Natural Scene Perception

PSY3280 - Week 10 Lecture (01 Oct 2018)

Rafik Hadfi
Zhao Hui Koh

Learning Objective - Natural Scene

Human?

Machine?

Images retrieved from https://www.planetminecraft.com/project/minecraft-poem-scene/ and https://www.expedia.com/pictures/usa/washington-state.d249/

Natural scene - Human perception

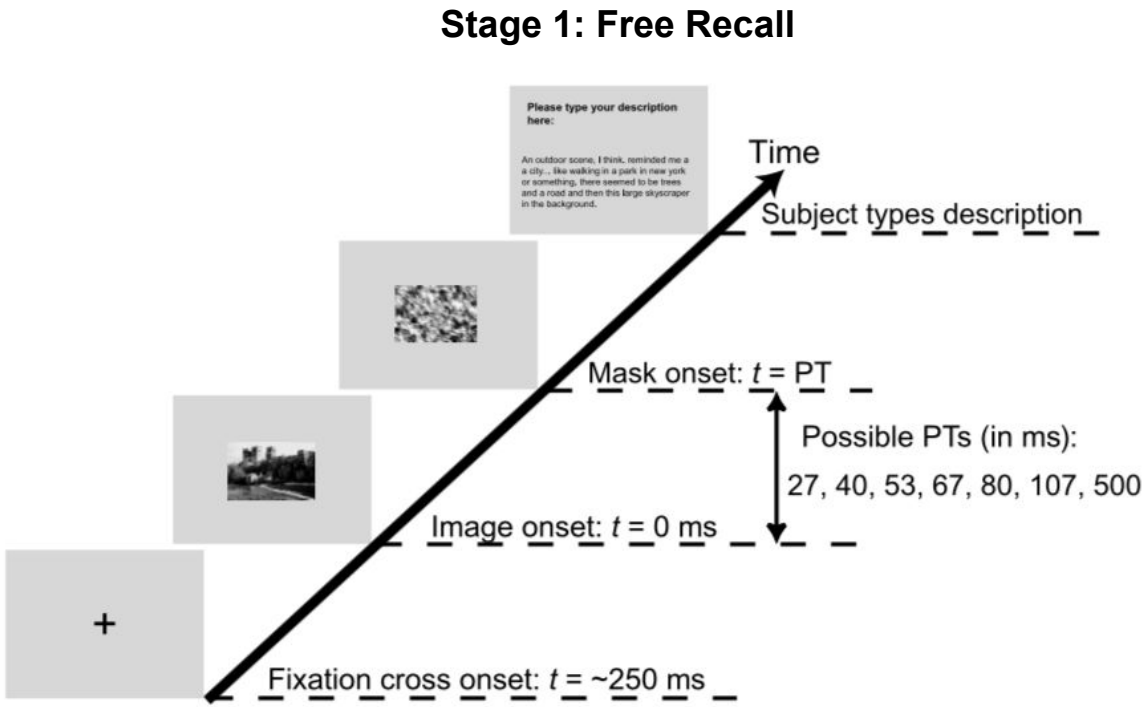
Natural Scene 1 - Words to describe the image?

Natural Scene 2 - Words to describe the image?

Natural Scene 3 - Words to describe the image?

Richness of a natural scene (Fei-Fei et al., 2007)

- The content/information of “gist”
- Objects, Physical Appearance, Spatial relations between objects
- Global semantic/context
- Hierarchical relationship (taxonomy) of object categories



Sample responses (Features and Semantic)

Easy

PT = 107 ms

This is outdoors. A black, furry dog is running/walking towards the right of the picture. His tail is in the air and his mouth is open. Either he had a ball in his mouth or he was chasing after a ball. (Subject EC)

PT = 500 ms

I saw a black dog carrying a gray frisbee in the center of the photograph. The dog was walking near the ocean, with waves lapping up on the shore. It seemed to be a gray day out. (Subject JB)

Complex

27 ms

Looked like something black in the center with four straight lines coming out of it against a white background. (Subject: AM)

500 ms

This looks like a father or somebod helping a little boy. The man had something in his hands, like a LCD screen or laptop. they looked like they were standing in a cubicle. (Subject: WC)

(Fei-Fei et al., 2007)

