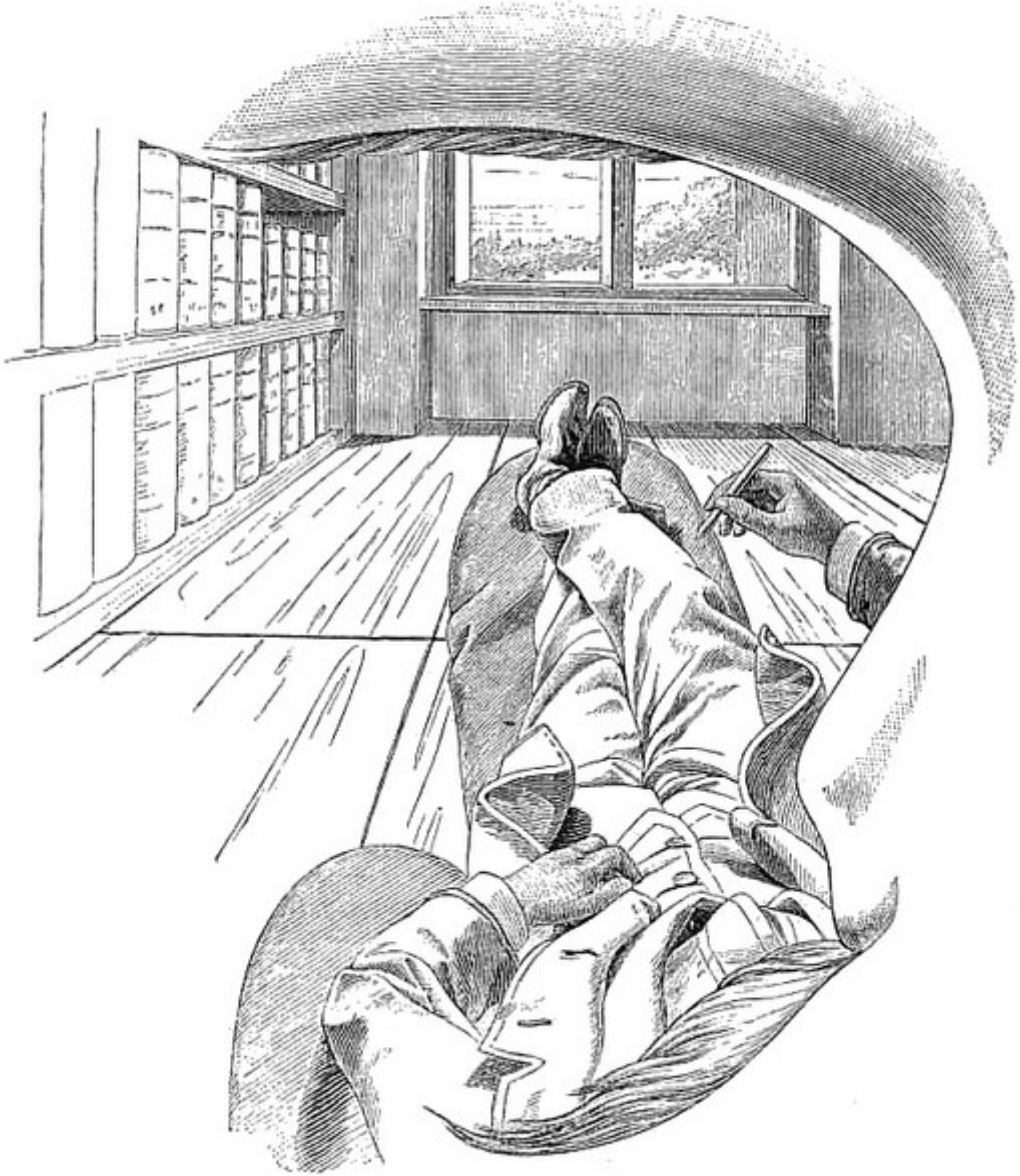
Convolutional neural network models of the first stages in biological vision

Lane McIntosh
with Niru Maheswaranathan, Aran Nayebi,
Surya Ganguli and Stephen Baccus

IEEE Signal Processing Society

June 16, 2016

Stanford University



Figur 1.

Self-portrait
Ernst Mach, 1886

Difficulties of vision

Objects can be difficult to detect

Difficulties of vision

Objects can be difficult to detect

Light levels change by 9 orders of magnitude

Difficulties of vision

Objects can be difficult to detect

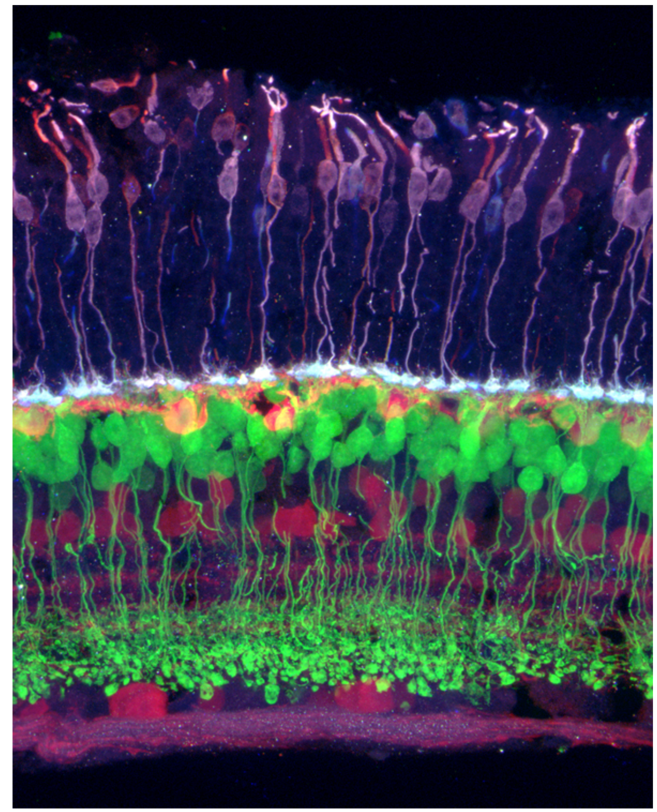
Light levels change by 9 orders of magnitude

The visual world is often low contrast

Further difficulties

The visual scene changes
with eye and body movements

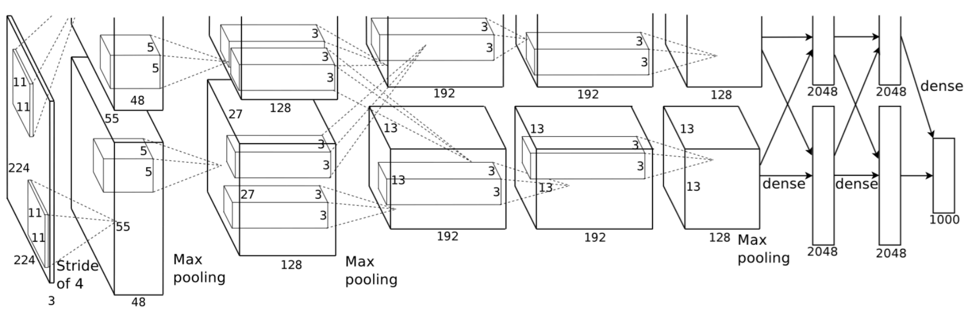
Further difficulties



The visual scene changes
with eye and body movements

Seen by a noisy,
biological system

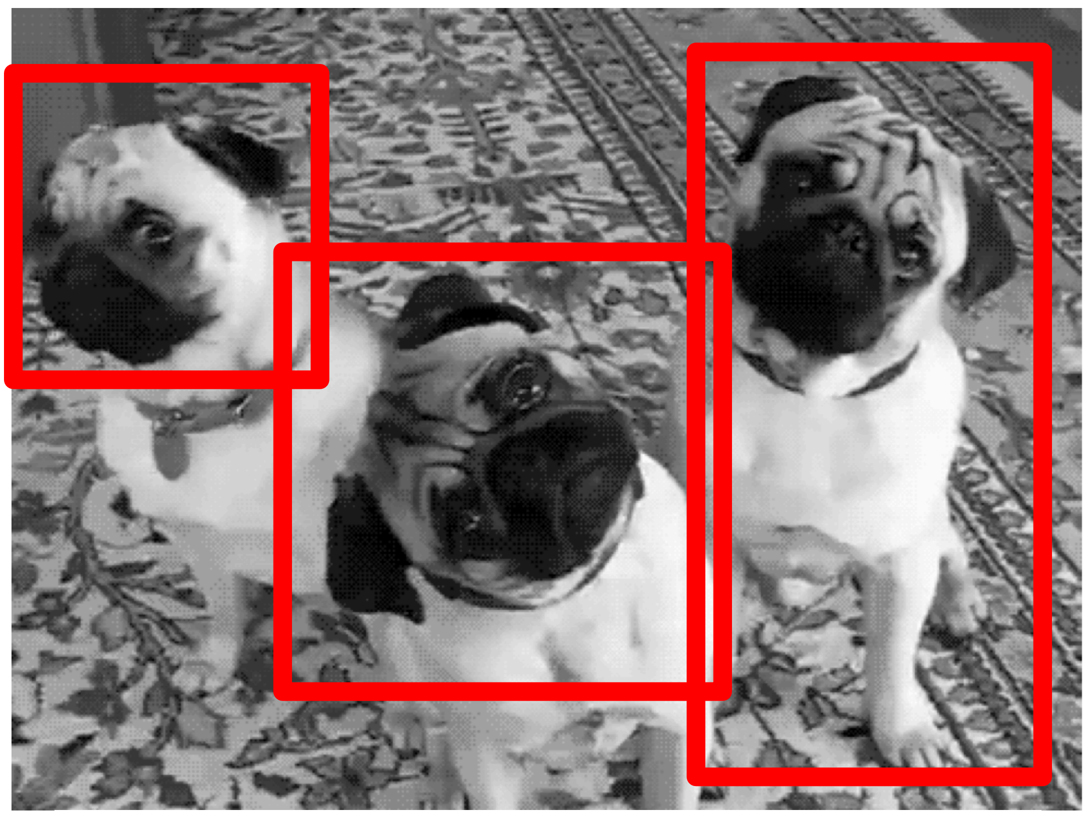
Recent advances in computer vision

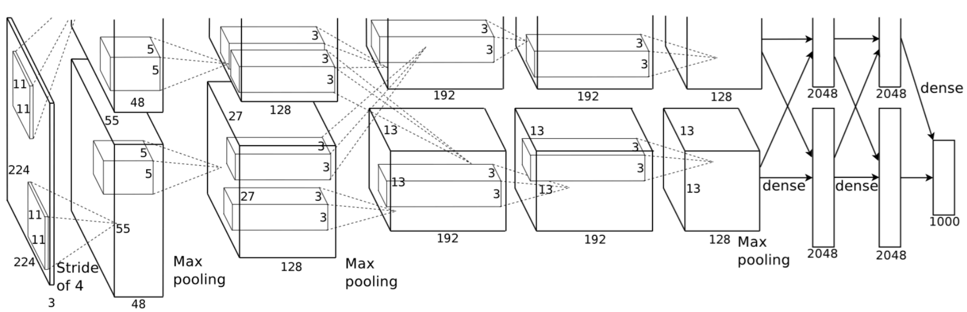


Top 5 classification:

Carpet
Sharpei
<u>Pug</u>
French Bulldog
Mastiff

Recent advances in computer vision





Top 5 classification:

Carpet

Sharpei

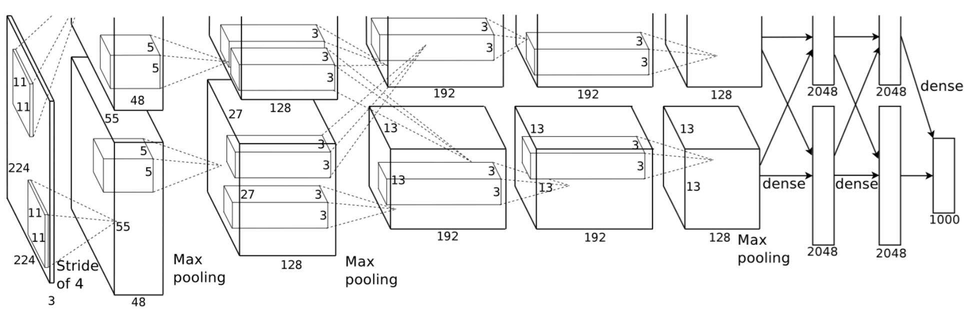
Pug

French Bulldog

Mastiff

Object localization

Recent advances in computer vision

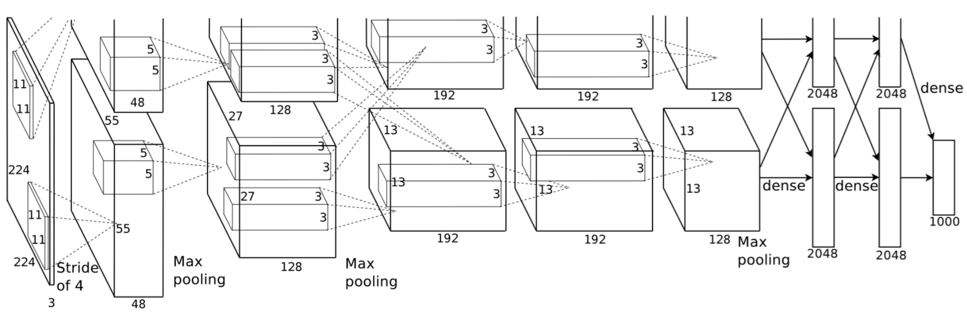


Top 5 classification:
Carpet
Sharpei
<u>Pug</u>
French Bulldog
Mastiff

Object localization

Image segmentation

Recent advances in computer vision



Top 5 classification:
Carpet
Sharpei
<u>Pug</u>
French Bulldog
Mastiff

Object localization

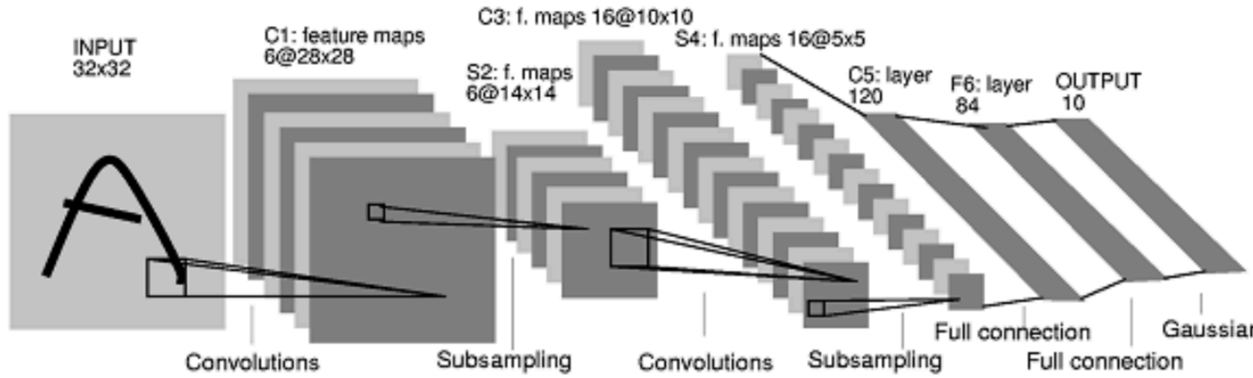
Image segmentation

Sentence captioning
"Three dogs on a carpet"

A brief history of deep learning

1998

LeCun et al.



# of transistors

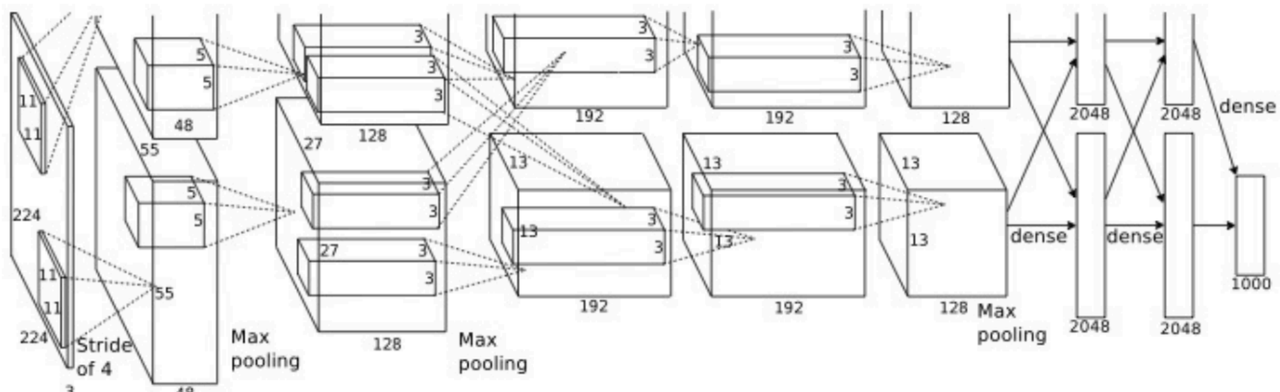
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors

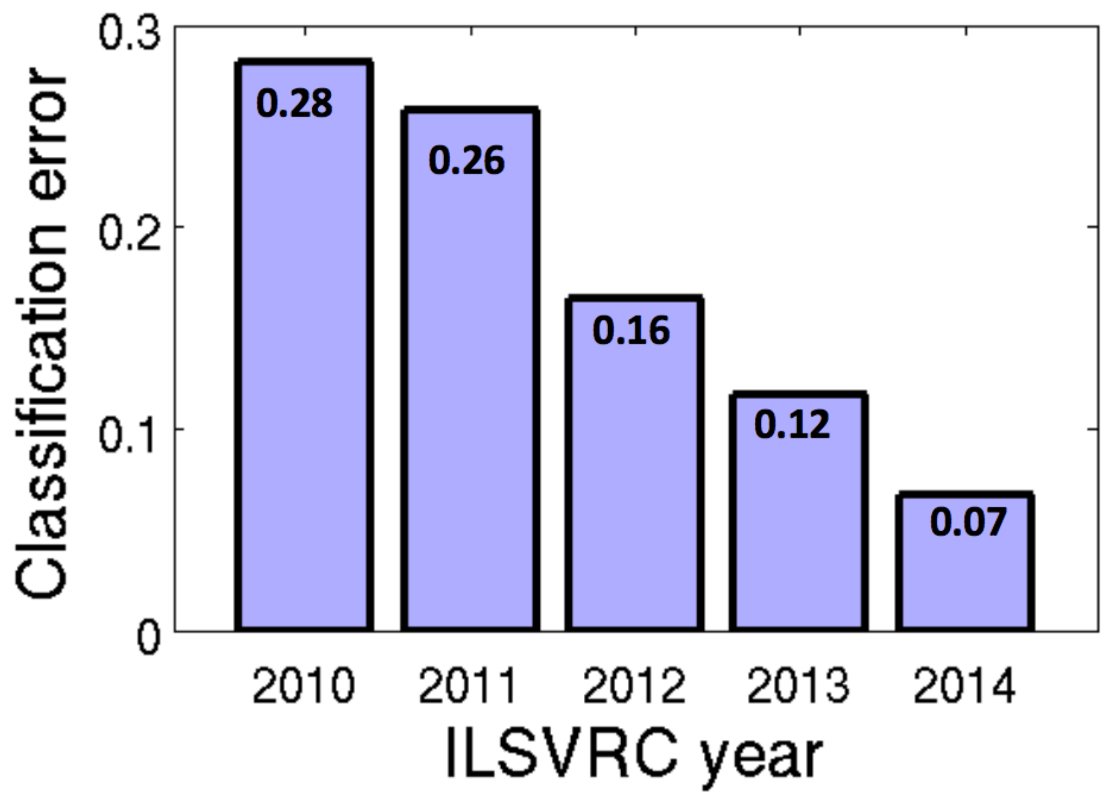
10^9

GPUs

# of pixels used in training

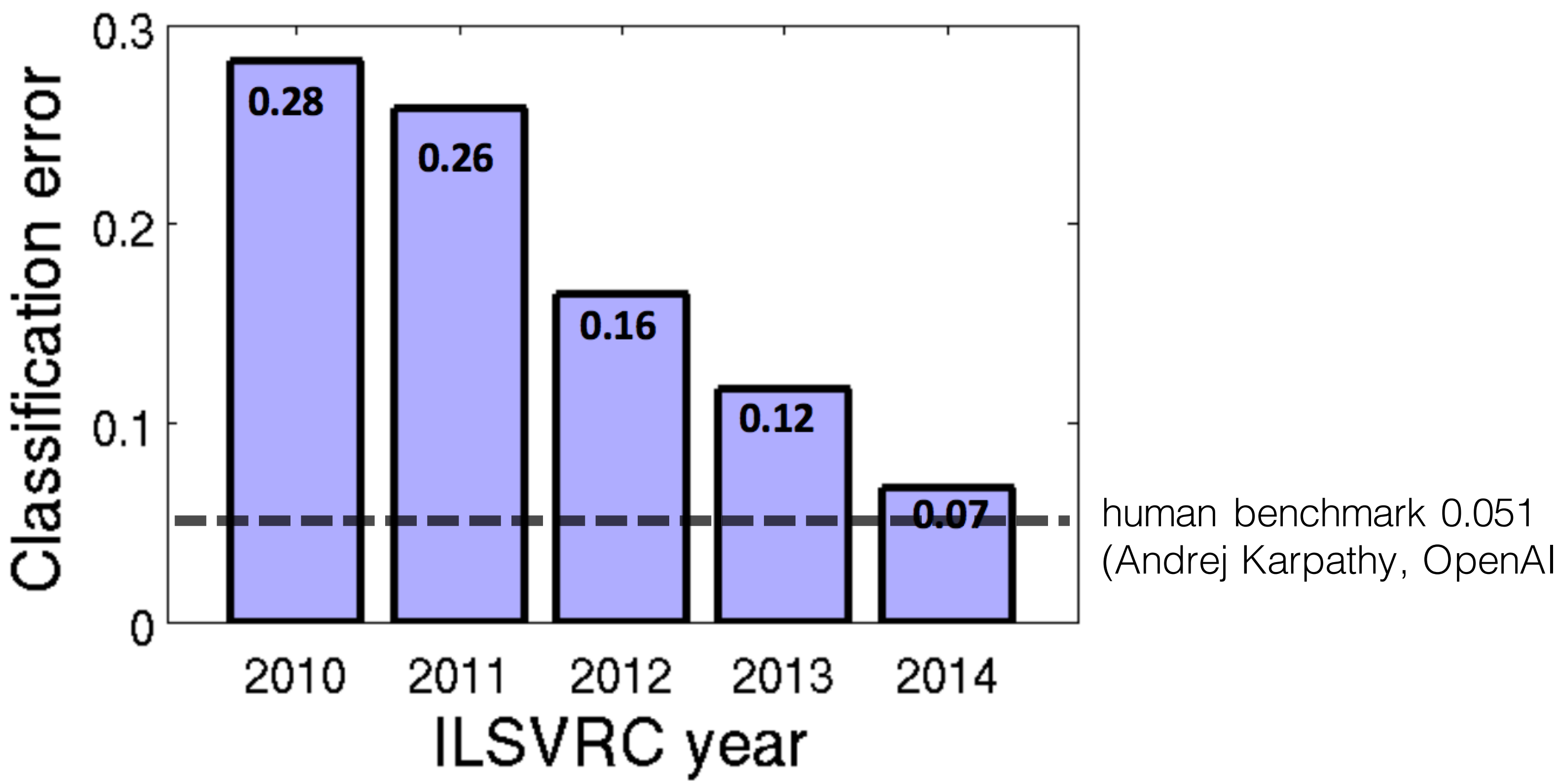
10^{14} IMAGENET

A brief history of deep learning



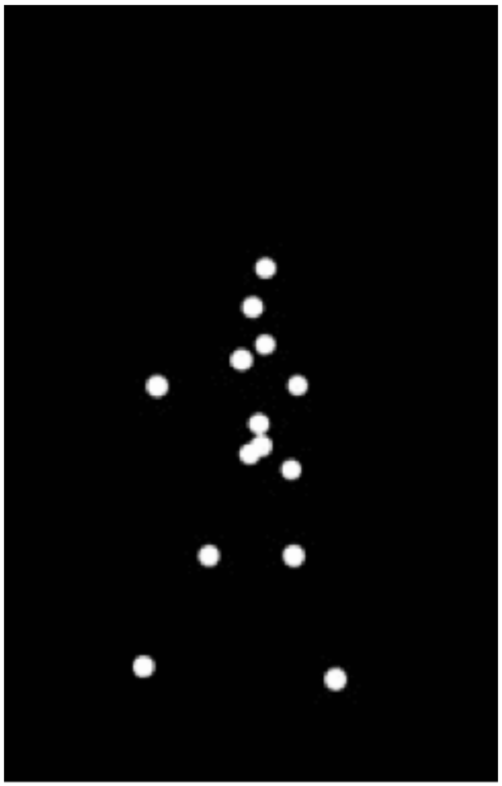
Russakovsky et al. arXiv, 2014

A brief history of deep learning



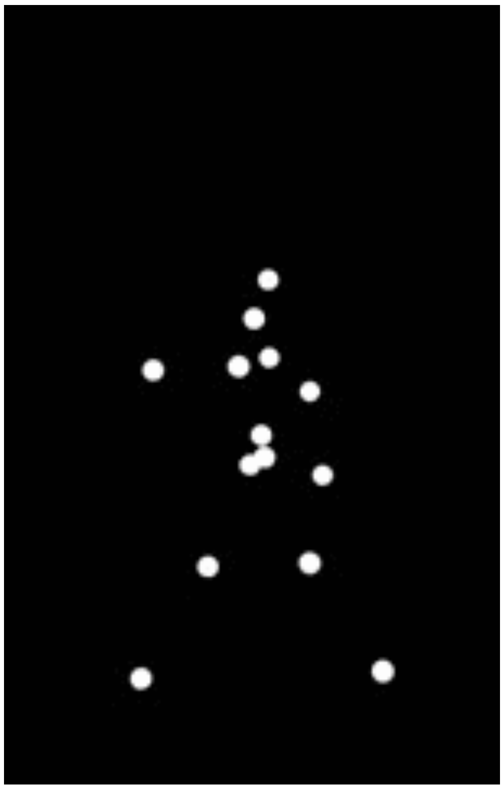
Russakovsky et al. arXiv, 2014

A small sample of what we can still learn from biology



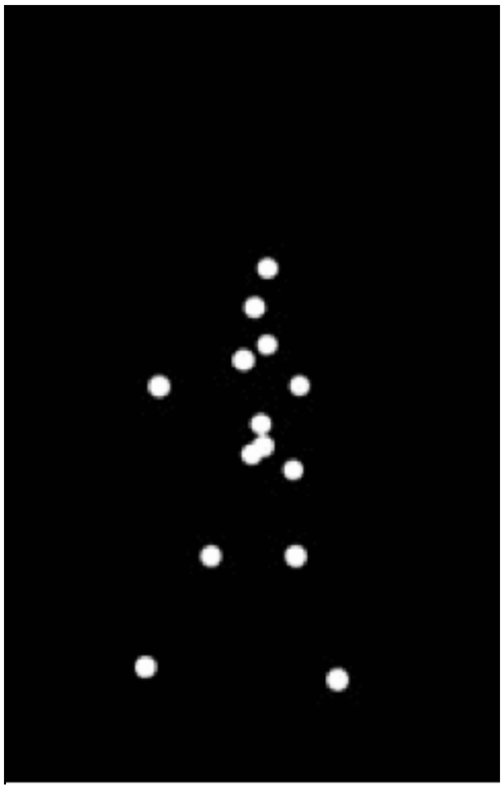
Working with
time

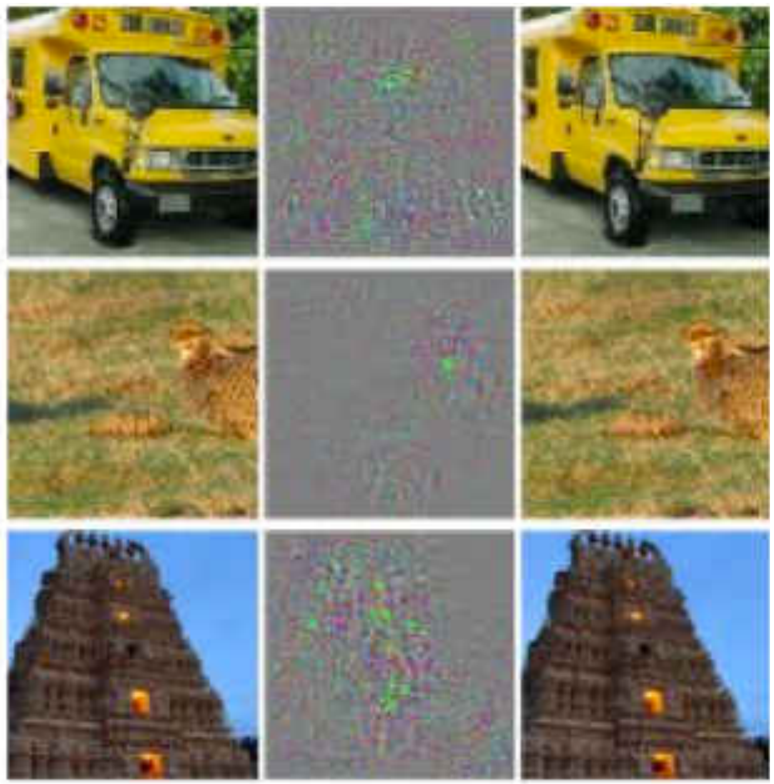
A small sample of what we can still learn from biology



Working with time

A small sample of what we can still learn from biology



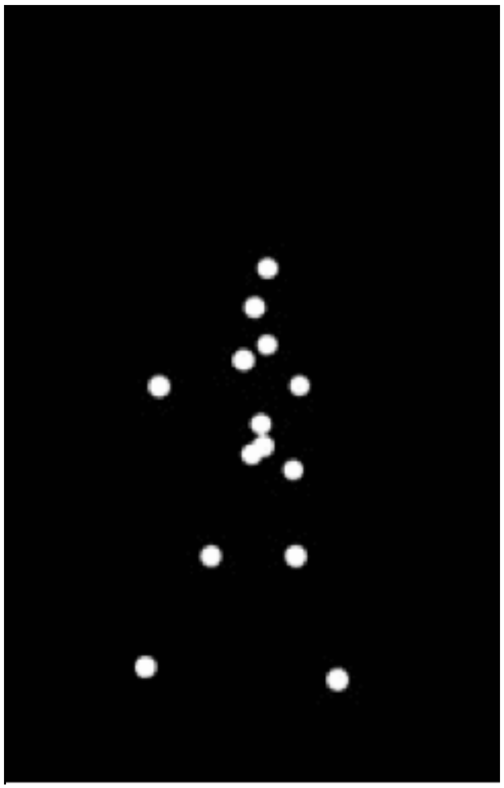


Szegedy et al., 2013

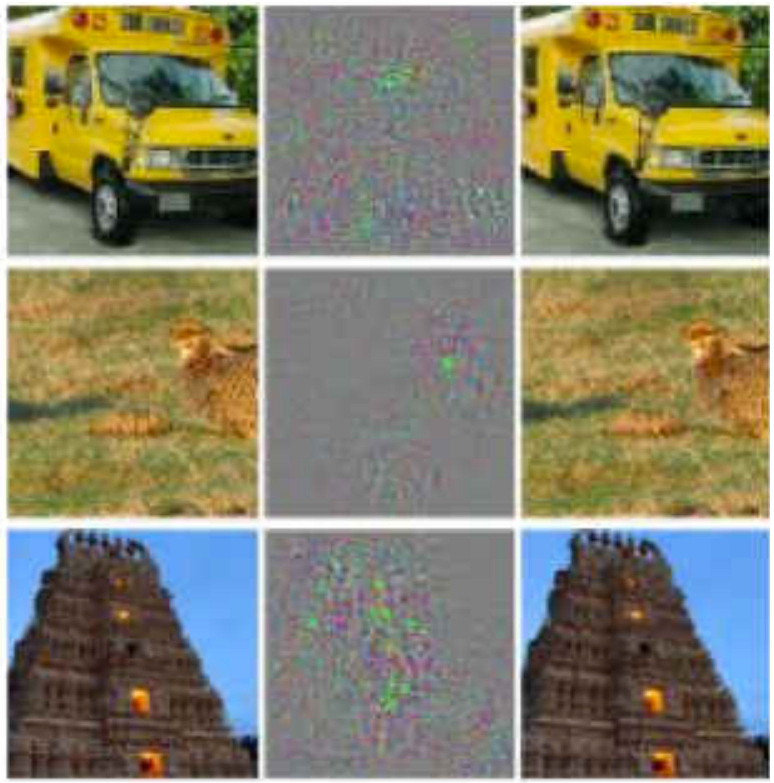
Working with time

Robustness to noise

A small sample of what we can still learn from biology

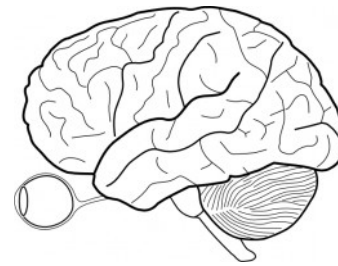


Working with time

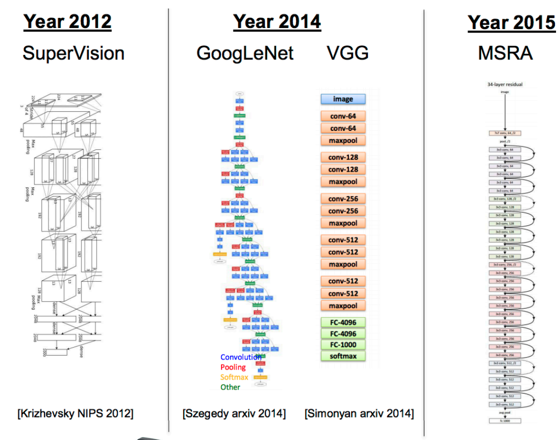


Szegedy et al., 2013

Robustness to noise



~20 W per brain

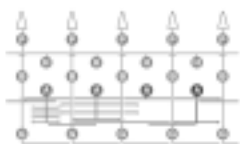


250 W per GPU

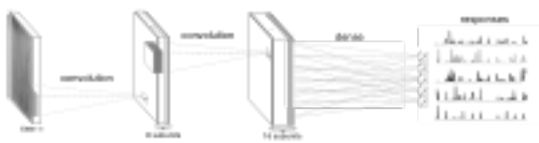
Efficiency

Outline

What we can still learn from biology?

