Computer Vision and Semantics

Fei-Fei Li
Justin Johnson, Andrej Karpathy
Ranjay Krishna, Michael Stark,
Li-Jia Li, David A. Shamma

Semantics
(Language)

Images

Image Retrieval



Semantics
(Language)

Images

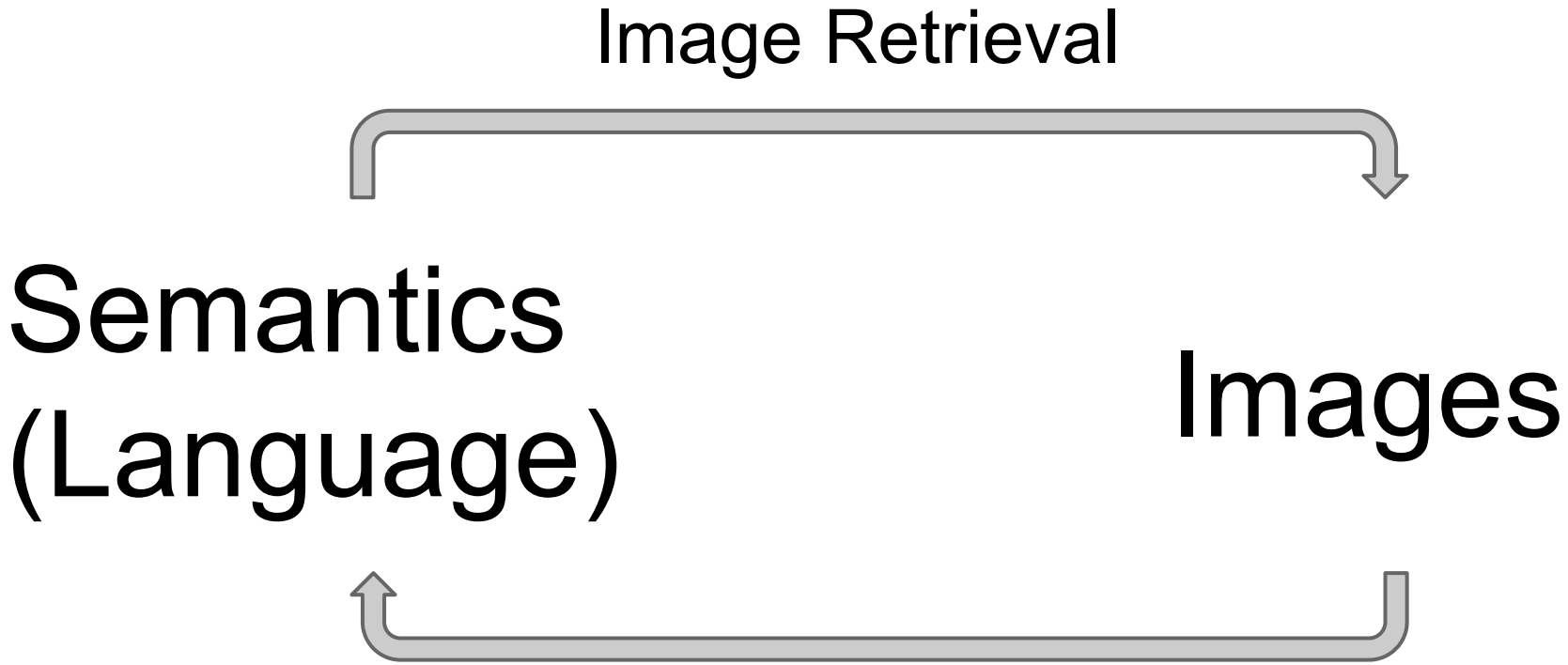

Image Retrieval
Semantics
(Language)
Images
Sentence Generation

Image Retrieval

“Image Retrieval using Scene Graphs”,
J Johnson, R Krishna, M Stark,
DA Shamma, LJ Li, Li FF,
CVPR 2015

Image Retrieval

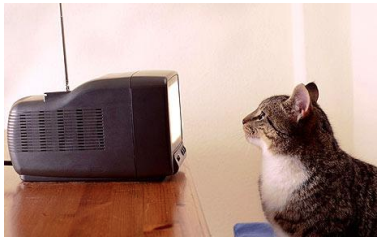
Input: Sentence query, set of images

Output: Matching images

“Cat watching television”







Google

man wearing suit holding cute dog

Justin

Web | Images | Shopping | Videos | News | More | Search tools | SafeSearch





















Google

man wearing suit holding cute dog

Justin

Web | Images | Shopping | Videos | News | More | Search tools | SafeSearch

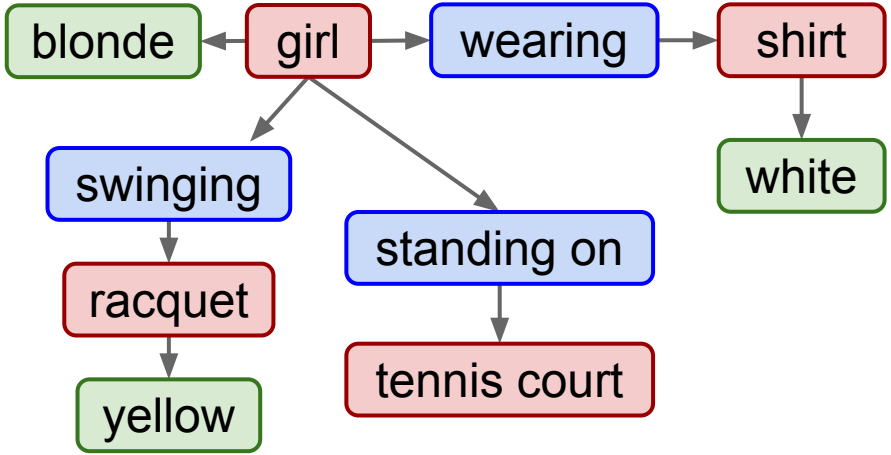


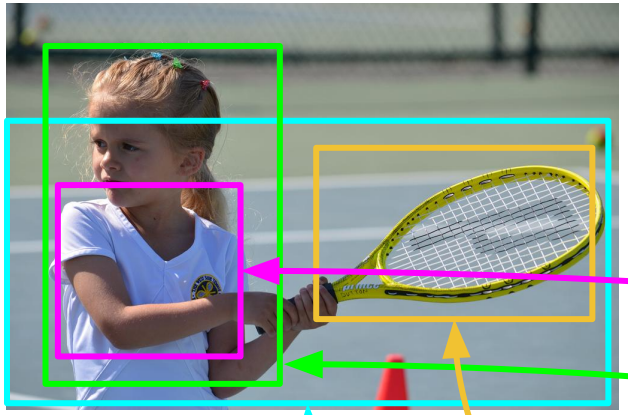


“A blonde girl wearing a white shirt standing on a tennis court and swinging a yellow racquet”

