Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Umut Güçlü and Marcel A. J. van Gerven

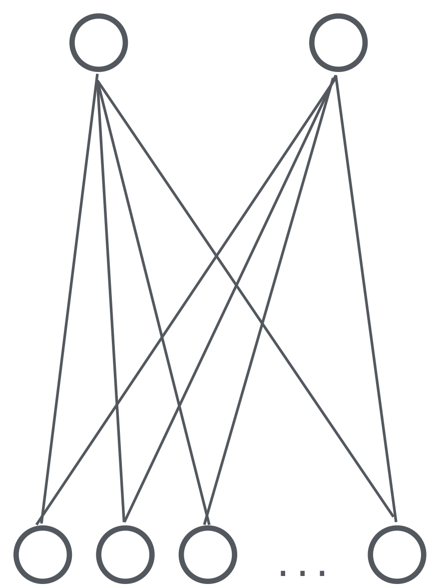
Article overview by
Ilya Kuzovkin

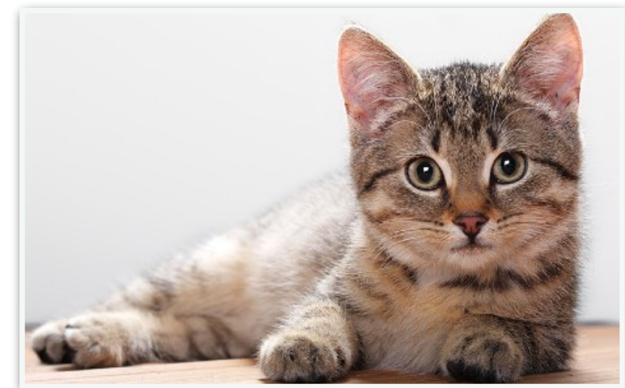
Computational Neuroscience Seminar
University of Tartu
2015

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

“spider”





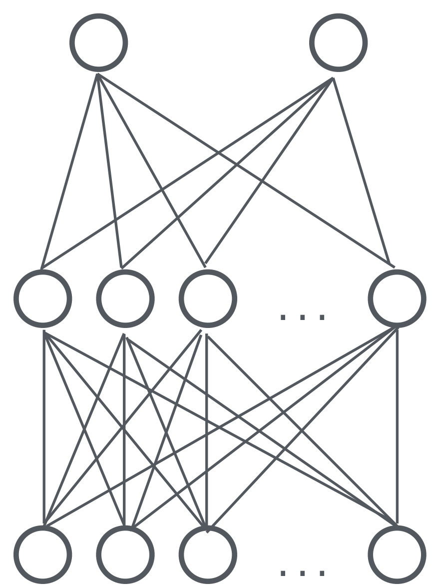
“cat”

Linear

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

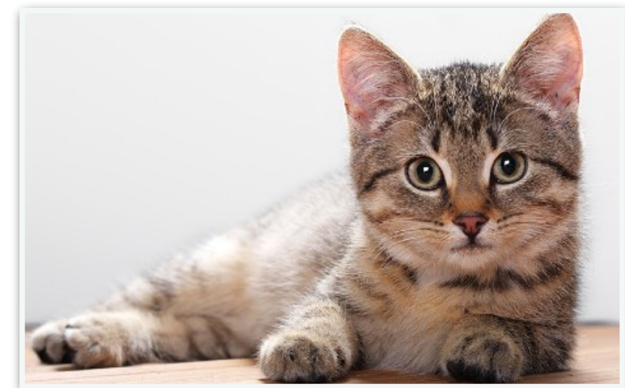
“spider”

classes



hidden layer

pixels



“cat”

Non-linear

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

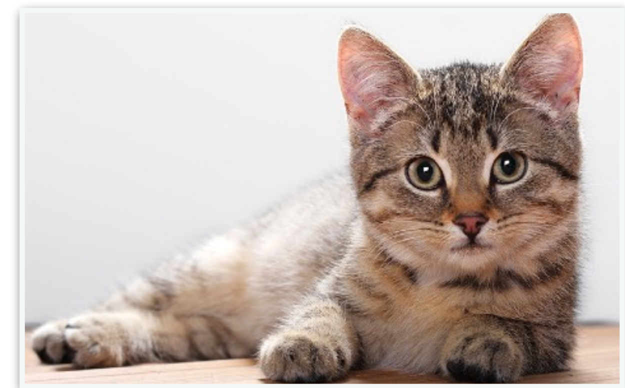
"spider"

classes

hidden layer

hidden layer

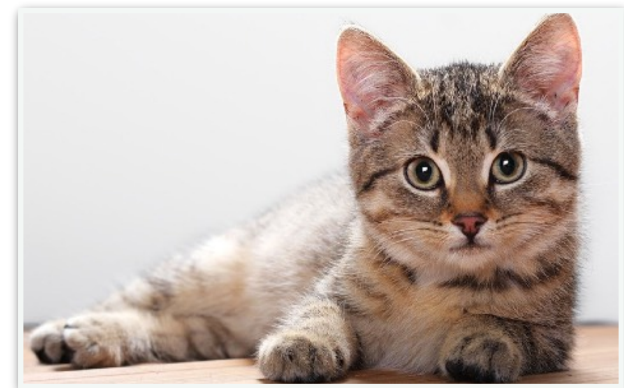
pixels



"cat"

Deep

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

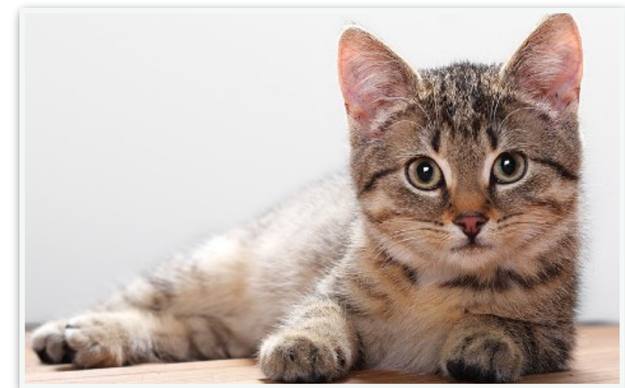


“cat”

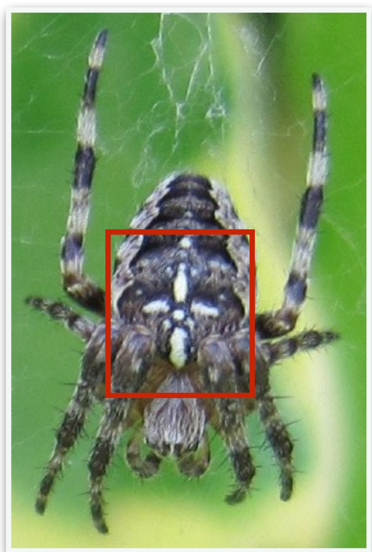
“spider”

important feature

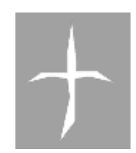
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



"cat"



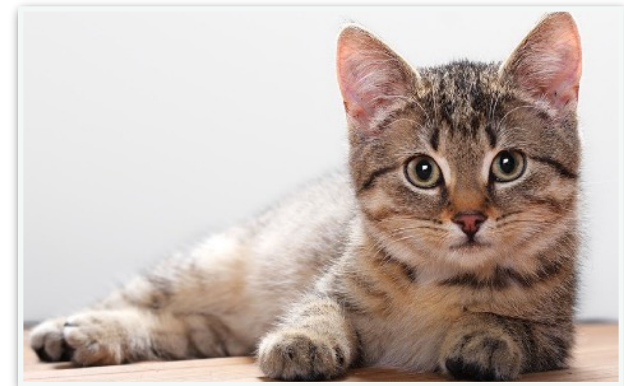
"spider"



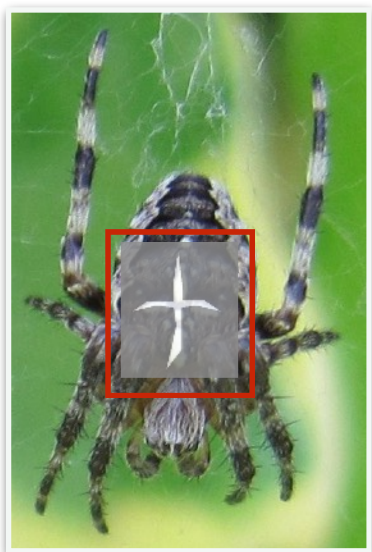
important feature

RUN!

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



"cat"

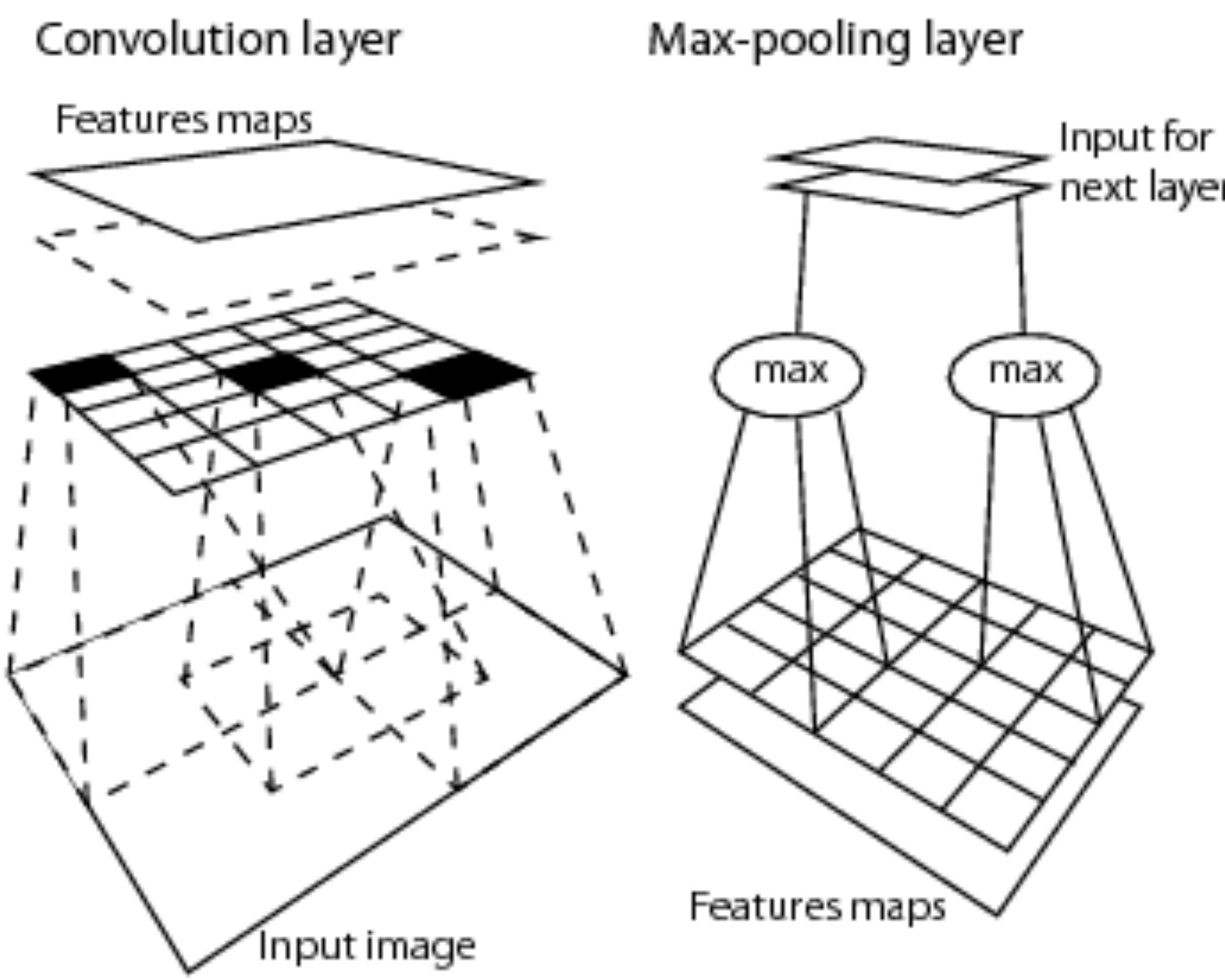


"spider"

important feature

RUN! Convolutional filter

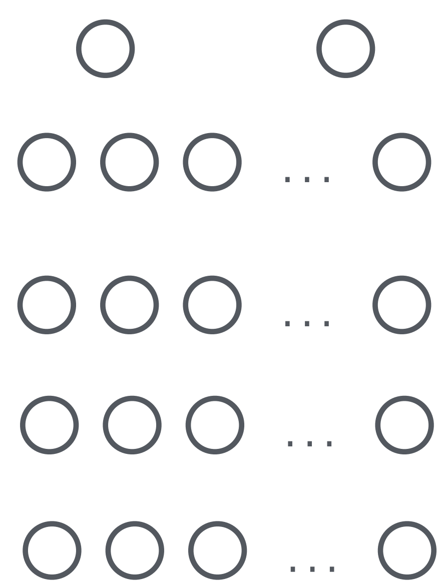
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