Findings

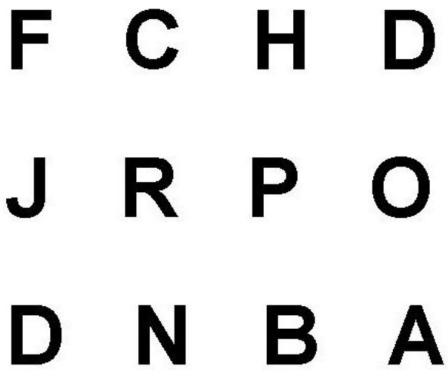
- Richness of perception is asymmetrical (object and scene recognition)
  - Preference of outdoor (vs indoor) if visual information is scarce (small PT)
- Seem to be able to recognise objects at a superordinate category level (e.g. vehicle) as well as basic category levels (e.g. train, plane, car)
- Single fixation is sufficient for recognition of most common scenes and activities
- Sensory information (shape recognition) vs higher level conceptual information (object identification, object/scene categorisation)

(Fei-Fei et al., 2007)

Quantify richness in visual experience

- Sperling's experiment (1960) - limited capacity of phenomenal vision
- Limitations of past studies on richness of visual experience (Haun et al., 2017)
    - Controlled experiments - what a participant can report on (high-level categorical response, binary choice)
- *"Participants were not asked"*

- Previous paradigms have underestimated the amount of information available for conscious report from brief exposures to the stimulus.



(Haun et al., 2017)

Images retrieved from http://psychologyexamrevision.blogspot.com/2012/01/sperling-1960.html

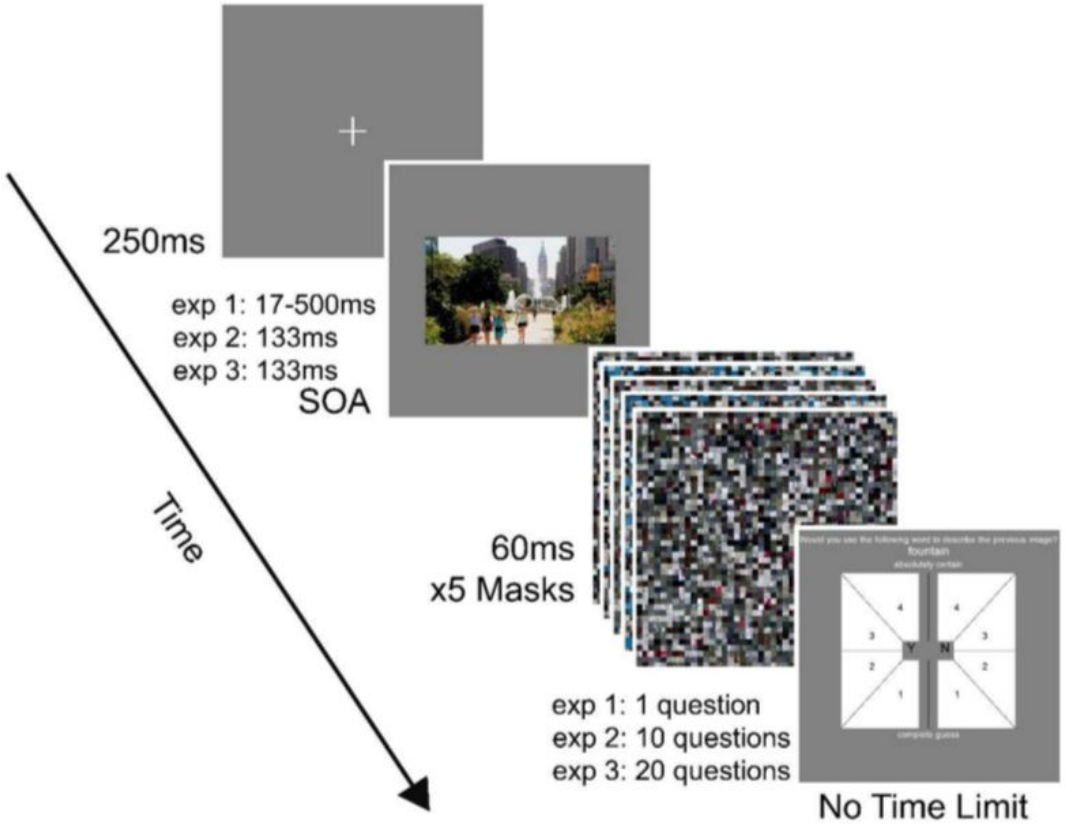
Richness - Bandwidth of Consciousness (BoC)