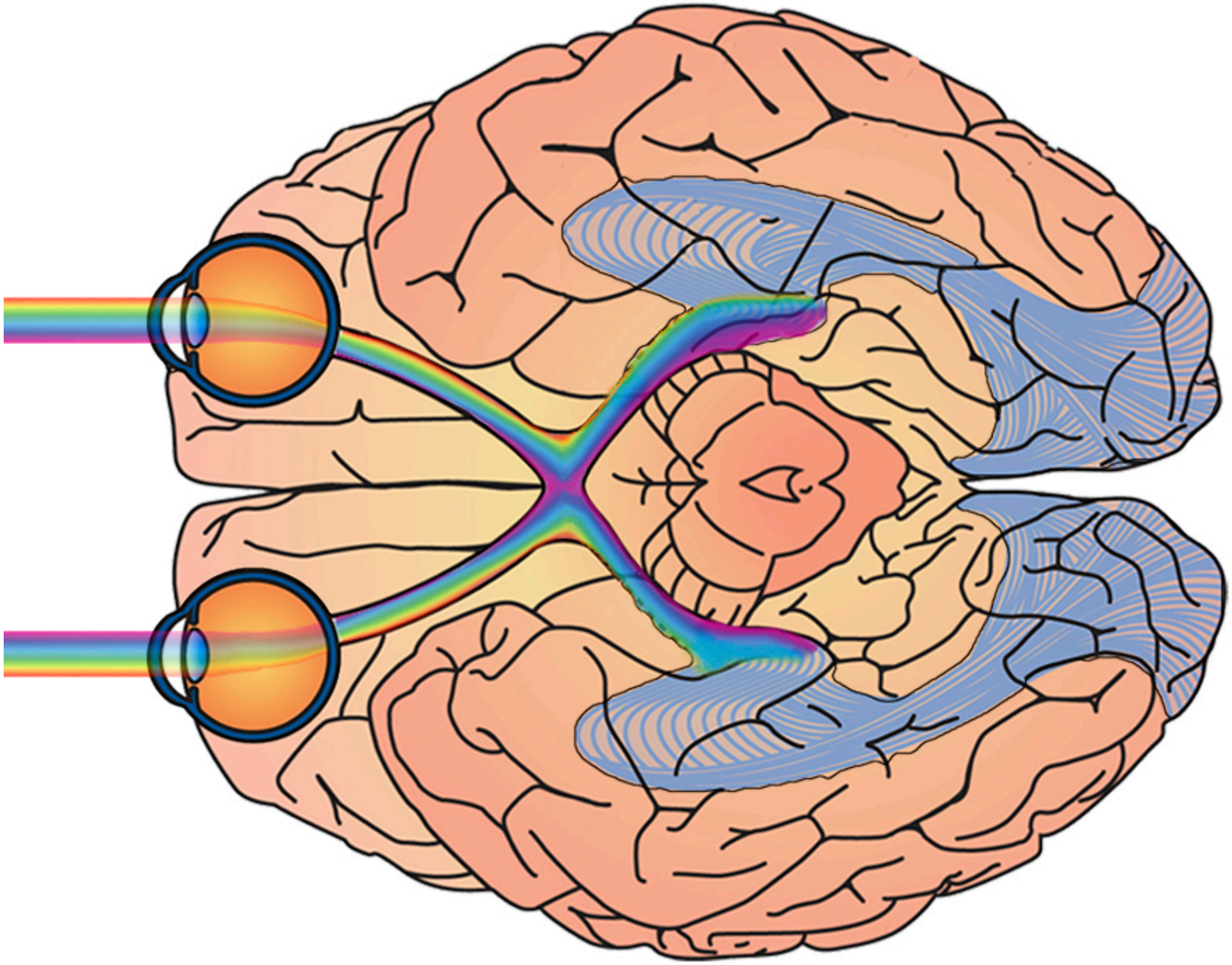


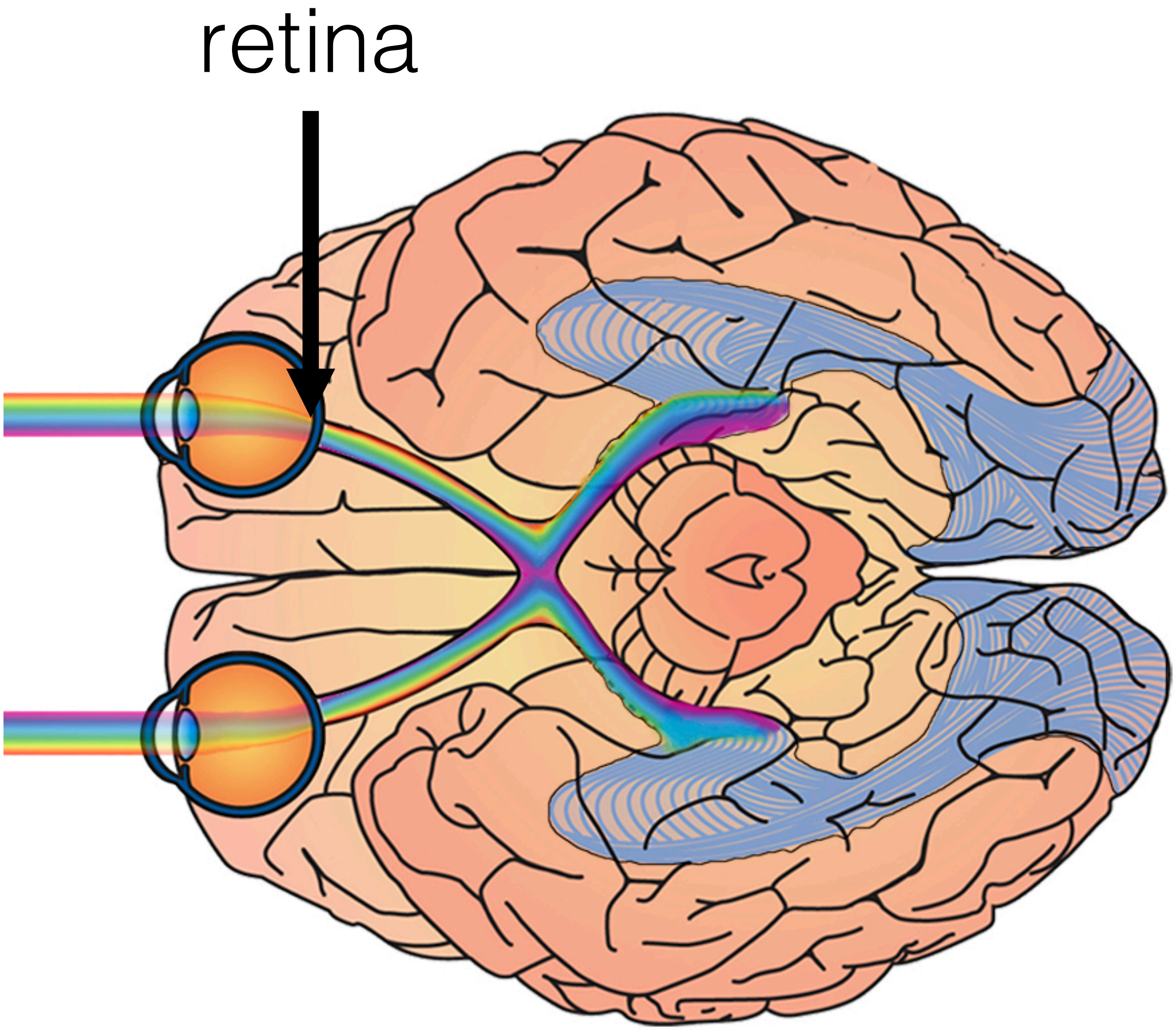
What is the retina and what does it do?



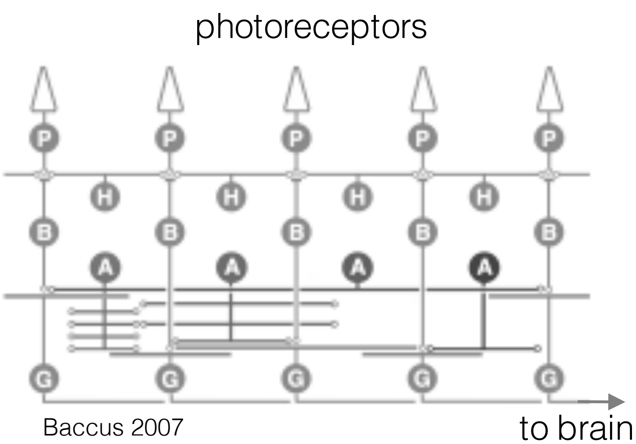
What can we learn from convolutional neural network models of the retina?

- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding




retina

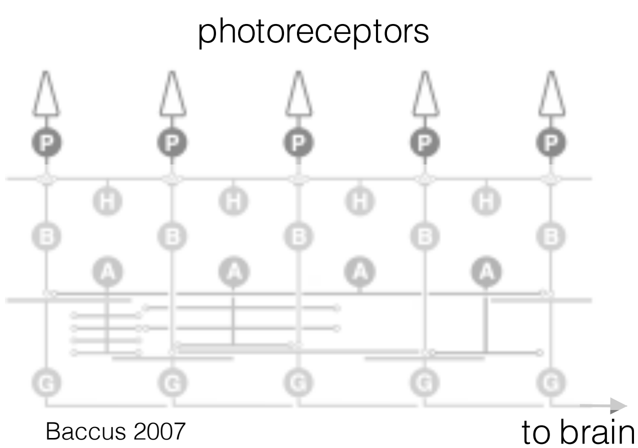
The retina is the first step of vision



Baccus 2007

The retina is the first step of vision

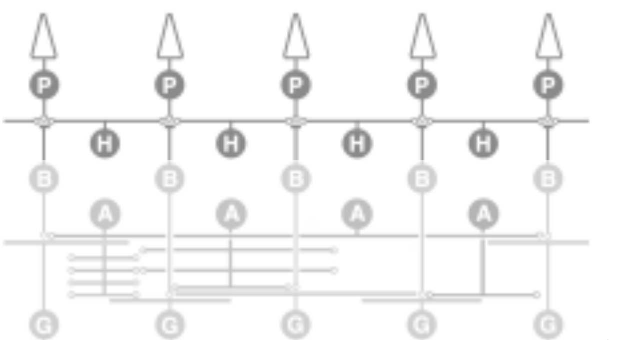
Masland 2001
rat, rabbit, cat, and monkey



Baccus 2007

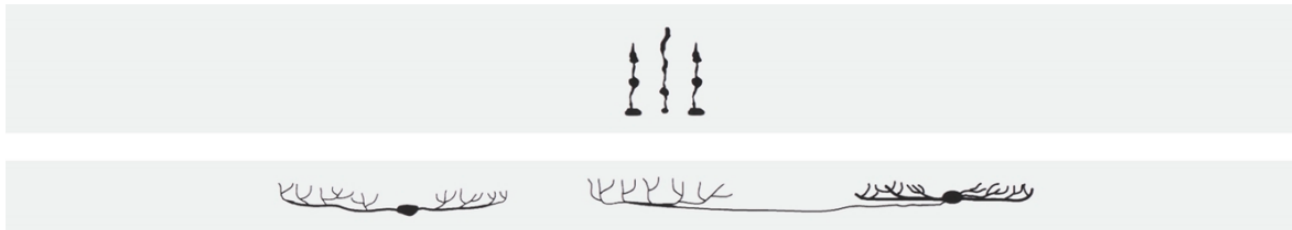
The retina is the first step of vision

photoreceptors



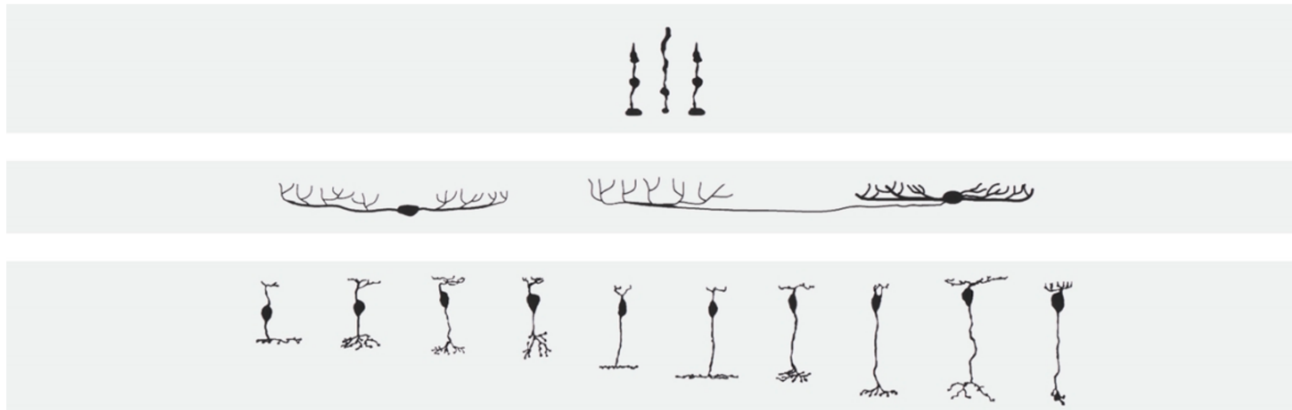
Baccus 2007

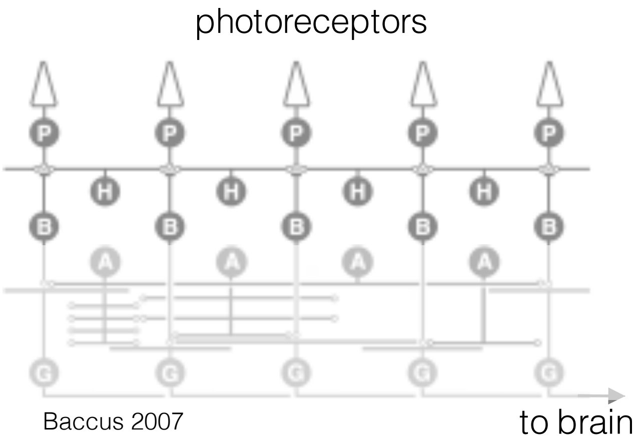
to brain



Masland 2001
rat, rabbit, cat, and monkey

The retina is the first step of vision

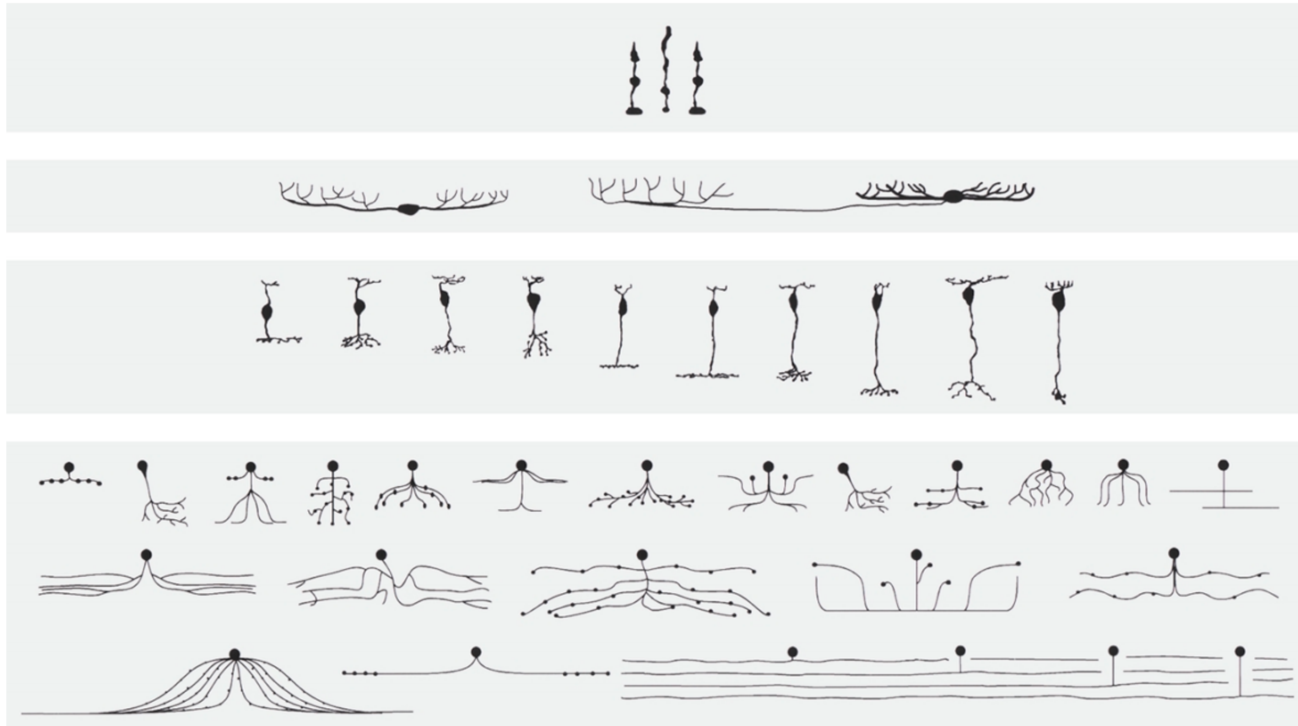




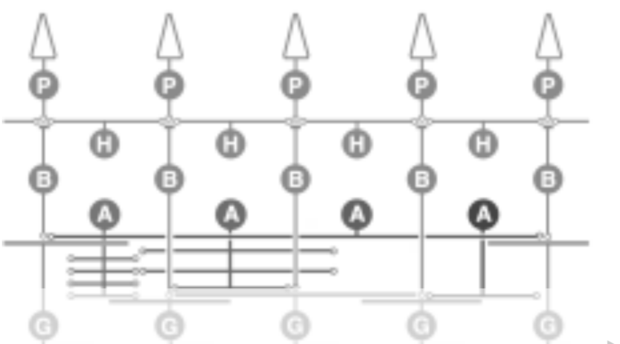
Baccus 2007

Masland 2001
rat, rabbit, cat, and monkey

The retina is the first step of vision



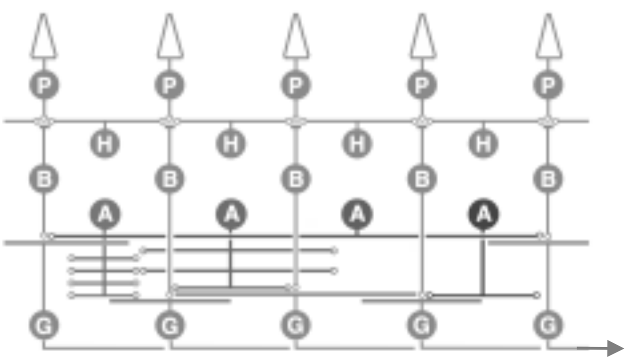
photoreceptors



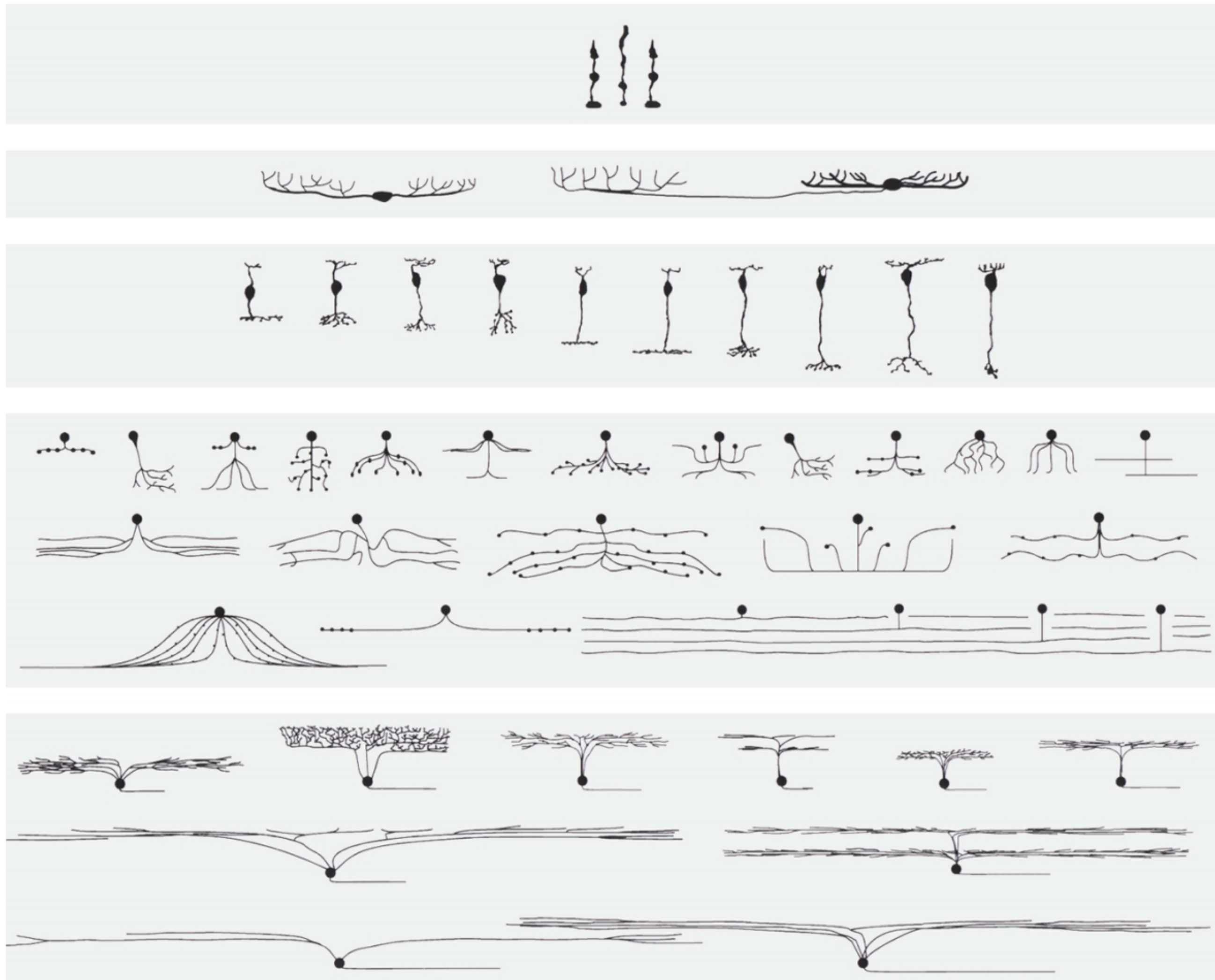
Baccus 2007

Masland 2001
rat, rabbit, cat, and monkey

The retina is the first step of vision



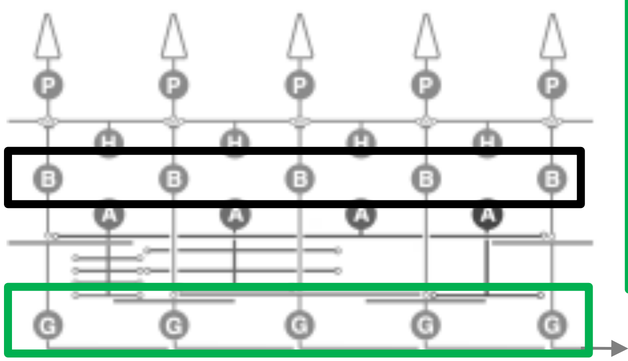
Baccus 2007



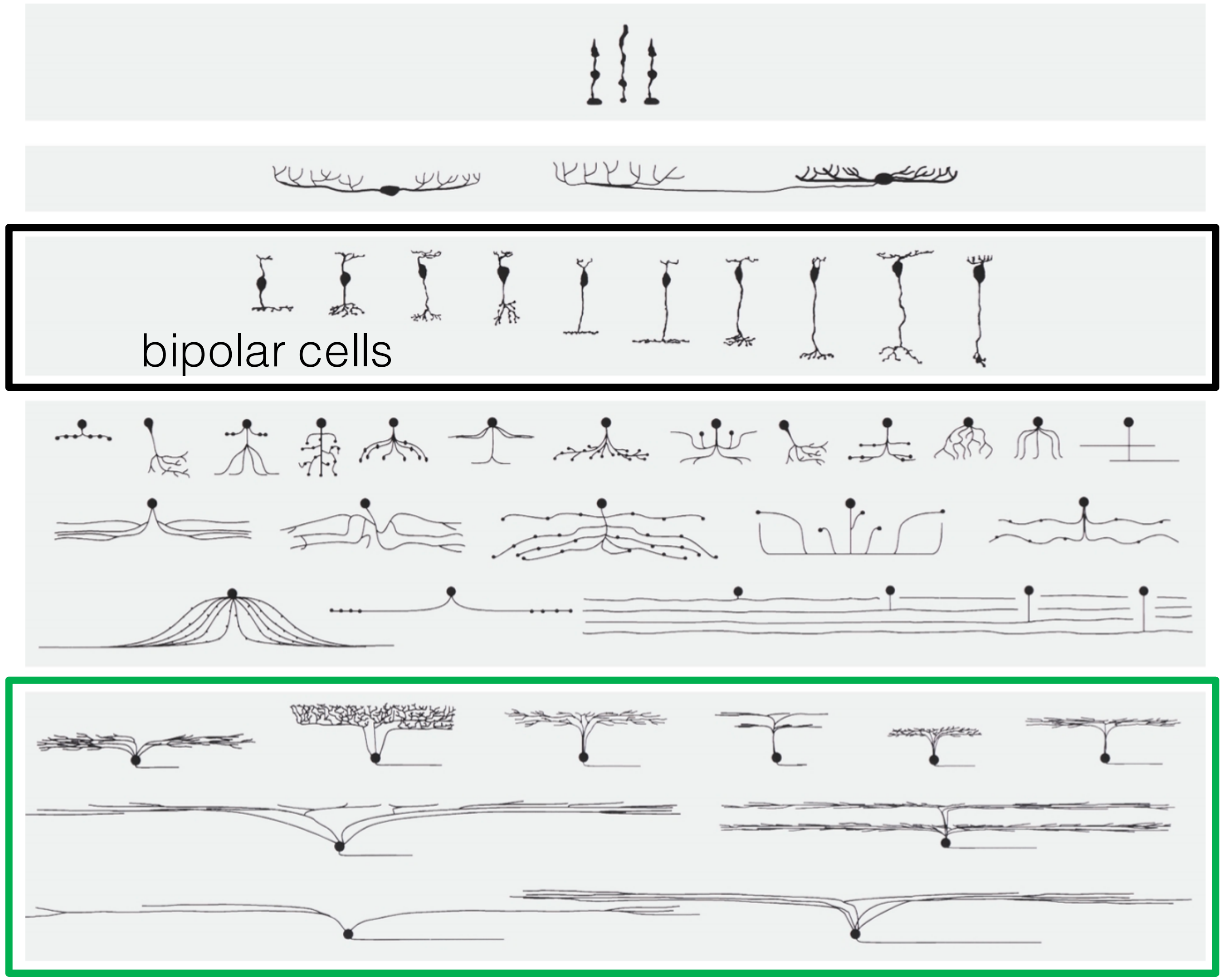
Masland 2001
rat, rabbit, cat, and monkey