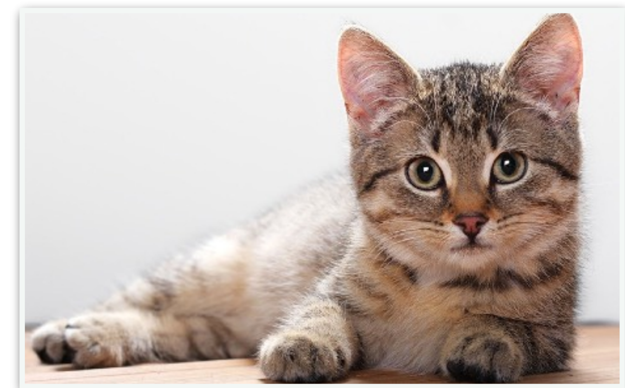


Convolutional (and pooling) layer

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

"spider"

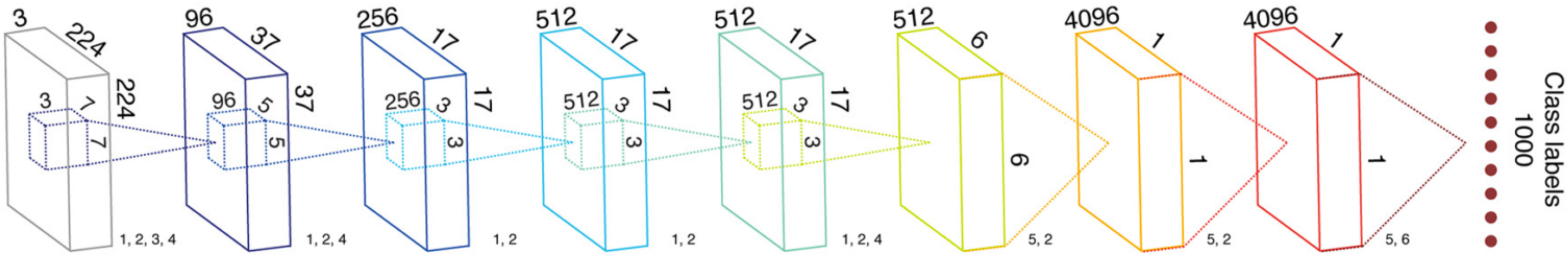




"cat"

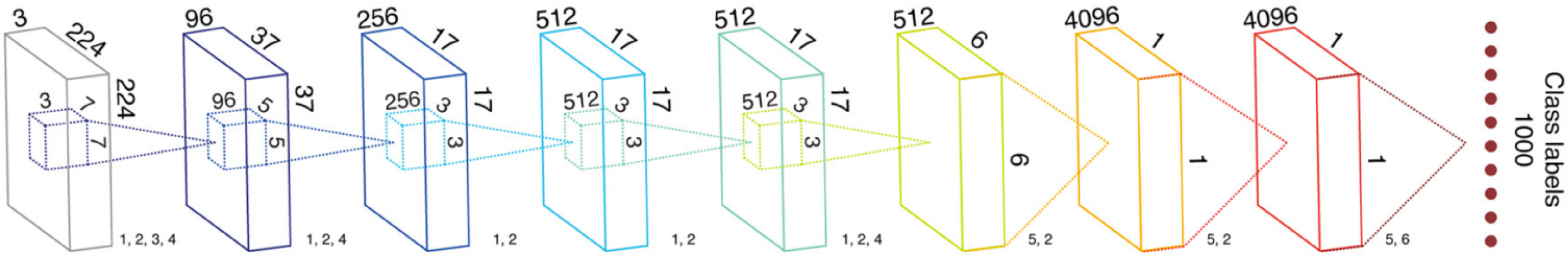
Deep Convolutional Neural Network

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



[1] convolution, [2] rectification,
[3] local response normalization,
[4] max pooling, [5] inner product, [6] softmax

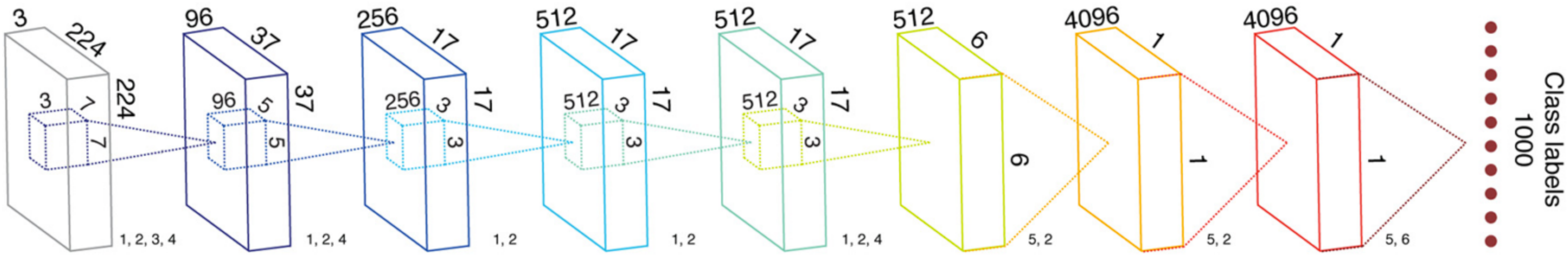
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



[1] convolution, [2] rectification,
[3] local response normalization,
[4] max pooling, [5] inner product, [6] softmax

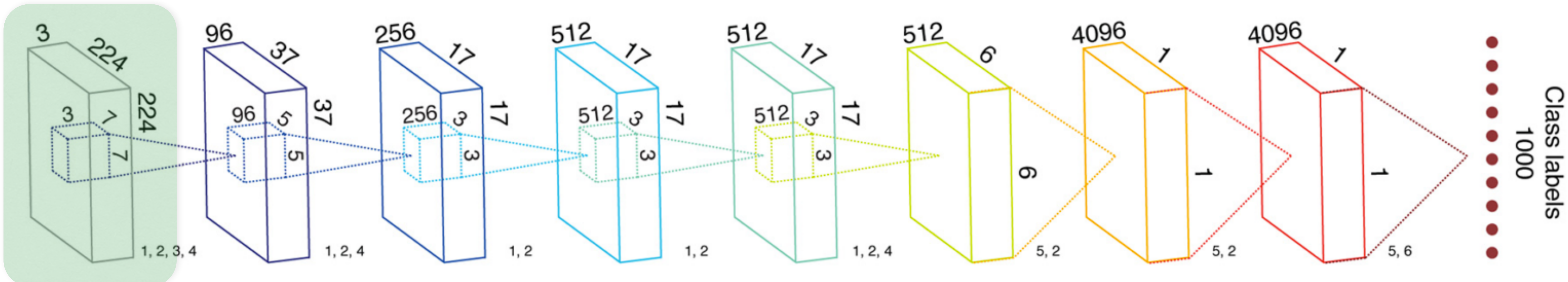
IMAGENET

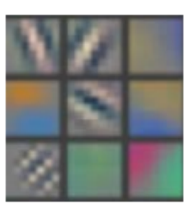
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



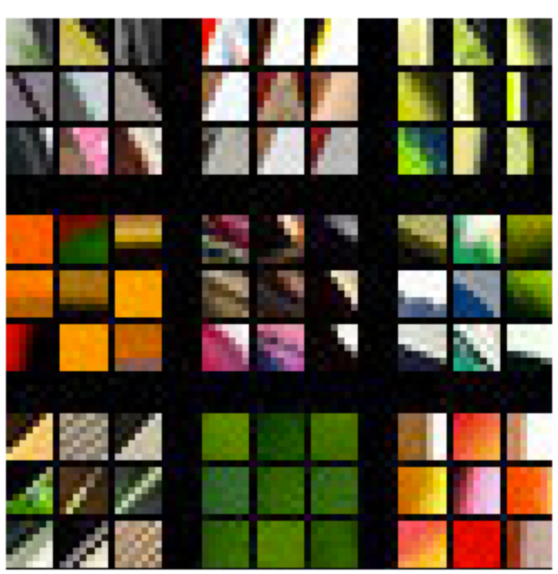
^{1} convolution, ^{2} rectification,
^{3} local response normalization,
^{4} max pooling, ^{5} inner product, ^{6} softmax