- IIT - Information axiom - Distinguishable from every other possible experience

- How bits are measured
  - Information Theory - quantify bits of information (reduction of uncertainty)
  - Yes/no question from an image (presented for 1 second) - 1 bit of information
  - Past research - We can perceived up to maximum of 44 bits/second (Pierce, 1980)

- Honours Student's Project - "A Moment of Conscious Experience is Very Informative" (Loeffler, Alon, 2017)
- Quantify the amount of information people can extract from brief exposure to a natural scene

Experiment

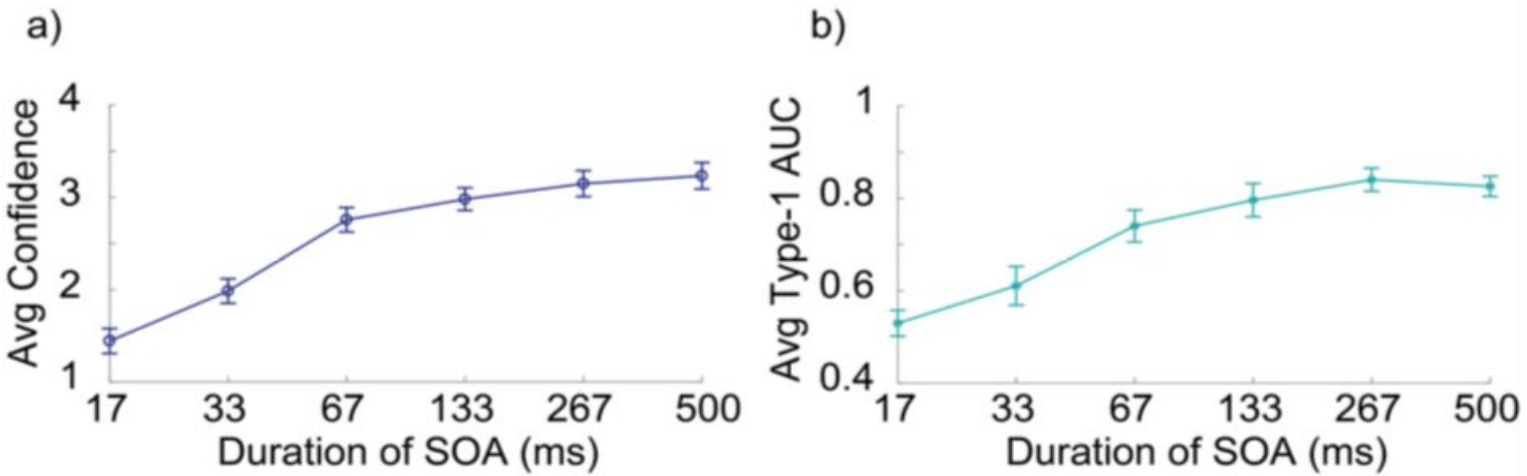


- Participants determined whether a word (descriptor) could describe the image **(present and absent)**
- Stimulus Onset Asynchrony (SOA - time between image onset and mask onset)
- Forced choice response (8 choices)
- Presence/Absence judgement + confidence rating

(Loeffler, Alon, 2017)