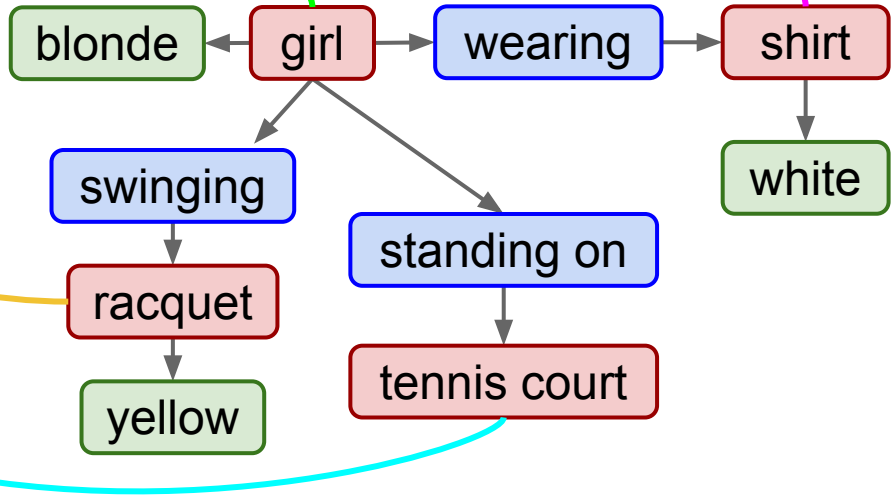


"A blonde girl wearing a white shirt standing on a tennis court and swinging a yellow racquet"





"A blonde girl wearing a white shirt standing on a tennis court and swinging a yellow racquet"



Conditional Random Field

“Man sitting on bench”

Only objects






Objects + attributes



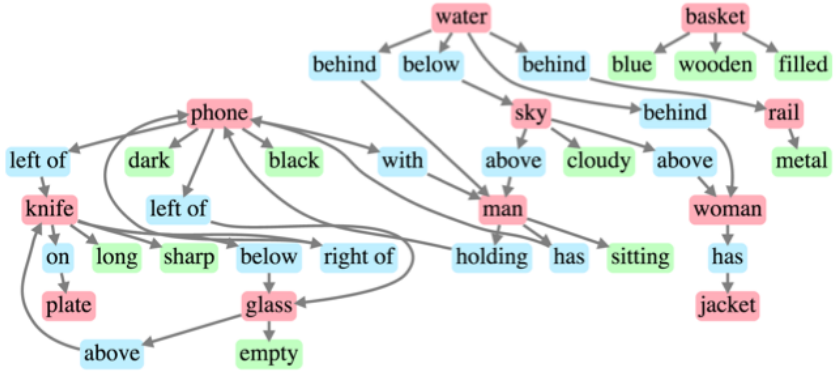


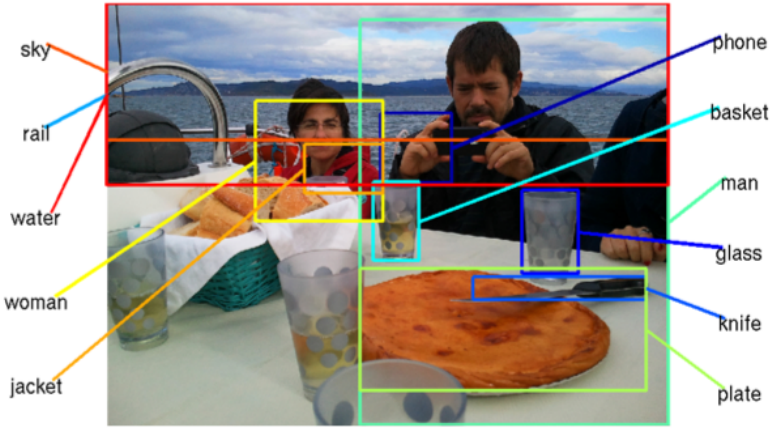

Ours



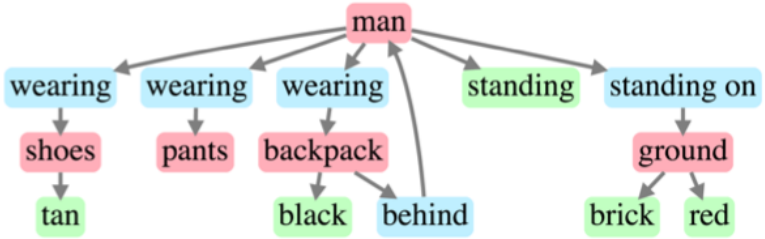


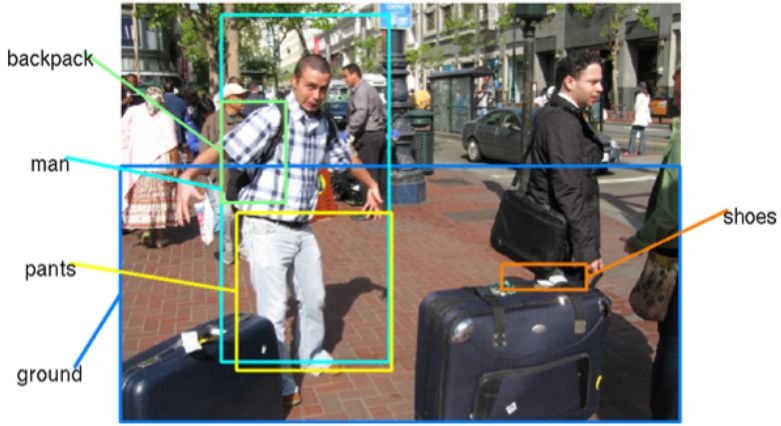

Sitting man holding a black phone to the left of a knife on a plate and an empty glass; a woman with a jacket is in front of a metal rail and water beneath a cloudy sky





Man standing on red brick ground wearing a black backpack and tan shoes





Sentence Generation

"Deep Visual-Semantic Alignments
for Generating Image Descriptions",
Andrej Karpathy, Li Fei-Fei,
CVPR 2015

Sentence Generation

Input: Image

Output: Sentence describing image





“Cat watching television”