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream





Layer 1

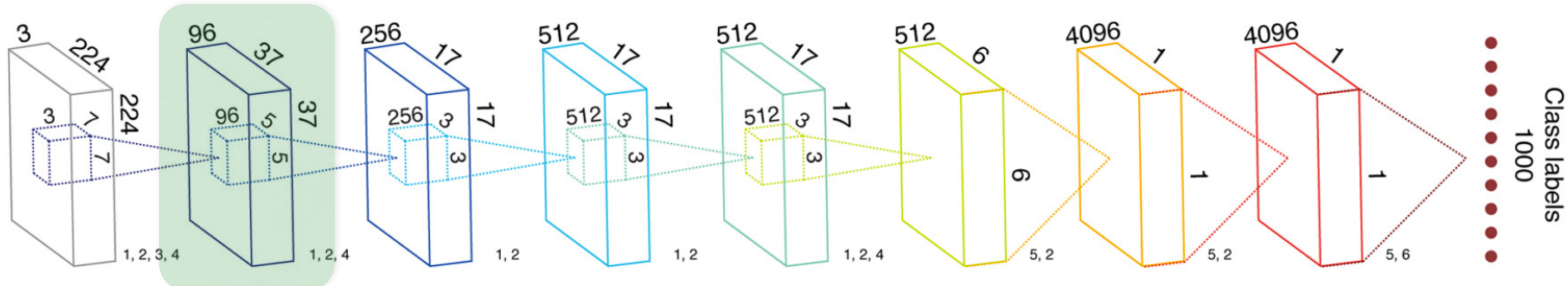


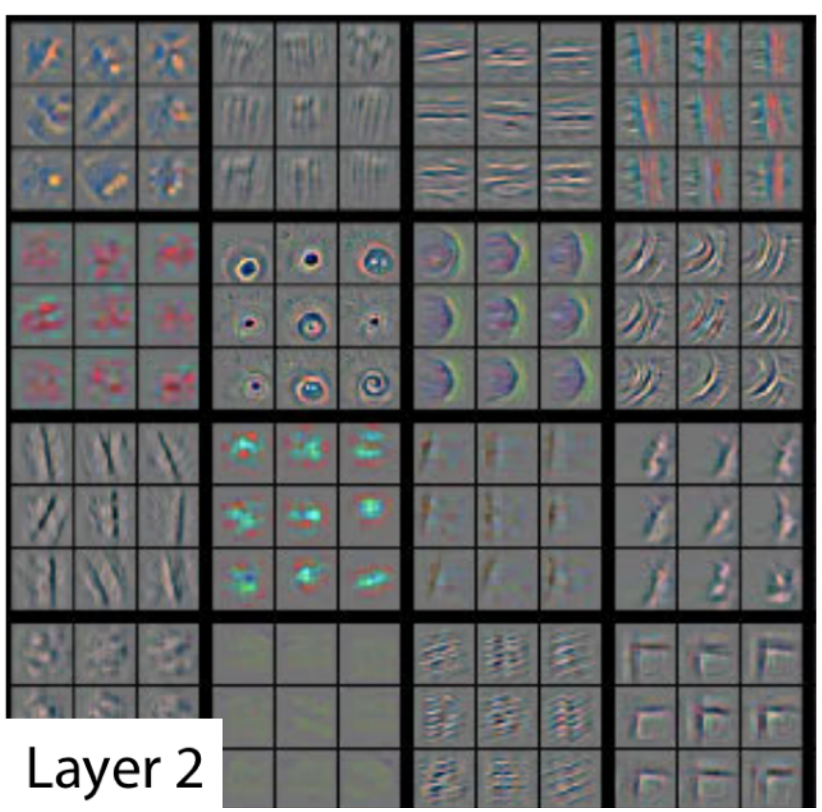
¹ convolution, ² rectification,
³ local response normalization,
⁴ max pooling, ⁵ inner product, ⁶ softmax

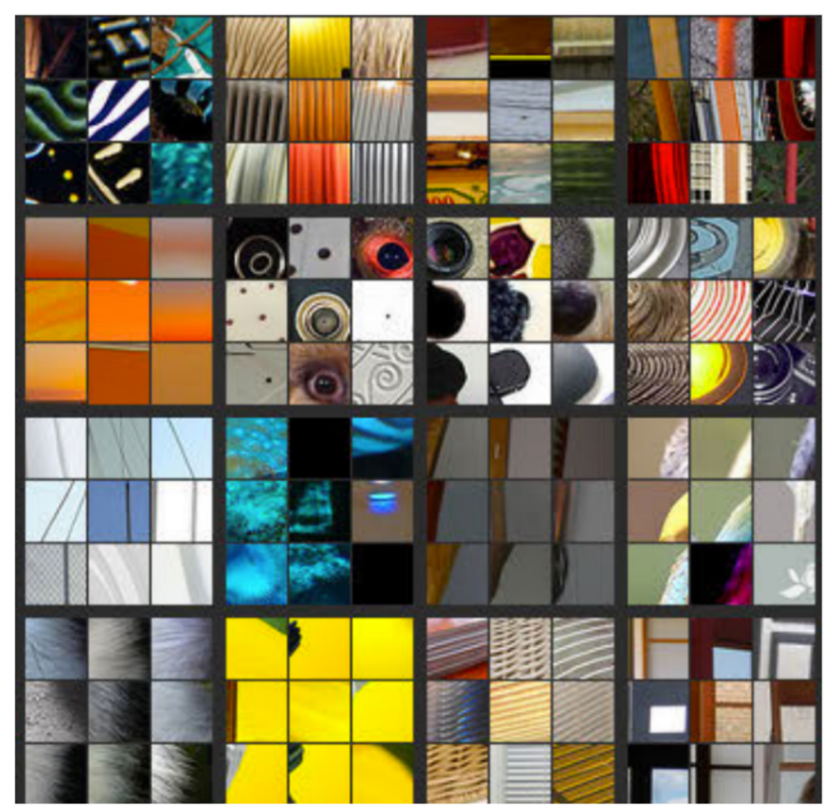
IMAGENET

Matthew D. Zeiler, Rob Fergus **Visualizing and Understanding Convolutional Networks** 2013

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream





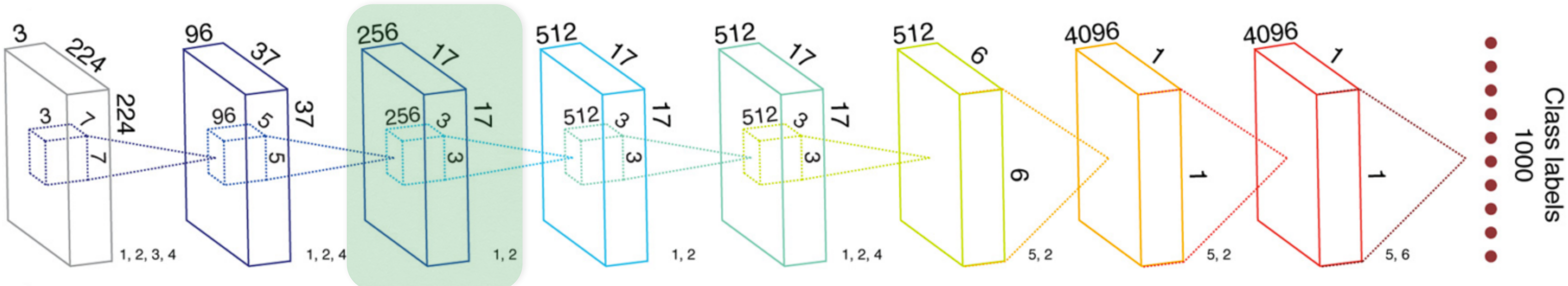


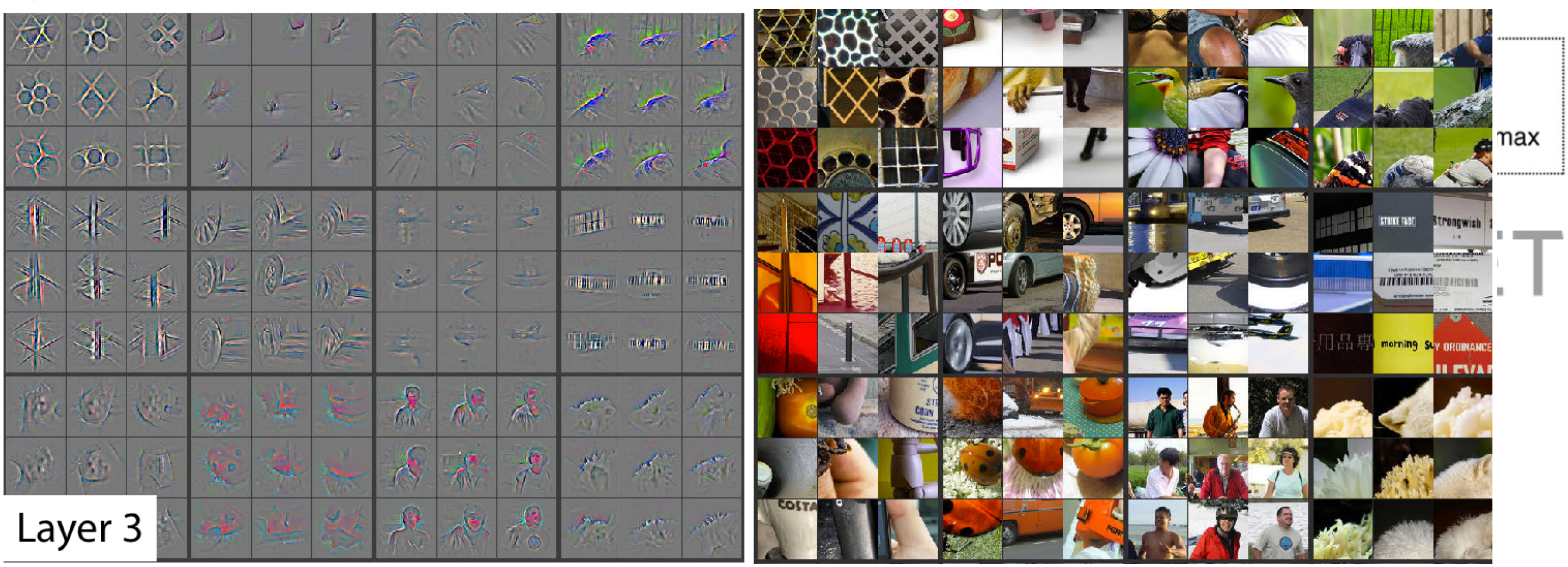
[1] convolution, [2] rectification,
[3] local response normalization,
[4] max pooling, [5] inner product, [6] softmax

IMAGENET

Matthew D. Zeiler, Rob Fergus **Visualizing and Understanding Convolutional Networks** 2013

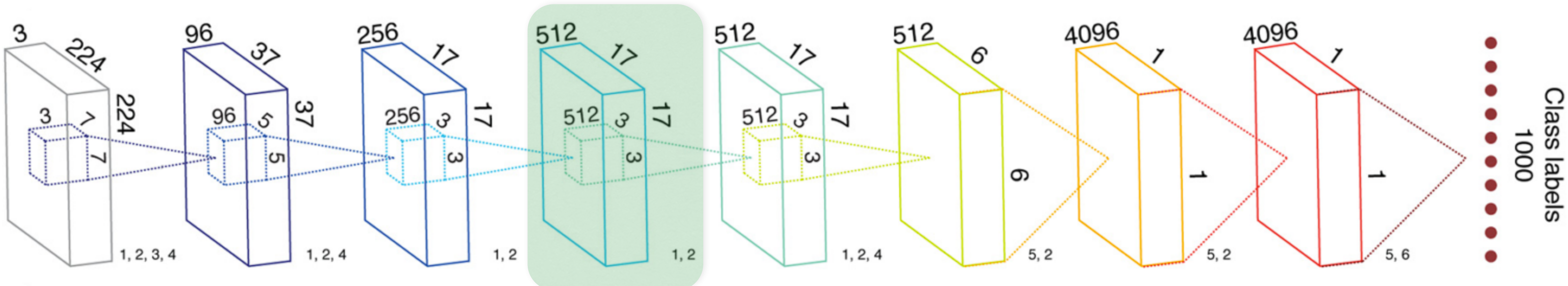
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

