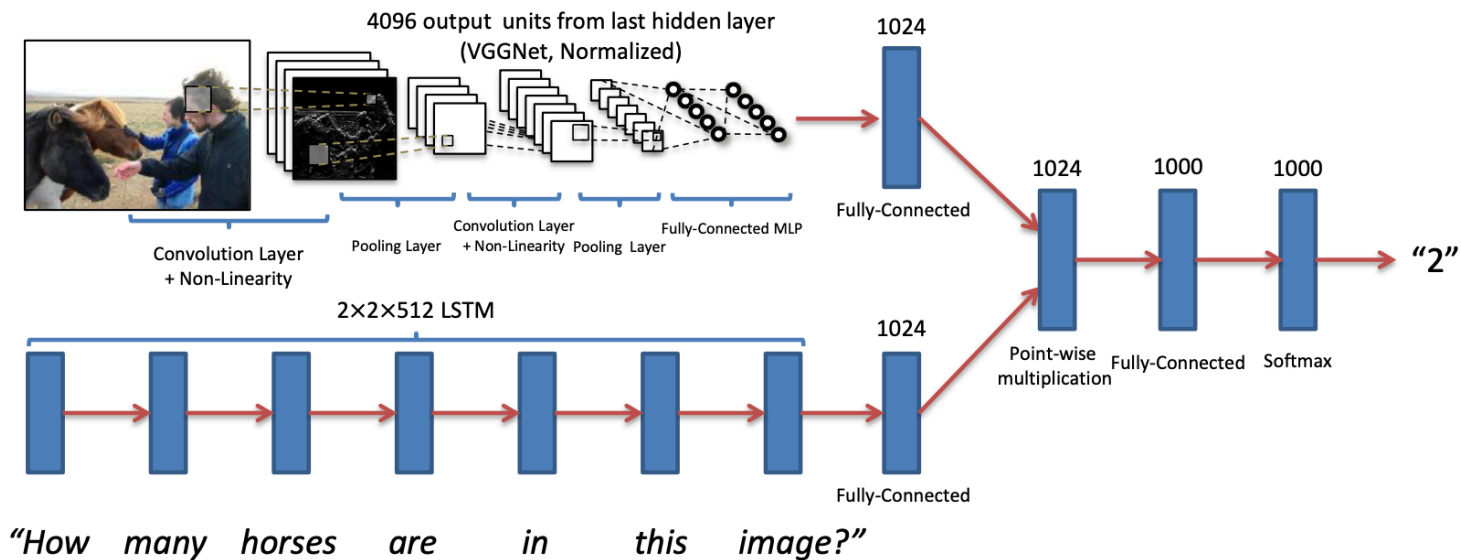


Is this person expecting company?
What is just under the tree?

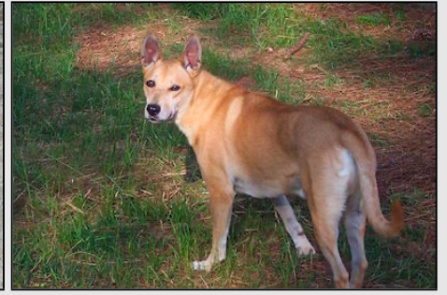


Does it appear to be rainy?
Does this person have 20/20 vision?

Visual QA



Visual Reasoning

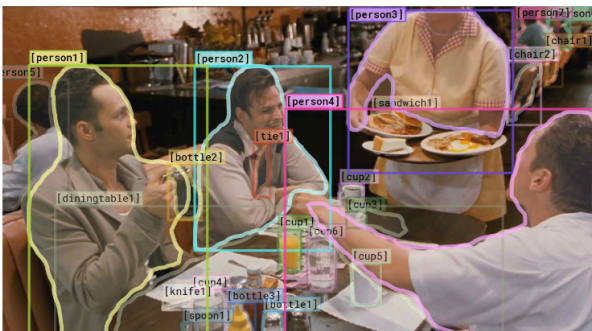


The left image contains twice the number of dogs as the right image, and at least two dogs in total are standing.



One image shows exactly two brown acorns in back-to-back caps on green foliage.

Visual Commonsense Reasoning

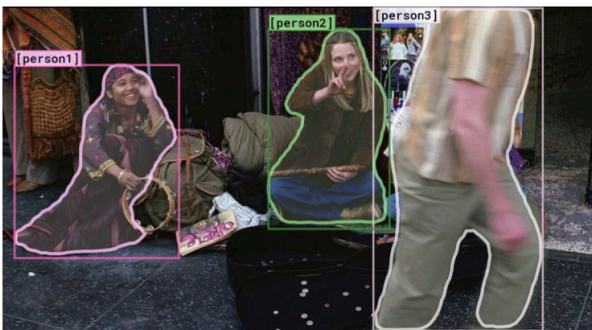


Why is [person4] pointing at [person1]?