test image

Slide credit: Andrej Karpathy

image

conv-64
conv-64
maxpool

conv-128
conv-128
maxpool

conv-256
conv-256
maxpool

conv-512
conv-512
maxpool

conv-512
conv-512
maxpool

FC-4096
FC-4096

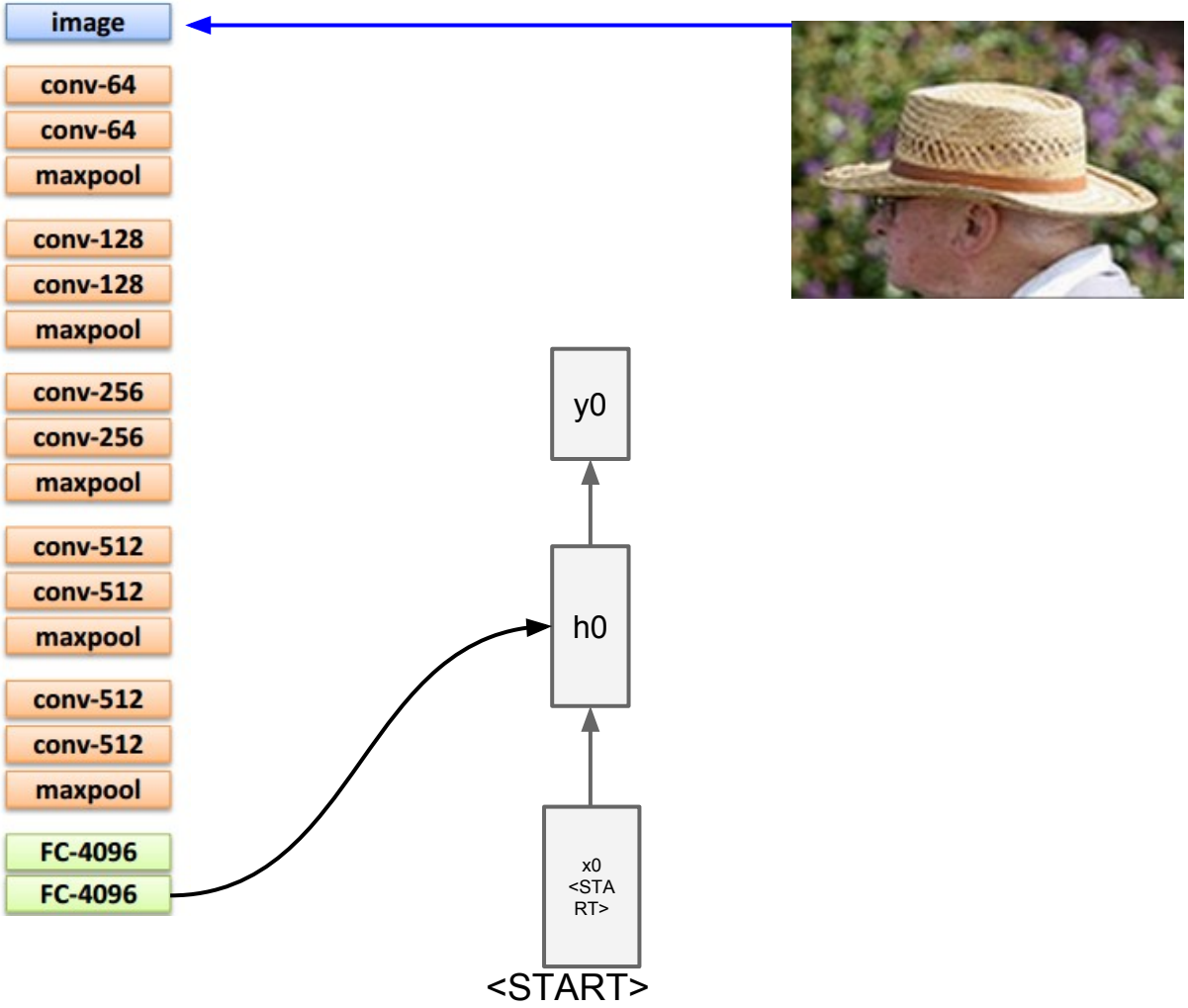


test image

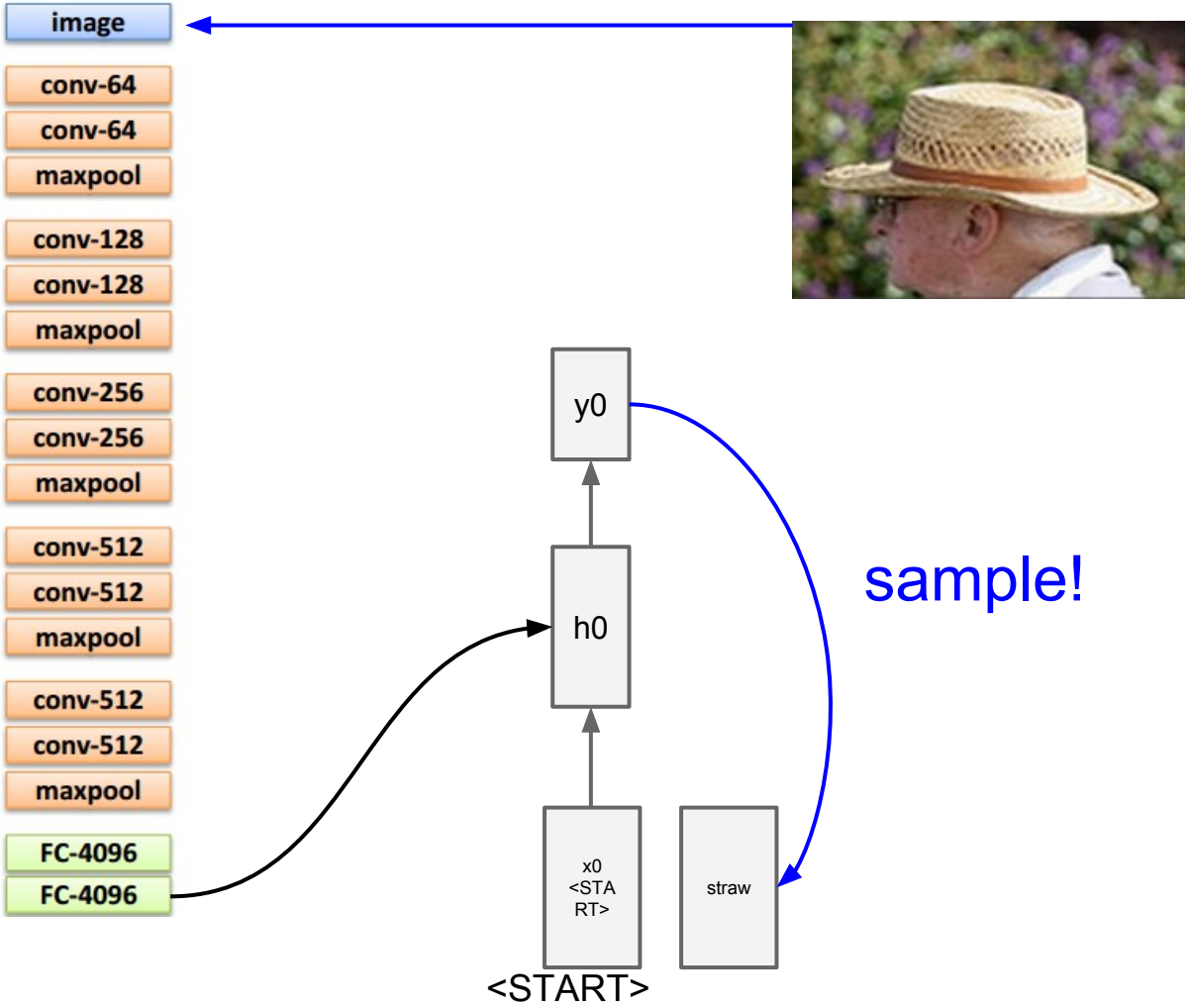
x0
<STA
RT>

<START>

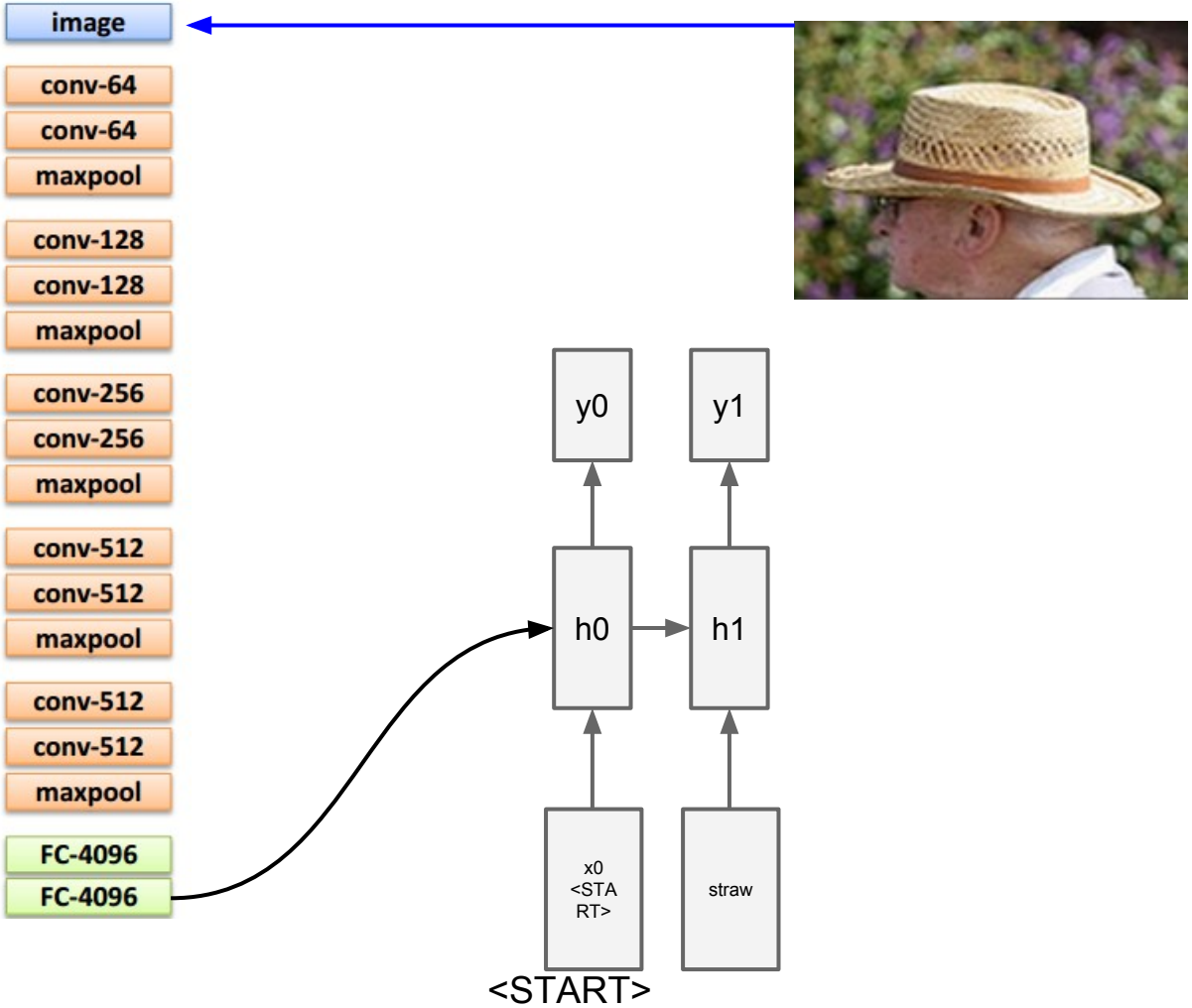
Slide credit: Andrej Karpathy



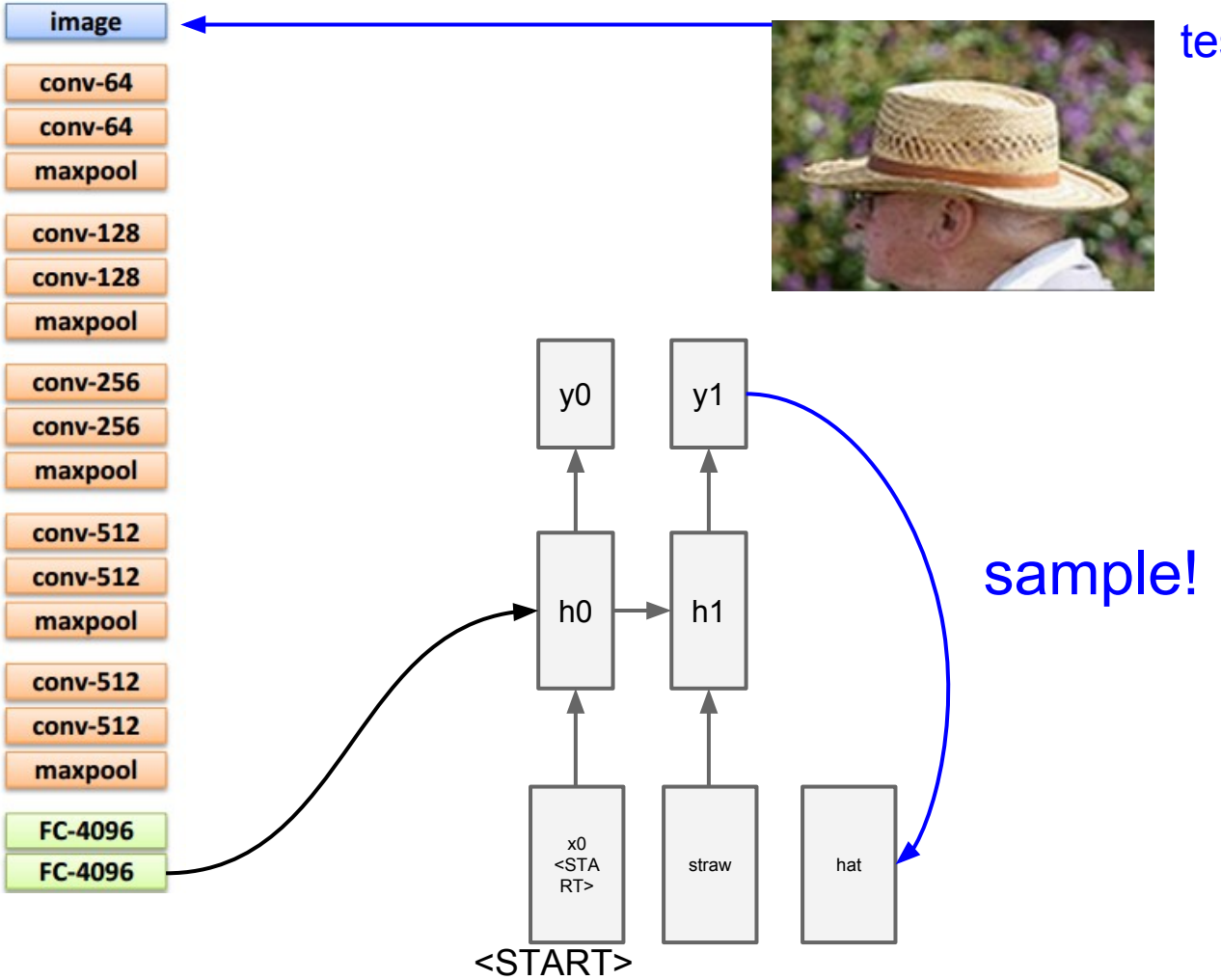
Slide credit: Andrej Karpathy