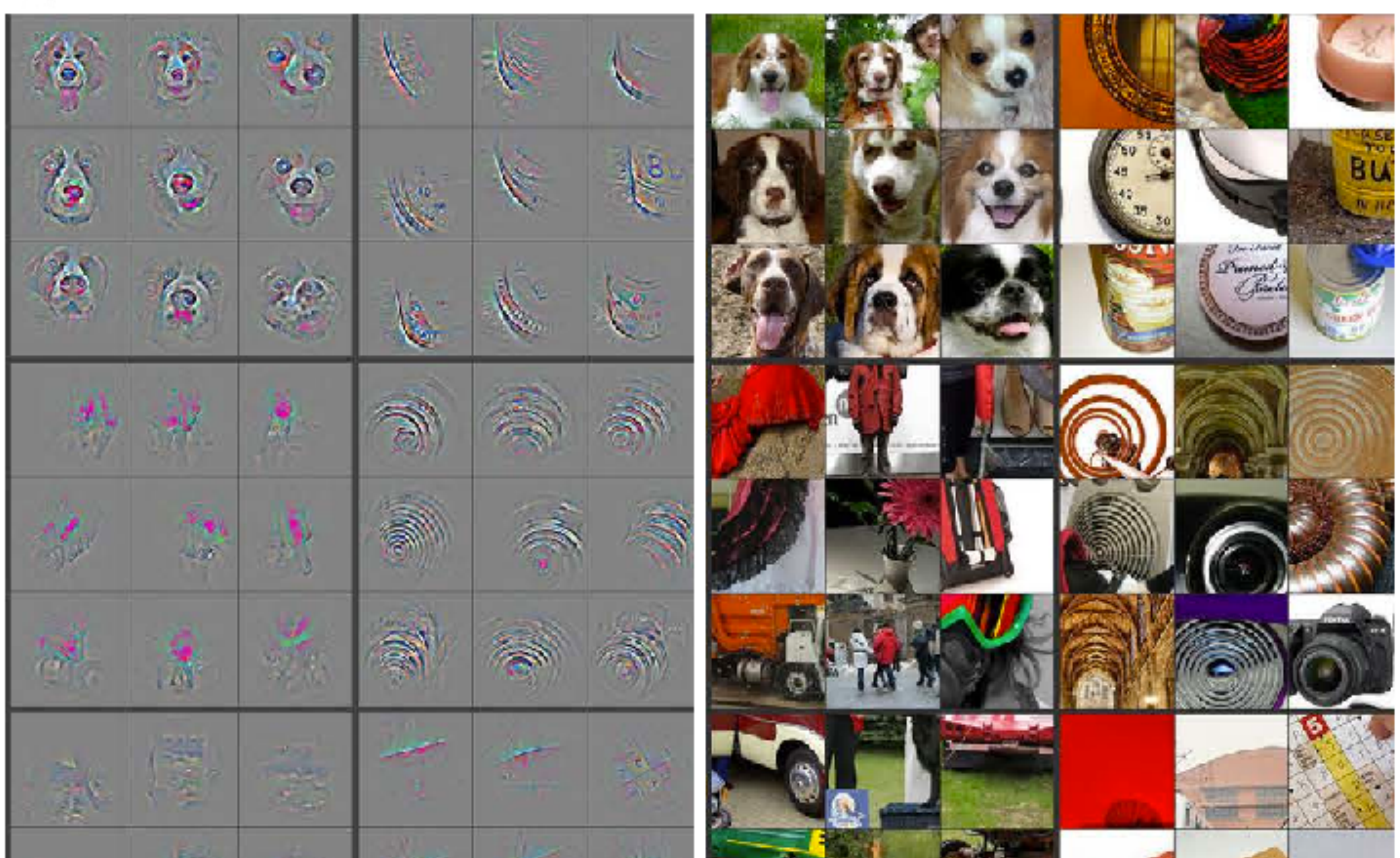




Matthew D. Zeiler, Rob Fergus **Visualizing and Understanding Convolutional Networks** 2013

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



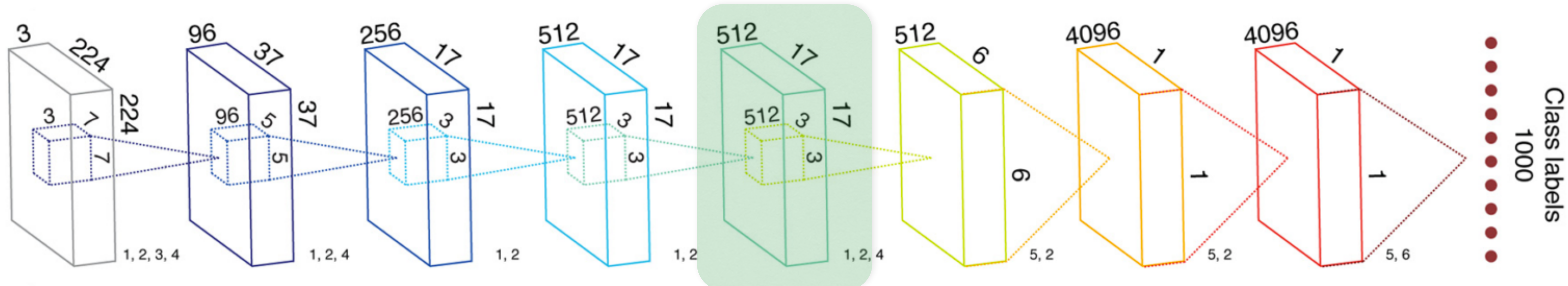


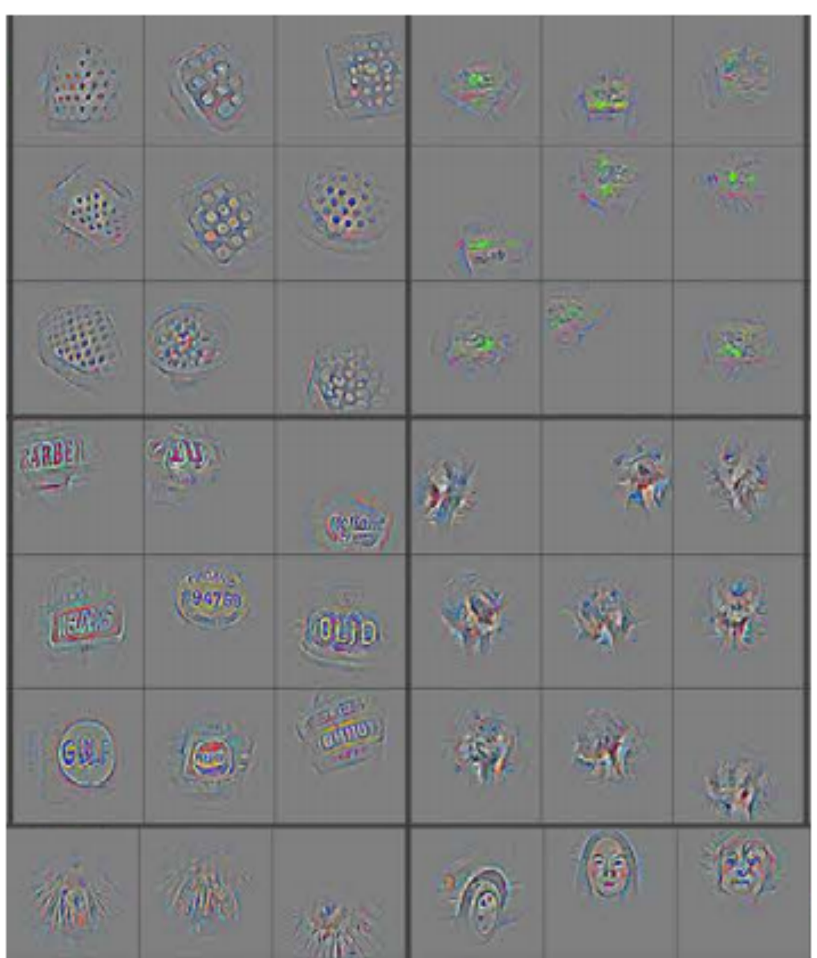
[1] convolution, [2] rectification,
[3] local response normalization,
[4] max pooling, [5] inner product, [6] softmax