The retina is the first step of vision

photoreceptors



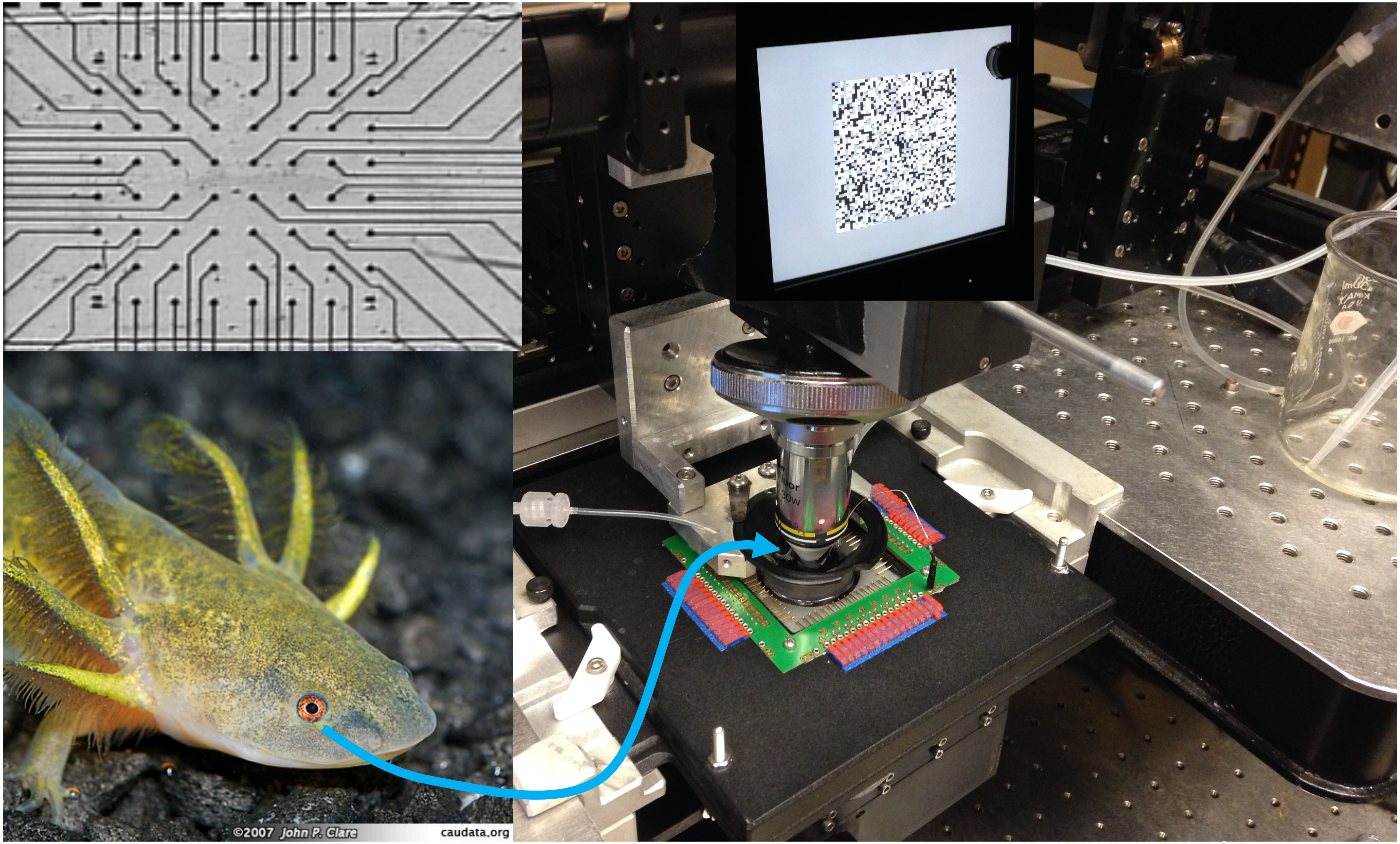
Baccus 2007



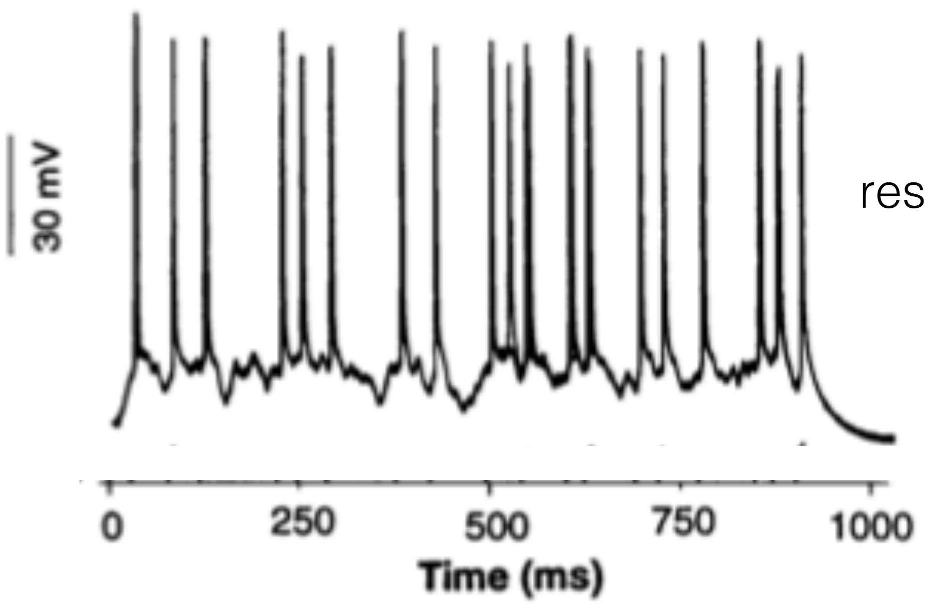
retinal ganglion cells

Masland 2001
rat, rabbit, cat, and monkey

Our experimental setup



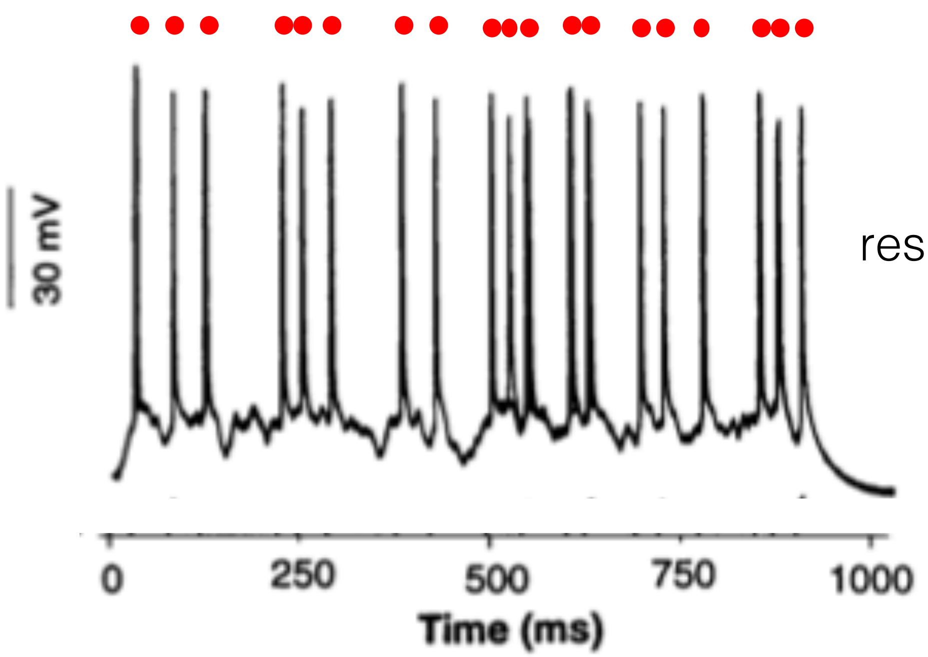
We can record the electrical activity of cells



response to one trial

Mainen and Sejnowski, 1995

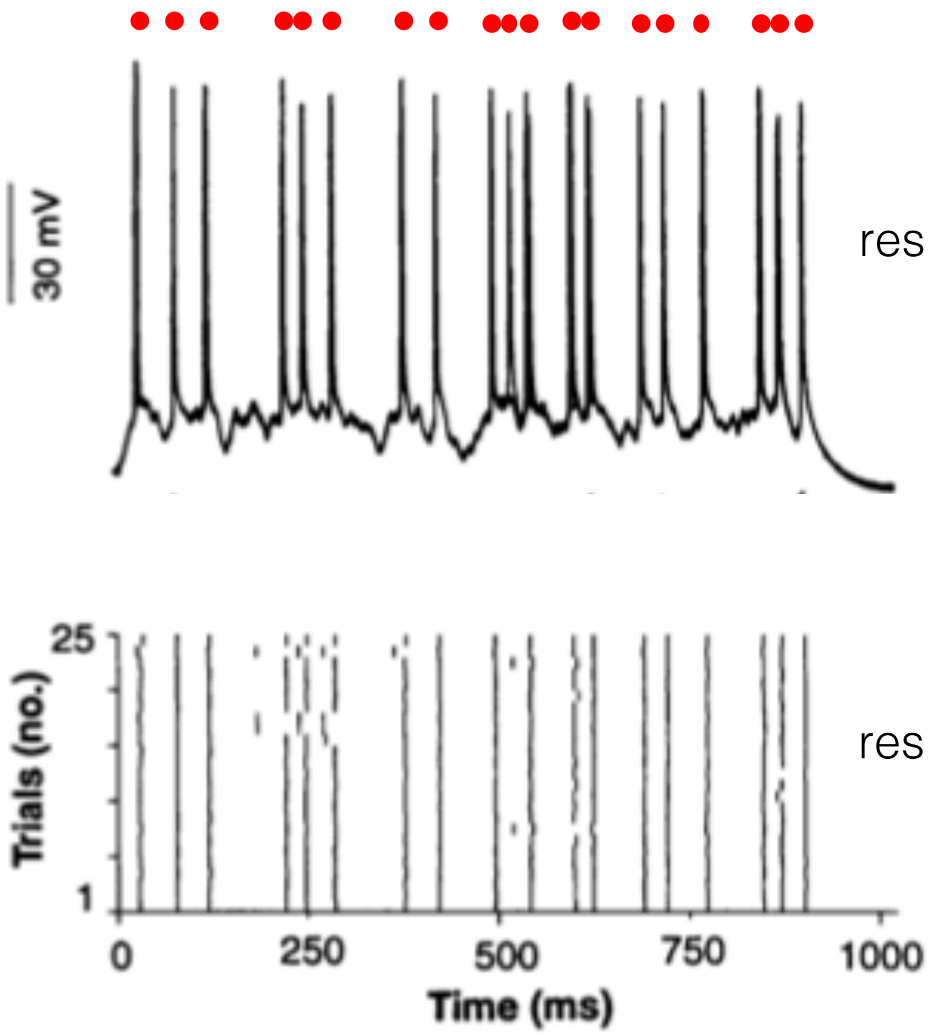
Each action potential is considered a binary event



response to one trial

Mainen and Sejnowski, 1995

Responses are very reliable for strong stimuli

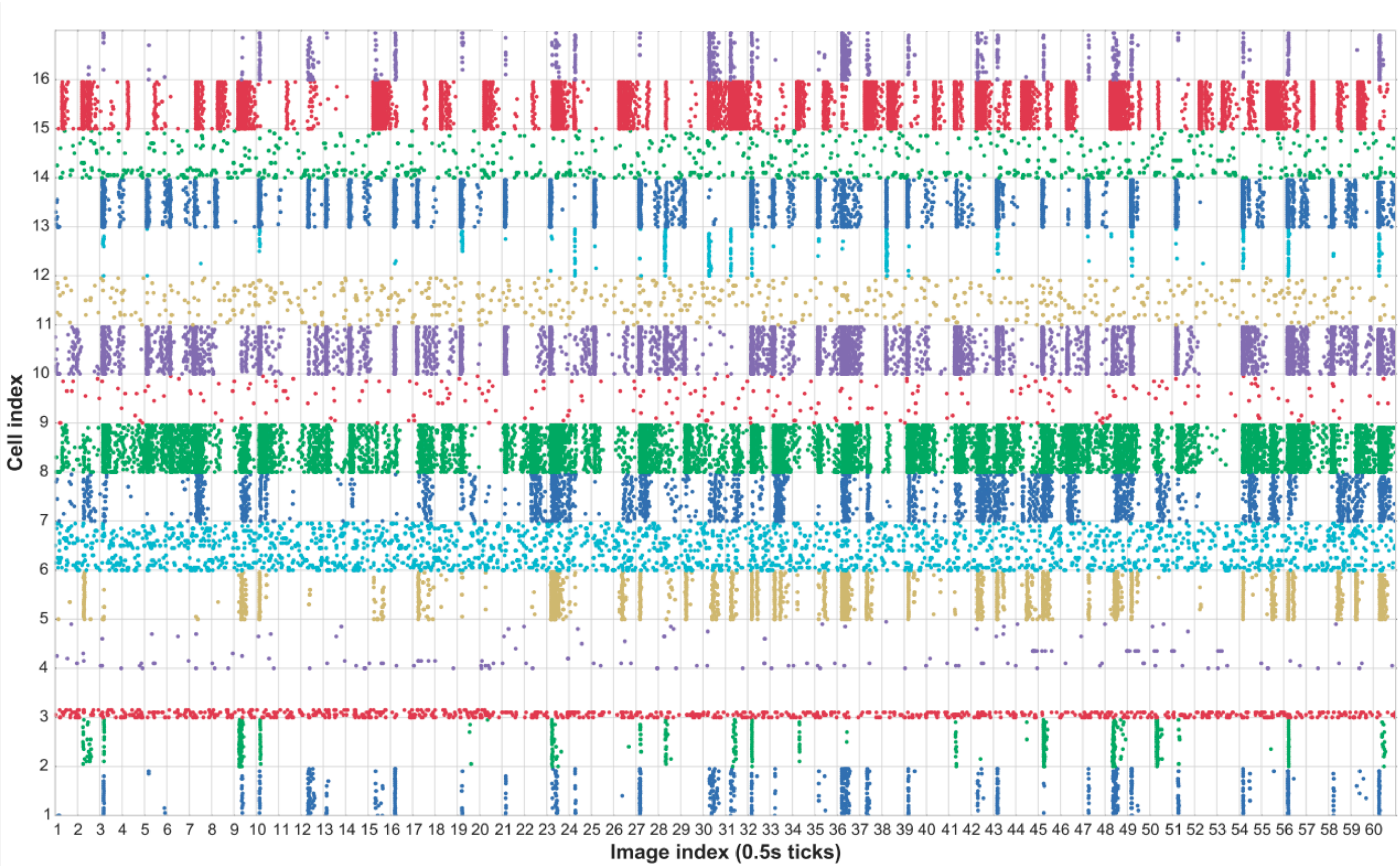


response to one trial

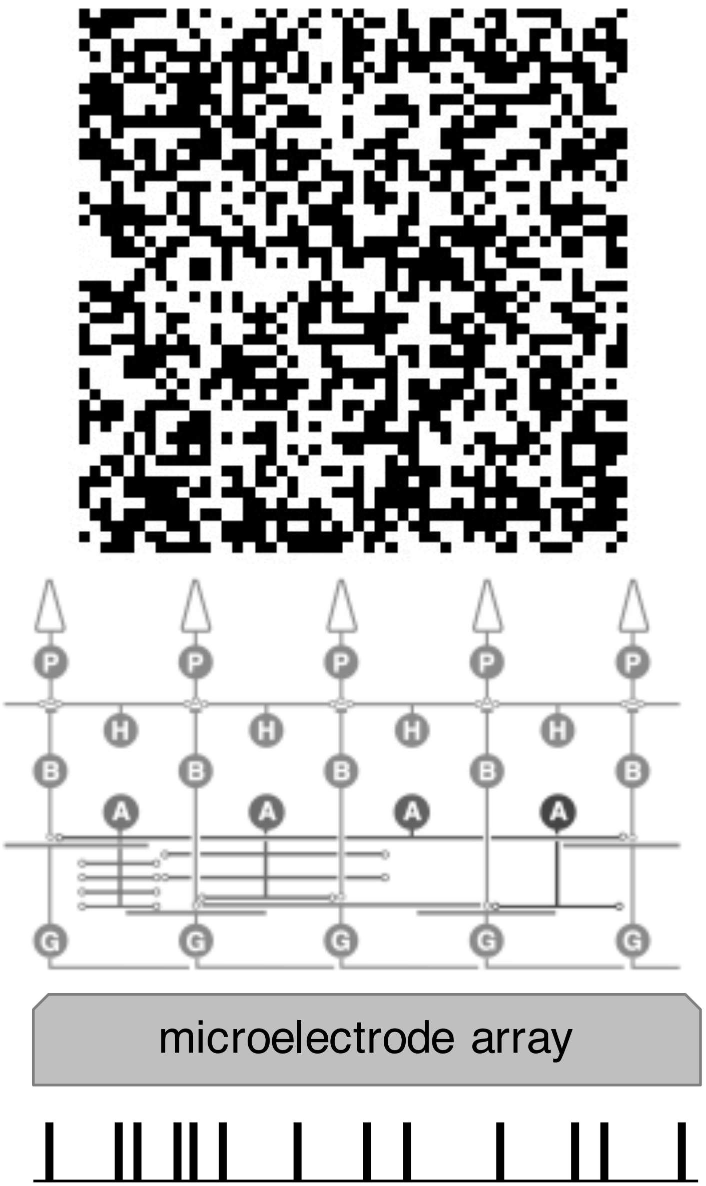
response over many trials

Mainen and Sejnowski, 1995

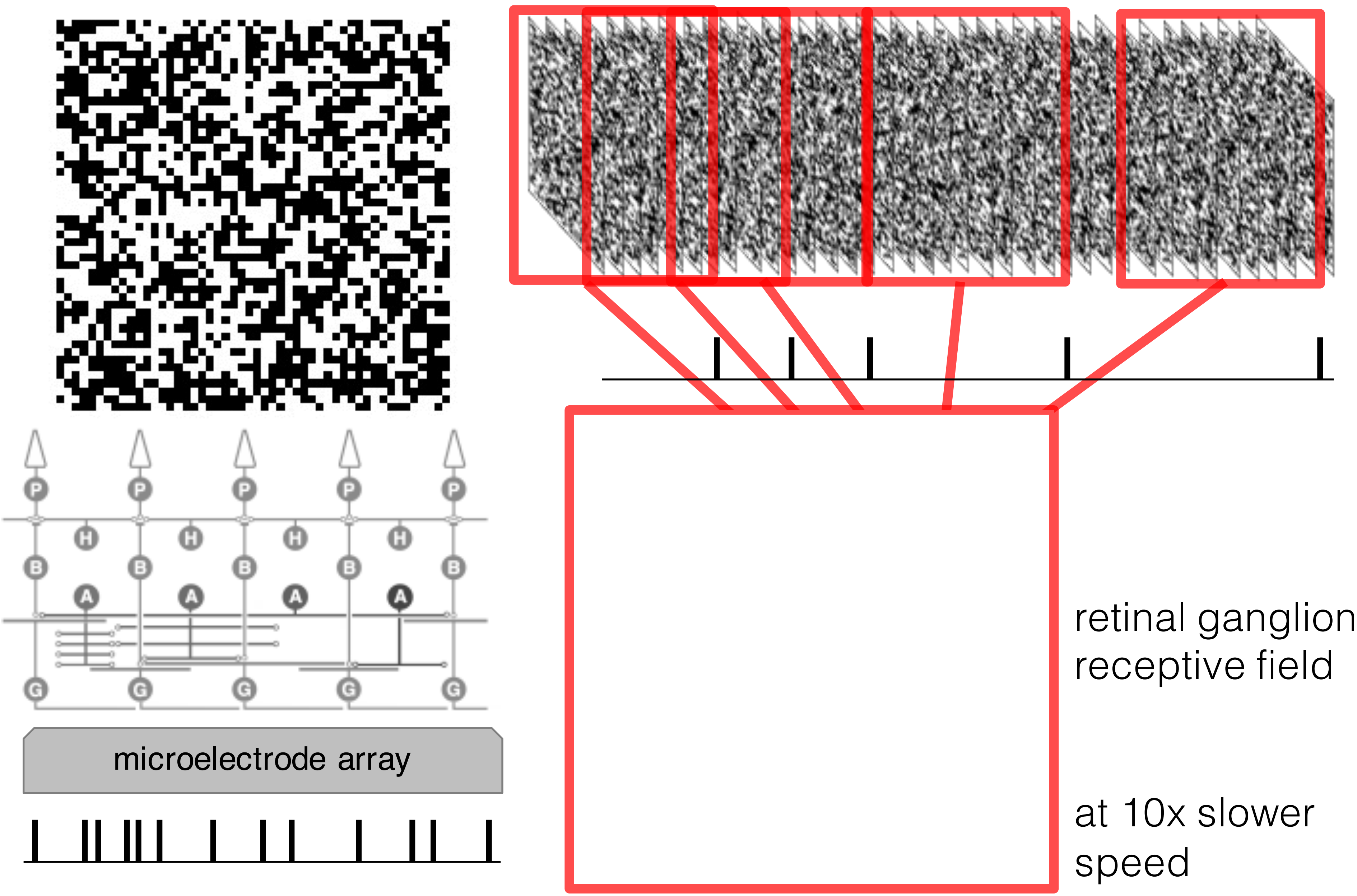
We can simultaneously record many cells



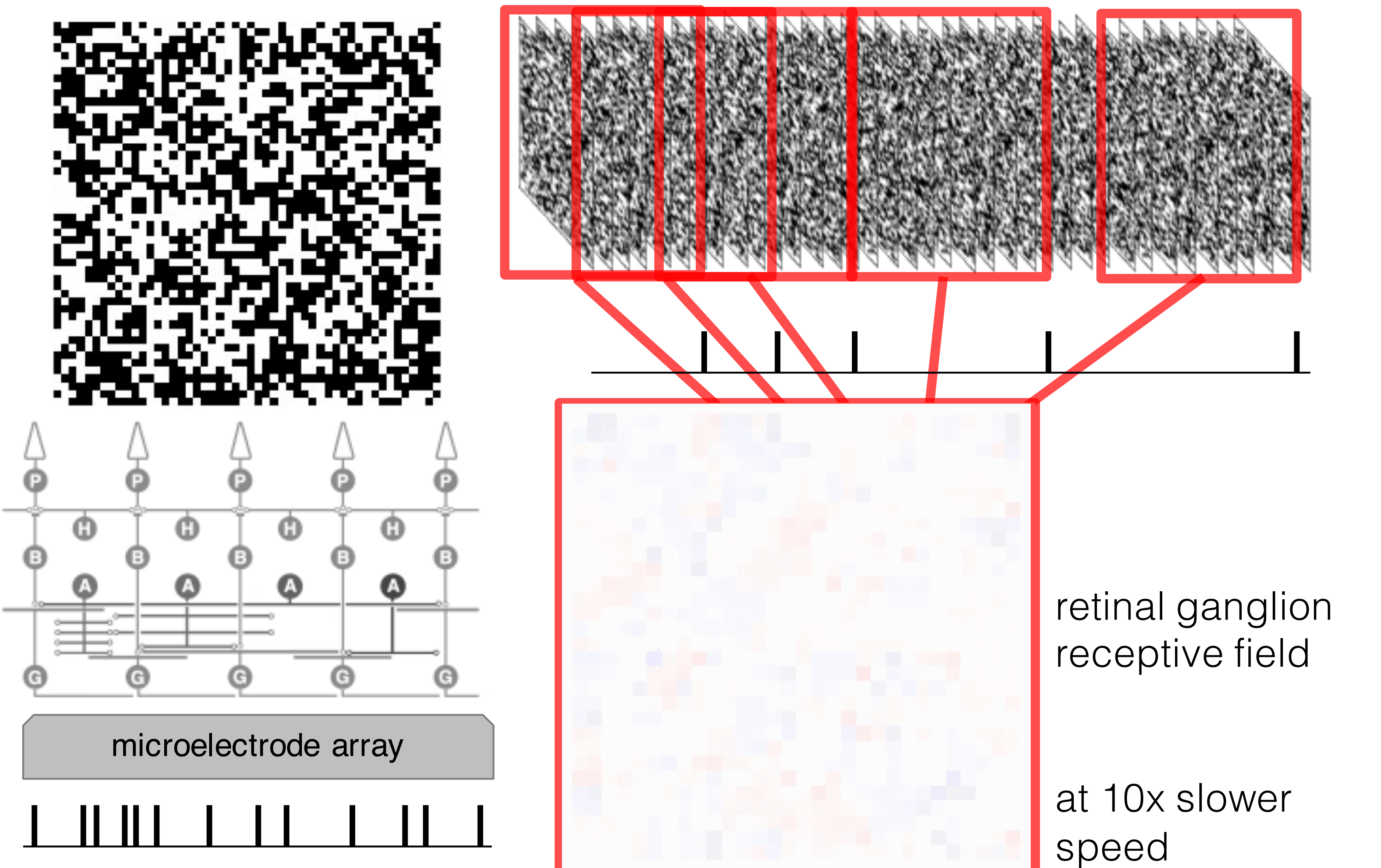
Neural computations under natural stimuli



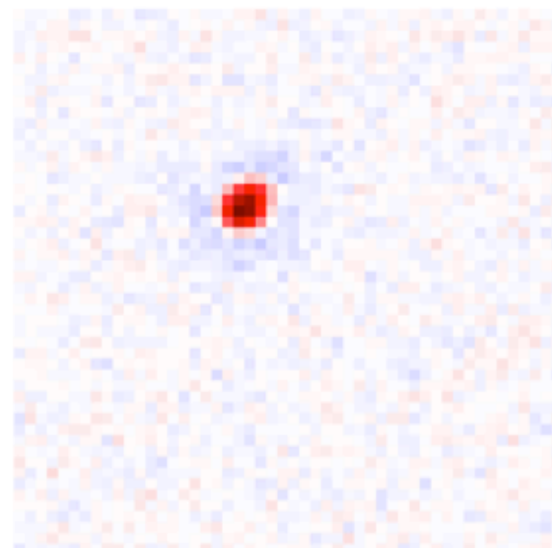
Neural computations under natural stimuli



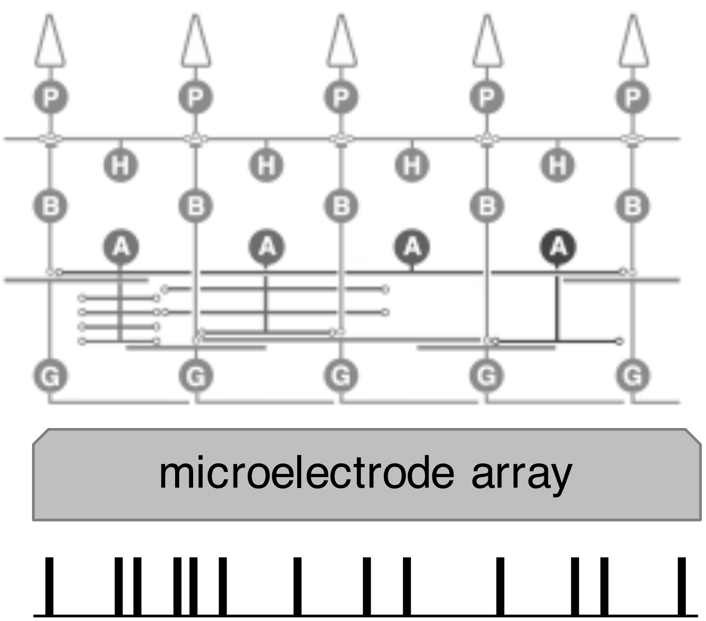
Neural computations under natural stimuli

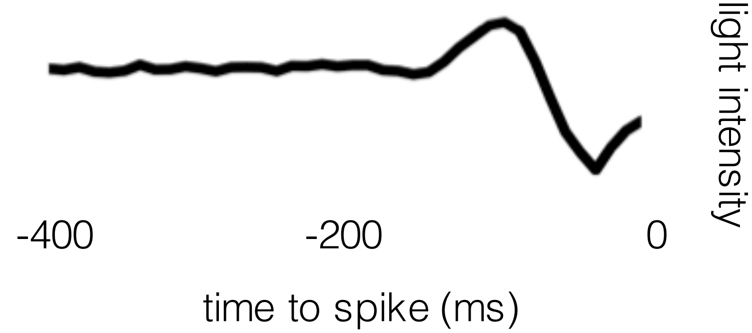


Neural computations under natural stimuli



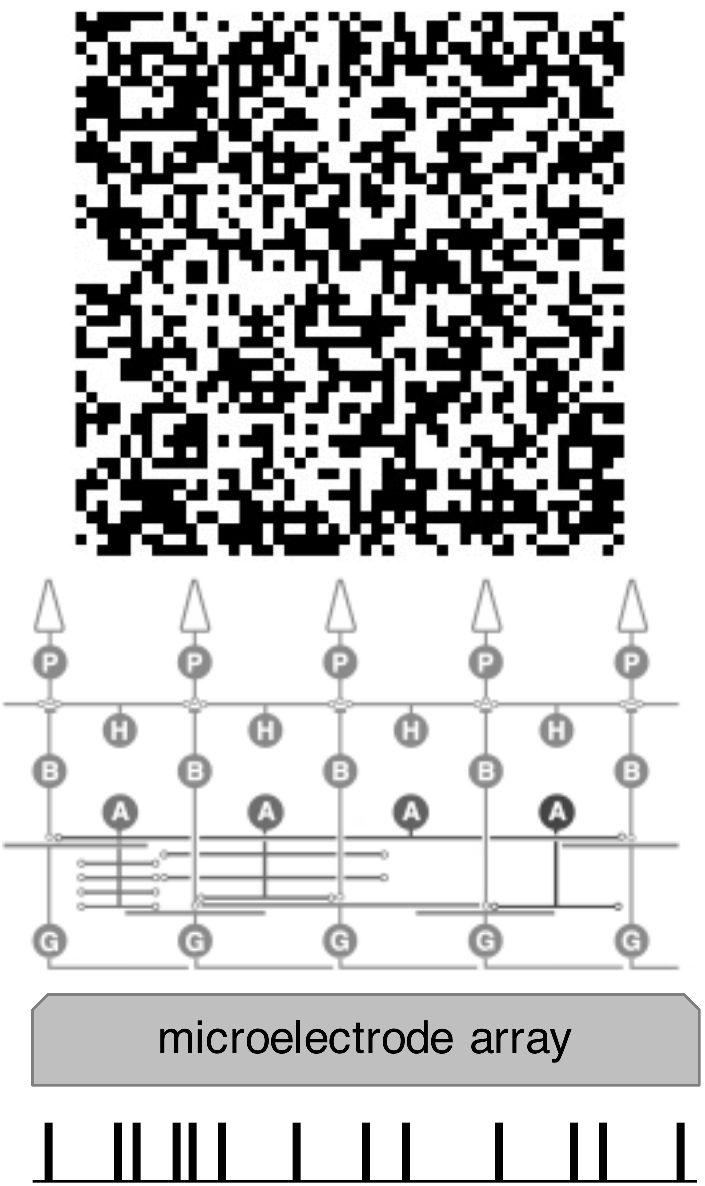
retinal ganglion
receptive field

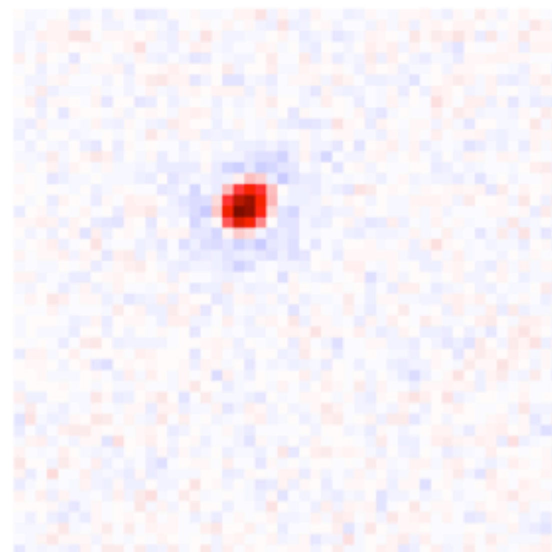




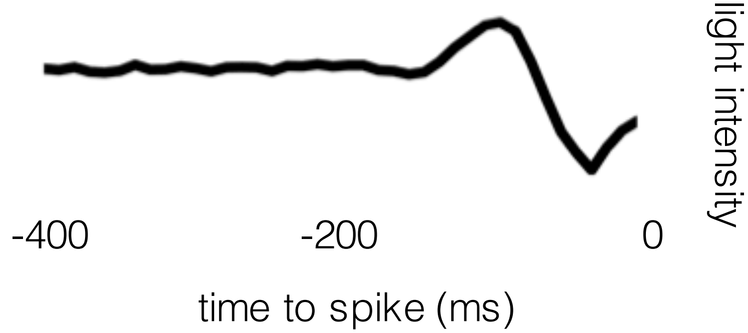
low rank spatio-temporal decomposition

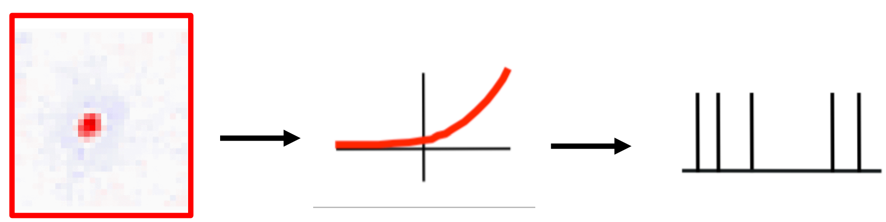
Neural computations under natural stimuli