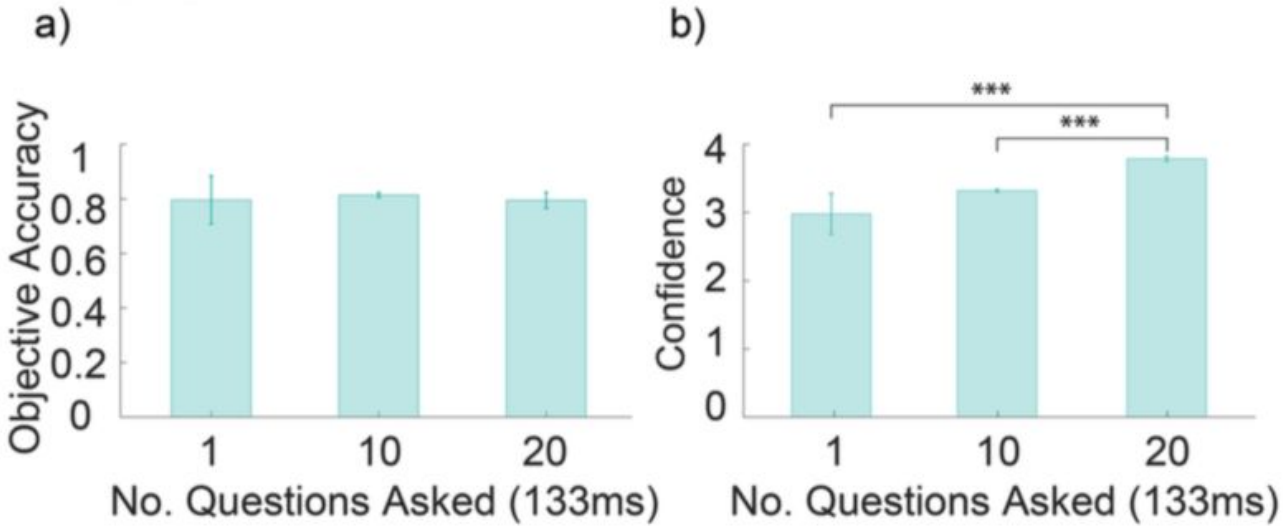
Findings

- Participants' feedback
  - Shorter SOA - bottom-up processing (features)
  - Longer SOA - top-down processing (semantic)



(Loeffler, Alon, 2017)

Findings (cont'd)