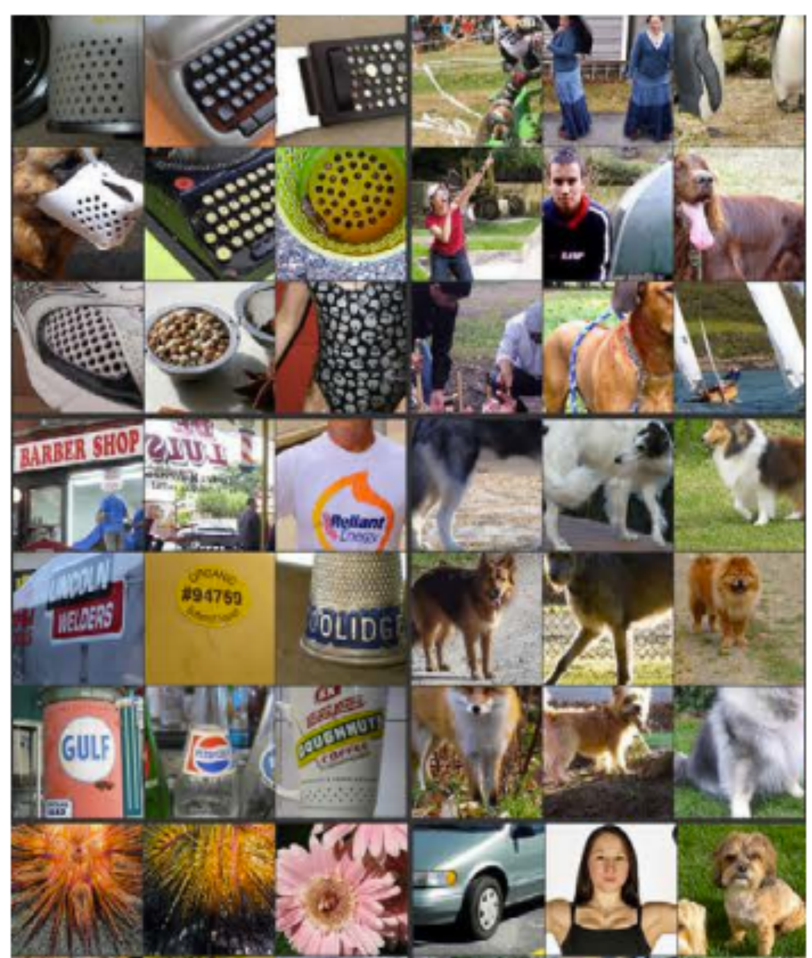
IMAGENET

izing and Understanding Convolutional Networks 2013

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream







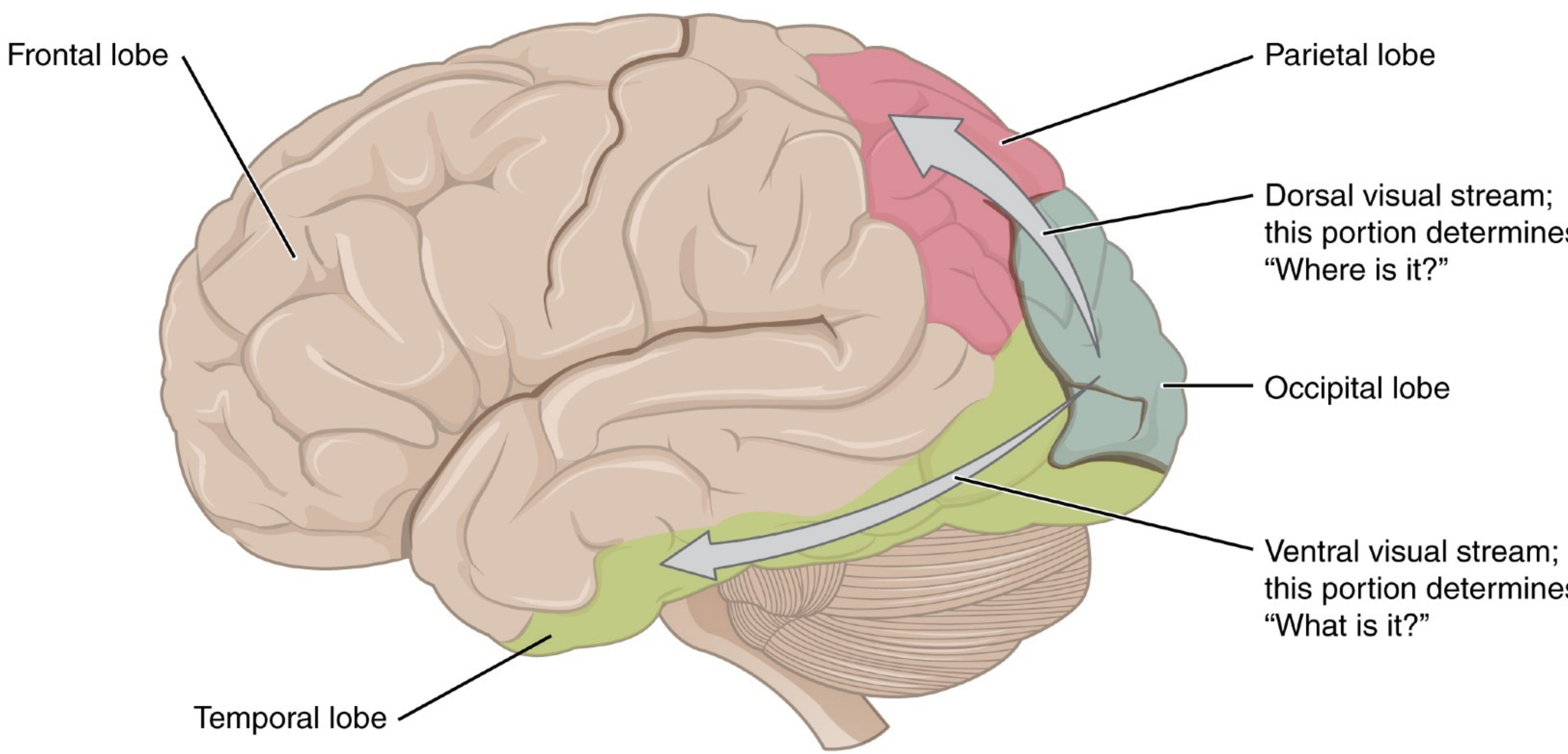
[1] convolution, [2] rectification,
[3] local response normalization,
[4] max pooling, [5] inner product, [6] softmax

IMAGENET

zing and Understanding Convolutional Networks 2013

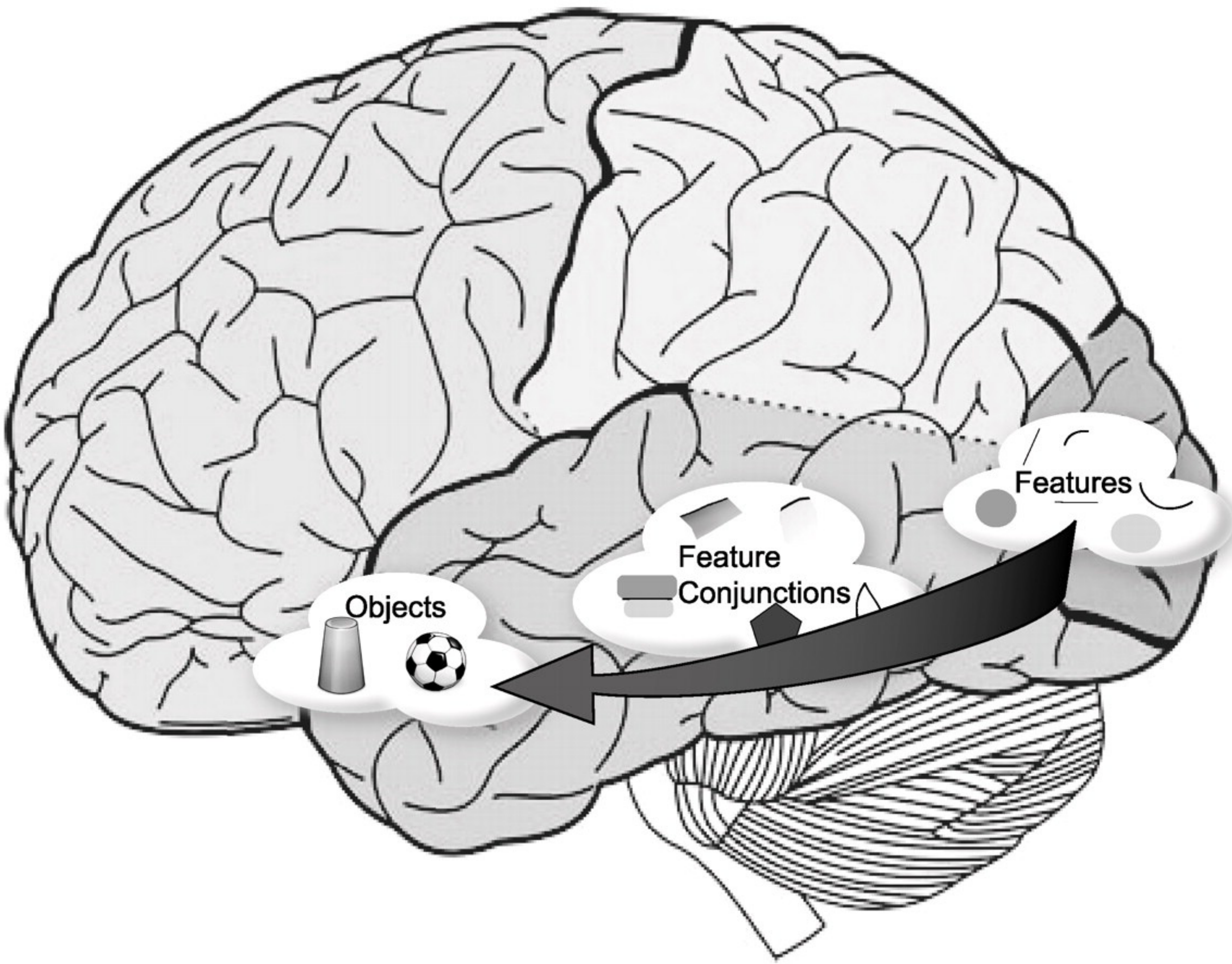
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