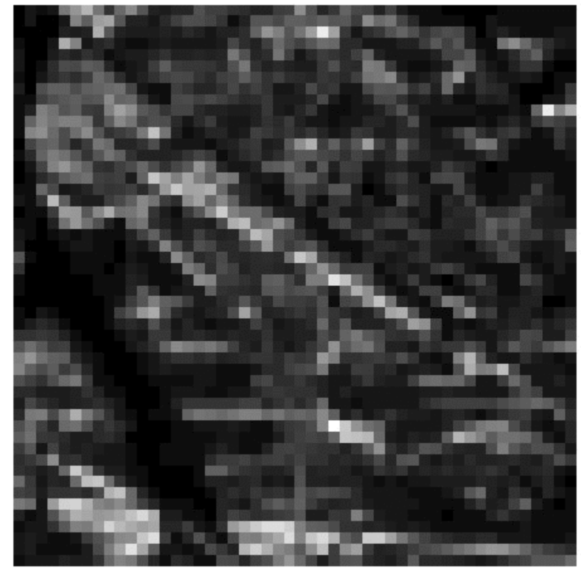


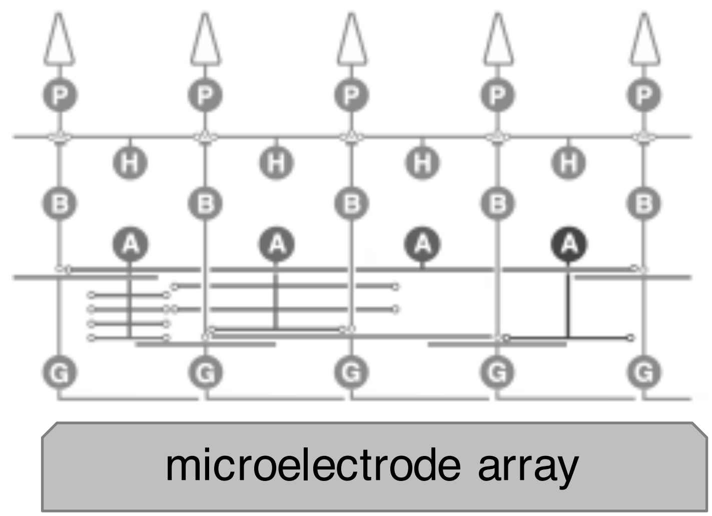
retinal ganglion
receptive field





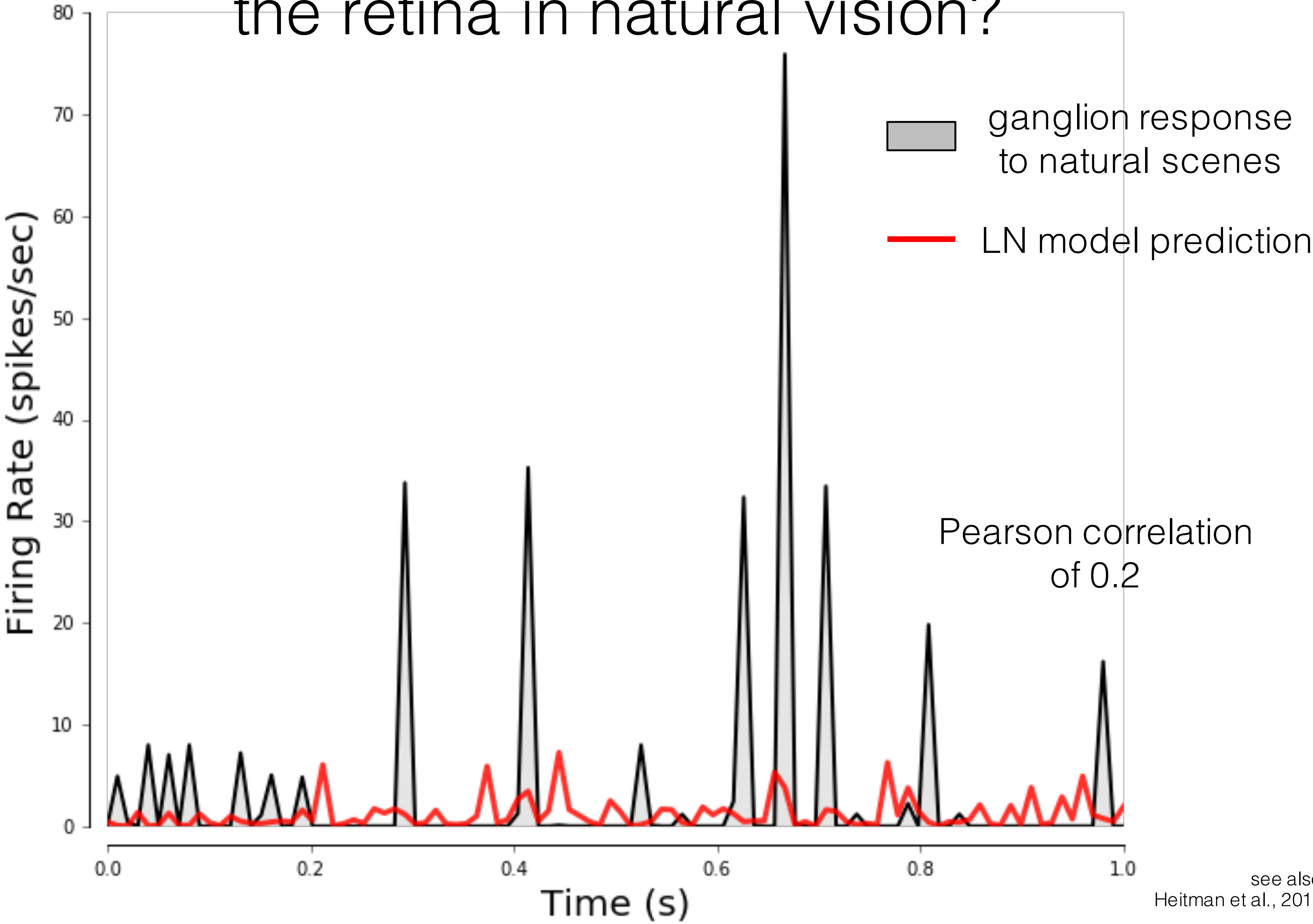
Neural computations under natural stimuli





?

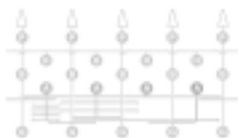
How well do linear-nonlinear models explain the retina in natural vision?



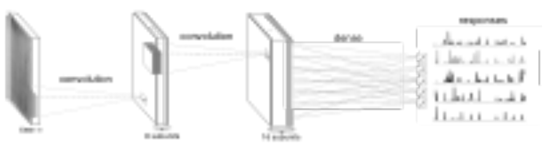
see also
Heitman et al., 2014

Outline

What we can still learn from biology?



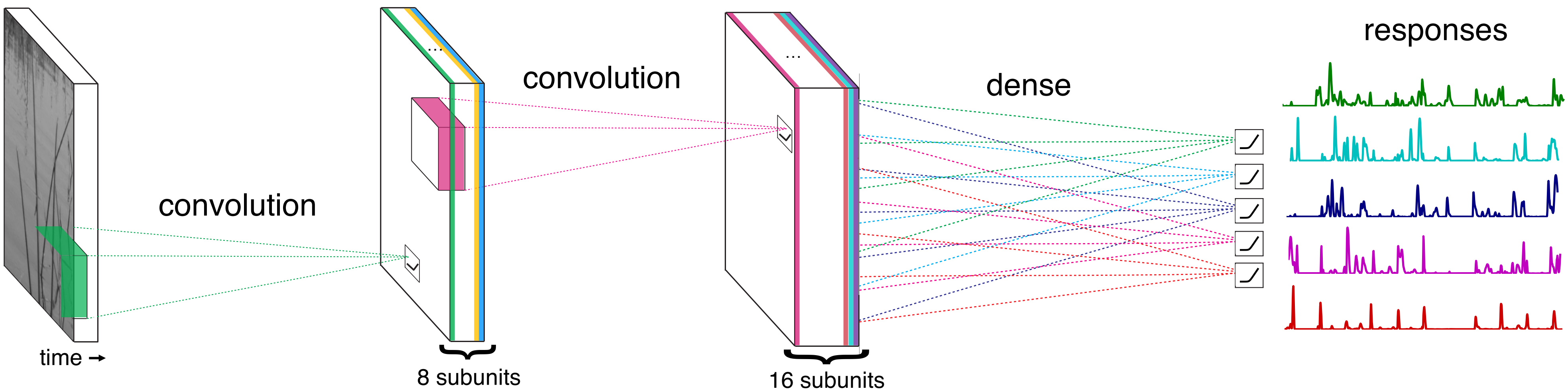
What is the retina and what does it do?



What can we learn from convolutional neural network models of the retina?

- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding

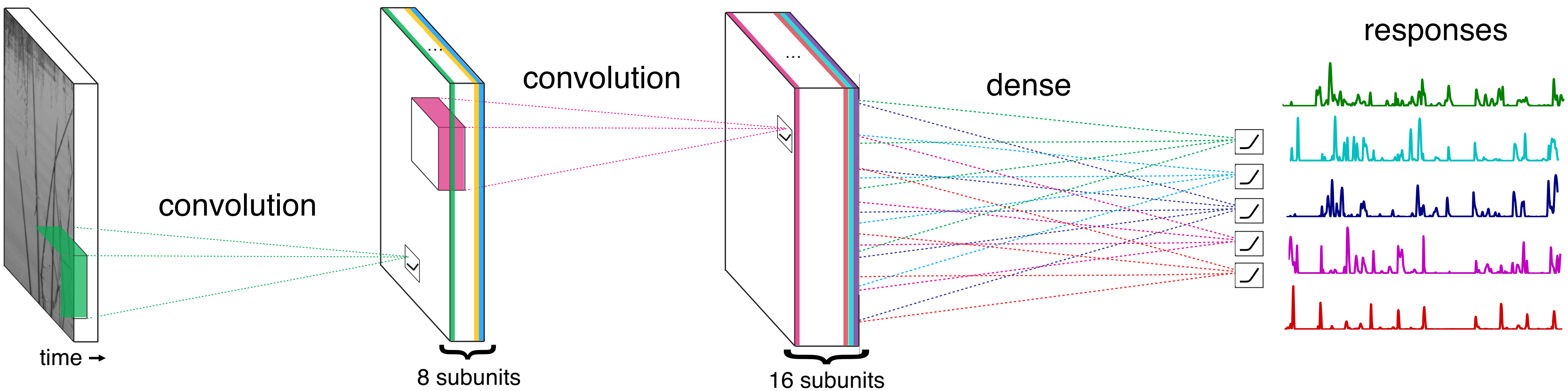
Convolutional neural network model

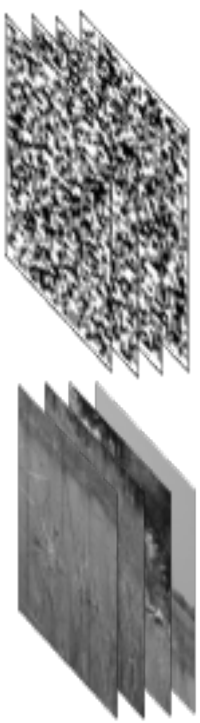


trained on
white noise

trained on
natural scenes

Convolutional neural network model





trained on white noise

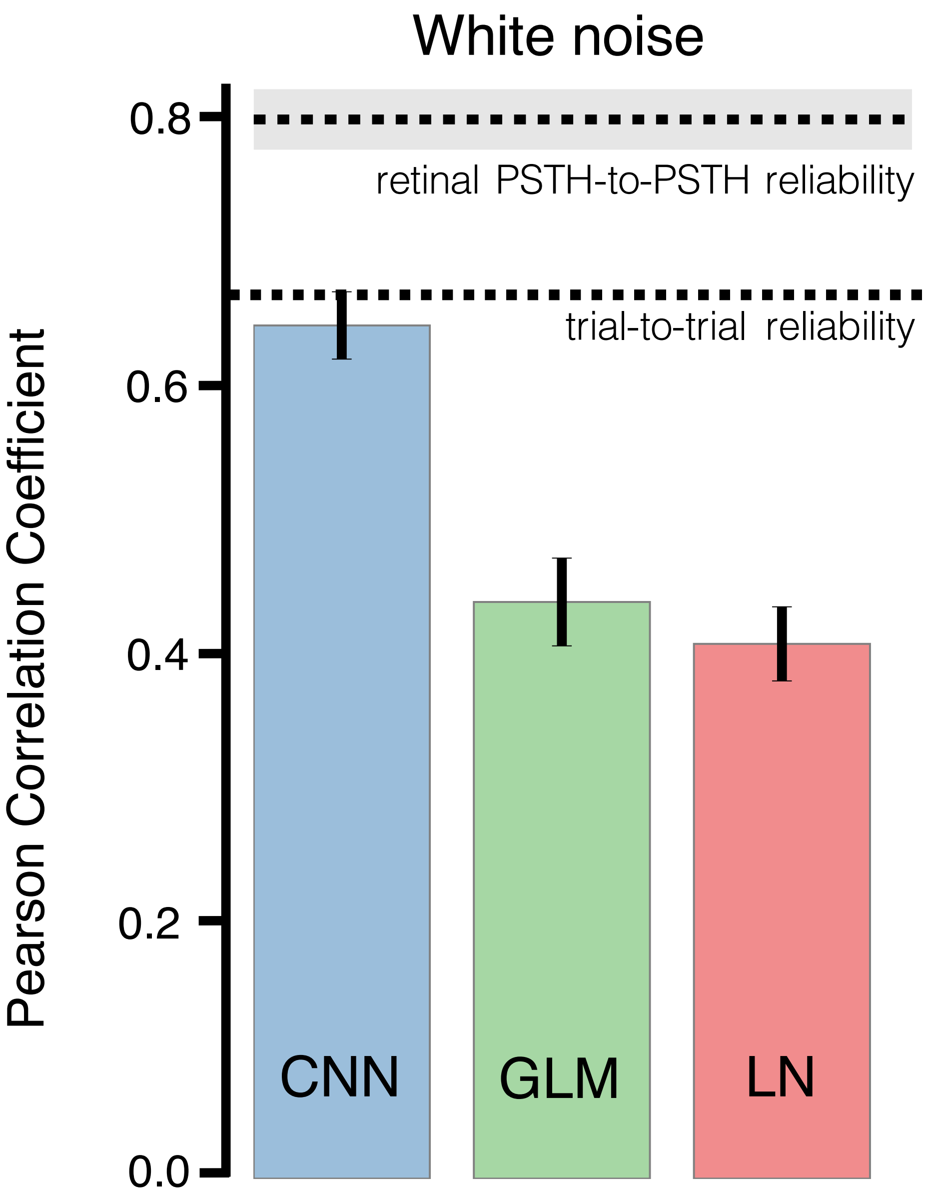
trained on natural scenes

$$\text{parameters} \approx 150{,}000$$

$$\frac{\text{data}}{\text{parameters}} \approx 3$$

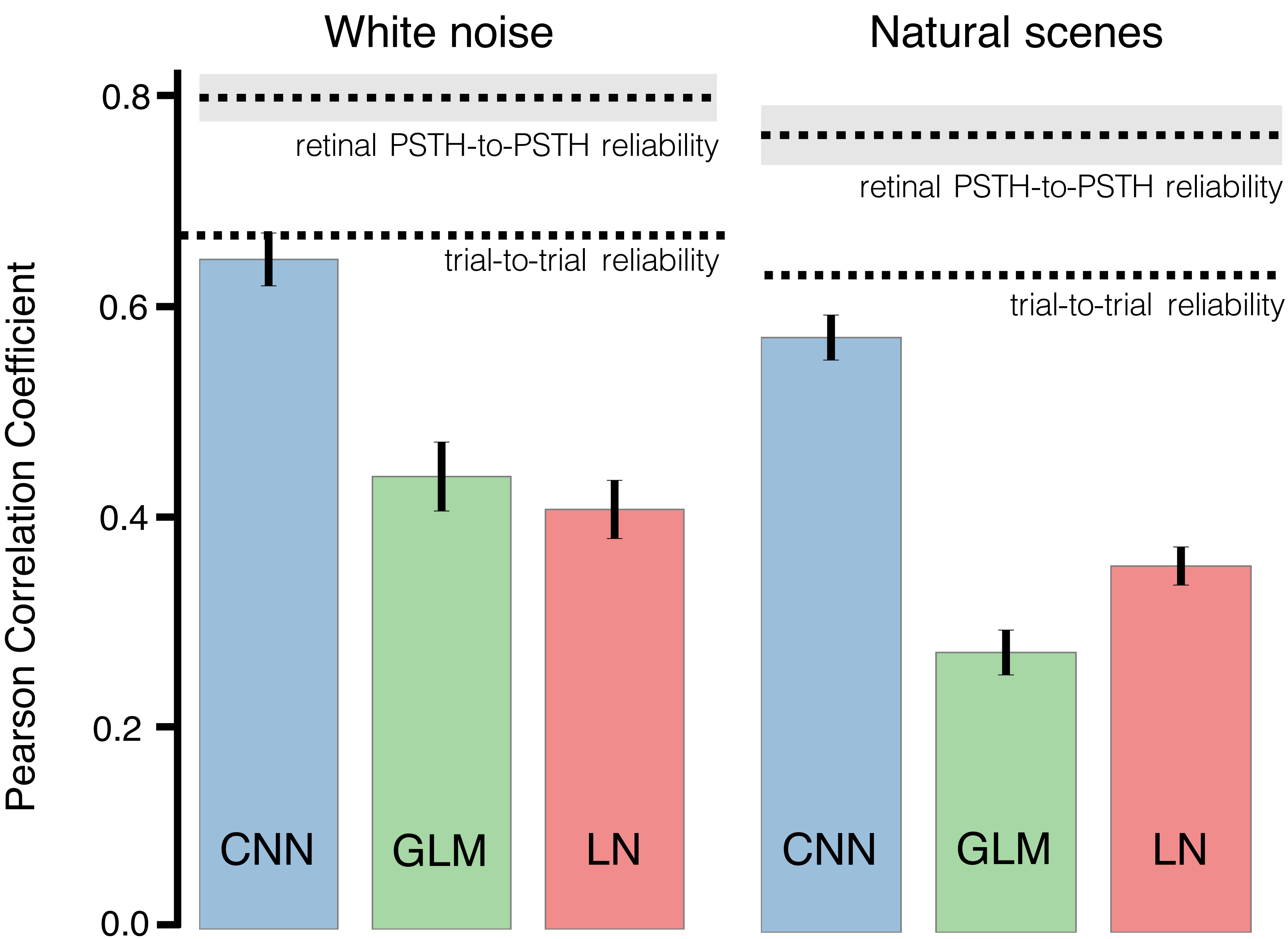
by comparison, the ganglion's receptive field for this stimulus is 100,000 parameters

CNNs approach retinal reliability



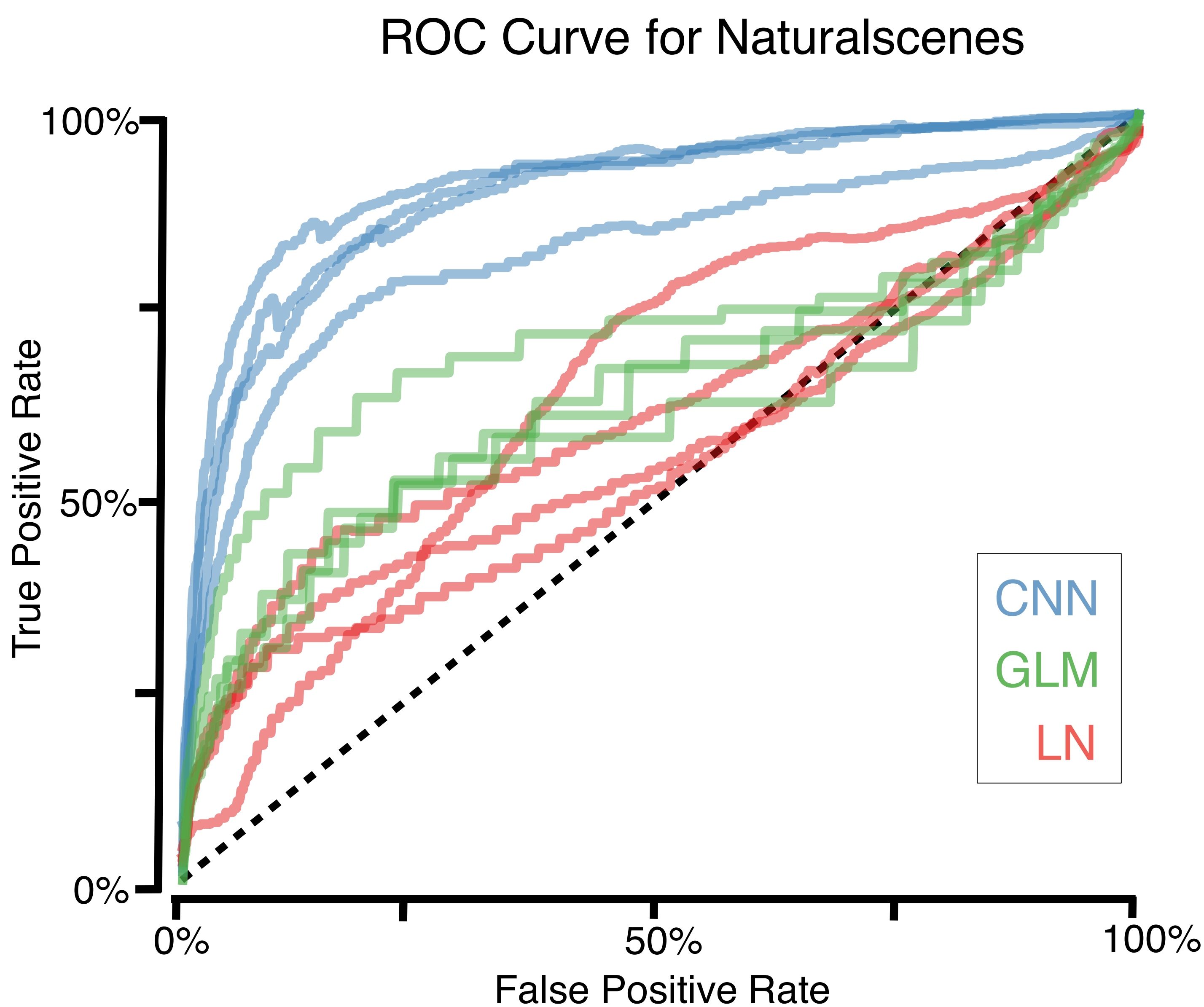
LN models:
Chichilnisky 2001
GLMs: Pillow et al. 2008

CNNs approach retinal reliability

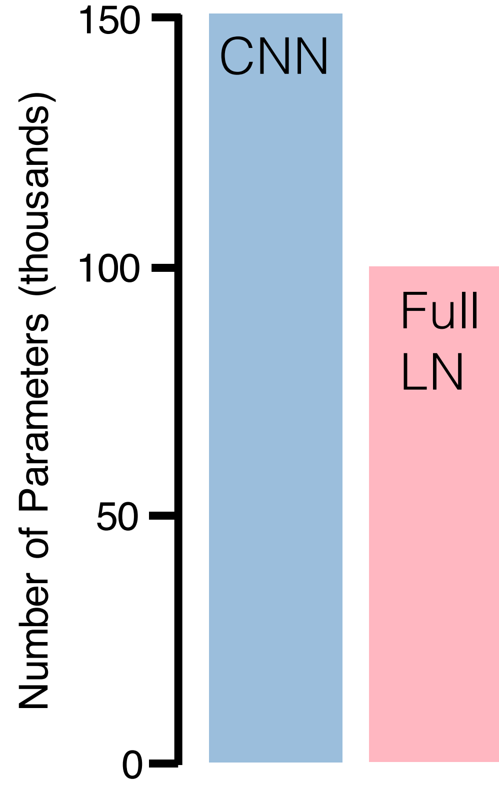


LN models:
Chichilnisky 2001
GLMs: Pillow et al. 2008

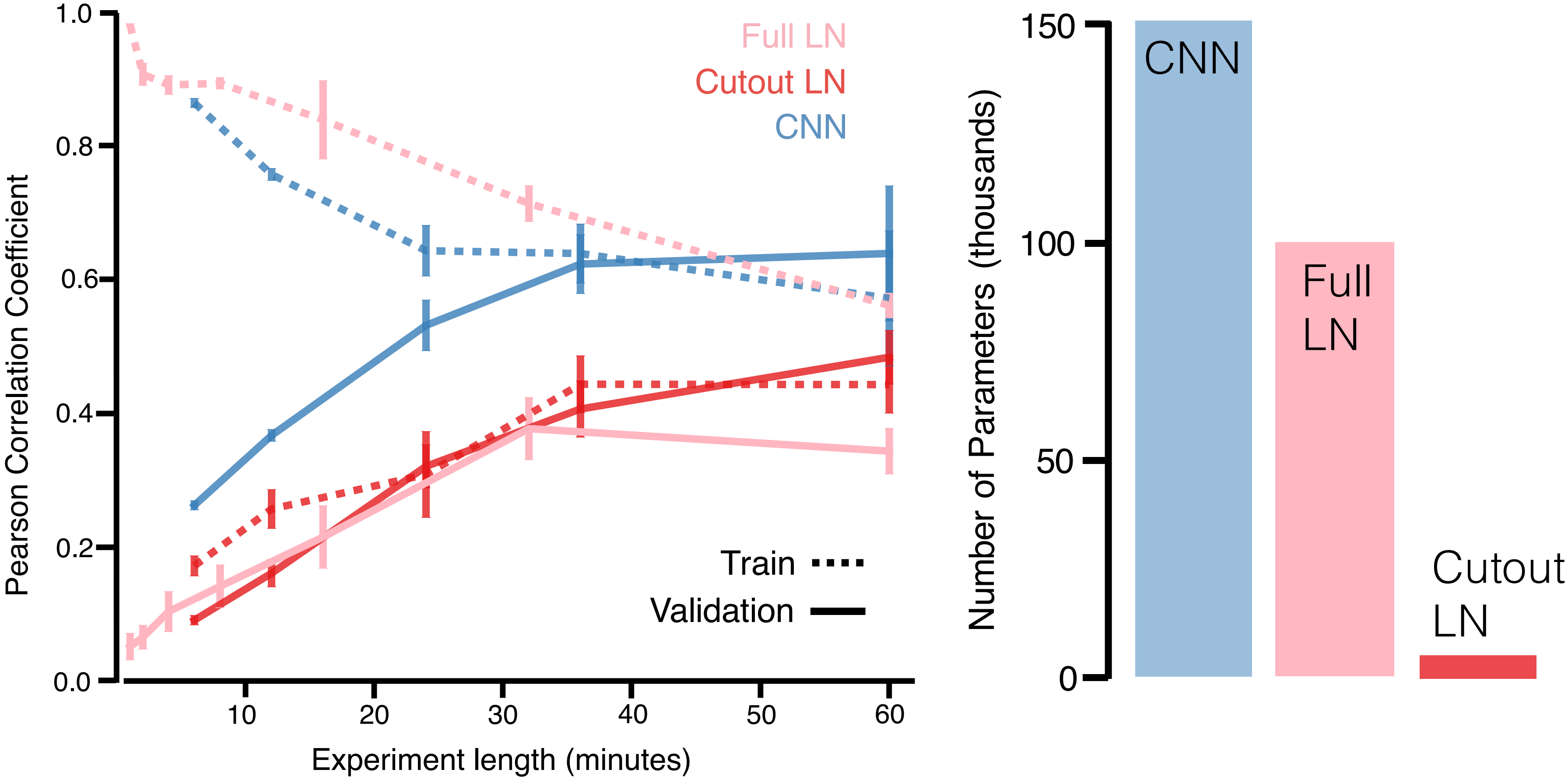
Event detection analysis



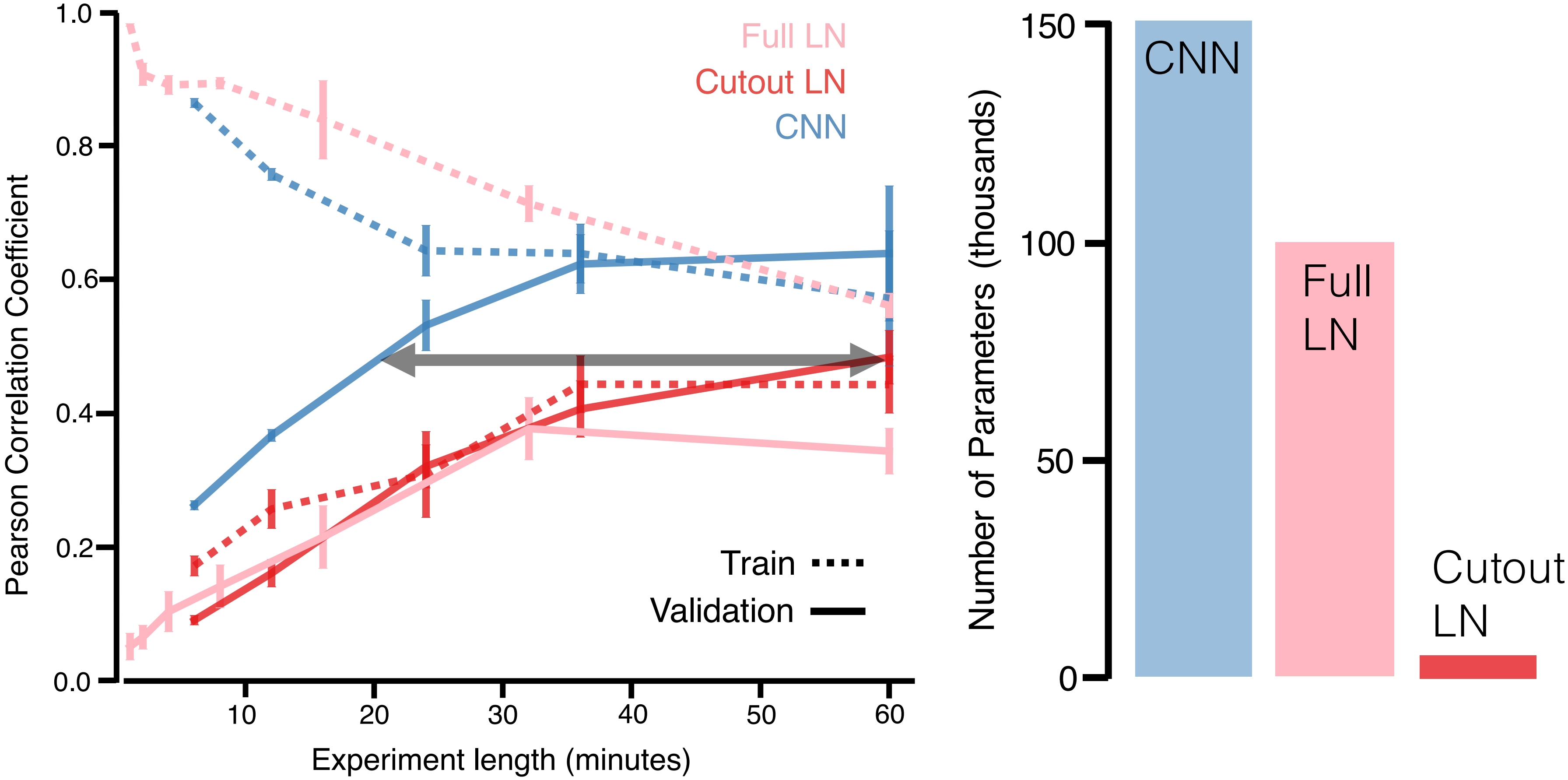
CNNs trained on less data outperform simpler models on more data



CNNs trained on less data outperform simpler models on more data

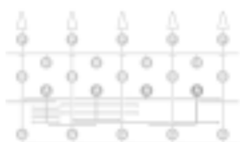


CNNs trained on less data outperform simpler models on more data

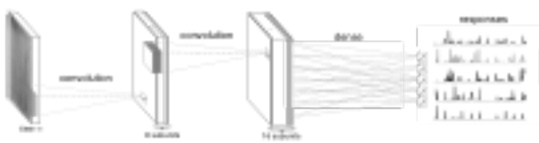


Outline

What we can still learn from biology?