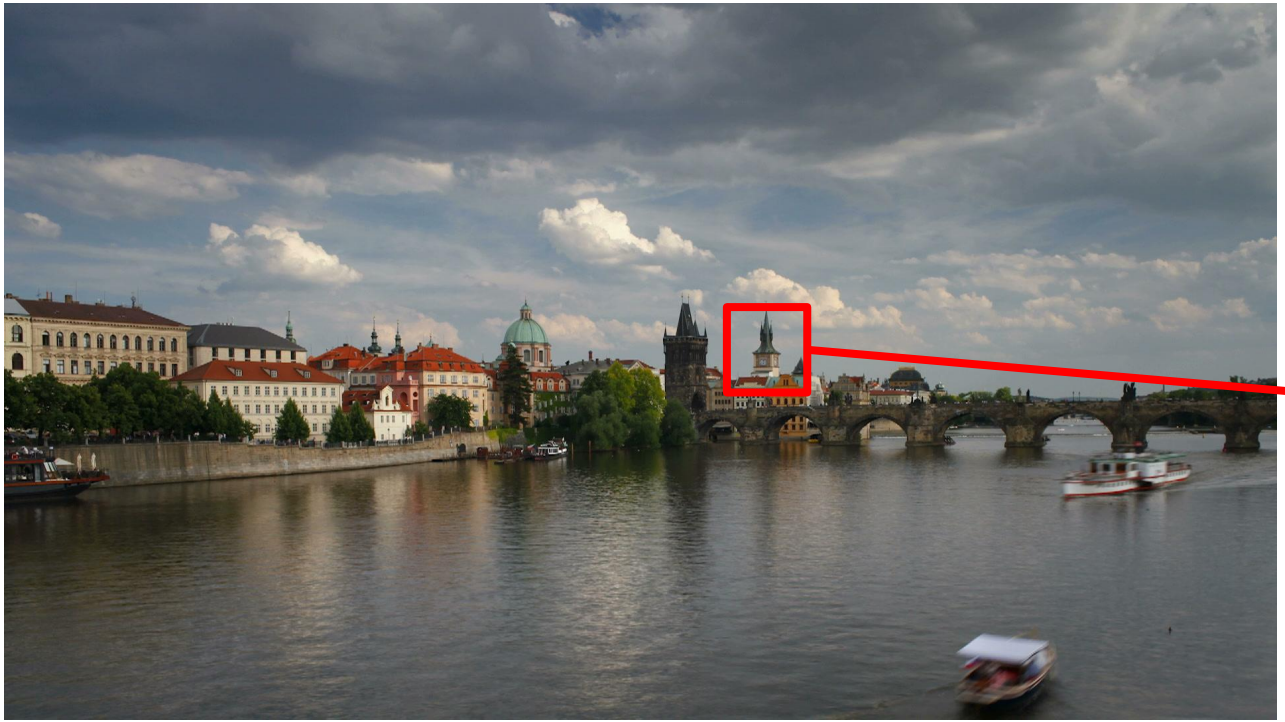
SOA: 133ms

Exp 2 (10 questions/image)	Exp 3 (20 questions/image)
52 bits/sec	100 bits/sec

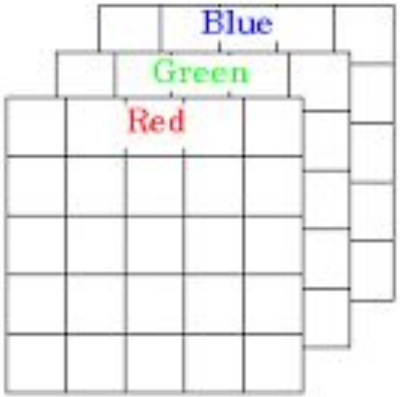
(Loeffler, Alon, 2017)

Natural scene - Machine perception

How machine sees image?



Pixel matrices with
RGB values

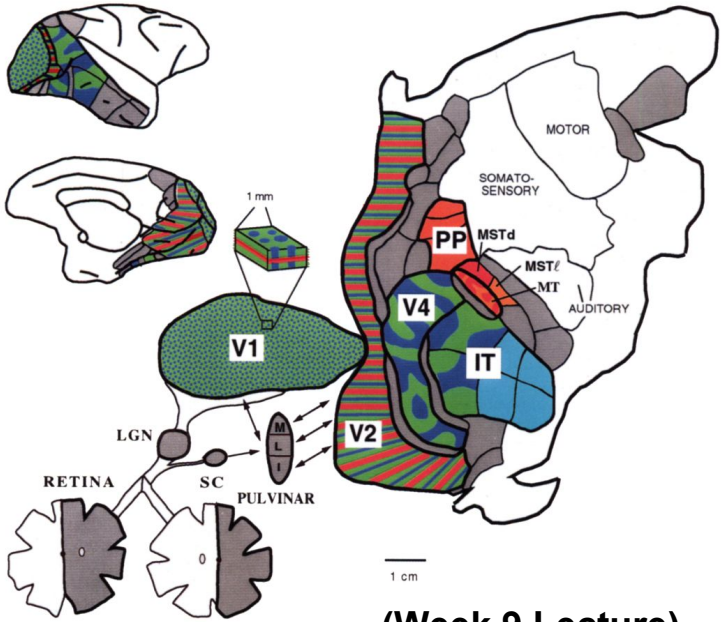


Images retrieved from Nishimoto (2015) and https://www.ini.uzh.ch/~ppyk/BasicsOfInstrumentation/matlab_help/visualize/coloring-mesh-and-surface-plots.html

Machine learning in image perception

- Convolutional Neural Network (Image recognition & classifications, object detection, face recognition, cameras, robots)