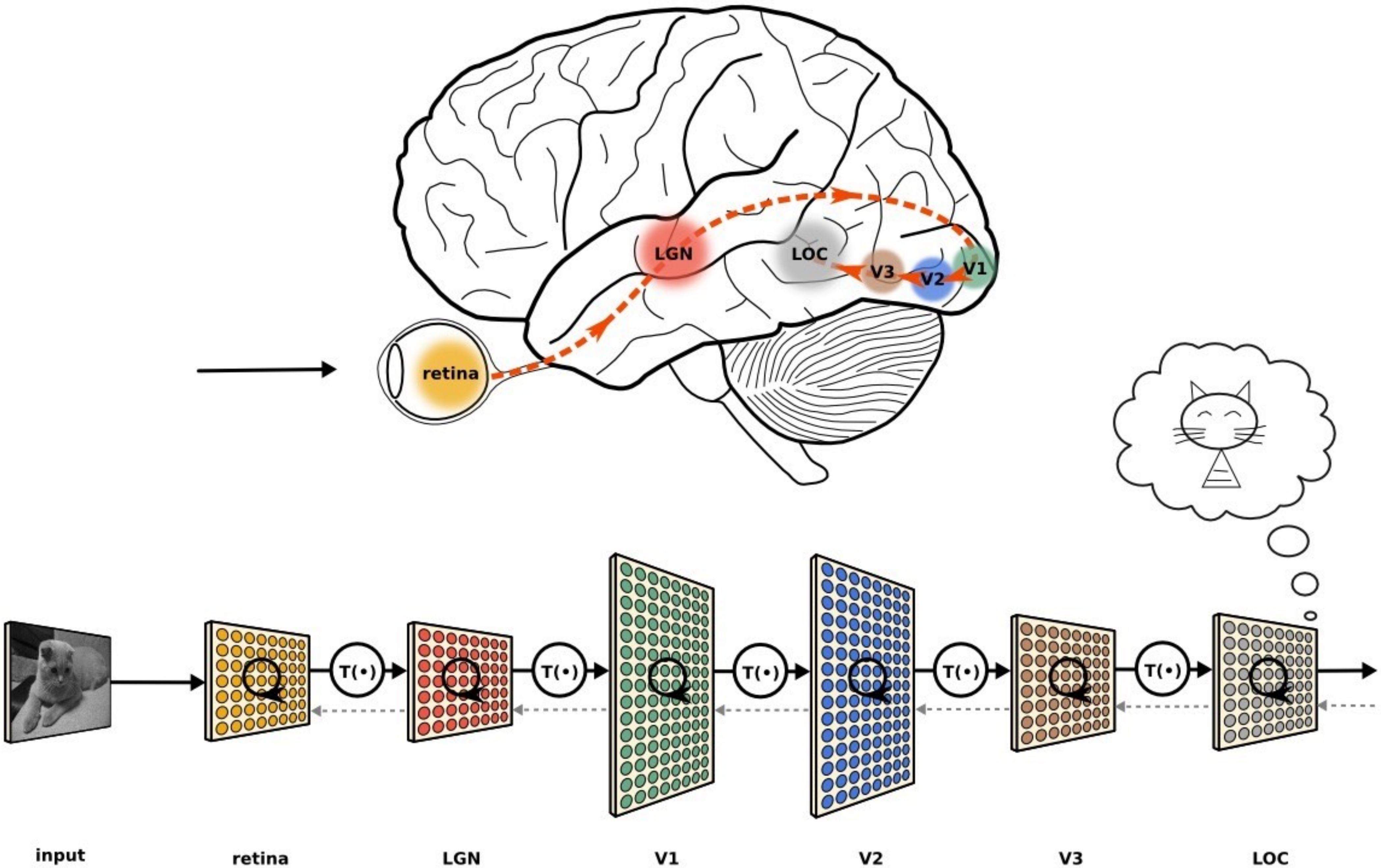


Two-stream hypothesis

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

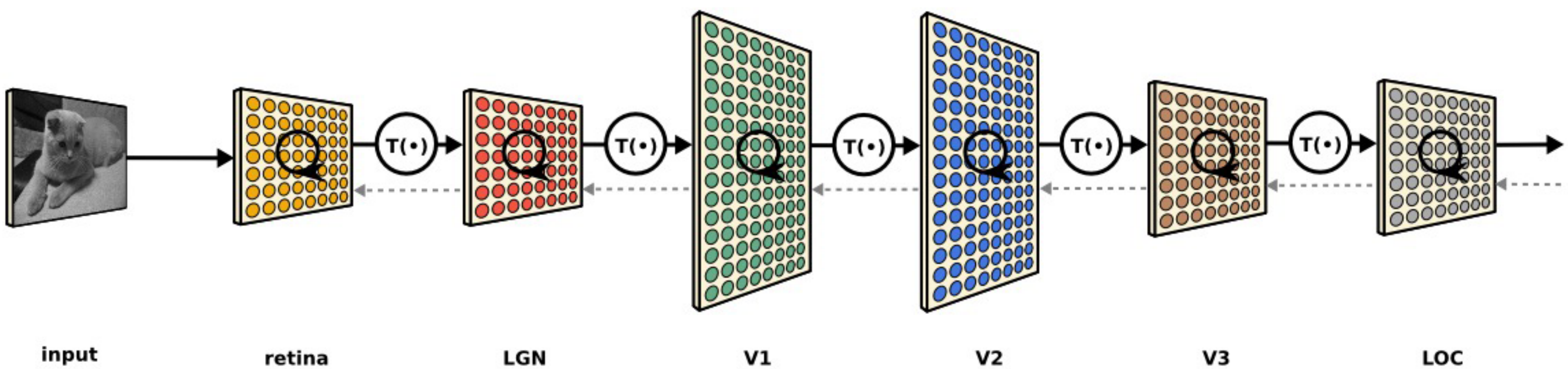


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

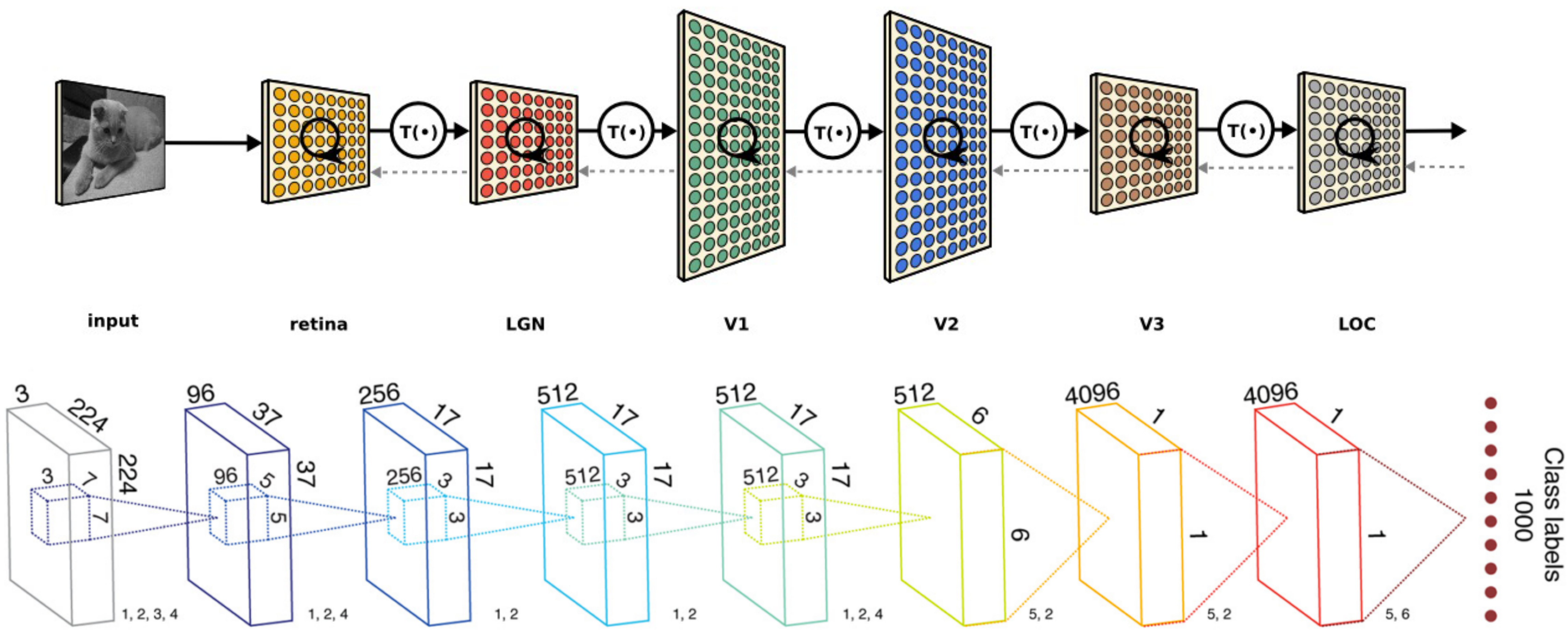


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

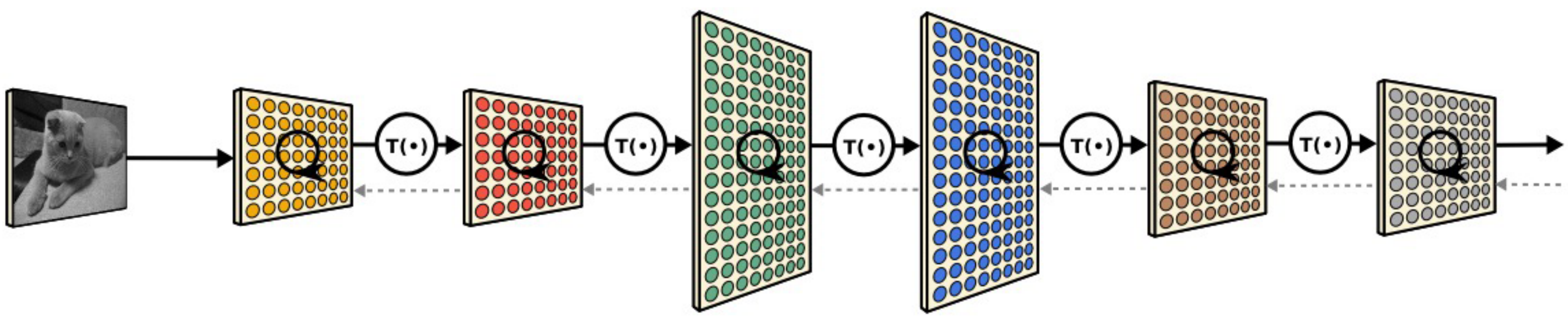
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

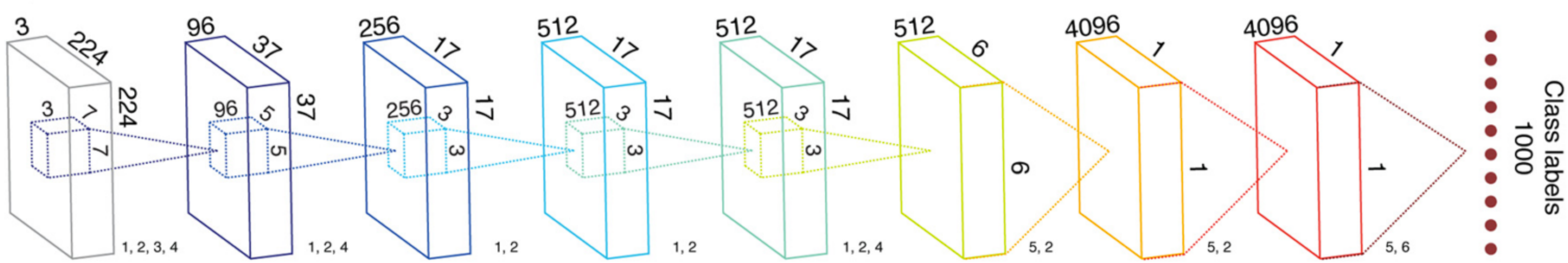


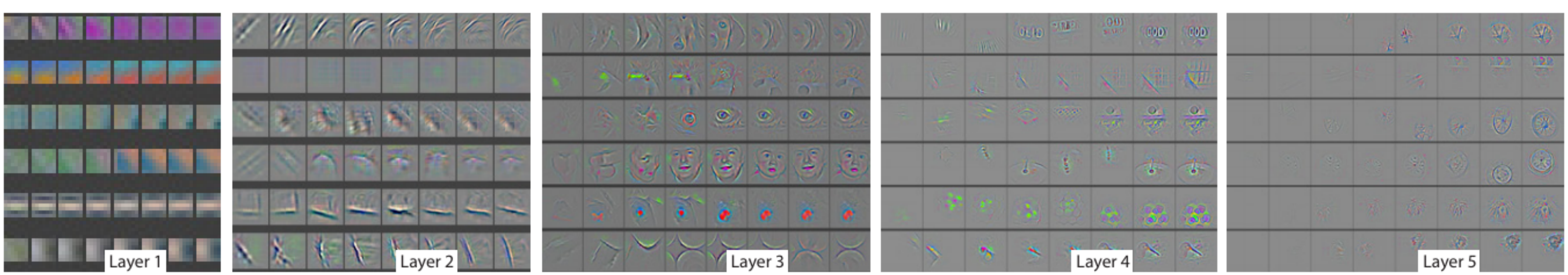
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



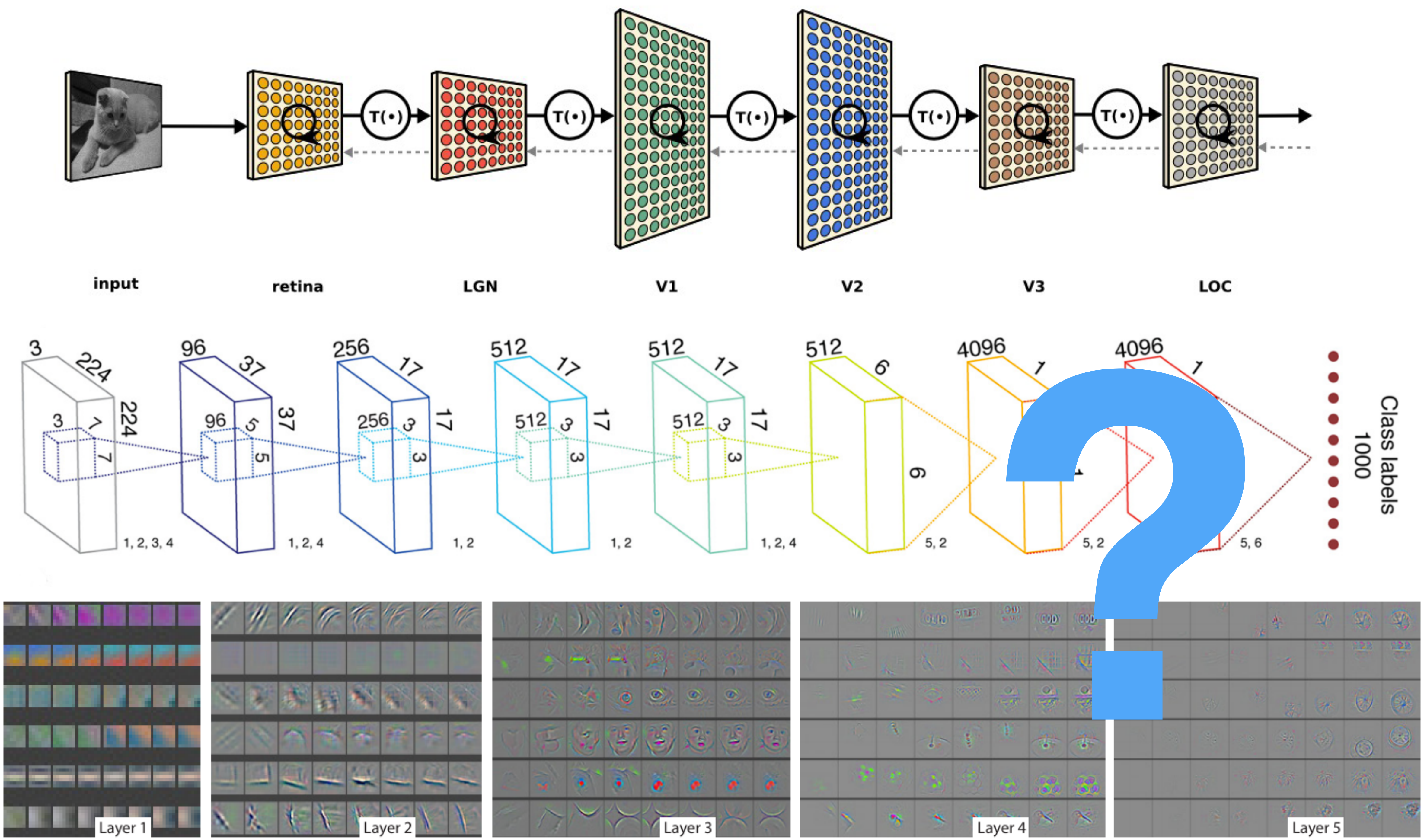
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream