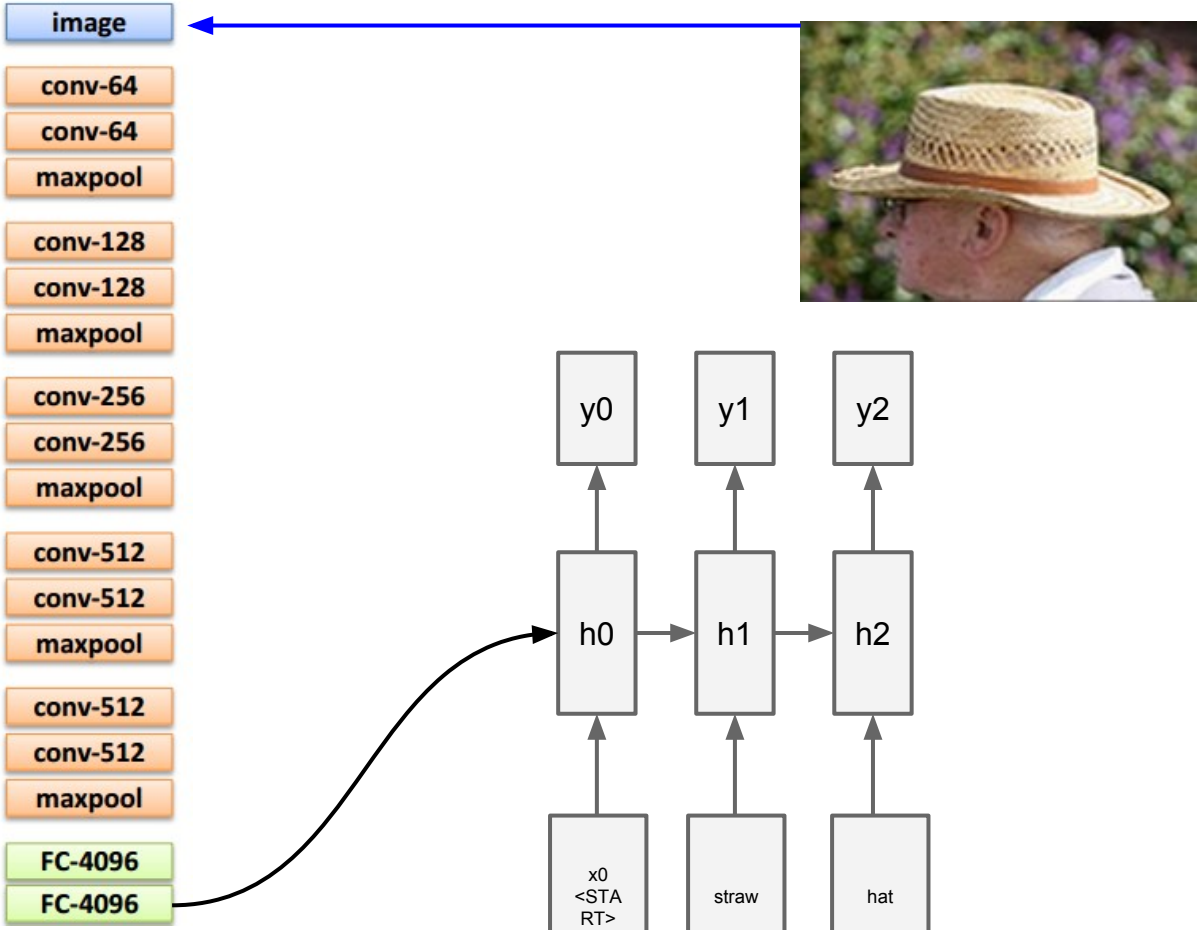
Slide credit: Andrej Karpathy



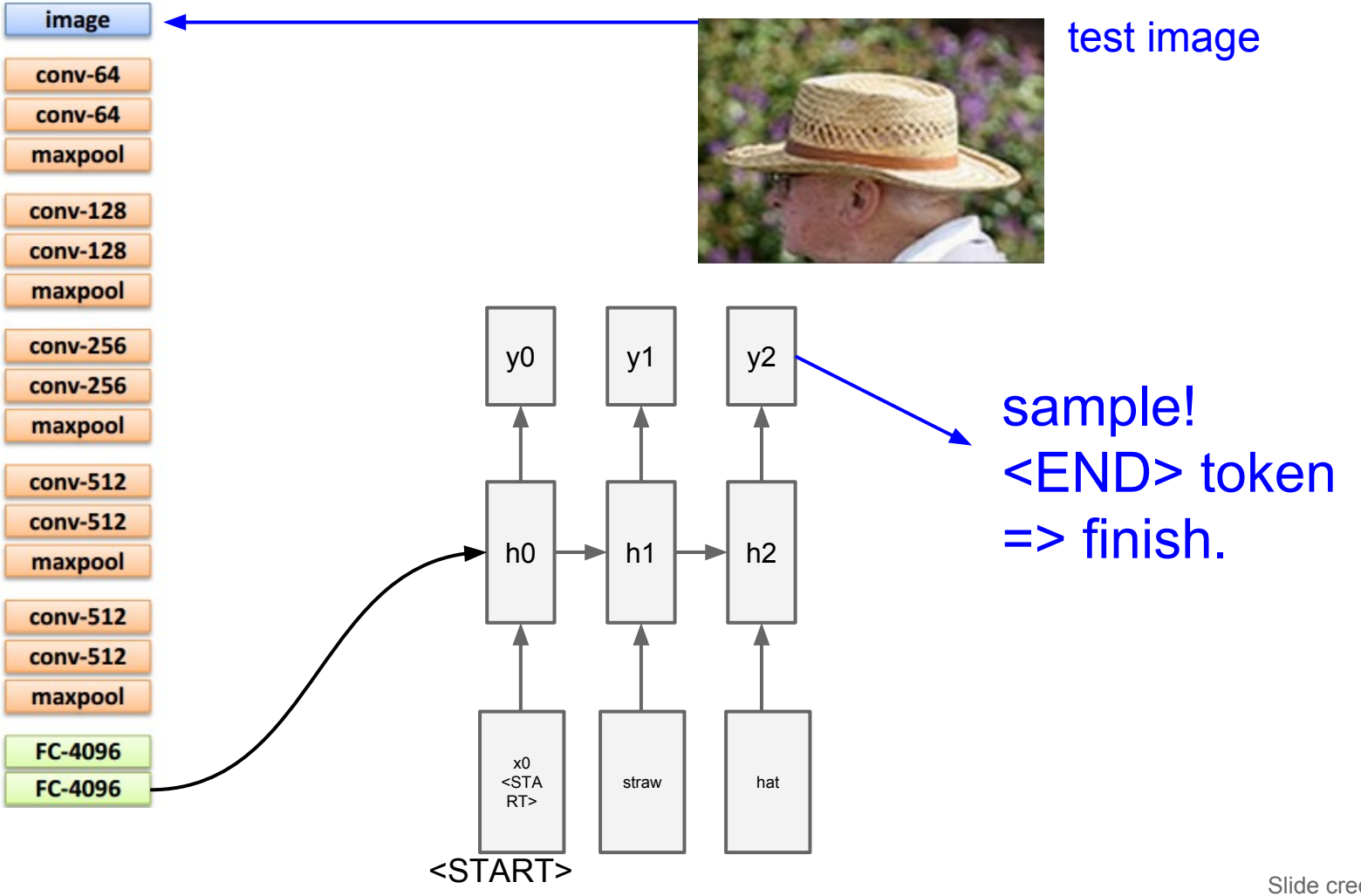
Slide credit: Andrej Karpathy



Slide credit: Andrej Karpathy



Slide credit: Andrej Karpathy



Slide credit: Andrej Karpathy



a group of people standing around a room with remotes
logprob: -9.17



a young boy is holding a baseball bat
logprob: -7.61



a cow is standing in the middle of a street
logprob: -8.84

Slide credit: Andrej Karpathy