Apple Face ID

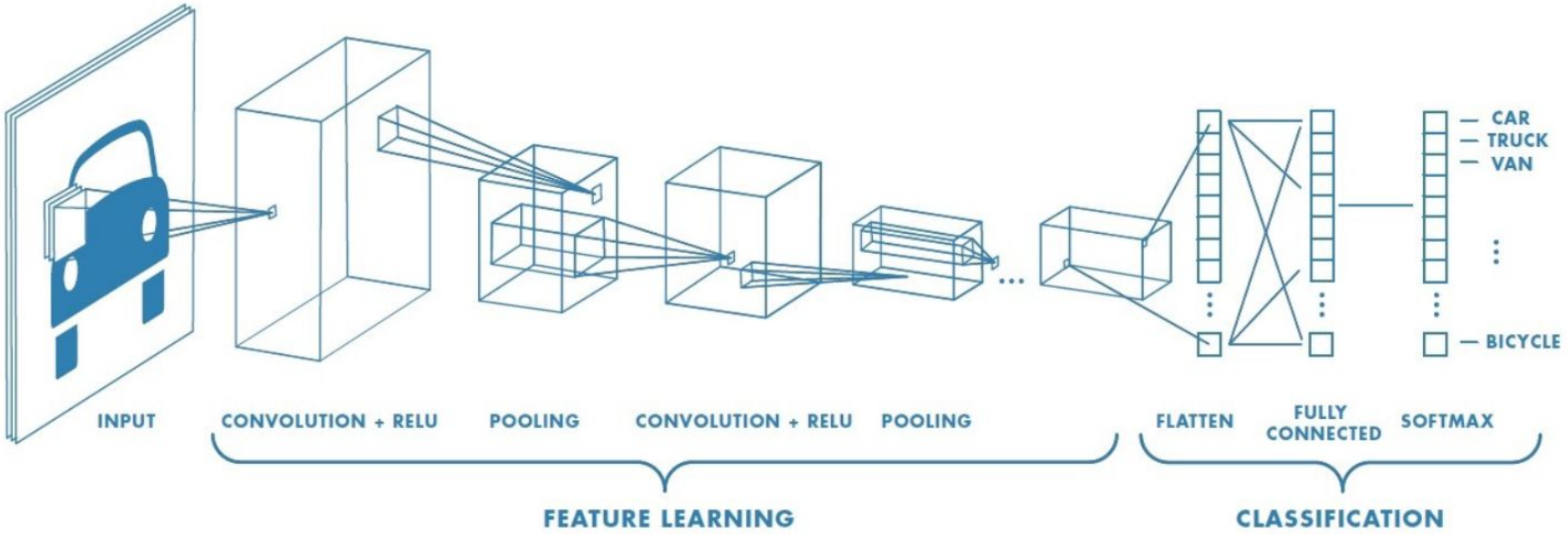


- Inspired by primate visual system

(Week 9 Lecture)

Images retrieved from https://support.apple.com/en-us/HT208109 and Van Essen, & Gallant (1994)

Convolutional Neural Network (ConvNet)



Retrieved from https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

Convolutional layer

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

1	0	1
0	1	0
1	0	1

Filter

$1_{\times 1}$	$1_{\times 0}$	$1_{\times 1}$	0	0
$0_{\times 0}$	$1_{\times 1}$	$1_{\times 0}$	1	0
$0_{\times 1}$	$0_{\times 0}$	$1_{\times 1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