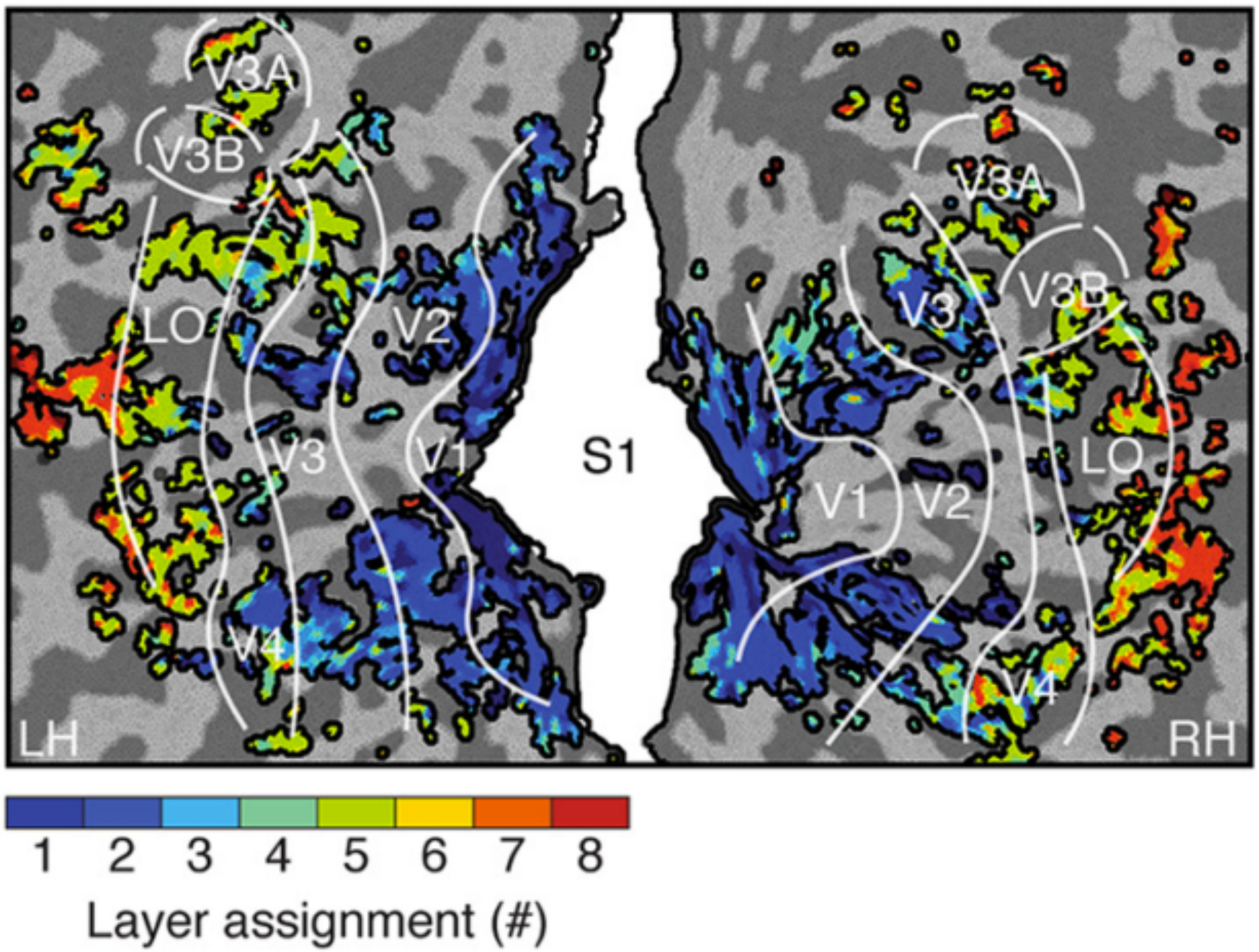


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

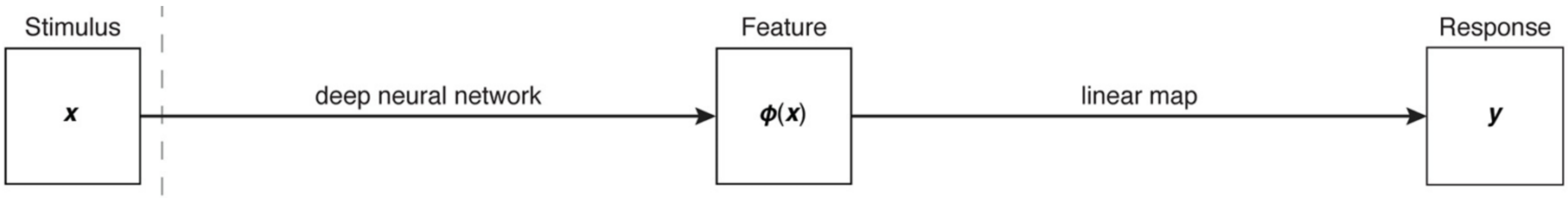


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

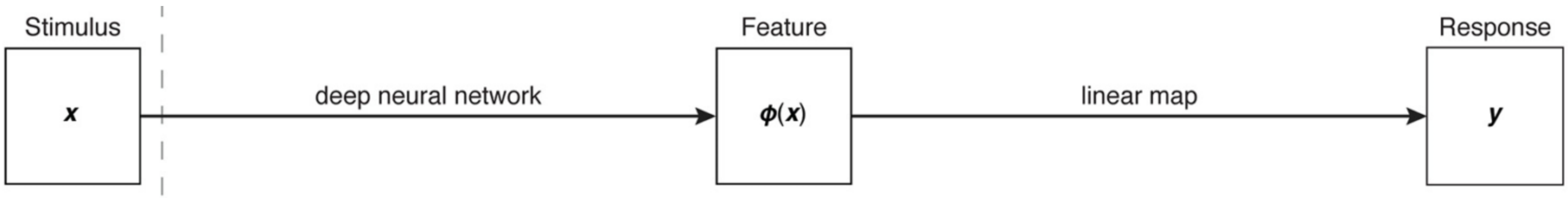


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



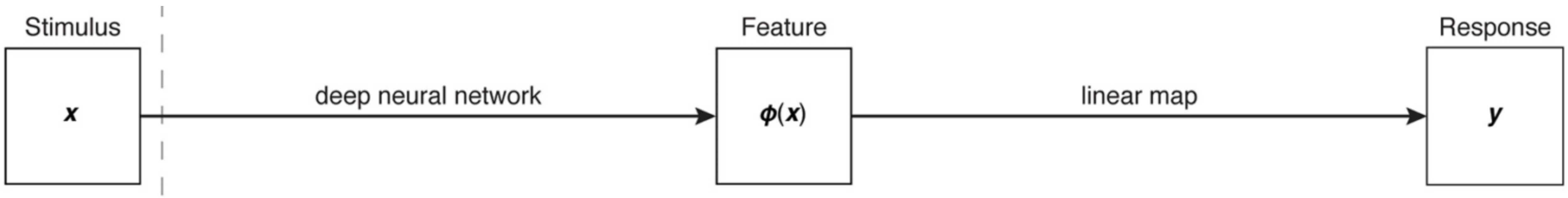
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations

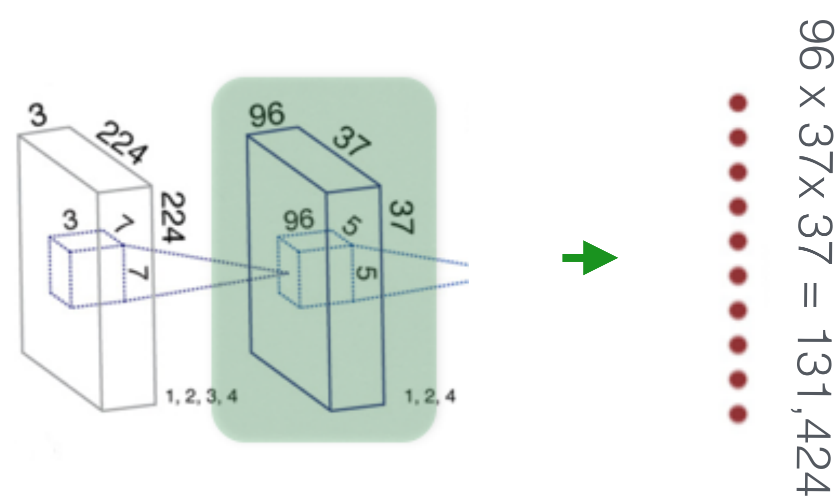


DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus ($\boldsymbol{x}$) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response ($\boldsymbol{y}$).

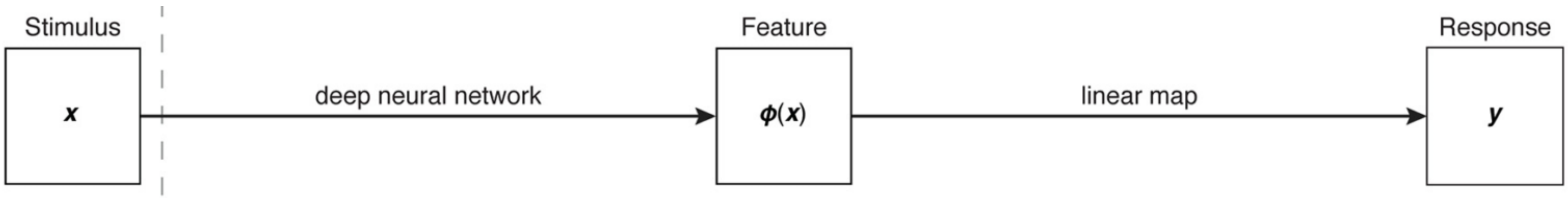
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



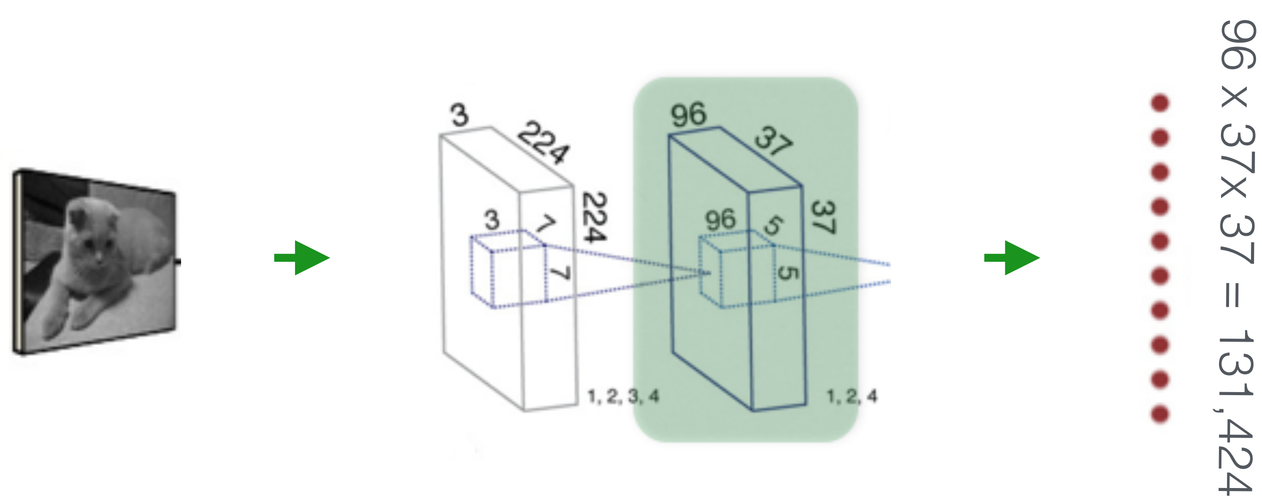
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