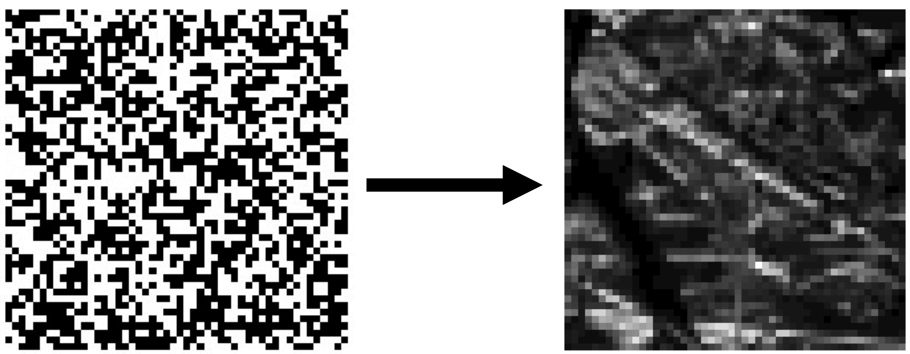
What is the retina and what does it do?



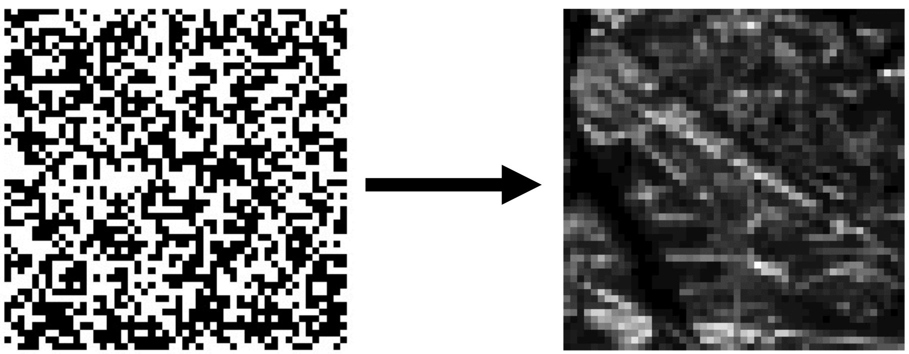
What can we learn from convolutional neural network models of the retina?

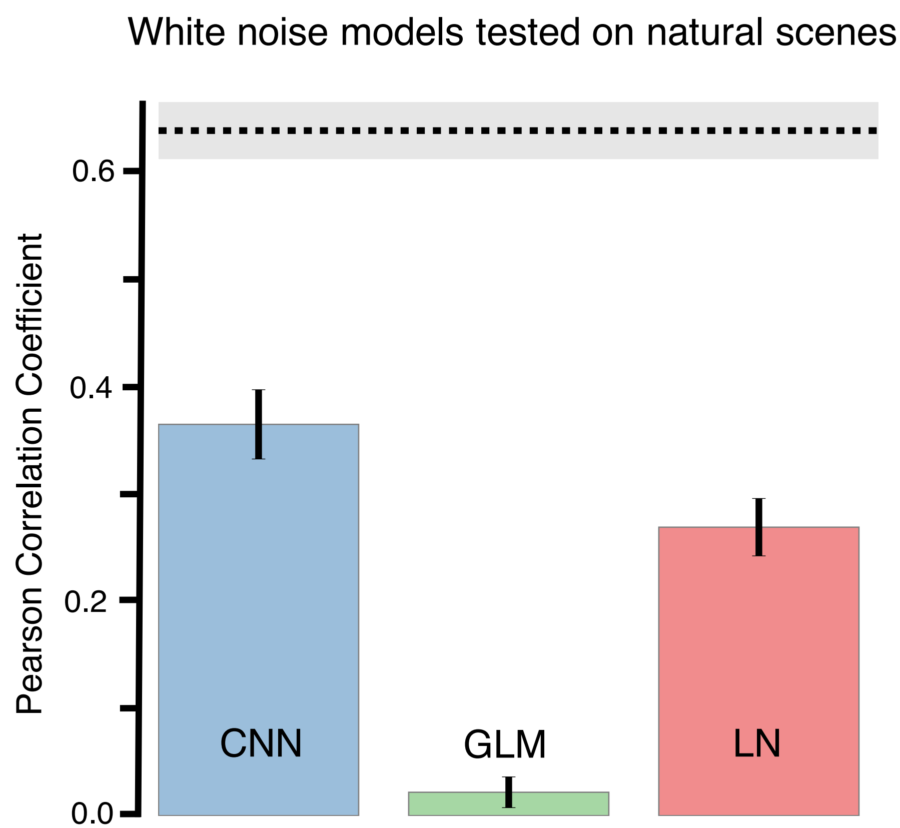
- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding

CNN models generalize better than simpler models



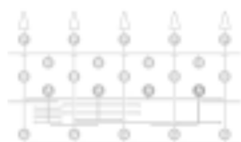
CNN models generalize better than simpler models



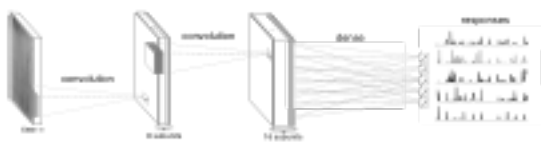


Outline

What we can still learn from biology?



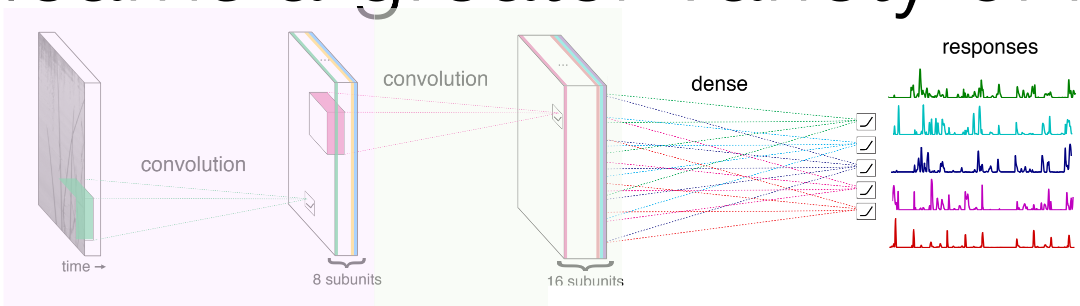
What is the retina and what does it do?



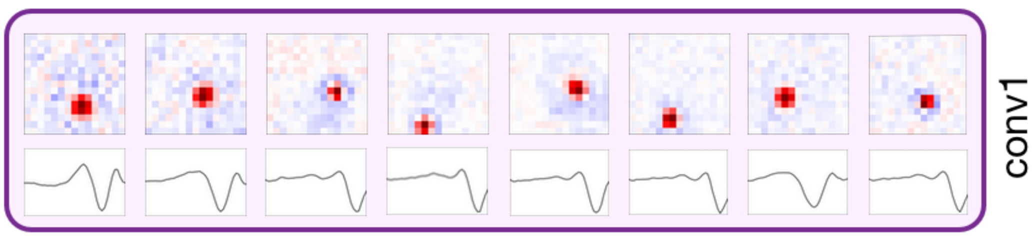
What can we learn from convolutional neural network models of the retina?

- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding

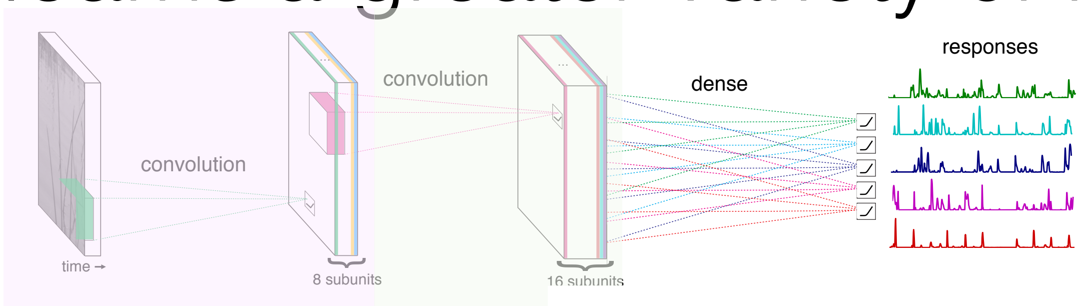
Model trained on responses to natural scenes learns a greater variety of features



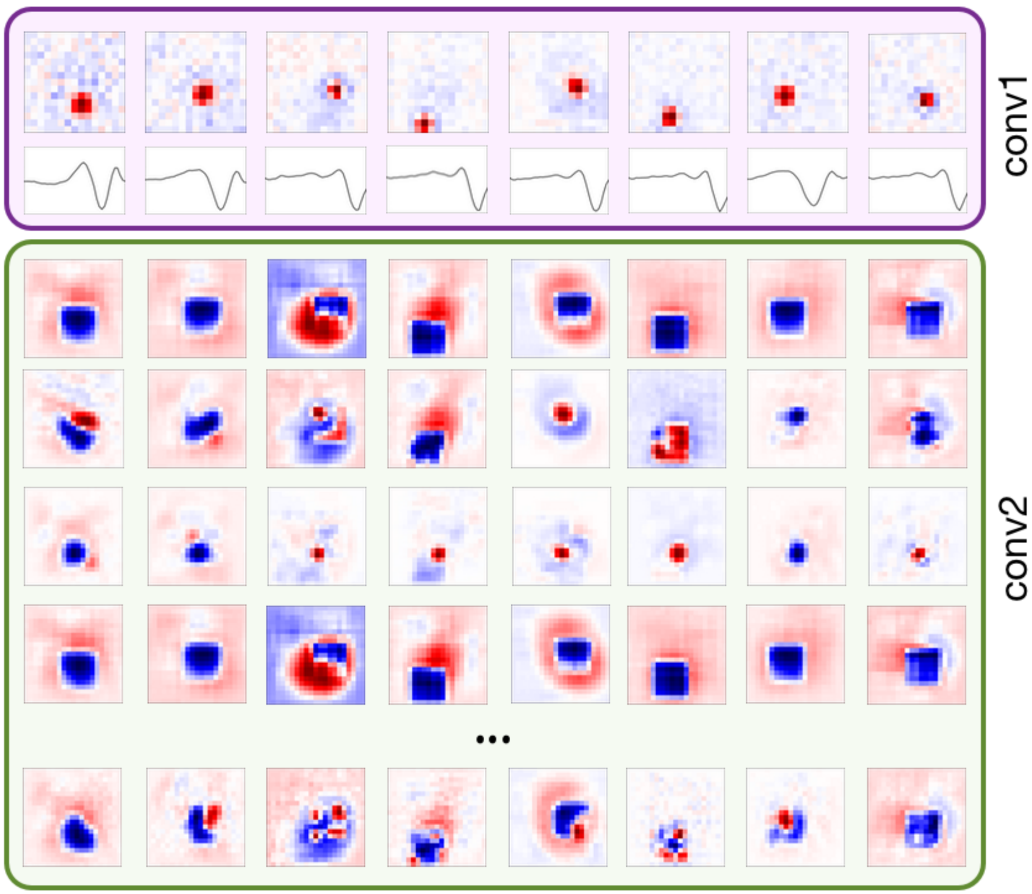
CNN filters trained on white noise



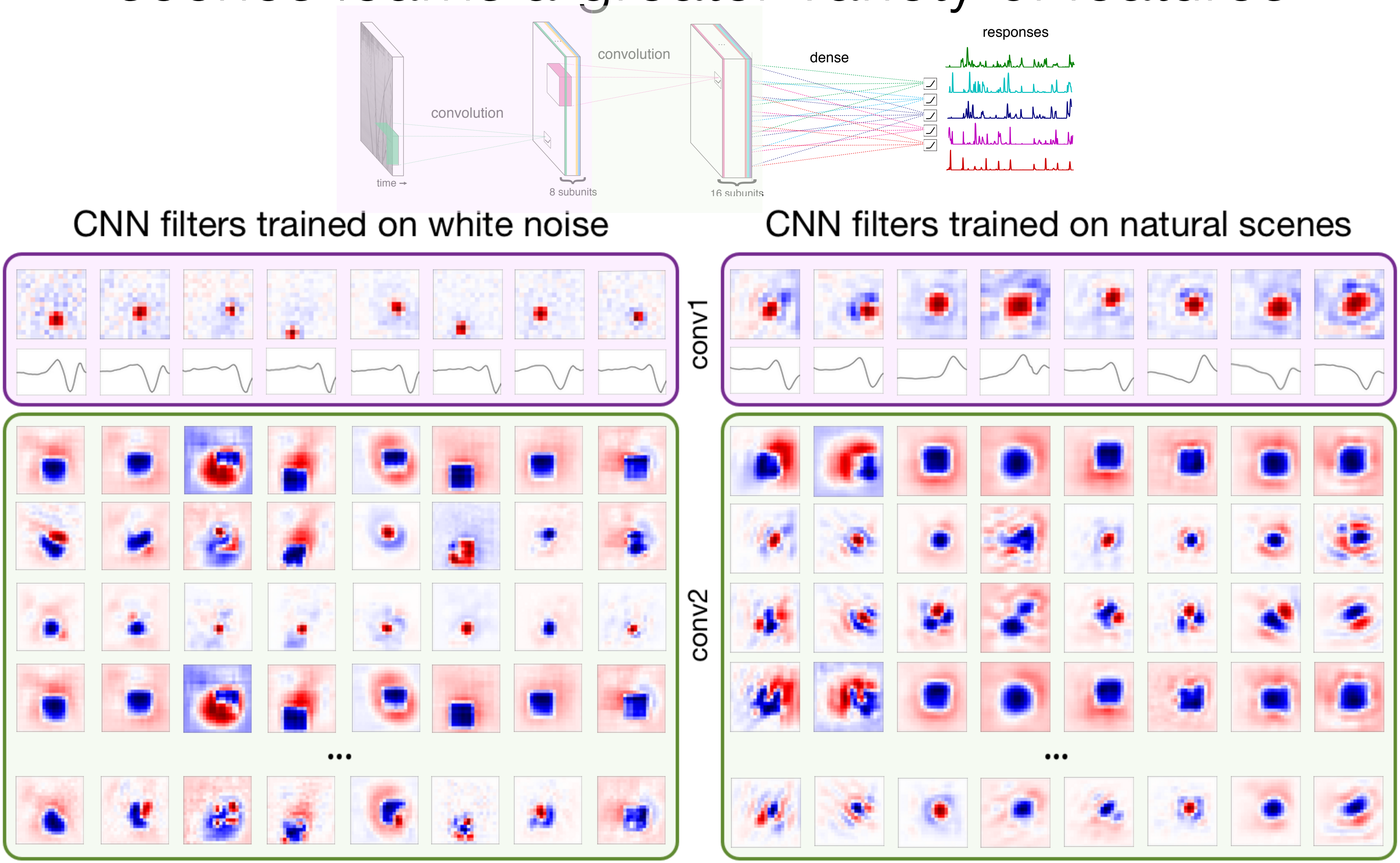
Model trained on responses to natural scenes learns a greater variety of features



CNN filters trained on white noise

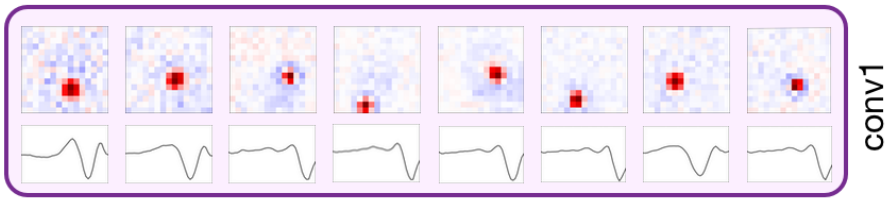


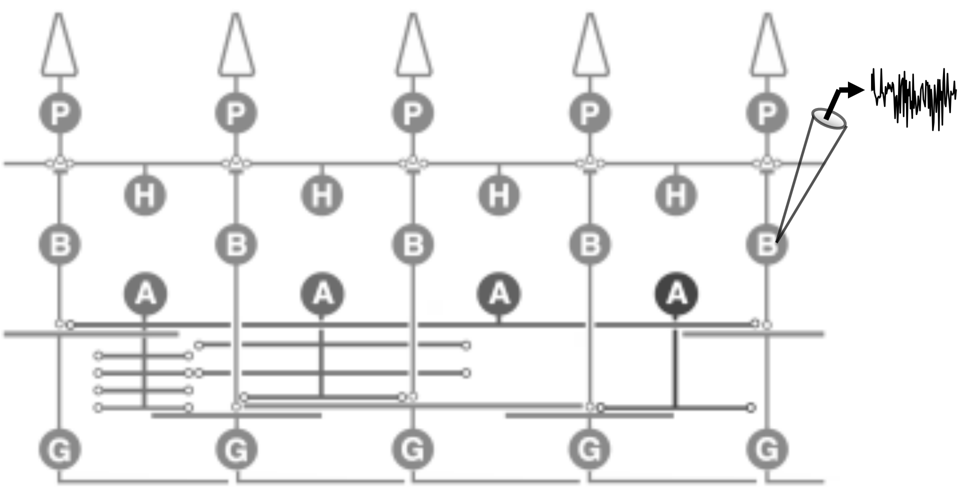
Model trained on responses to natural scenes learns a greater variety of features



Features bear striking resemblance to unobserved structure in retina

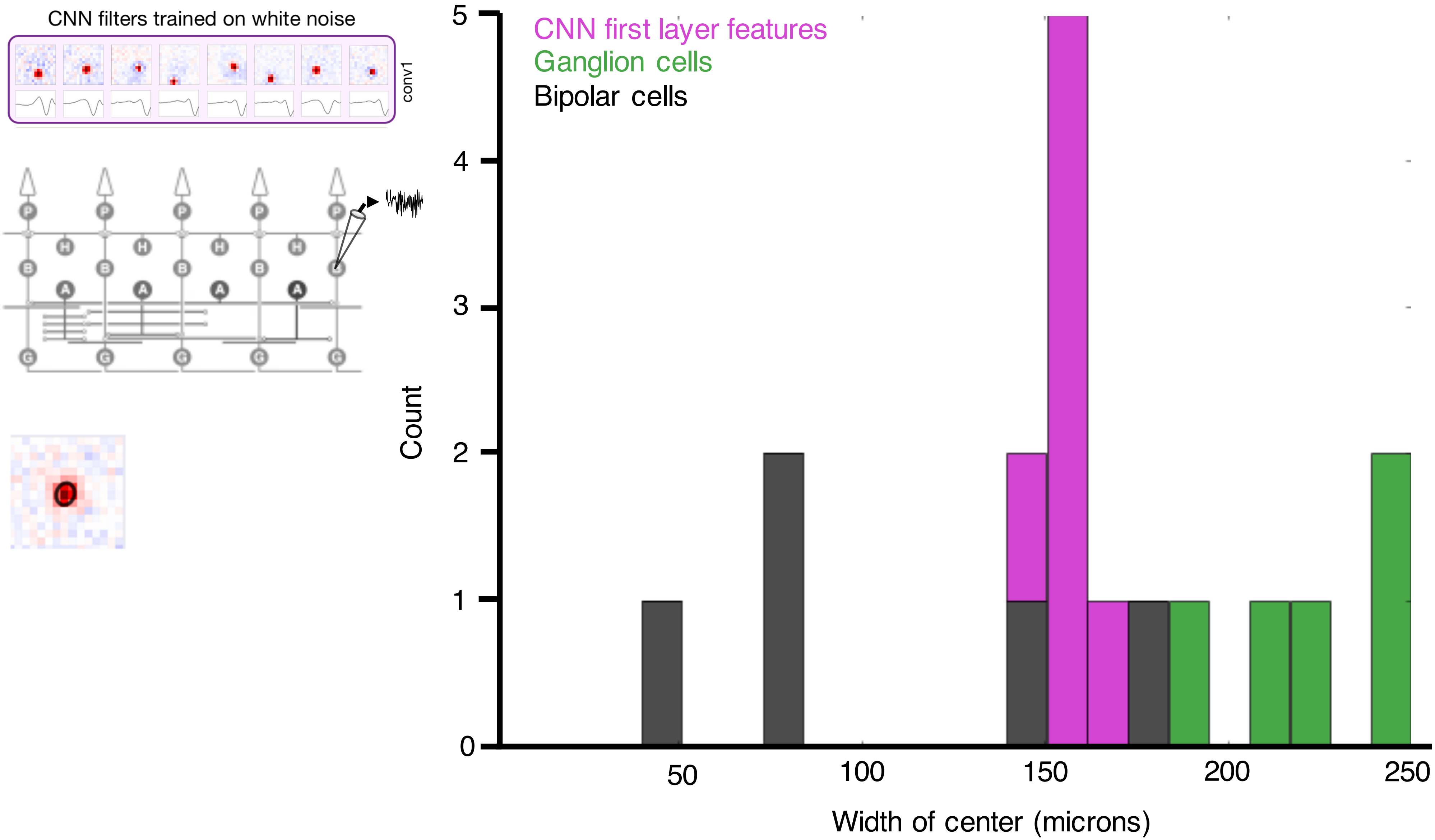
CNN filters trained on white noise





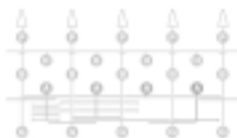
Data courtesy of David Kastner

Features bear striking resemblance to unobserved structure in retina

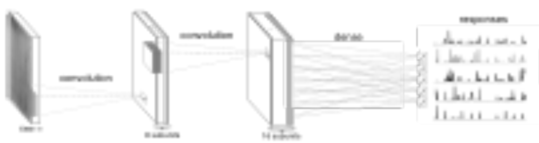


Outline

What we can still learn from biology?



What is the retina and what does it do?



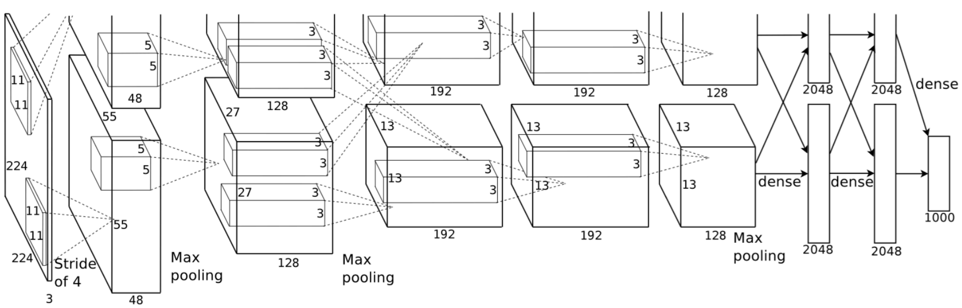
What can we learn from convolutional neural network models of the retina?

- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding

How certain are we about our predictions?

training data

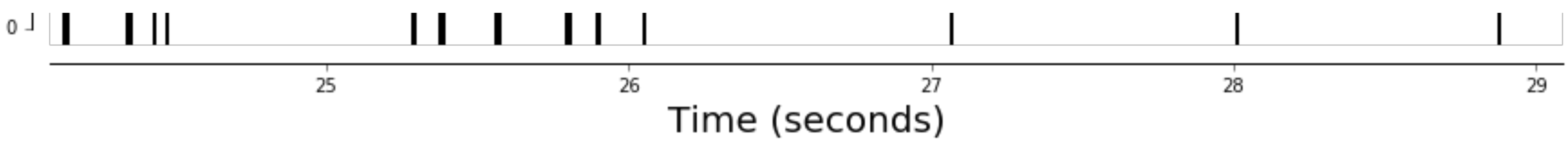
model



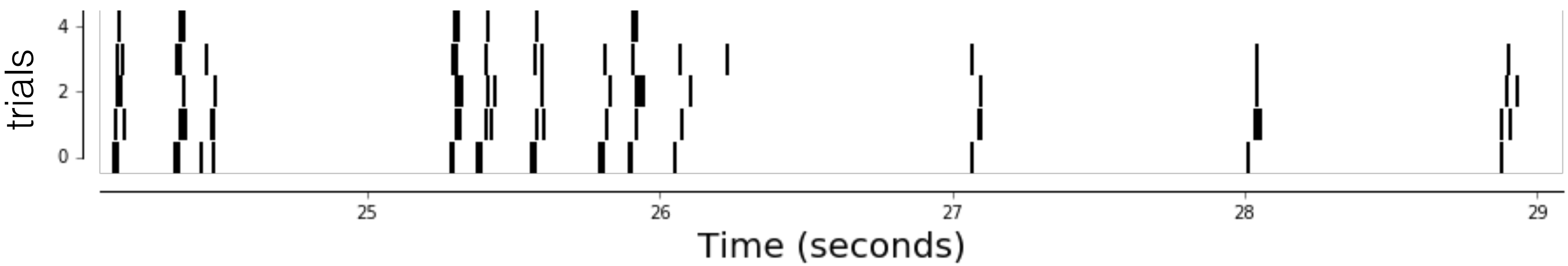
cat or dog?

test data

How certain are we of these spikes?

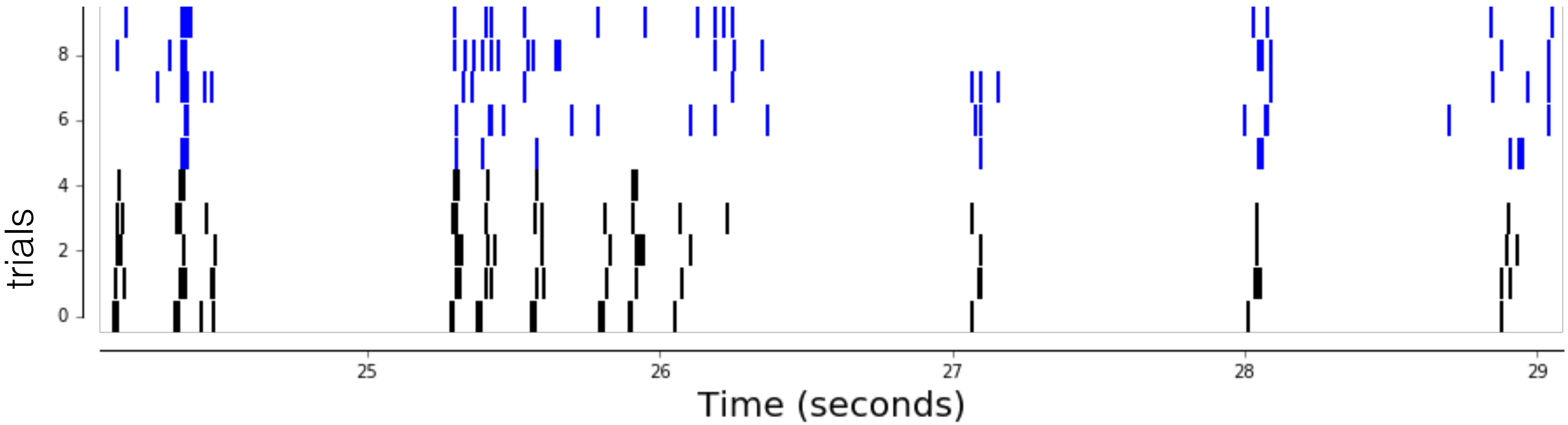


How certain are we of these spikes?