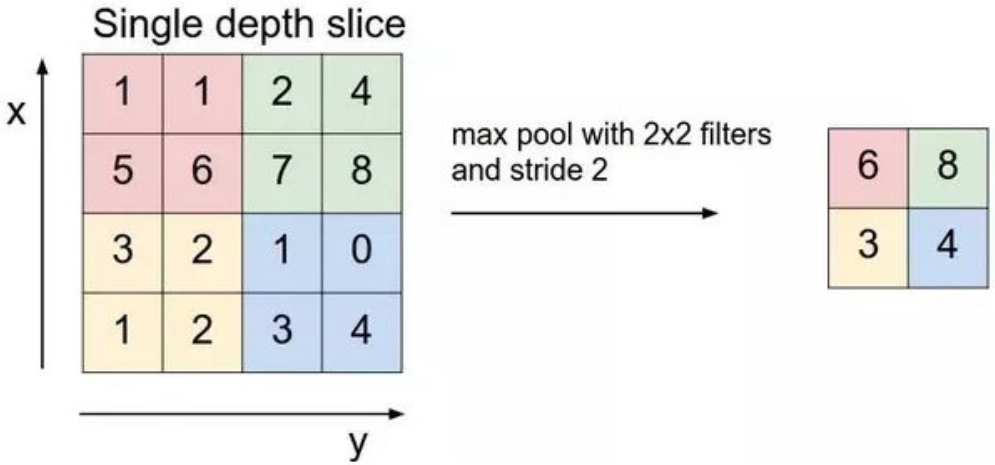
Convolved Feature

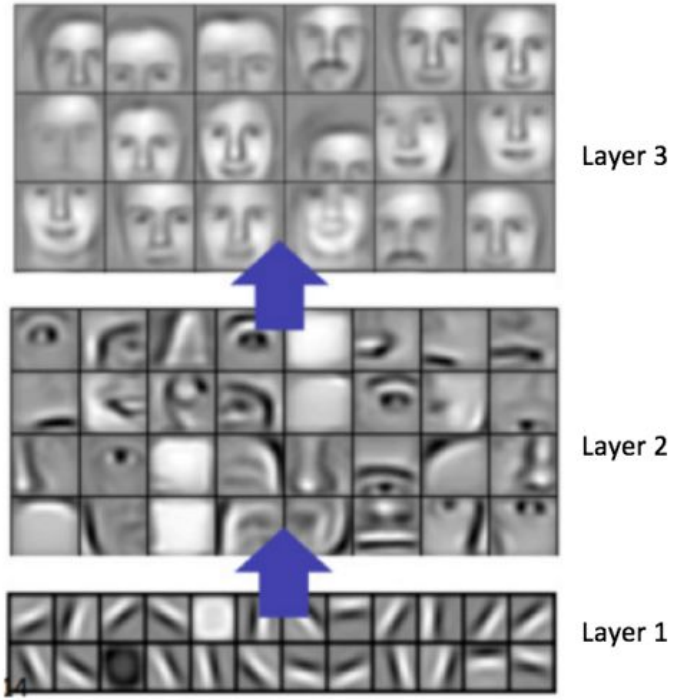
(Activation map/Feature Map)

Images retrieved from https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Pooling layer

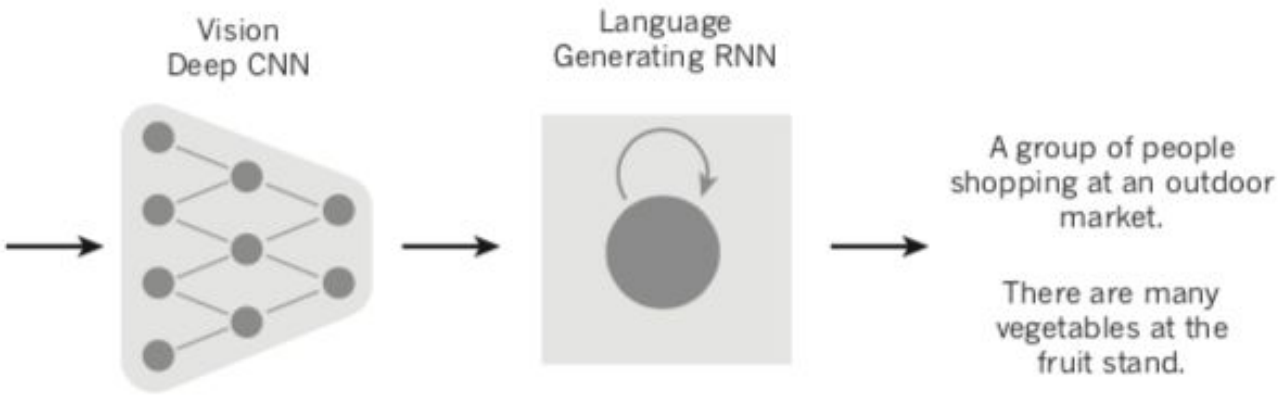
- Spatial reduction





Images retrieved from https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

“Show and Tell” - Natural scene captions



Captions?

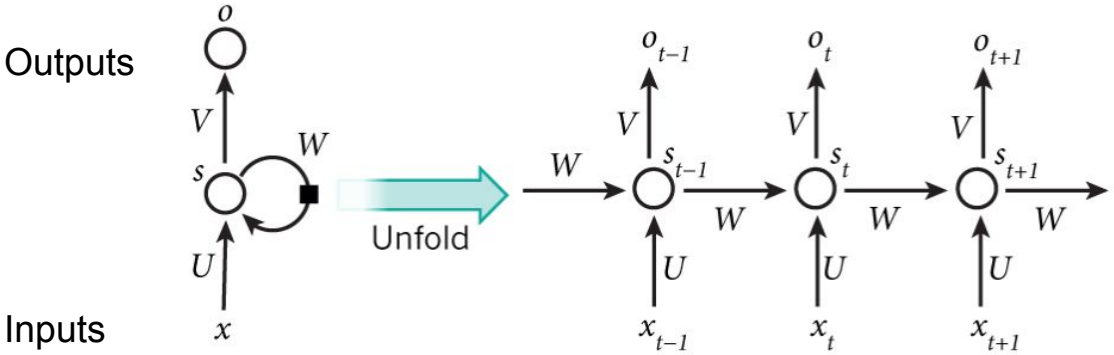
Encoder

Decoder

(Vinyals et al., 2015; LeCun et al., 2015)