a cat is sitting on a toilet seat
logprob: -7.79



a display case filled with lots of different types of donuts
logprob: -7.78



a group of people sitting at a table with wine glasses
logprob: -6.71

Slide credit: Andrej Karpathy



a baby laying on a bed with a stuffed bear
logprob: -8.66

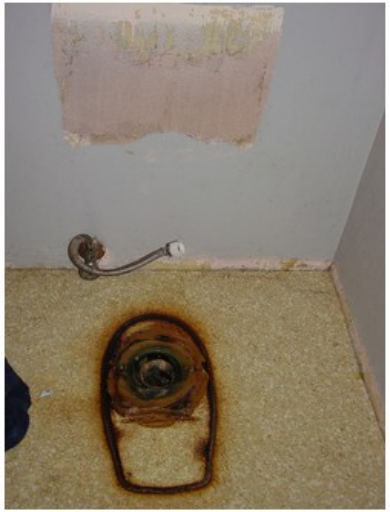


a table with a plate of food and a cup of coffee
logprob: -9.93



a young boy is playing frisbee in the park
logprob: -9.52

Slide credit: Andrej Karpathy



a toilet with a seat up in a bathroom
logprob: -13.44



a woman holding a teddy bear in front of a mirror
logprob: -9.65



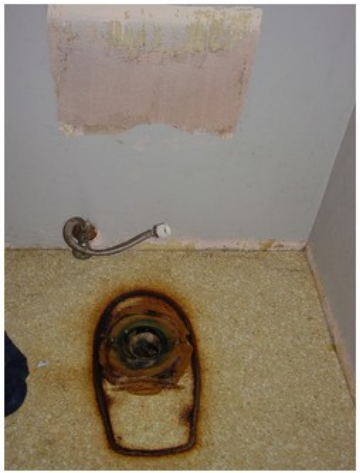
a horse is standing in the middle of a road
logprob: -10.34

Slide credit: Andrej Karpathy

Future Work

Context

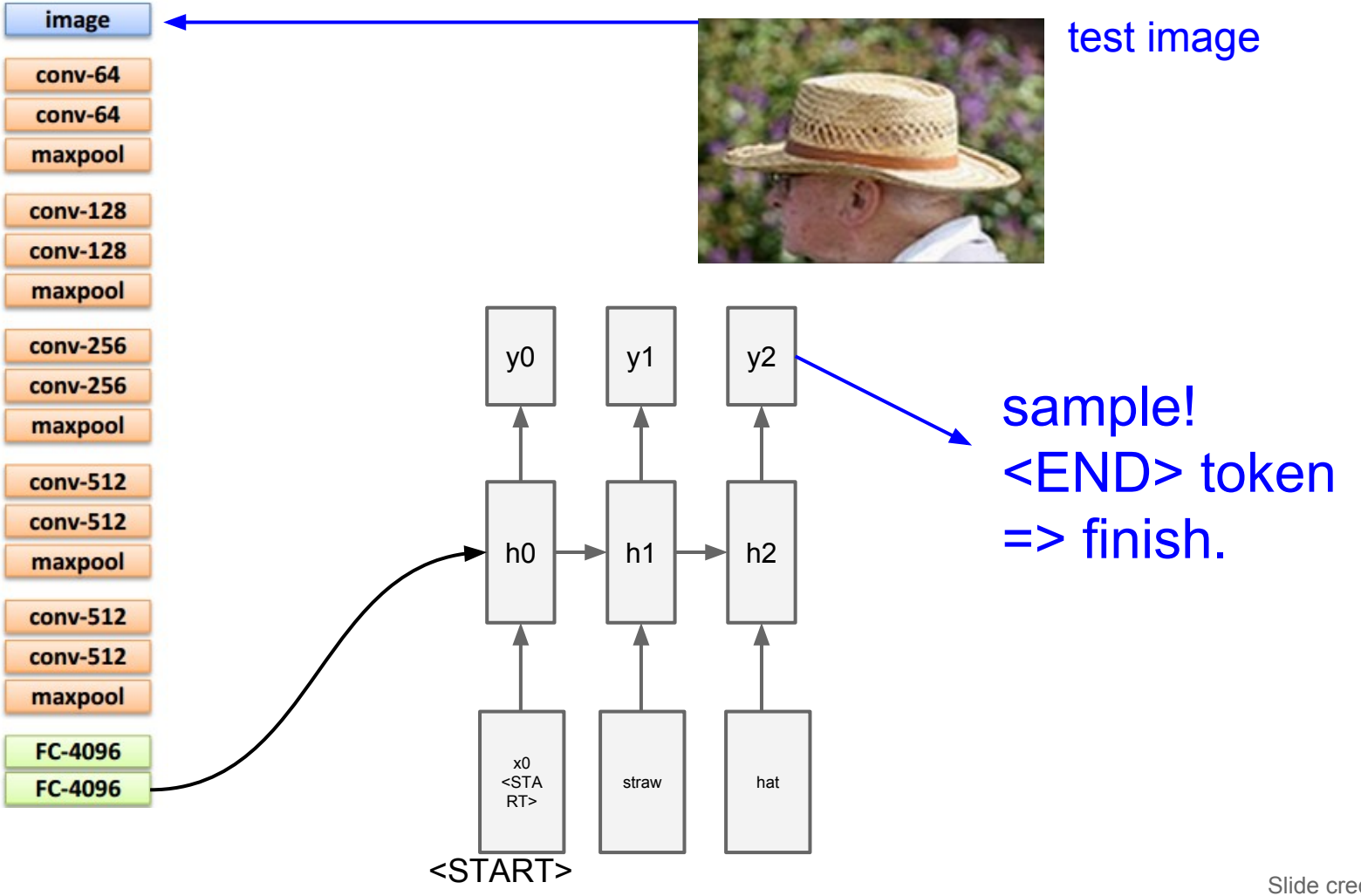
- Generation model relies heavily on context
- Model can fail when context is misleading
- Solve with feedback?