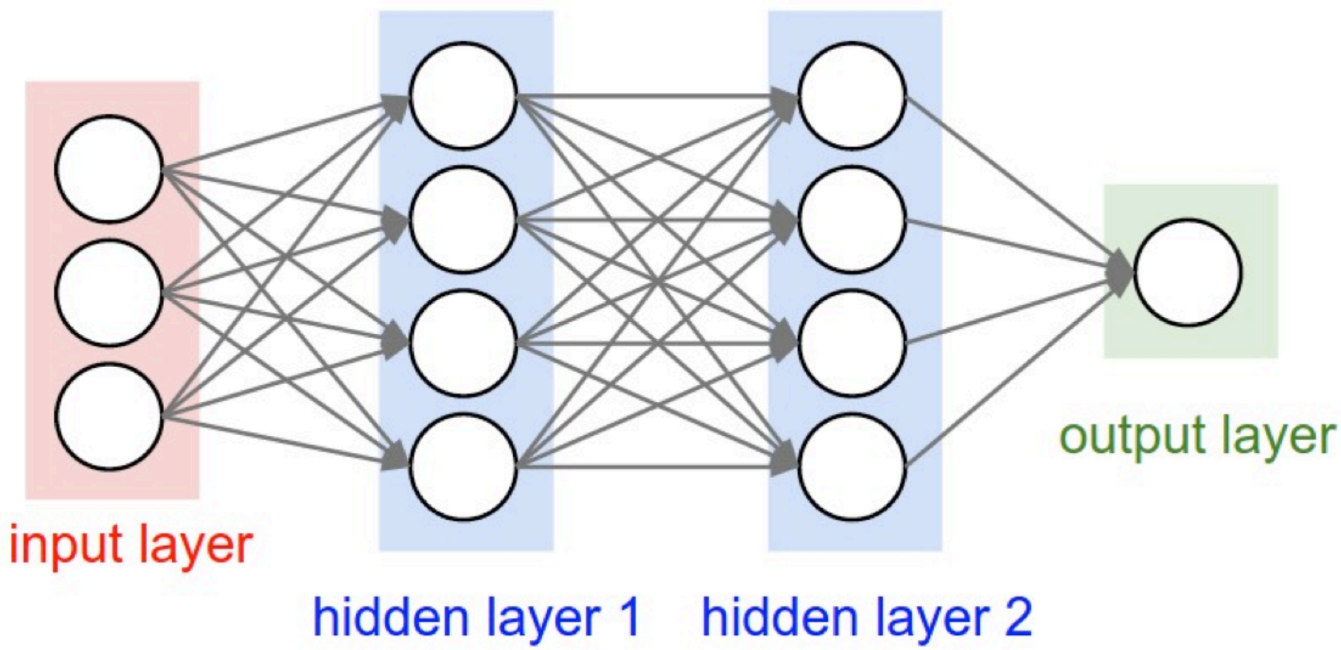


How certain are we of these spikes?

a good model of the retina should capture both the mean response and its variance



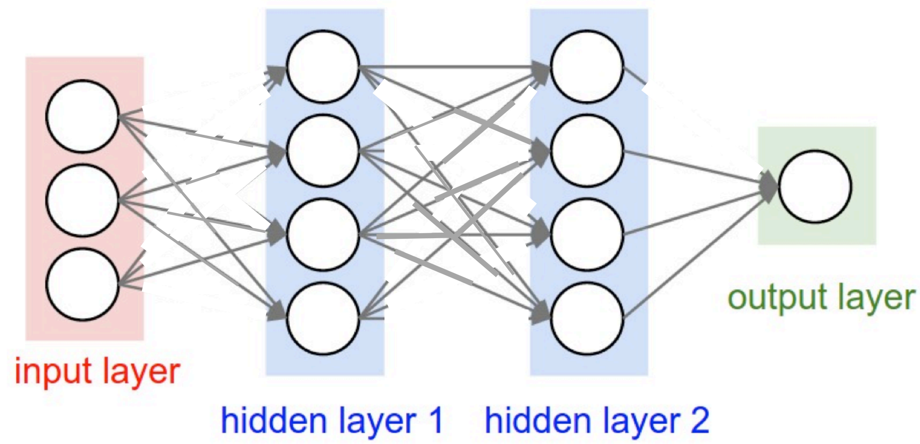
But how can deep learning models capture uncertainty?



We do lots of probabilistic tricks during training however…

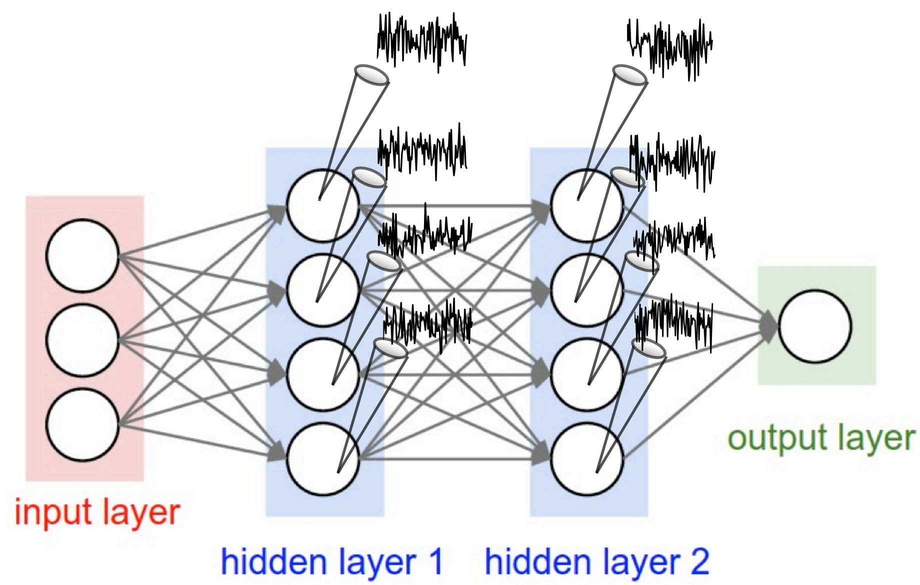
dropout

drop connection with probability p

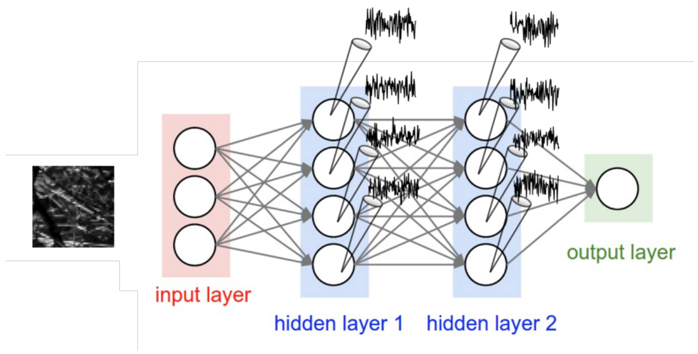


Gaussian noise injection

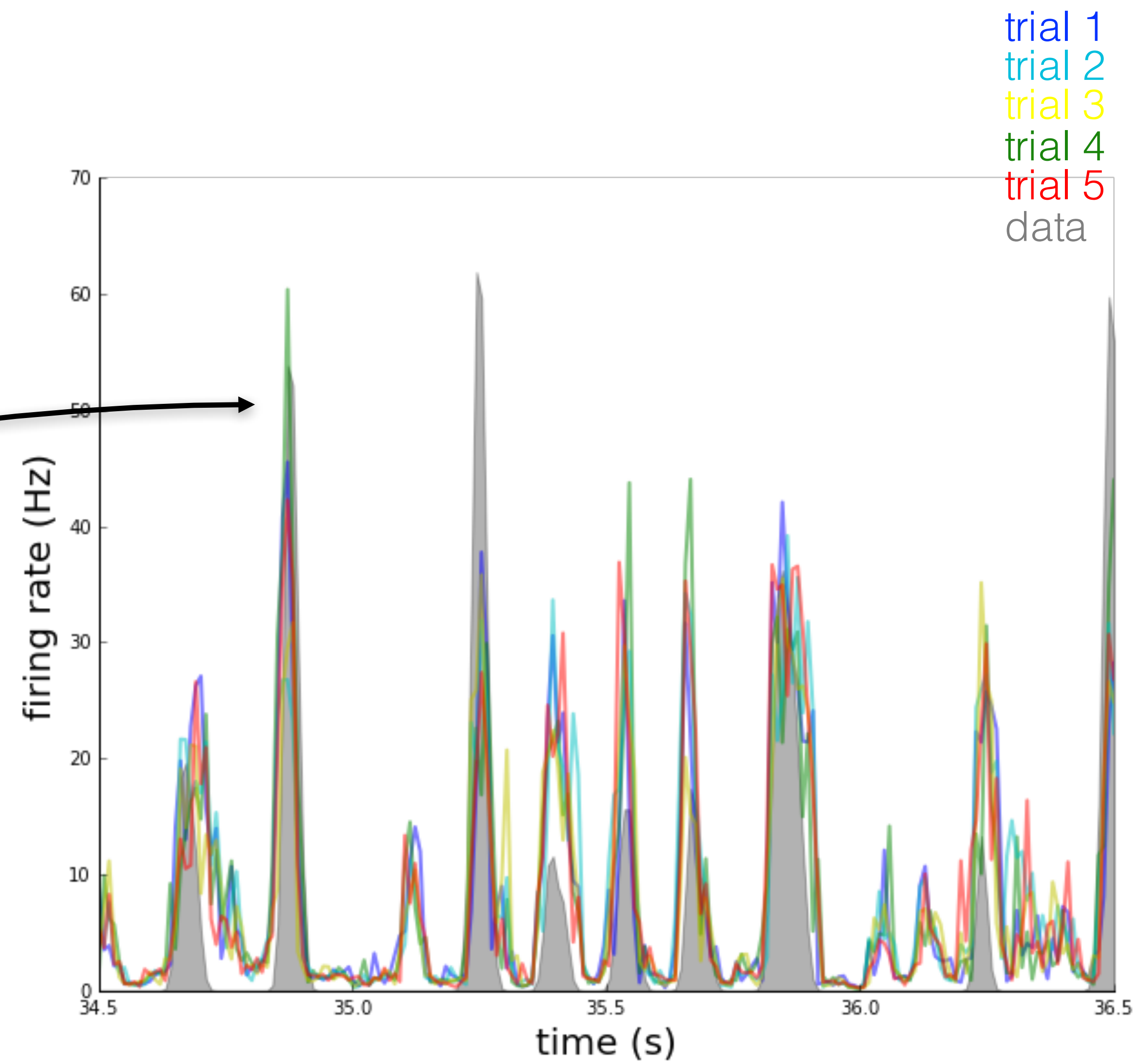
inject hidden units with noise of variance σ^2



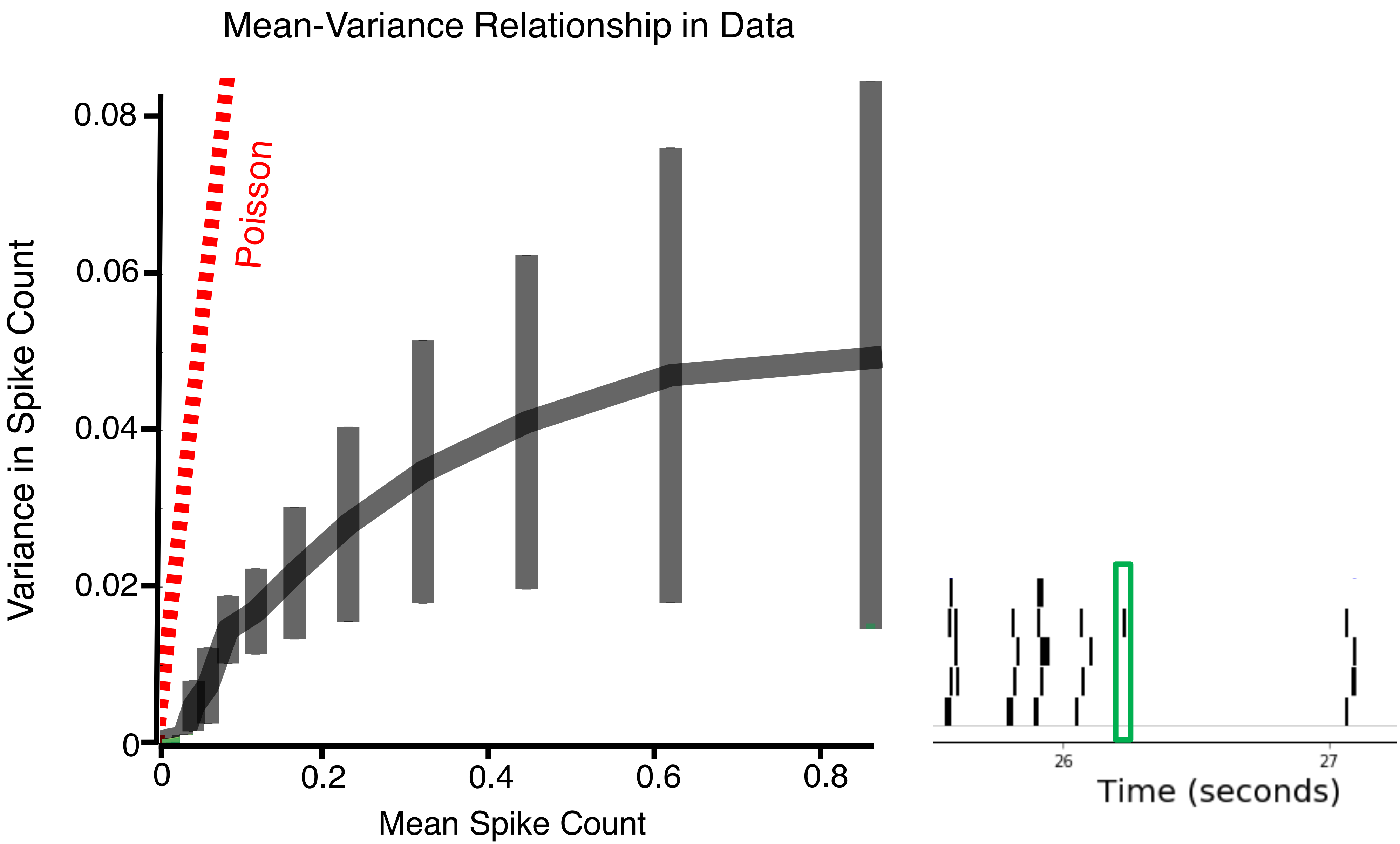
We now have deep probabilistic models



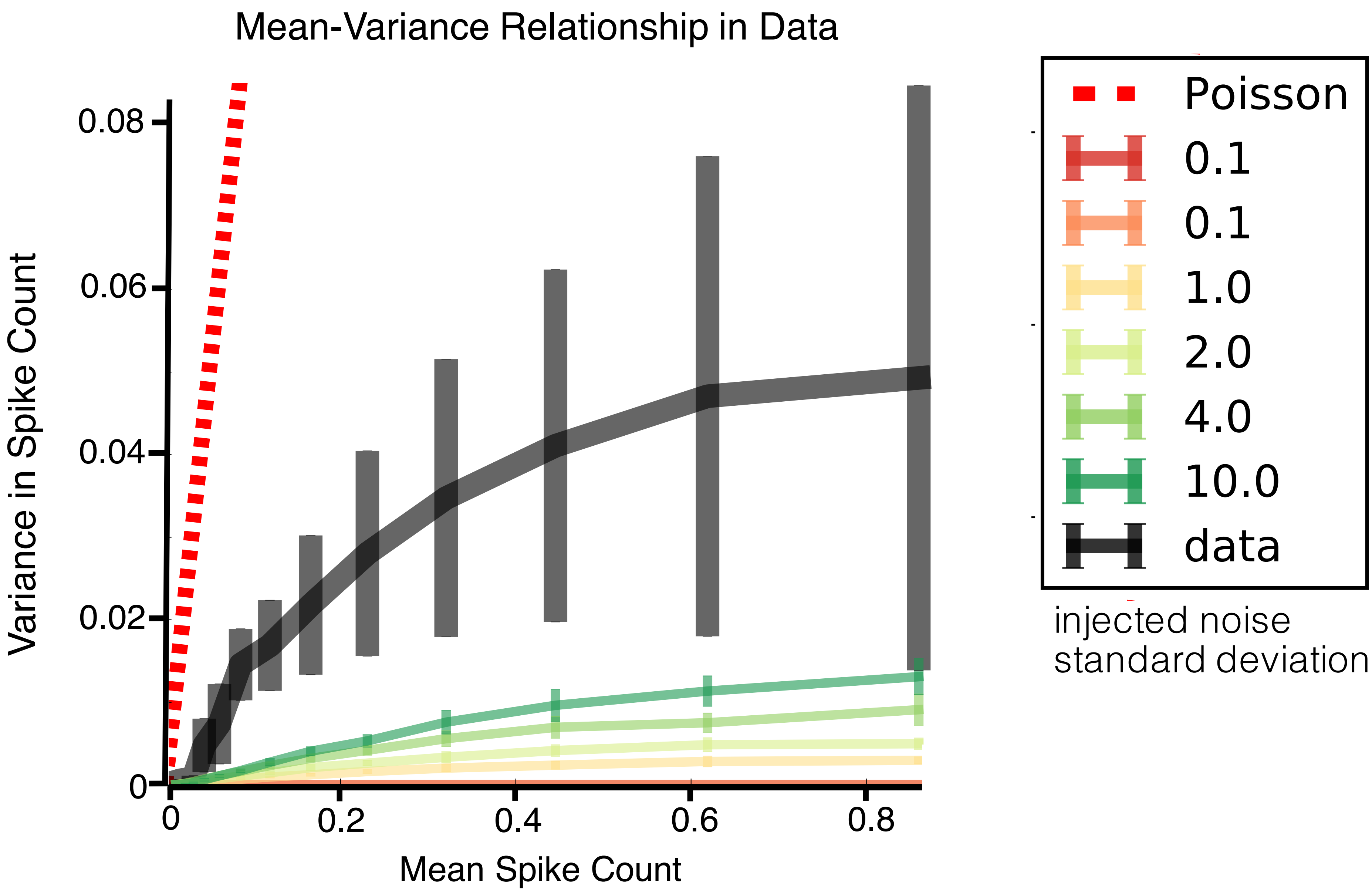
we control the noisiness by adjusting standard deviation of noise injection



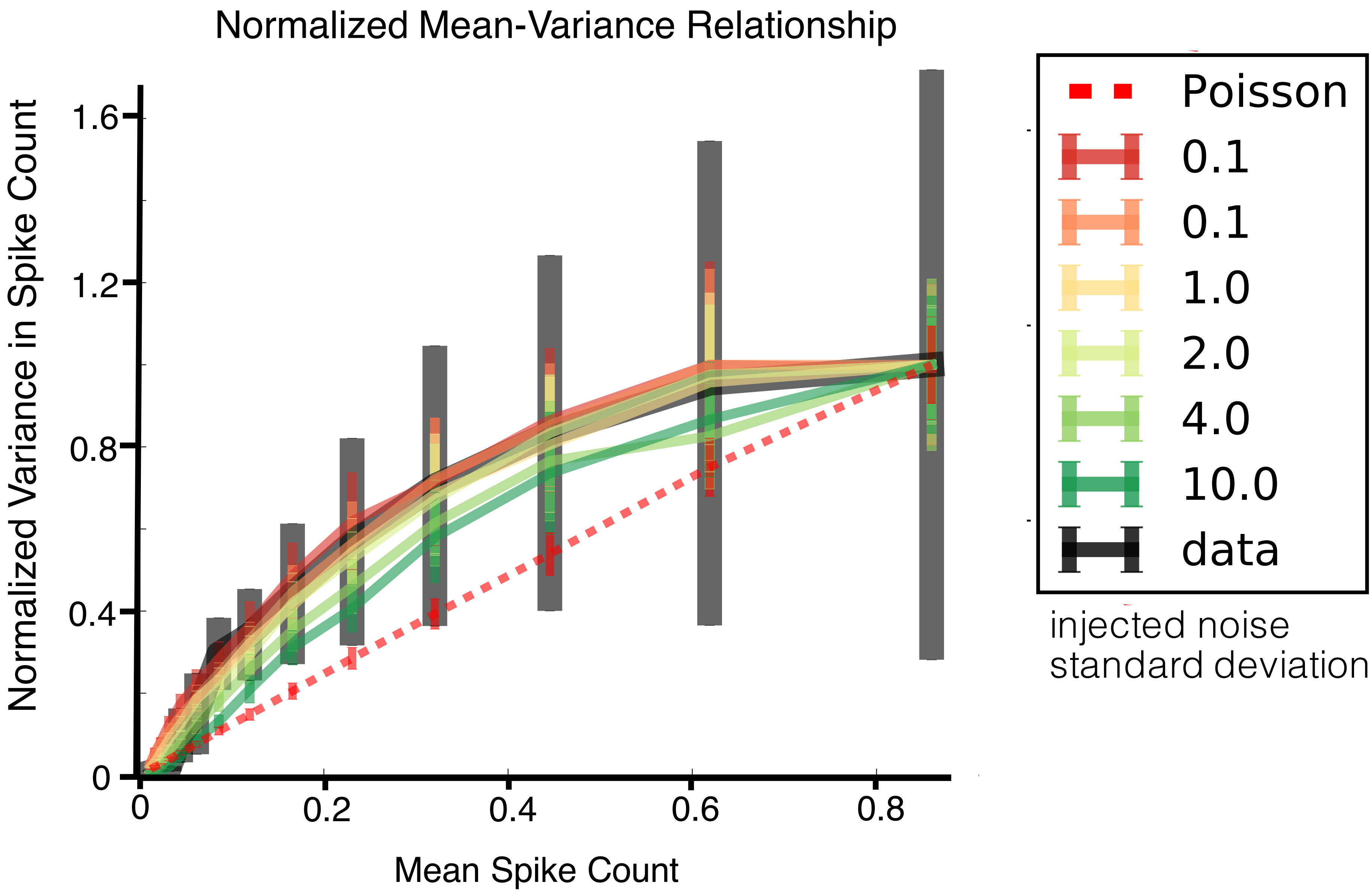
Mean-variance relationships



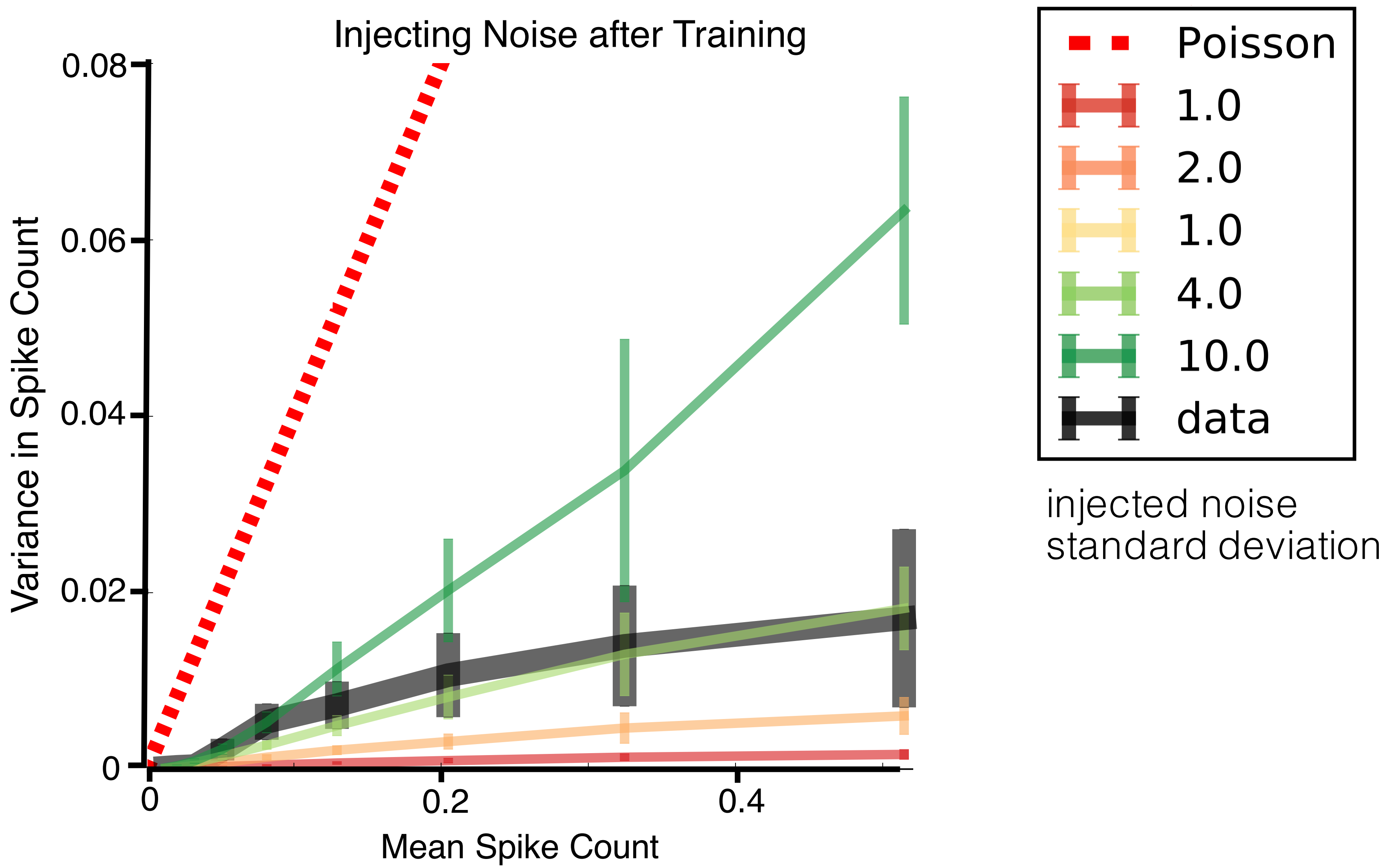
Model has lower variance than data



However model uncertainty has same scaling relationship as the retina

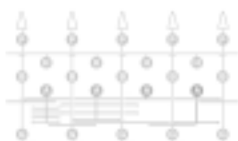


This is not true when noise is injected after training

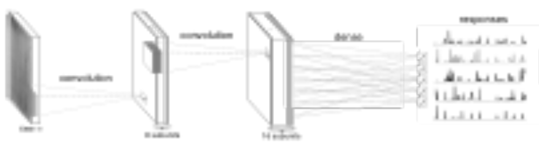


Outline

What we can still learn from biology?



What is the retina and what does it do?



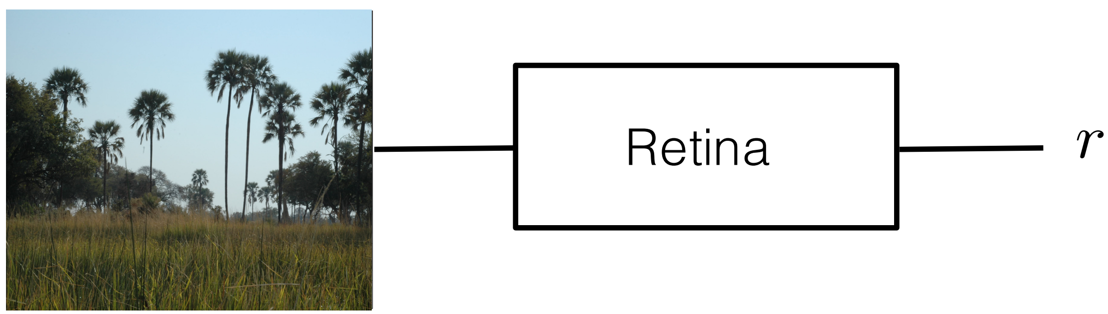
What can we learn from convolutional neural network models of the retina?