Recurrent neural networks

- Best for **sequential** input tasks - speech and language
- Process one element at a time and use hidden units to keep past history (feedback/recurrent)



- Machine translation (encoder + decoder)
  - English -> French
  - Image -> Caption

(LeCun et al., 2015)

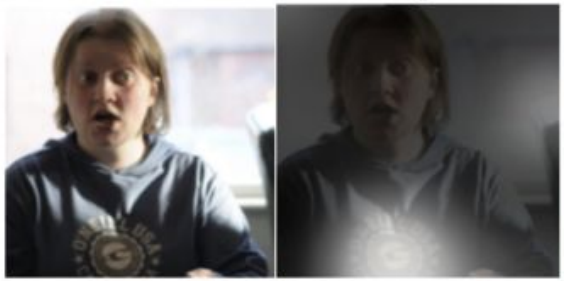
“Show, Attend and Tell” - Attention based

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

A large white bird standing in a forest.



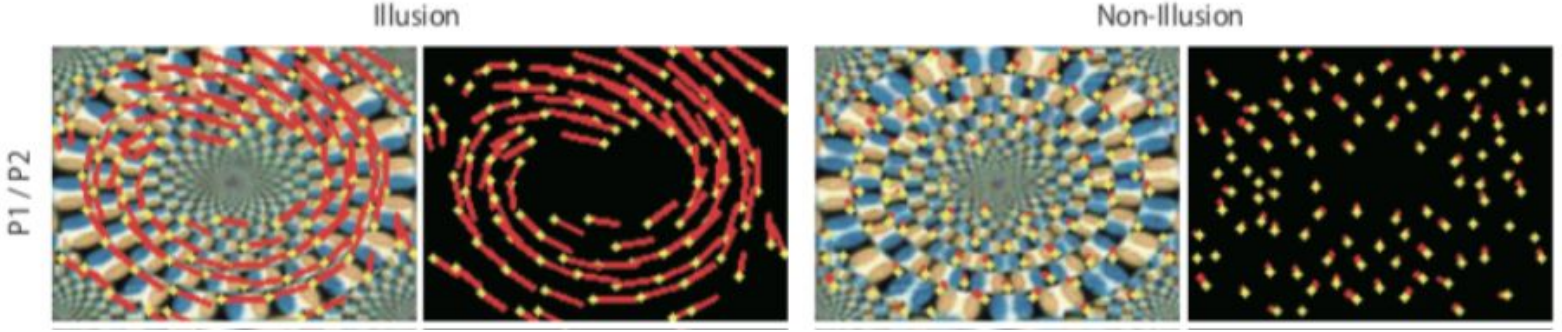
A woman holding a clock in her hand.

A man wearing a hat and a hat on a skateboard.

(Kelvin et al., 2016; LeCun et al., 2015)

Discussion

- Can an artificial neural network (e.g. ConvNet) experience visual illusion, change blindness, binocular rivalry?
  - PredNet (Watanabe et al., 2018) Rotating Snake Illusion



- Is an artificial neural network conscious?

References

- Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Computational Biology*, *10*(12), e1003963–18. http://doi.org/10.1371/journal.pcbi.1003963
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10–29. http://doi.org/10.1167/7.1.10
- Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017). Are we underestimating the richness of visual experience?, *2017*(1), 817–4. http://doi.org/10.1093/nc/niw023
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. http://doi.org/10.1038/nature14539
- Loeffler, Alon (2017). *A Moment of Conscious Experience is Very Informative* (Honour's thesis). Monash University, Melbourne, Australia.
- Nishimoto, S. (2015). CiNet VideoBlocks movie library. Unpublished dataset.
- Pierce, J. R. (1980). *Introduction to Information Theory - Symbols, Signals and Noise* (2nd Ed.). Mineola, NY: Dover Publications.
- Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M., & Tanaka, K. (2018). Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction. *Frontiers in Psychology*, *9*, 1143–12. http://doi.org/10.3389/fpsyg.2018.00345
- Van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, *13*(1), 1-10.
- Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator (pp. 3156–3164). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. http://doi.org/10.1109/CVPR.2015.7298935
- Xu, K., Ba, Jimmy L, Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel., R., Bengoi., Y. (2015) Show, attend and tell: Neural image caption generation with visual attention. *Jmlr.org*