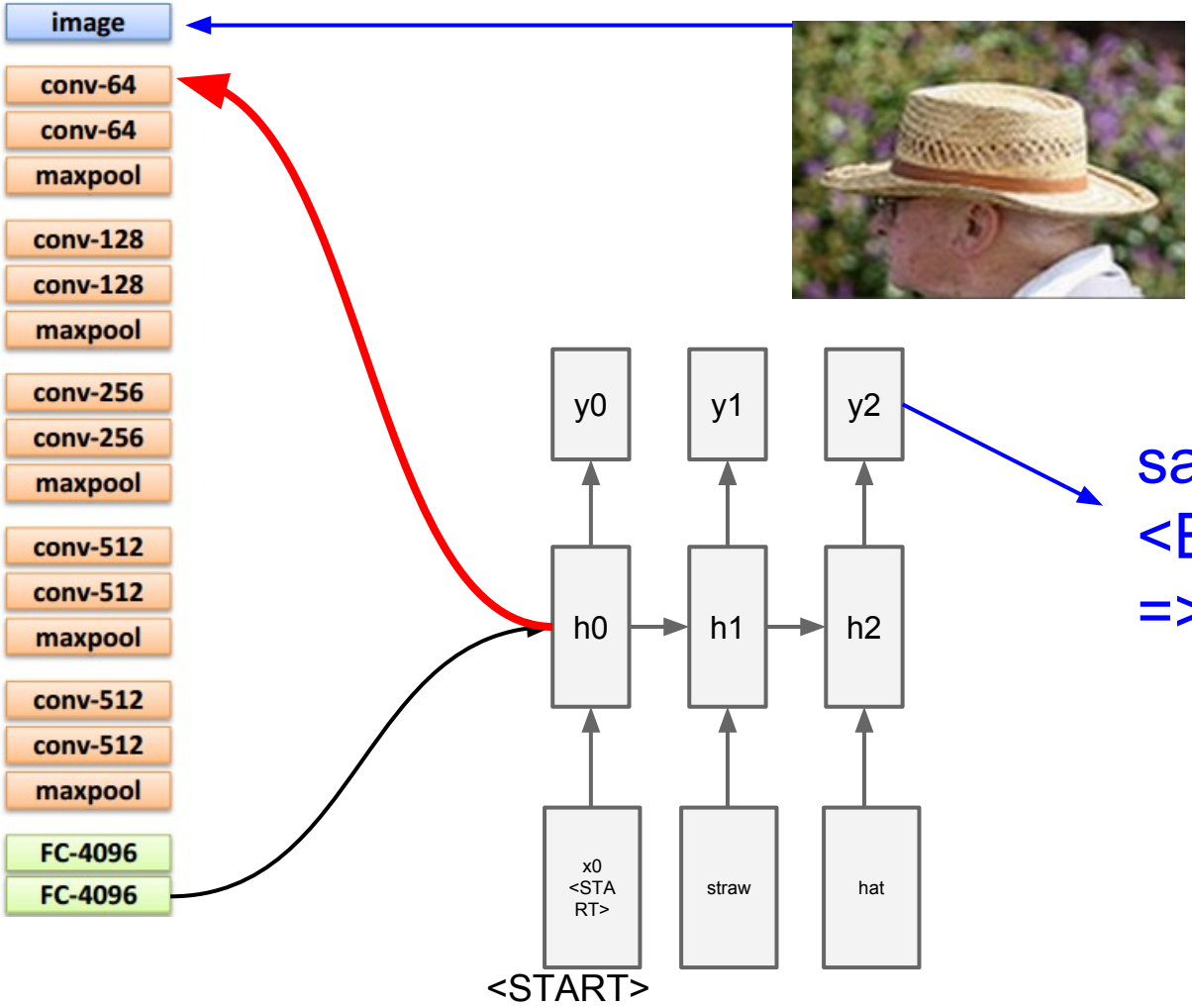
a toilet with a seat up in a bathroom
`logprob: -13.44`



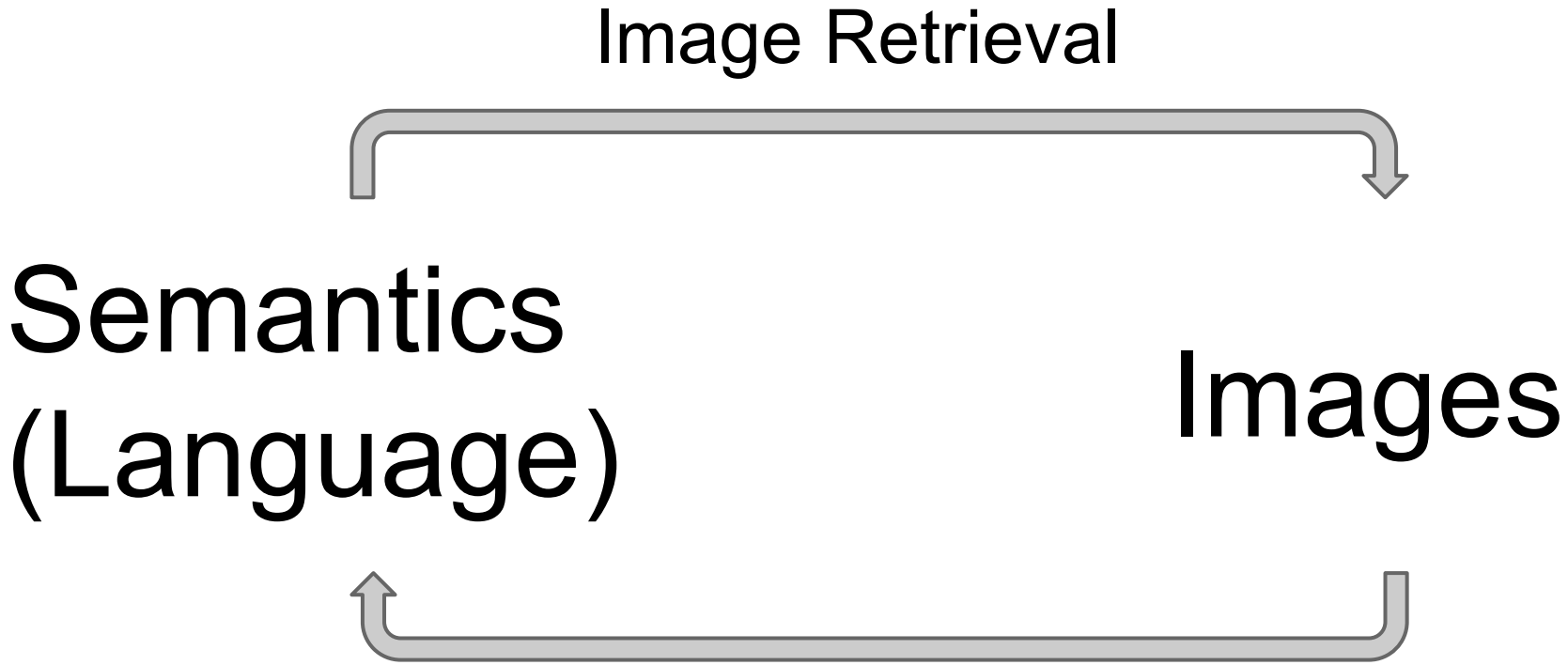
a table with a plate of food and a cup of coffee
`logprob: -9.93`