- architecture
- performance
- generalization
- model features
- capturing uncertainty
- efficient coding

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints



Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints

1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

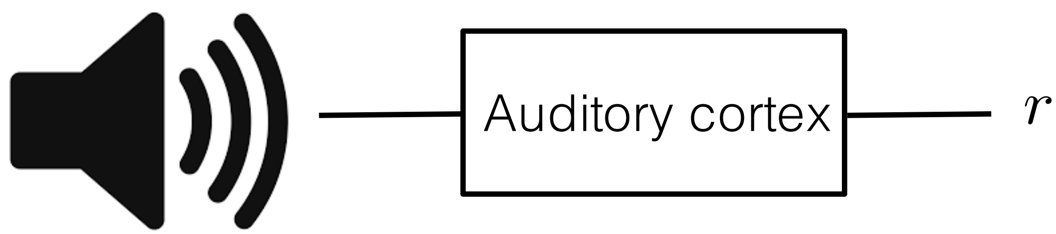
2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints



1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

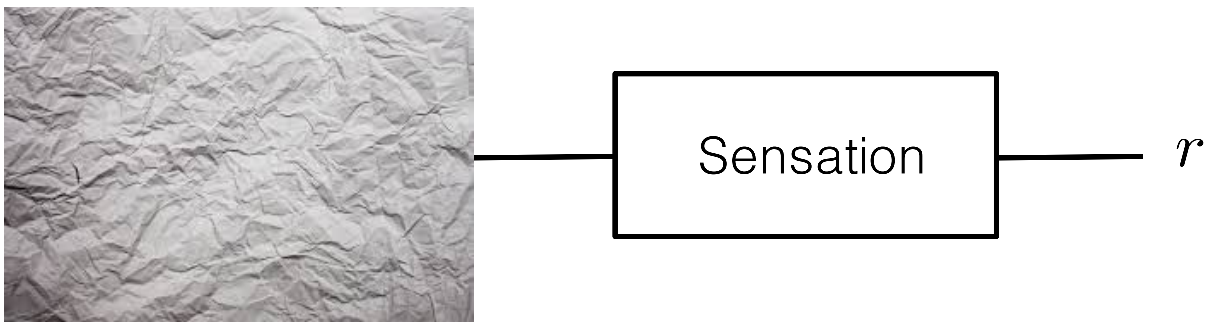
2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints



1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

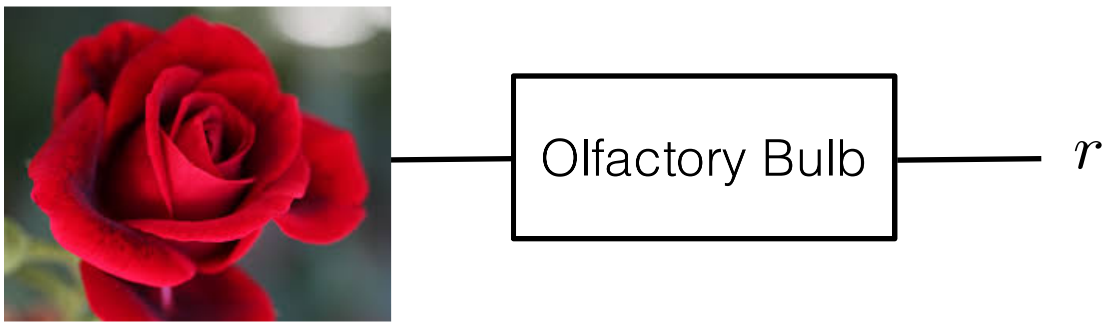
2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints



1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

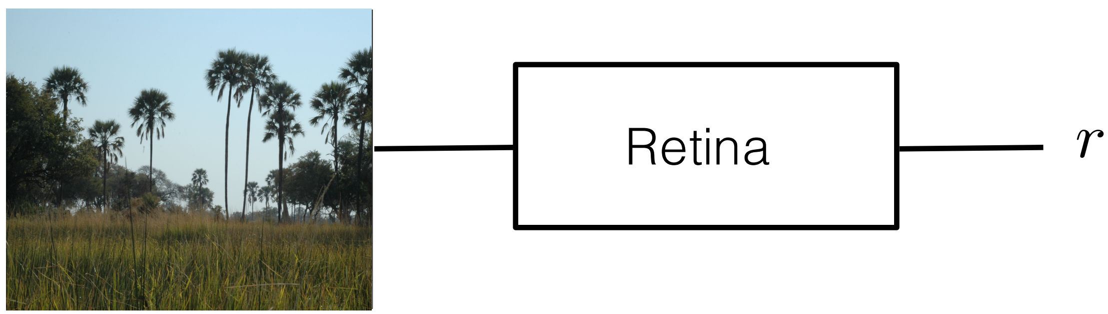
2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints



1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints

Retina — r

1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints

Retina

r

1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints

Retina

r

1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

2010s

Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as communication channel

1950 - present

Maximize information about the visual world,
subject to biological constraints

Retina

r

1950s

Barlow 1952
Attneave 1954

1960s

Barlow 1961

1980s

Laughlin 1981

1990s

Atick and Redlich 1990
Atick 1991
Van Hateren 1992
Linsker 1992
Dan et al. 1996
Olshausen and Field 1996

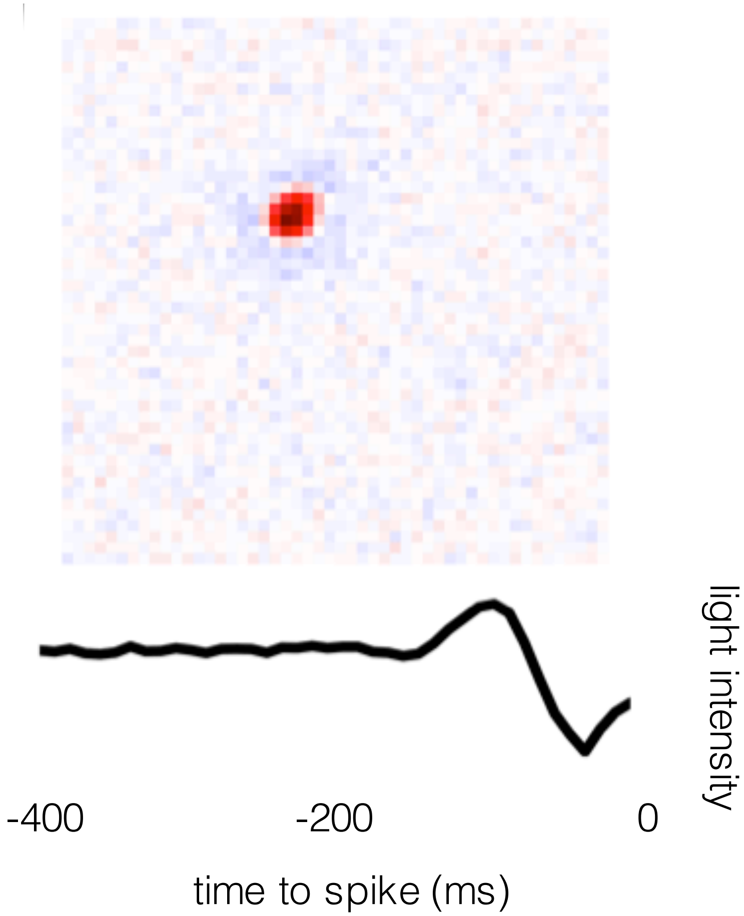
2000s

Bell 2001
Fairhall et al. 2001
Schwartz and Simoncelli 2001
Balasubramanian et al. 2002
Lewicki 2002
Puchalla et al. 2005
Hosoya et al. 2005
Smith and Lewicki 2006
Chelaru et al. 2008
Kostal et al. 2008
Vig et al. 2009

2010s

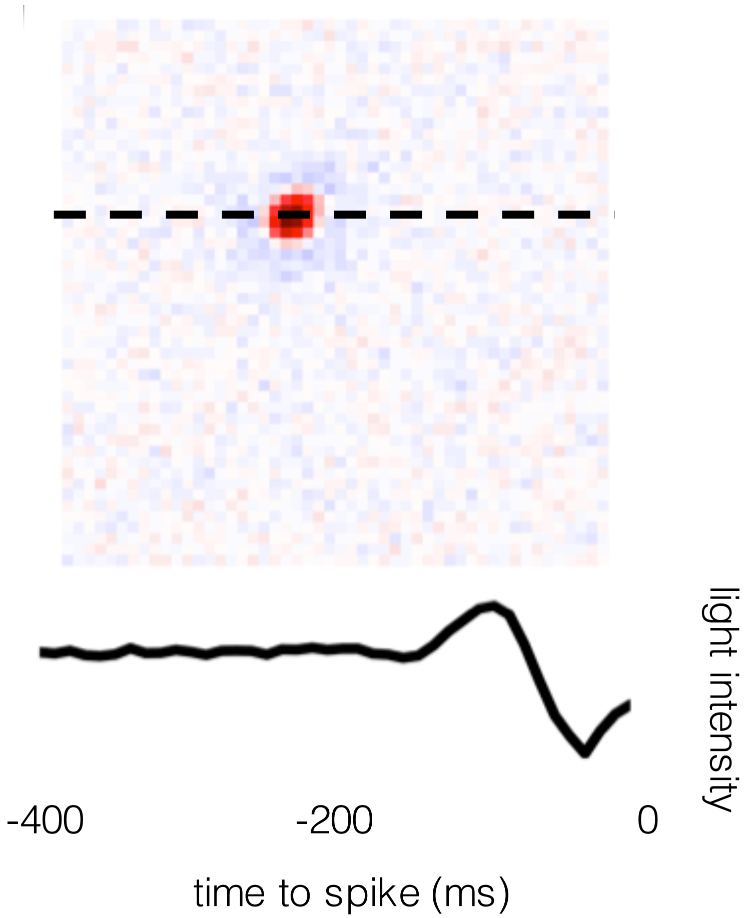
Friston 2010
Tkacik et al. 2010
de-Wit et al. 2010
Stilp et al. 2010
Adibi et al. 2011
Karklin and Simoncelli 2011
Chacron et al. 2011
Pitkow and Meister 2012
Doi et al. 2012
Gjorgjieva et al. 2014
Kastner et al. 2015
Segal et al. 2015

Retina as decorrelating visual scene



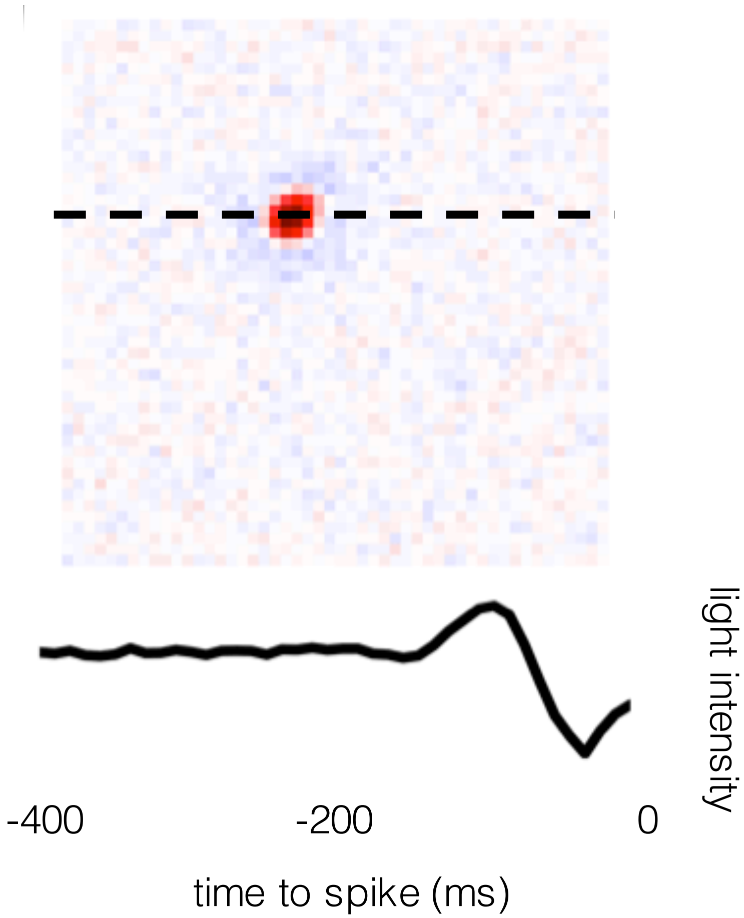
retinal ganglion
receptive field

Retina as decorrelating visual scene

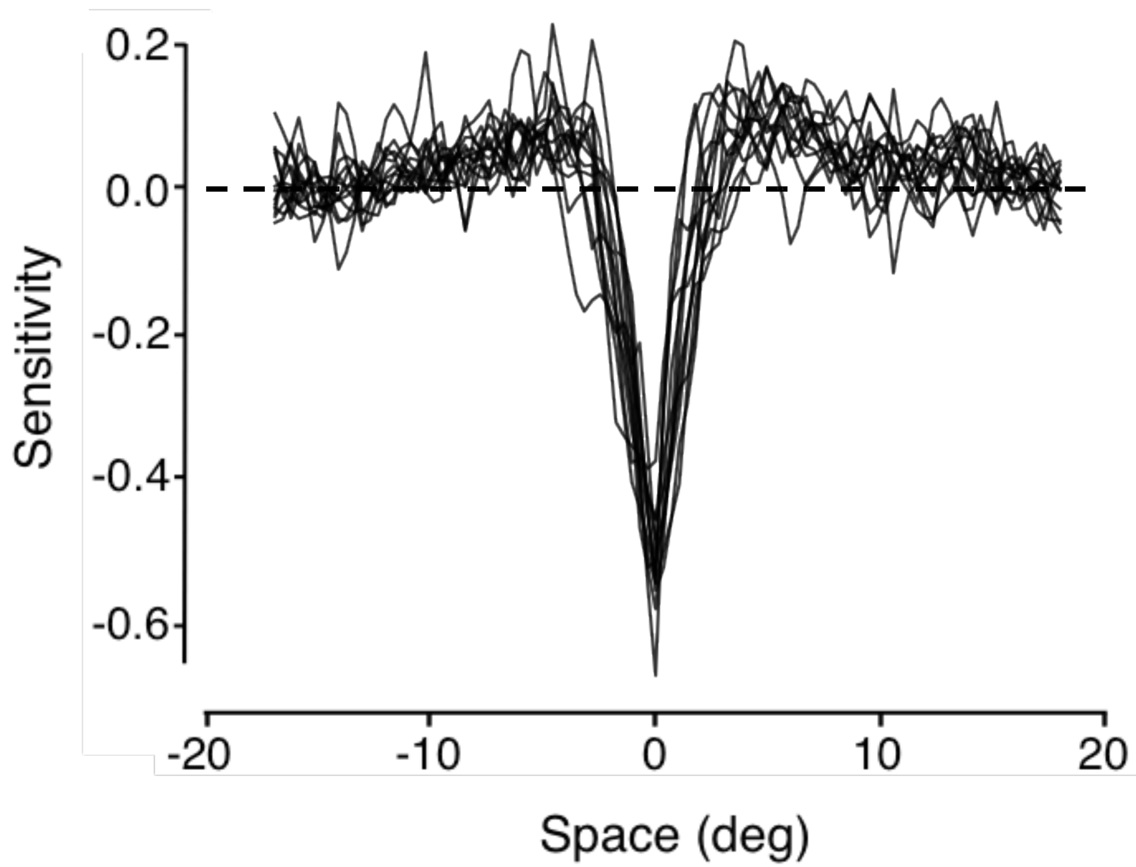


retinal ganglion
receptive field

Retina as decorrelating visual scene

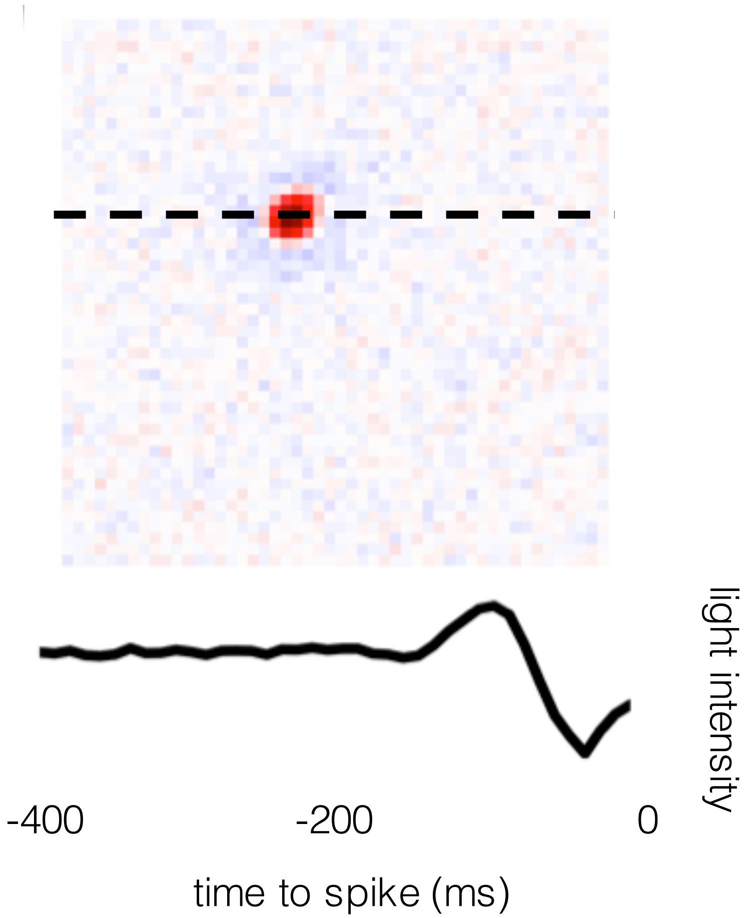


retinal ganglion
receptive field

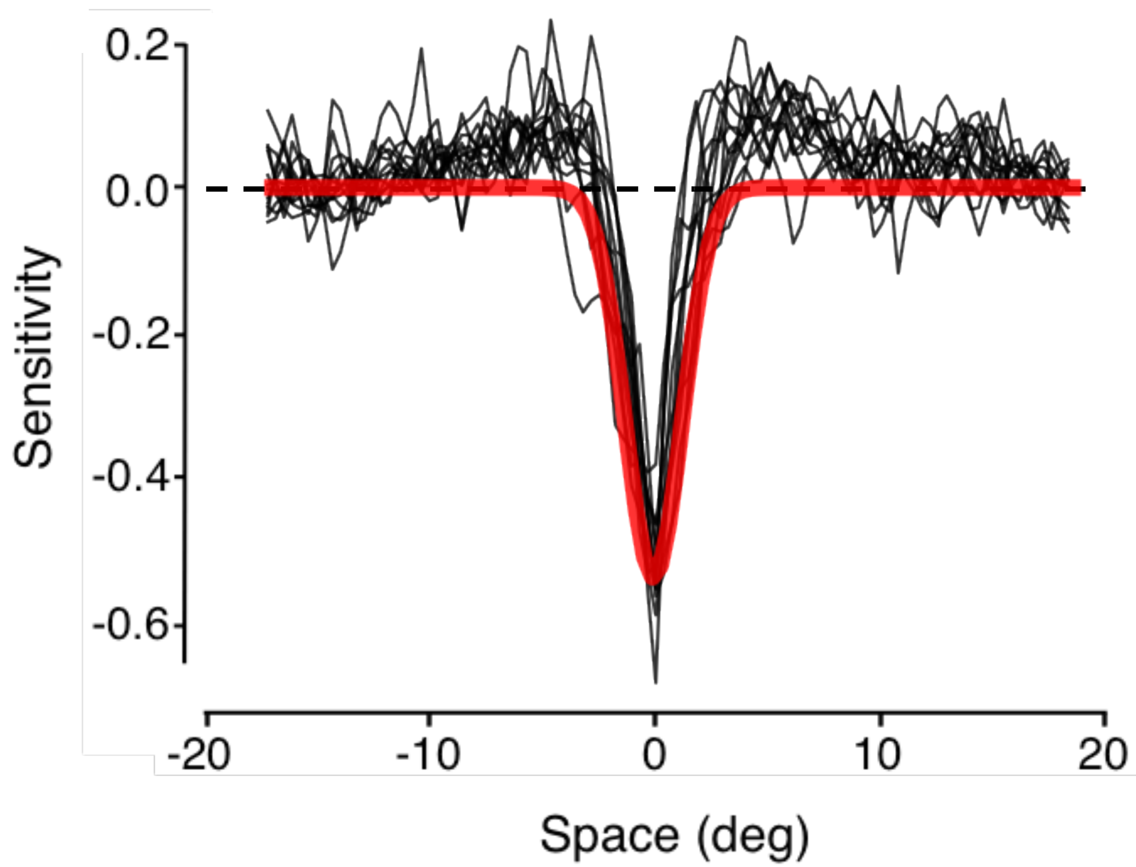


Ganglion cells, n = 13

Retina as decorrelating visual scene

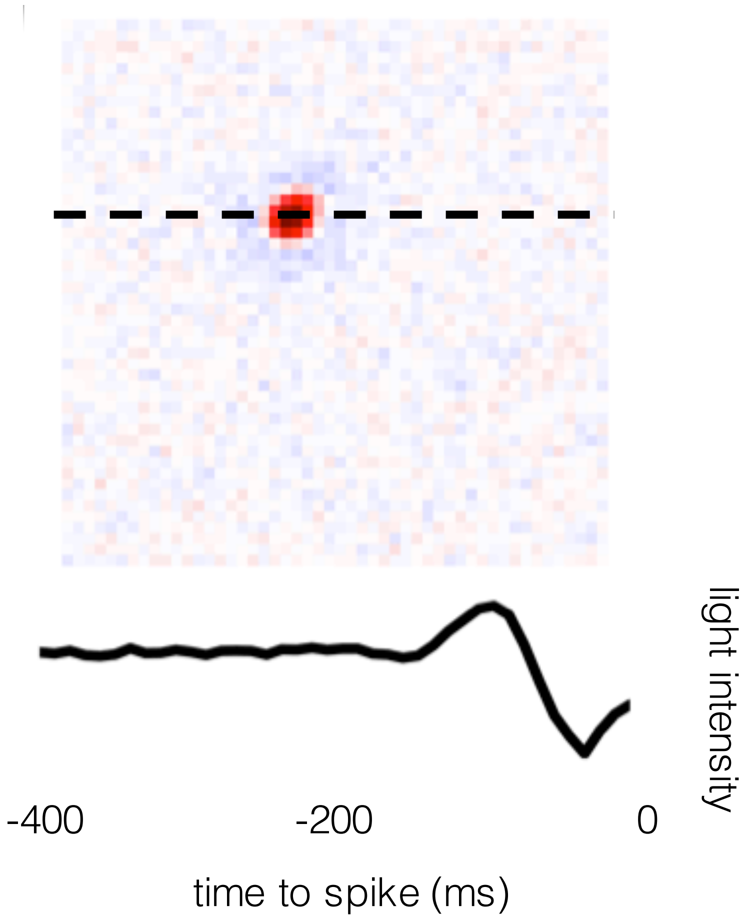


retinal ganglion
receptive field

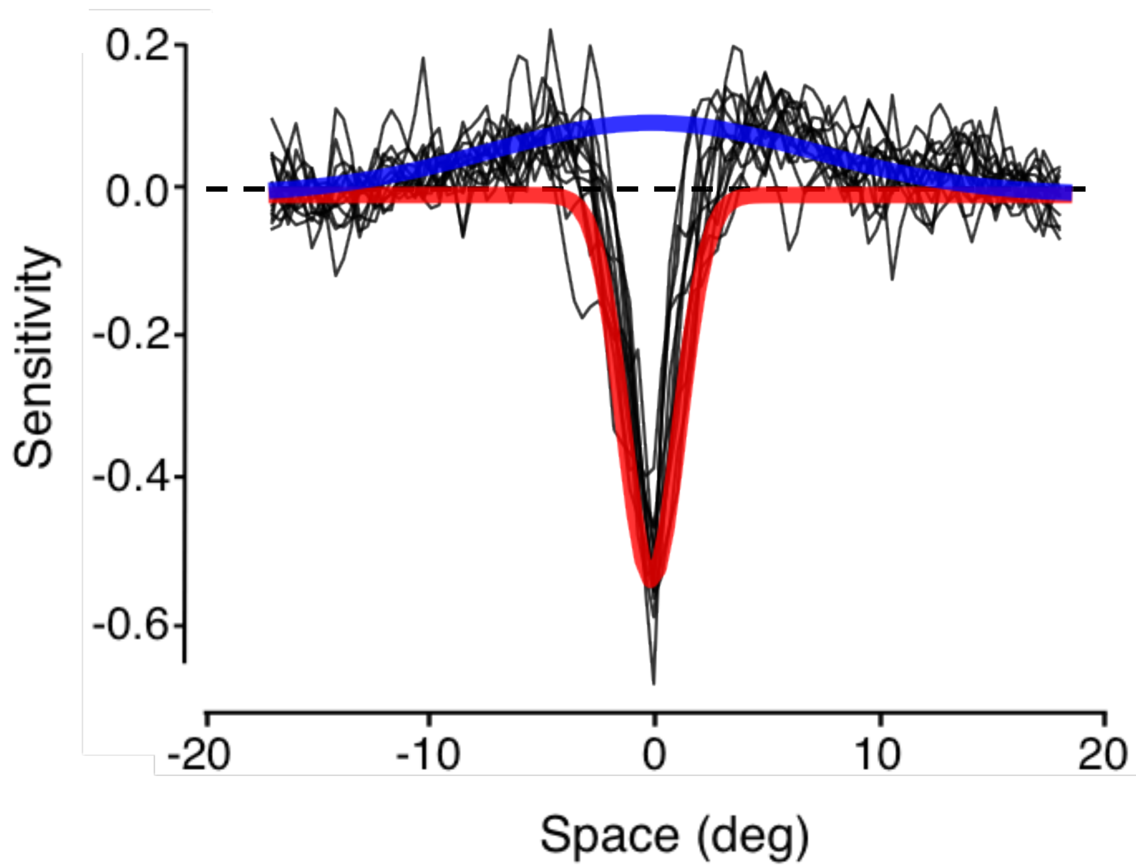


Ganglion cells, n = 13
Excitatory center

Retina as decorrelating visual scene

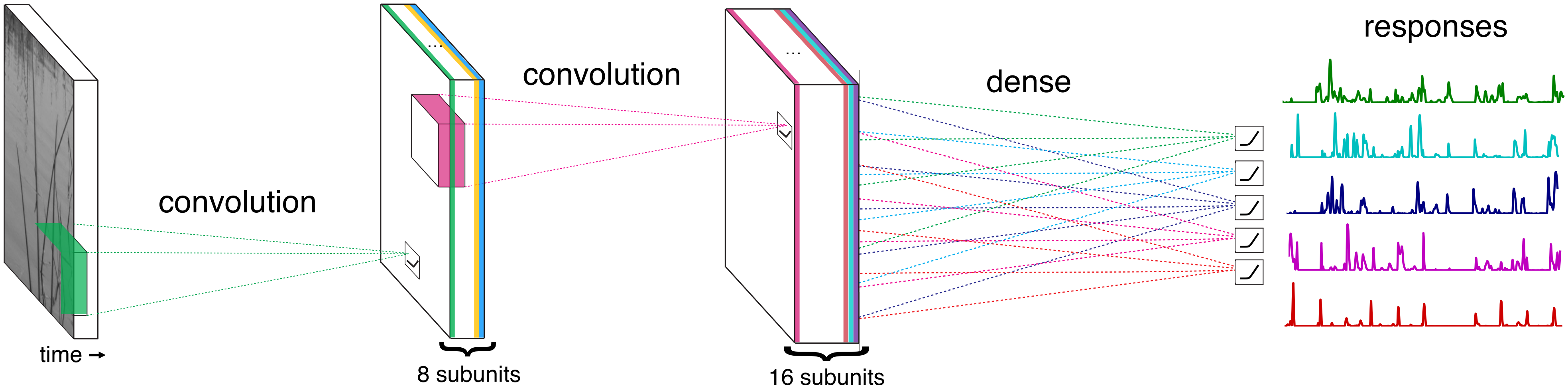


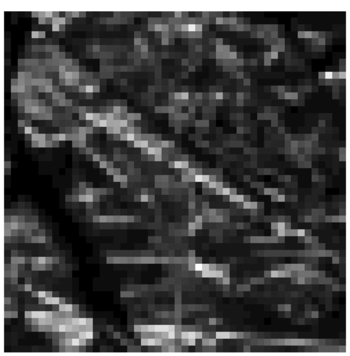
retinal ganglion
receptive field



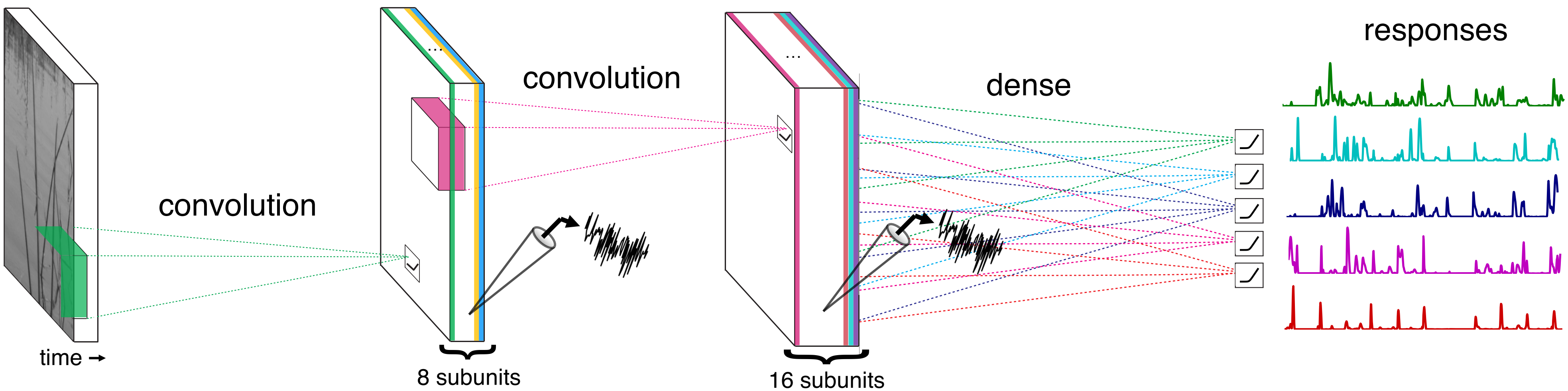
Ganglion cells, n = 13
Excitatory center
Inhibitory surround

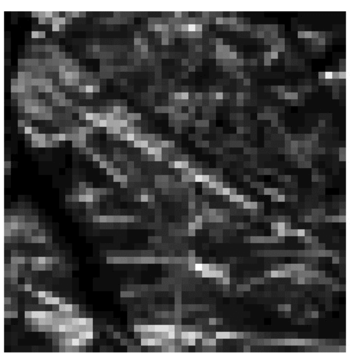
How do signals flow in a CNN?



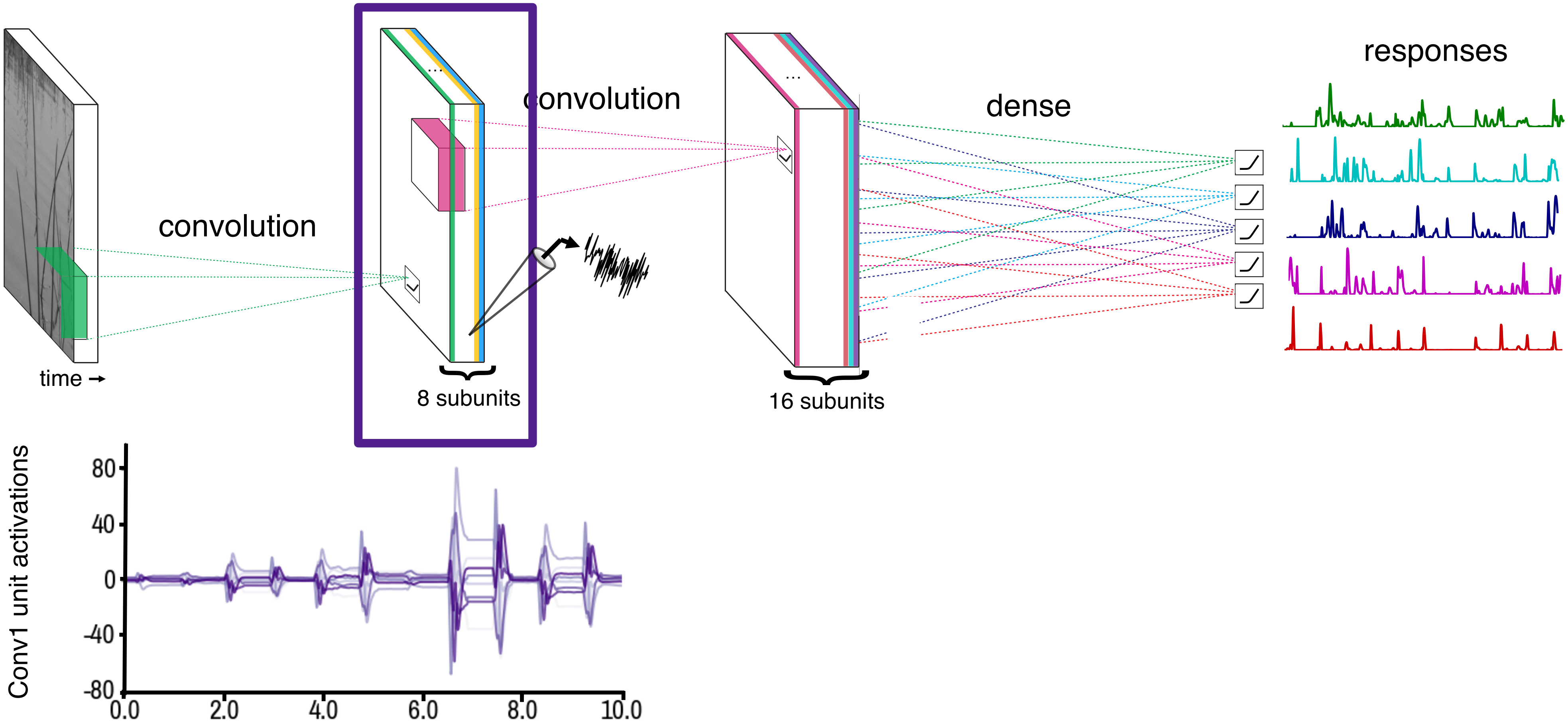


How do signals flow in a CNN?