- a) He is telling [person3] that [person1] ordered the pancakes.
- b) He just told a joke.
- c) He is feeling accusatory towards [person1].
- d) He is giving [person1] directions.

I chose a)
because...

- a) [person1] has the pancakes in front of him.
- b) [person4] is taking everyone's order and asked for clarification.
- c) [person3] is looking at the pancakes and both she and [person2] are smiling slightly.
- d) [person3] is delivering food to the table, and she might not know whose order is whose.



How did [person2] get the money that's in front of her?

- a) [person2] is selling things on the street.
- b) [person2] earned this money playing music.
- c) She may work jobs for the mafia.
- d) She won money playing poker.

I chose b)
because...

- a) She is playing guitar for money.
- b) [person2] is a professional musician in an orchestra.
- c) [person2] and [person1] are both holding instruments, and were probably busking for that money.
- d) [person1] is putting money in [person2]'s tip jar, while she plays music.

Text-to-Image Generation

- DALL-E (Ramesh et al. 2022)
- Transformer over streams of text+image tokens trained on Conceptual Captions (3.3 M text/image pairs)

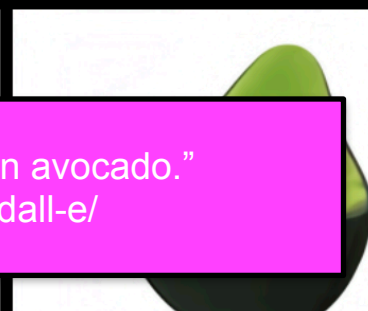
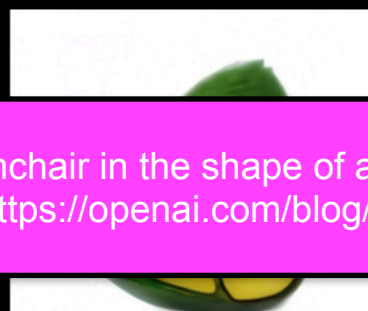
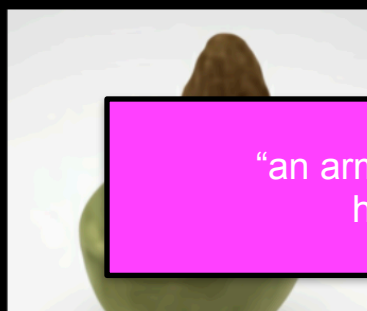
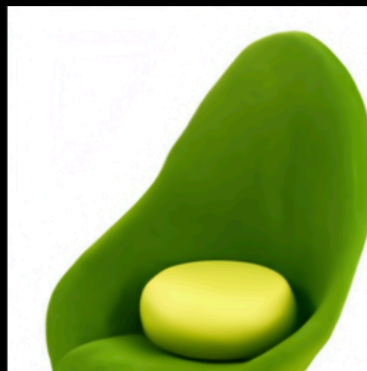


**(a) a tapir made of accordion.
a tapir with the texture of an
accordion.**



Week	Date	Topic
1	1/18	Introduction [slides]
	1/20	Lexical semantics/static word embeddings [slides]
2	1/25	Text classification 1: Logistic regression [slides]
	1/27	Text classification 2: MLP and CNN [slides]
3	2/1	Text classification 3: Attention and transformers [slides]
	2/3	Annotation [slides]
4	2/8	Language modeling 1 [slides]
	2/10	Language modeling 2 [slides]
5	2/15	Language modeling 3: Contextual embeddings [slides]
	2/17	Language modeling 4: Few-shot learning and prompting methods [slides]
6	2/22	Sequence labeling: POS tagging; HMM [slides]
	2/24	Neural sequence labeling [slides]

Text is data.



“an armchair in the shape of an avocado.”
<https://openai.com/blog/dall-e/>



The importance of being on twitter

by Jerome K. Jerome
London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter." "Ah!" I said, "I know about that. But what is Twitter?"

"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."



R. N.
Loves
L.A.



1770



5005

Handwritten manuscript page from the Voynich manuscript, showing six lines of text written in the characteristic Voynich script. The text is densely packed and appears to be a continuous passage.



Course evaluations!

<https://course-evaluations.berkeley.edu/berkeley/>