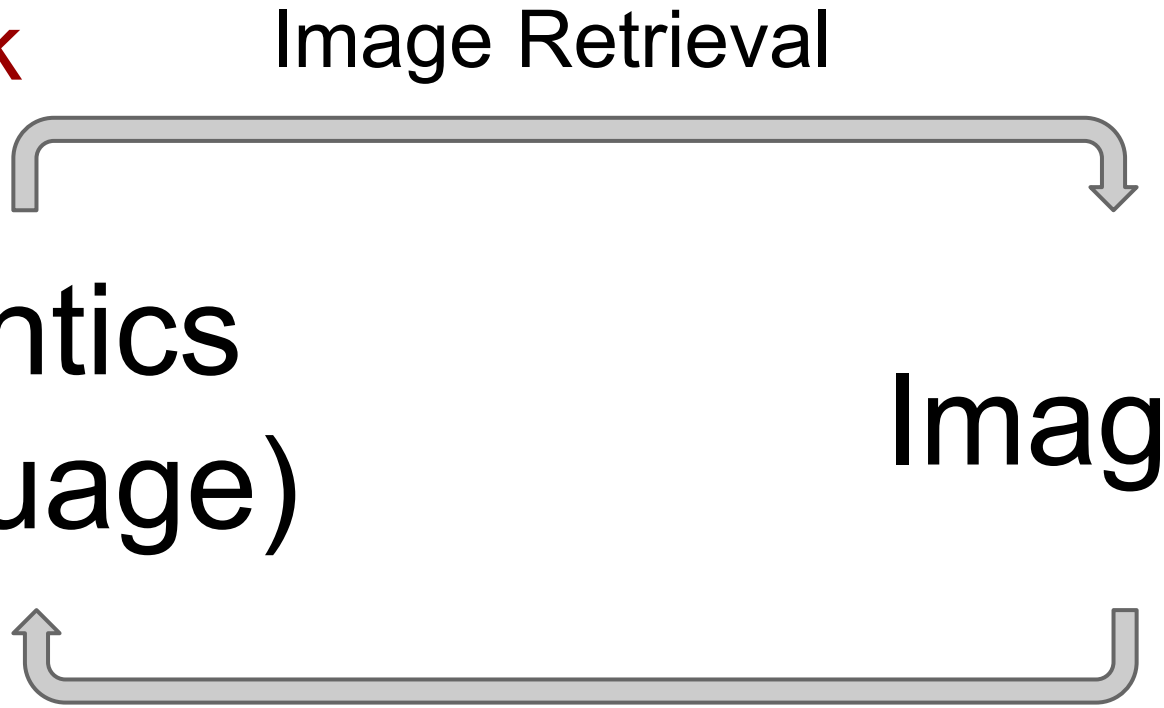
Slide credit: Andrej Karpathy



Slide credit: Andrej Karpathy


Image Retrieval
Semantics
(Language)
Images
Sentence Generation

Cross-task feedback



extra slides



“straw hat”

training example

Slide credit: Andrej Karpathy

image

conv-64
conv-64
maxpool

conv-128
conv-128
maxpool

conv-256
conv-256
maxpool

conv-512
conv-512
maxpool

conv-512
conv-512
maxpool

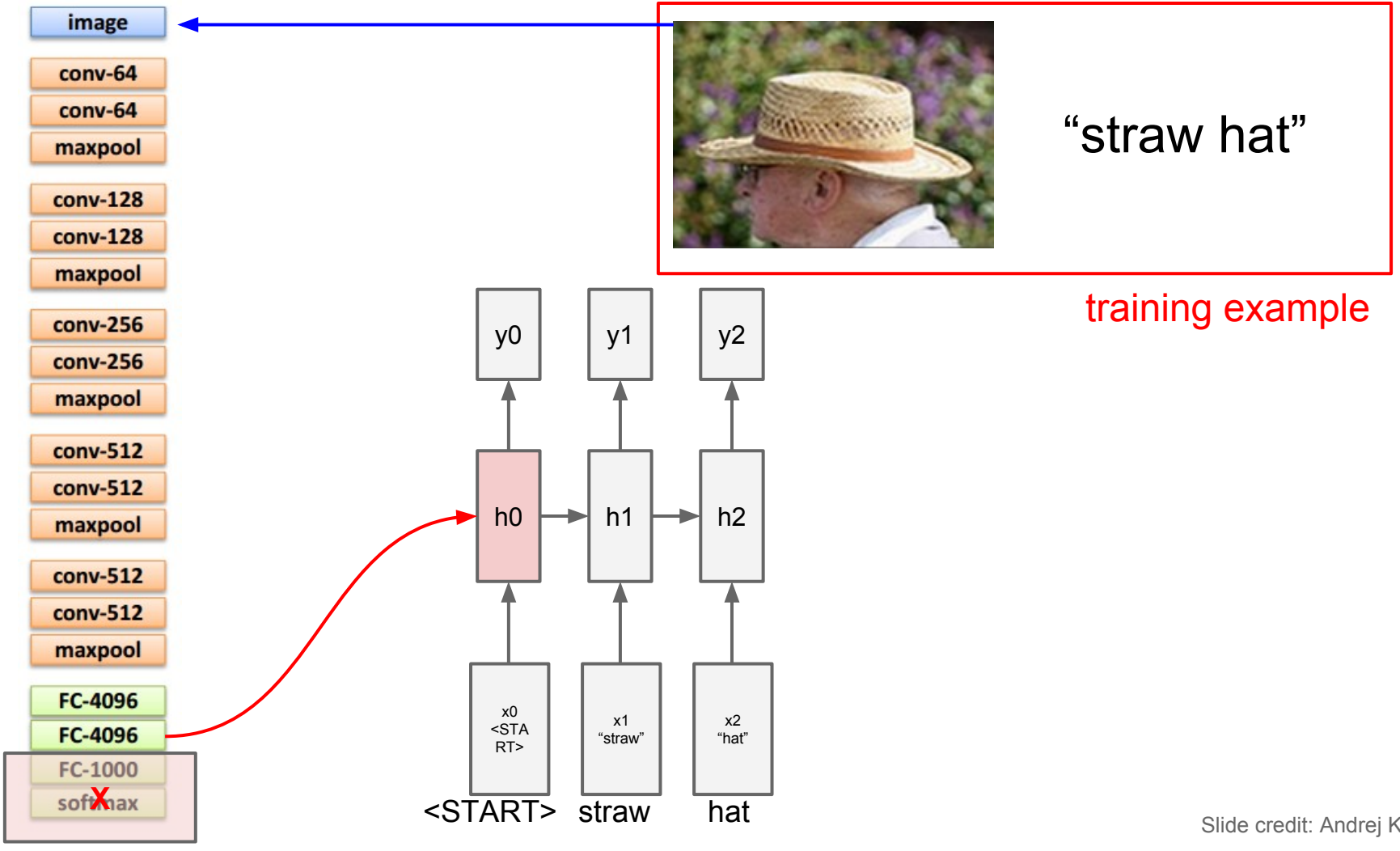
FC-4096
FC-4096
FC-1000
softmax



“straw hat”

training example

Slide credit: Andrej Karpathy



Slide credit: Andrej Karpathy



a man standing next to a clock on a wall
`logprob: -10.08`



a young boy is holding a baseball bat
`logprob: -7.65`



a cat is sitting on a couch with a remote control
`logprob: -12.45`

Slide credit: Andrej Karpathy