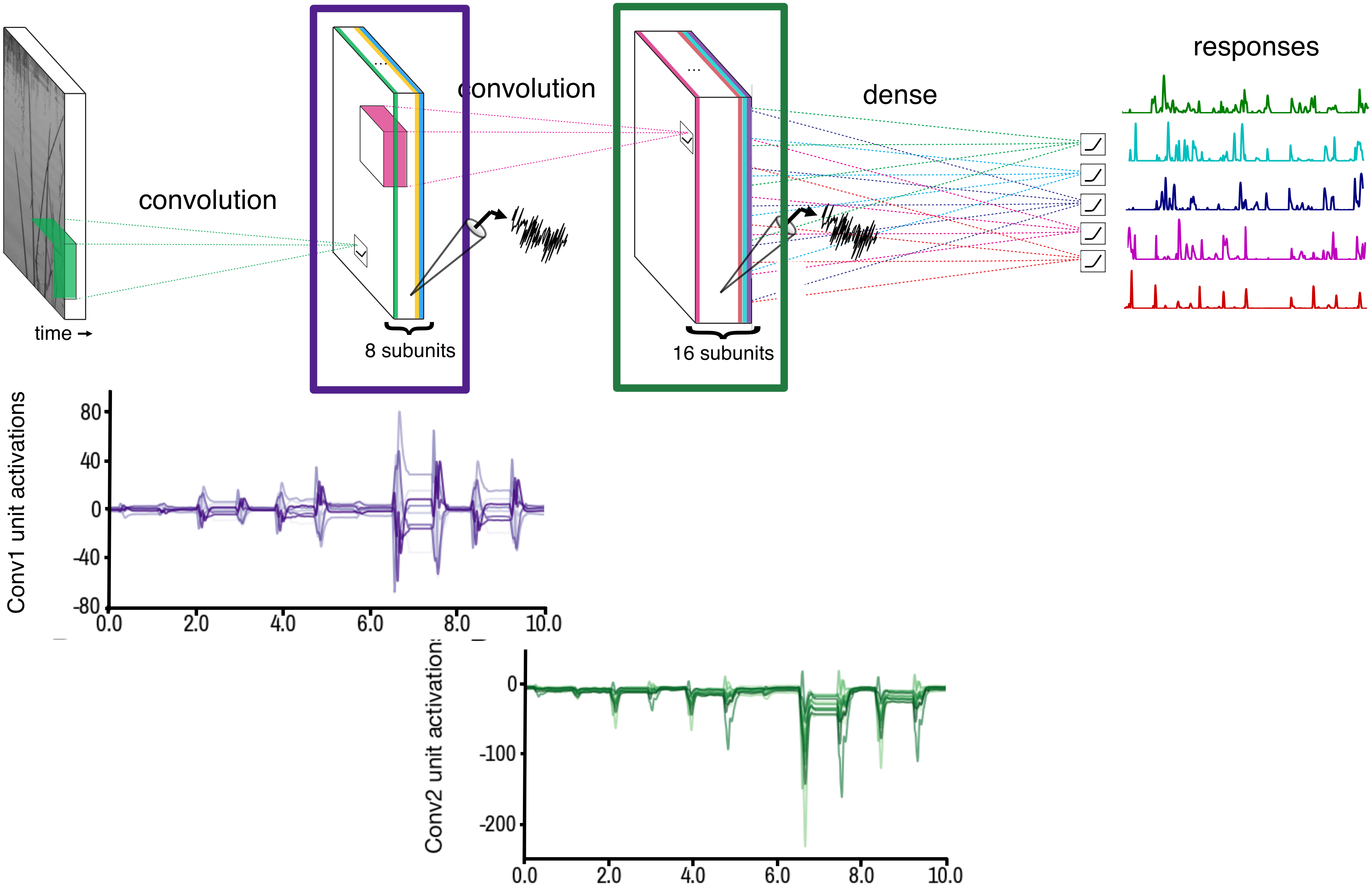




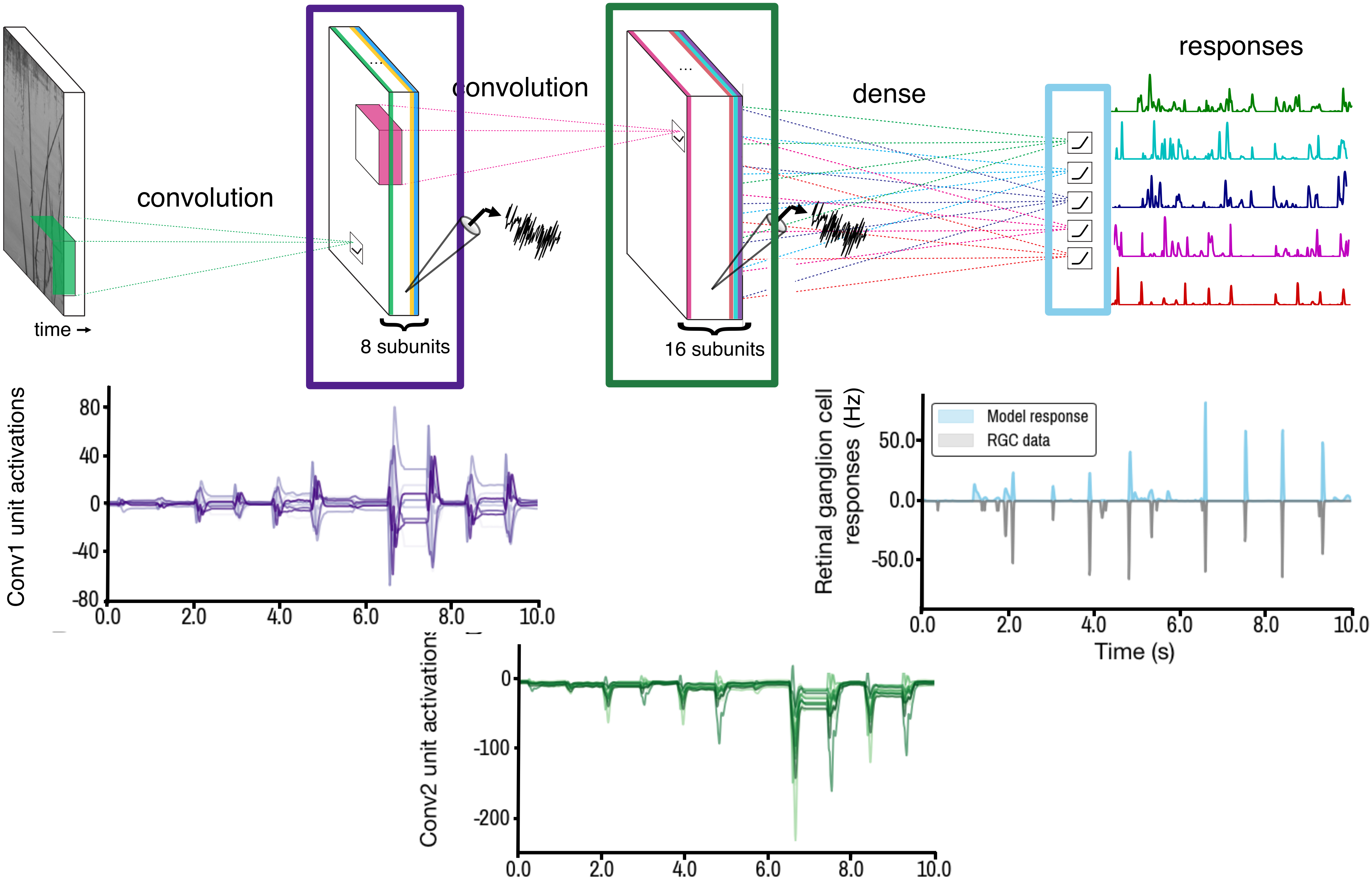
How do signals flow in a CNN?

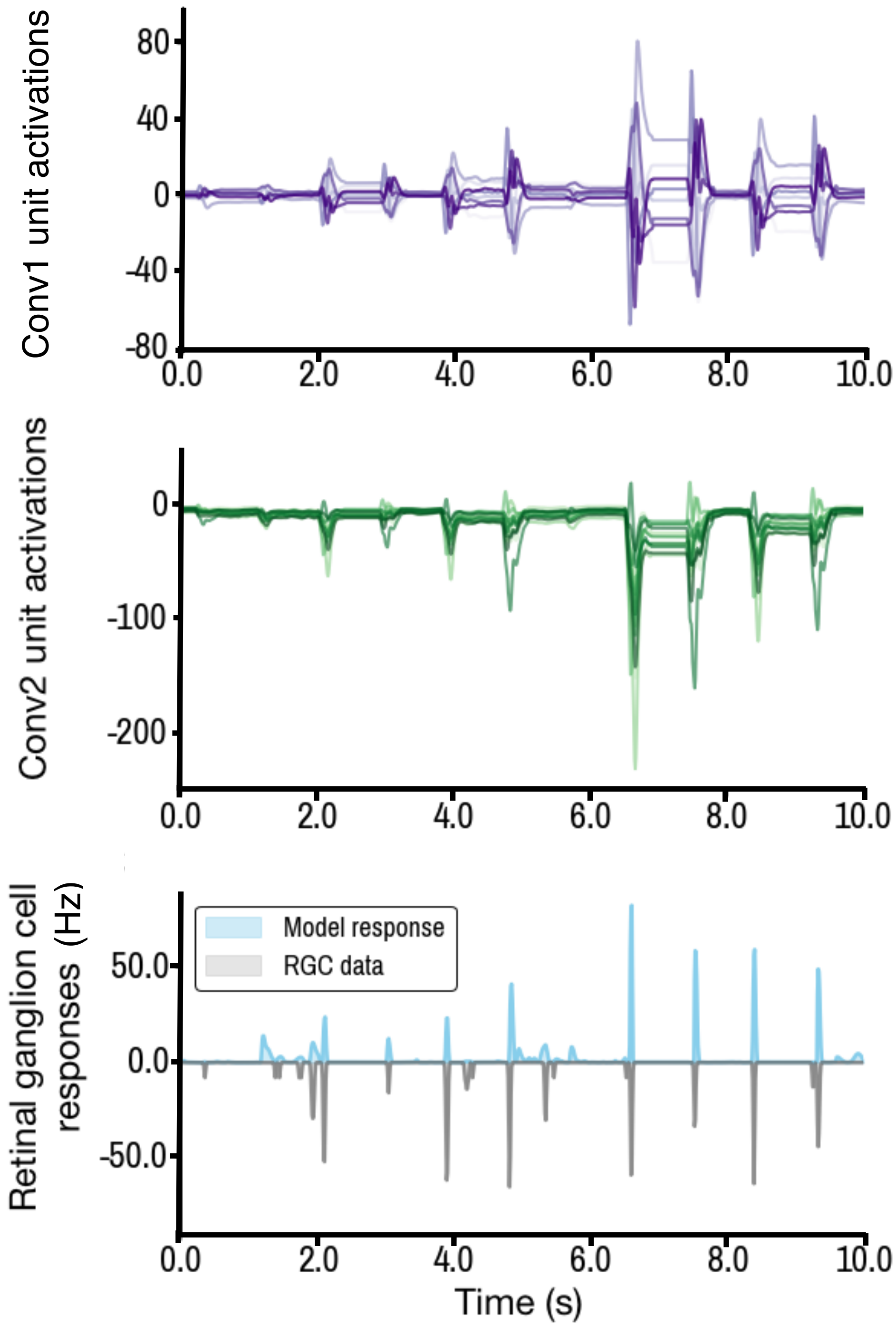


How do signals flow in a CNN?

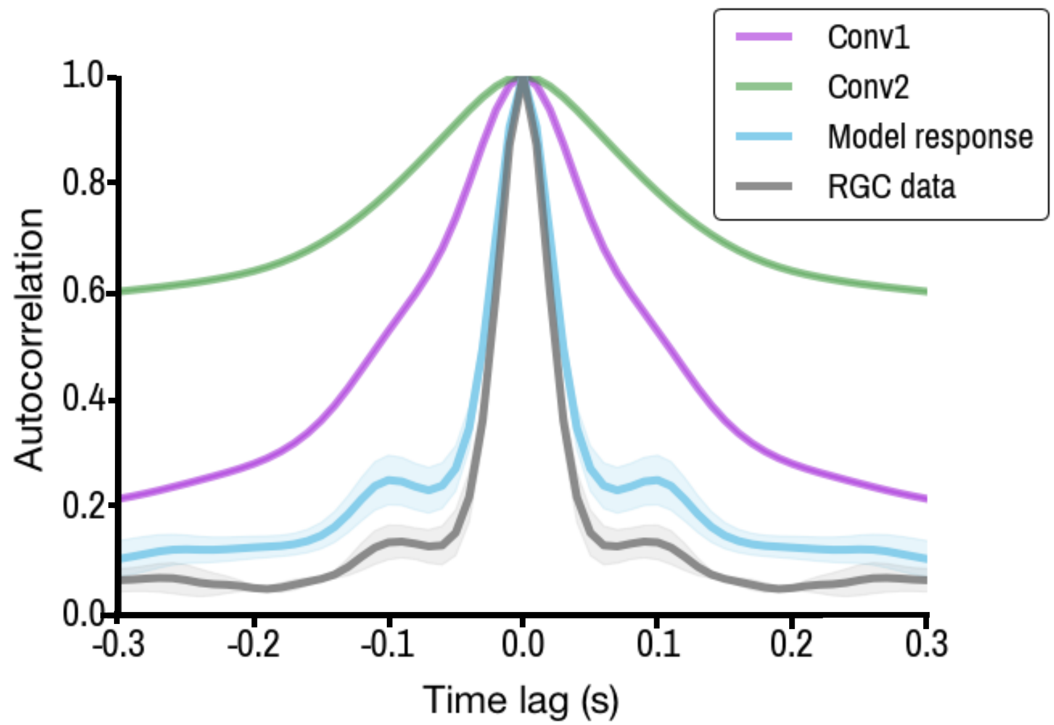


How do signals flow in a CNN?

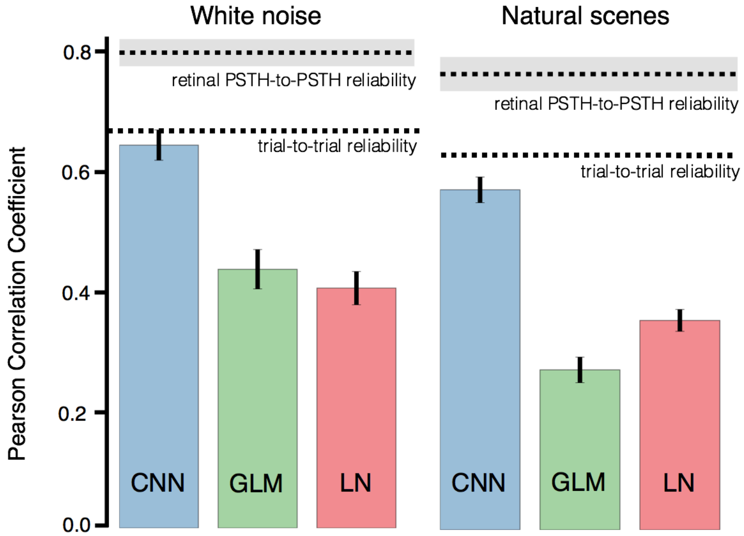




Temporal autocorrelation of signals

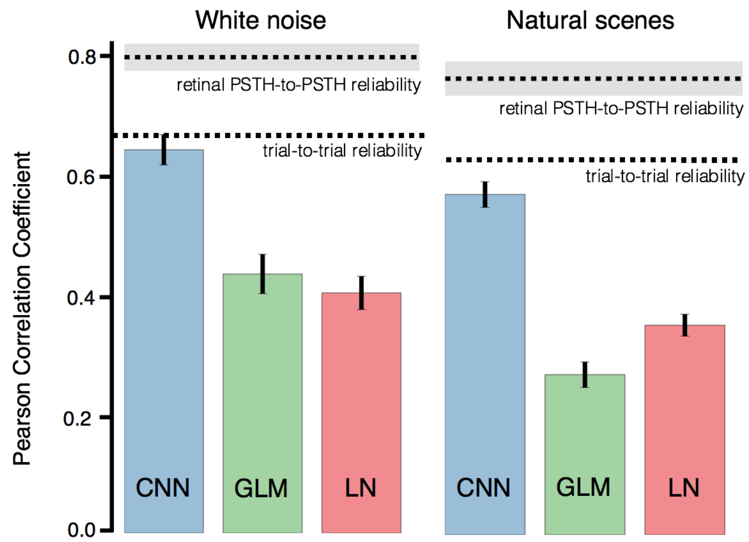


Summary



CNNs capture substantially more retinal responses than previous models.

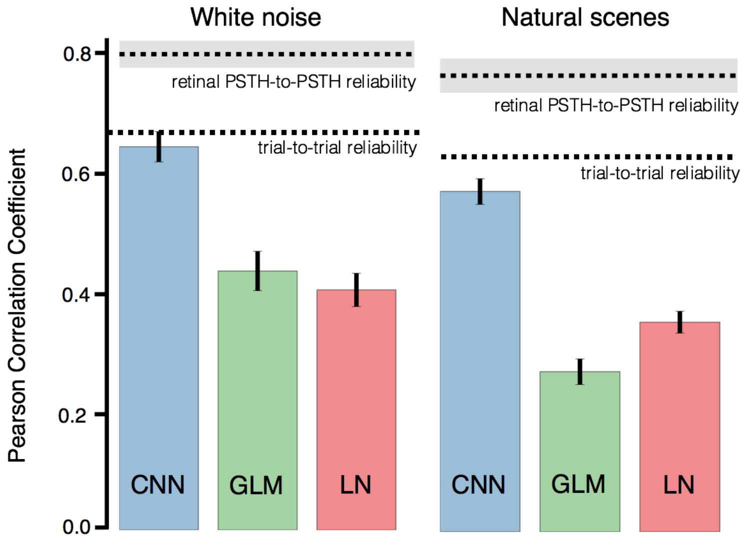
Summary



CNNs capture substantially more retinal responses than previous models.

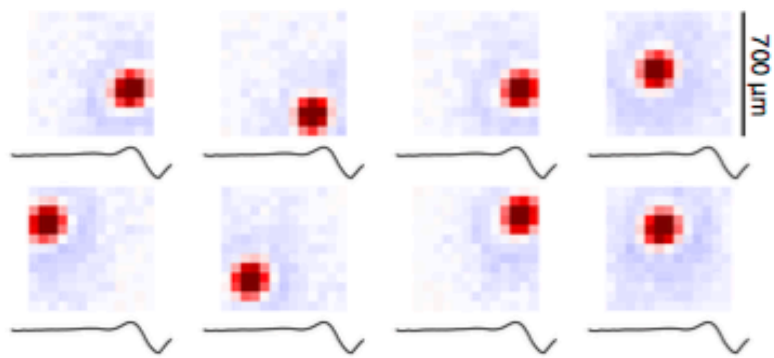
CNNs also generalize better to different stimuli classes.

Summary



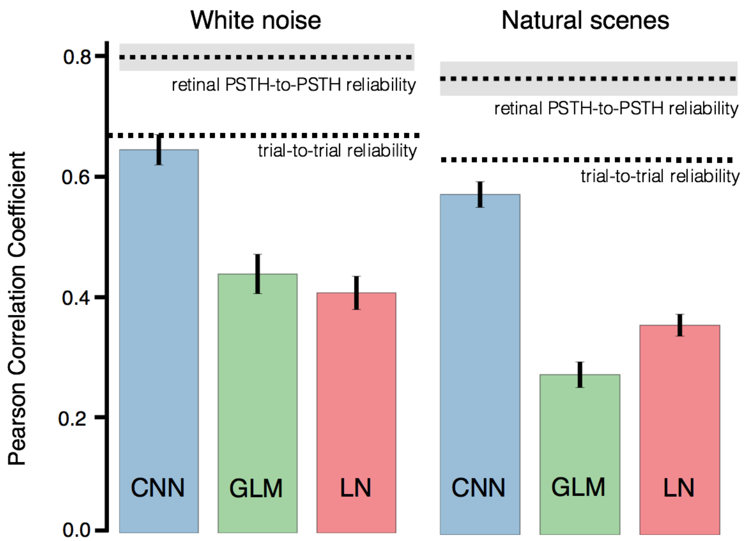
CNNs capture substantially more retinal responses than previous models.

CNNs also generalize better to different stimuli classes.



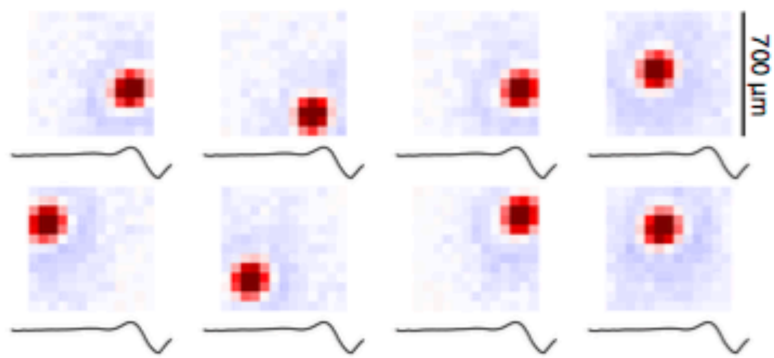
CNNs learn the internal, unobserved nonlinear structure of the retina

Summary

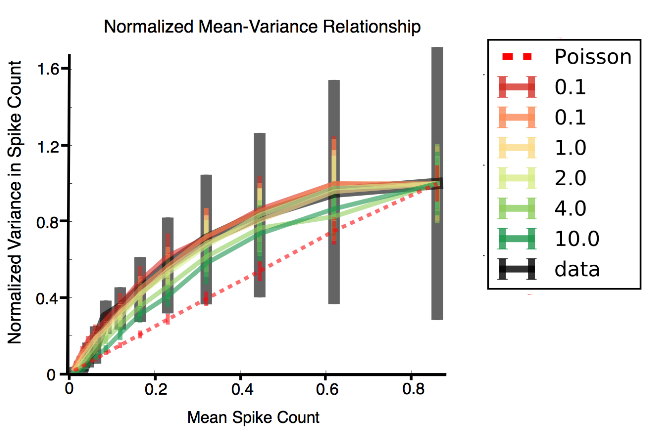


CNNs capture substantially more retinal responses than previous models.

CNNs also generalize better to different stimuli classes.

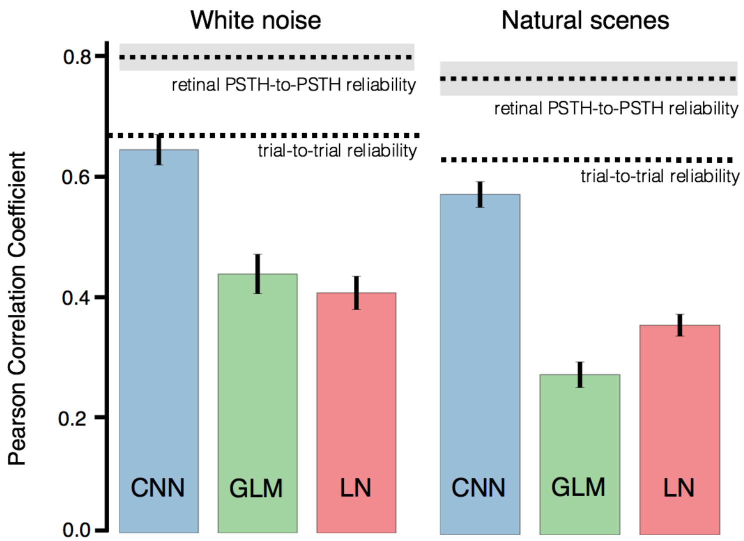


CNN learn the internal, unobserved nonlinear structure of the retina



We can capture not only the mean response of the retina well, but also its noise distribution… from single trial data

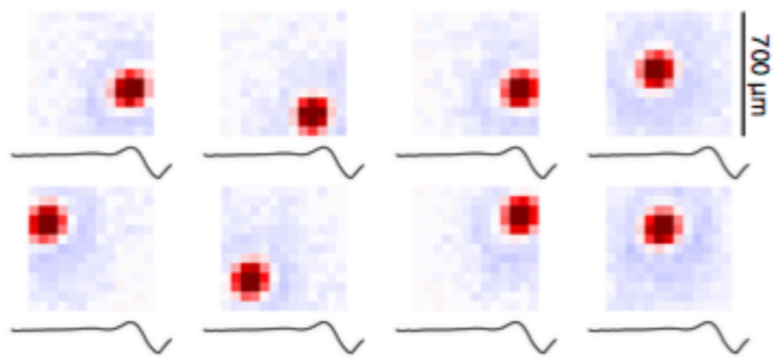
Summary

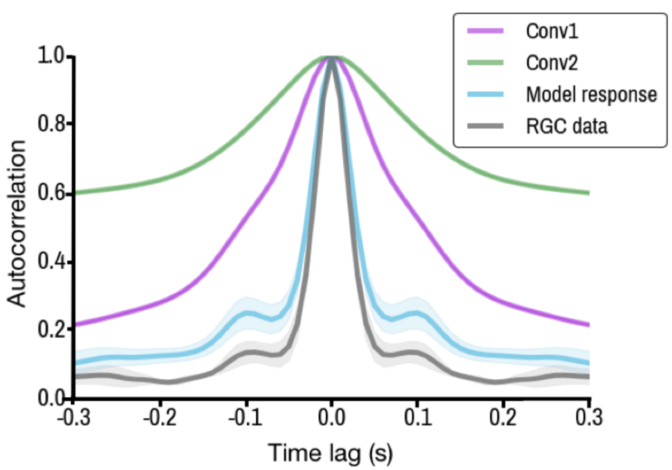


CNNs capture substantially more retinal responses than previous models.

CNNs also generalize better to different stimuli classes.

CNN learn the internal, unobserved nonlinear structure of the retina

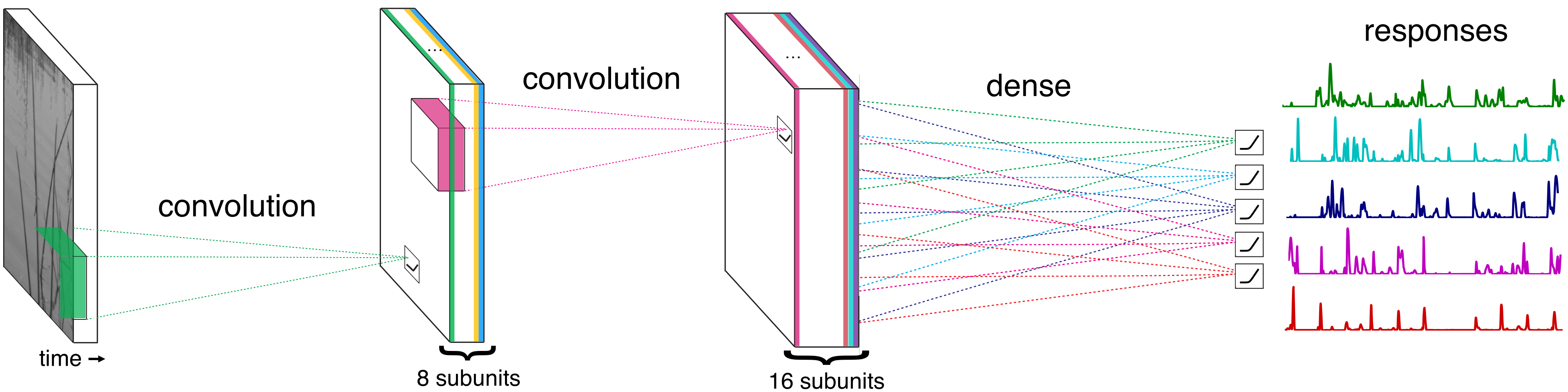


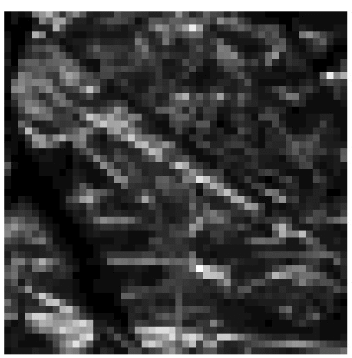


We can capture not only the mean response of the retina well, but also its noise distribution… from single trial data

Our CNN models reproduce principles of signal processing in the retina, decorrelating the visual world

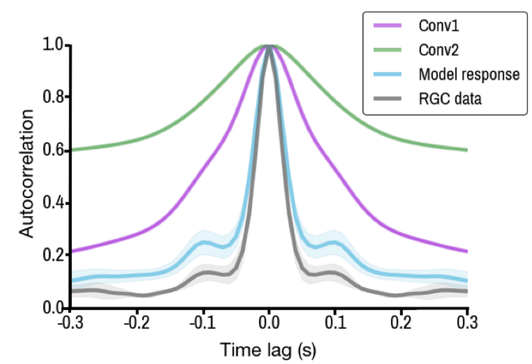
Conclusion





Systems identification in nonlinear, natural environments

Quantify neuroscience design principles for use in computer vision



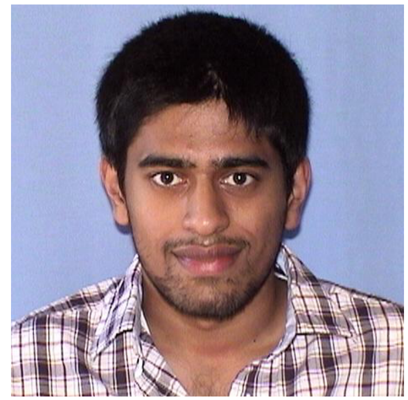
Acknowledgements

Baccus Lab
(Neurobiology)

Stephen Baccus
Mike Menz
Ju Young Kim
Pablo Jadzinsky
Benjamin Naecker
Bongsoo Suh
Dongsoo Lee
Aran Nayebi

Ganguli Lab
(Applied Physics)

Surya Ganguli
Subhaneil Lahiri
Friedemann Zenke
Niru Maheswaranathan
Madhu Advani
Peiran Gao
Kiah Hardcastle
Ben Poole
Alex Williams



Niru Maheswaranathan

Aran Nayebi

Contact

lmcintosh@stanford.edu
lanemcintosh.com

Funding

NSF IGERT Mind, Brain, and Computation Graduate Fellowship
NIH Ruth L. Kirschstein National Research Award
NVIDIA Titan X GPU hardware award

STANFORD
NEUROSCIENCE