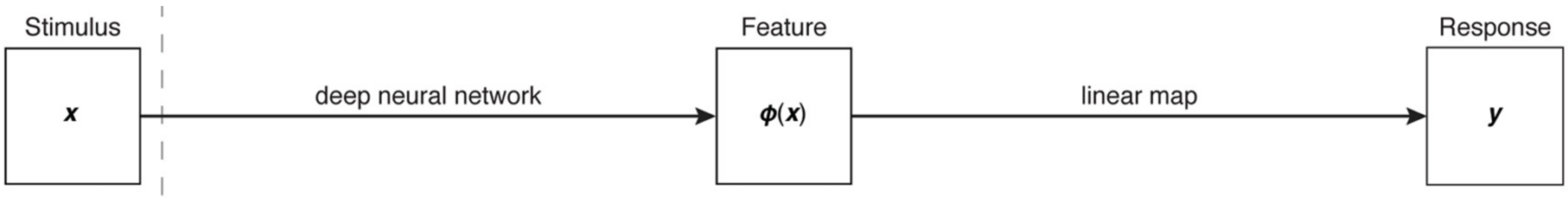
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



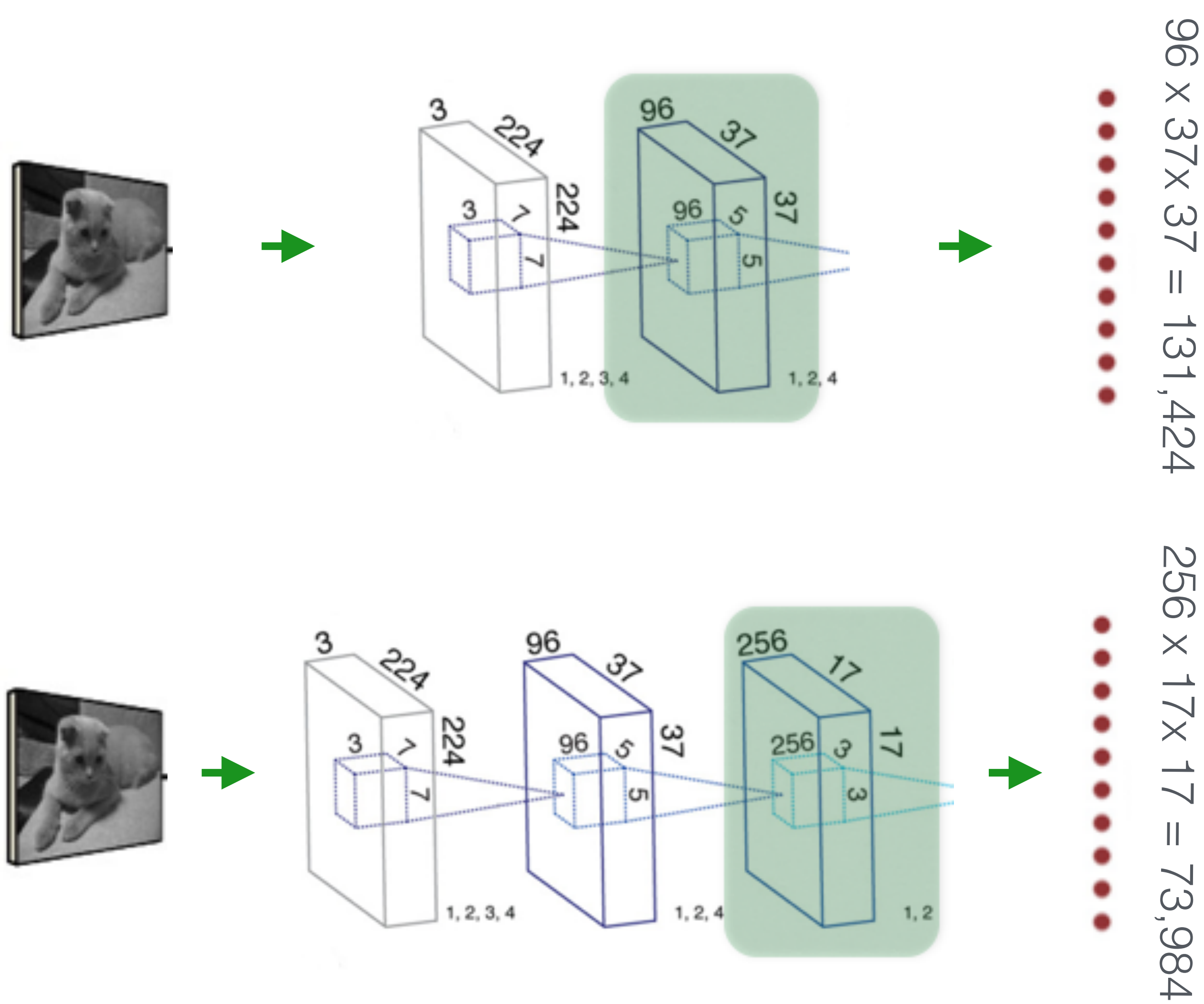
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



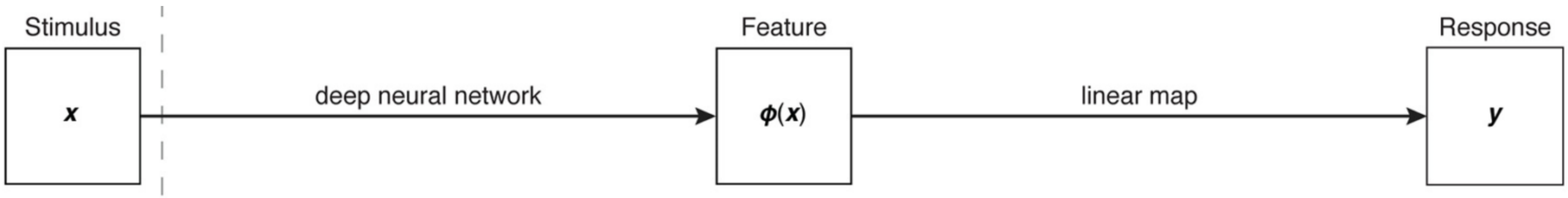
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



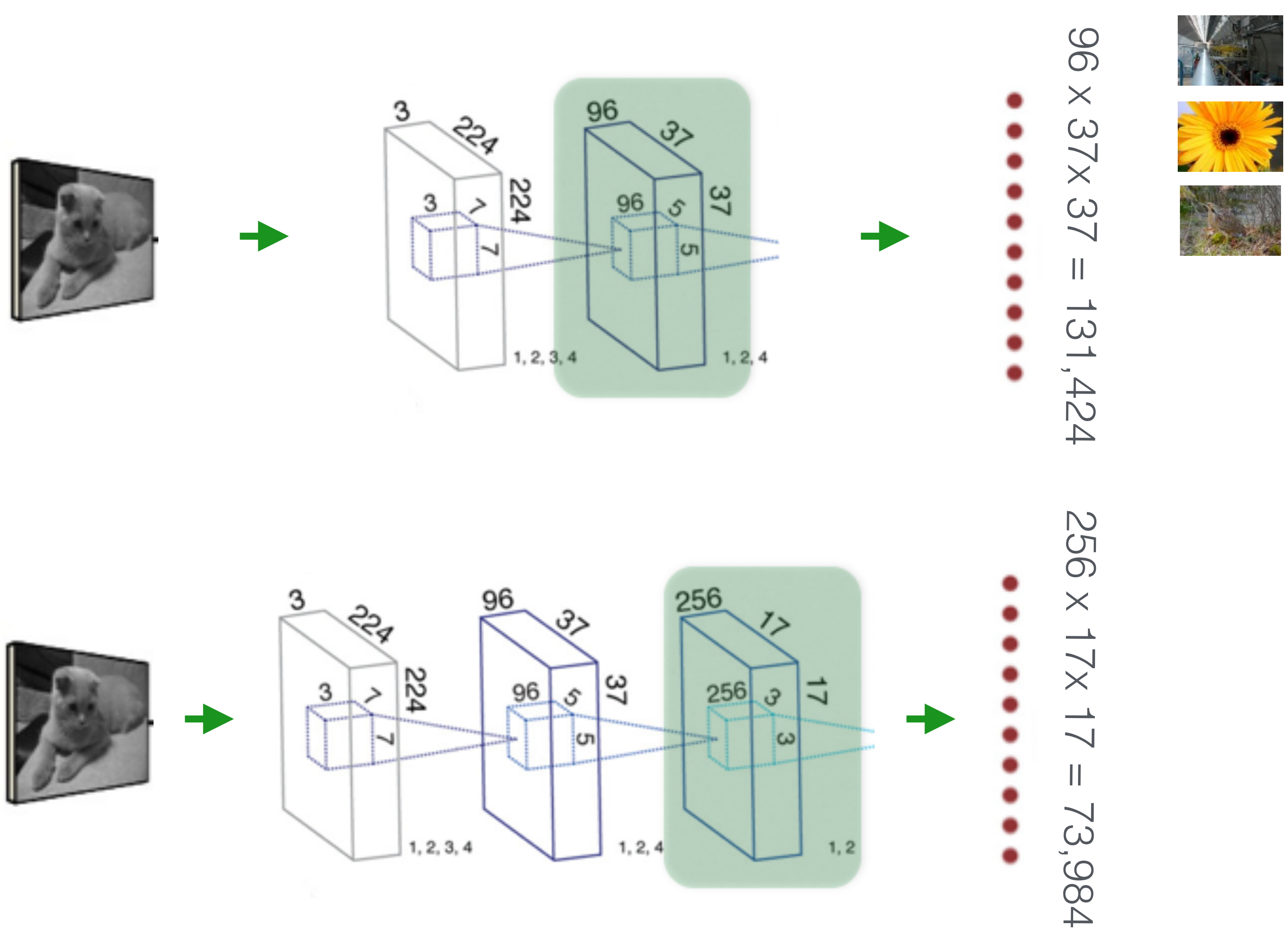
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



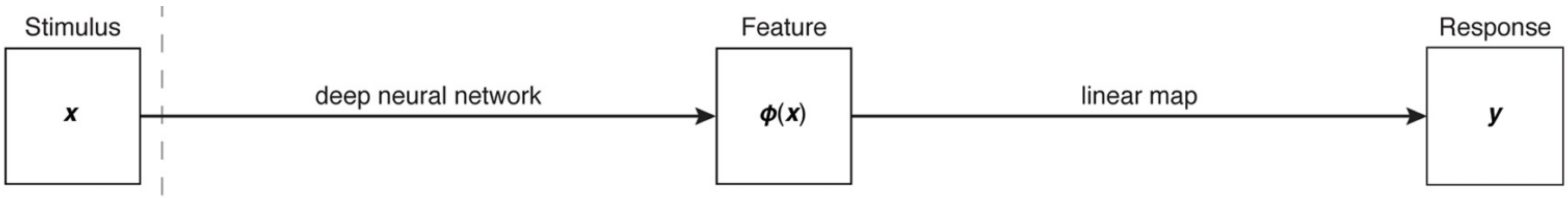
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



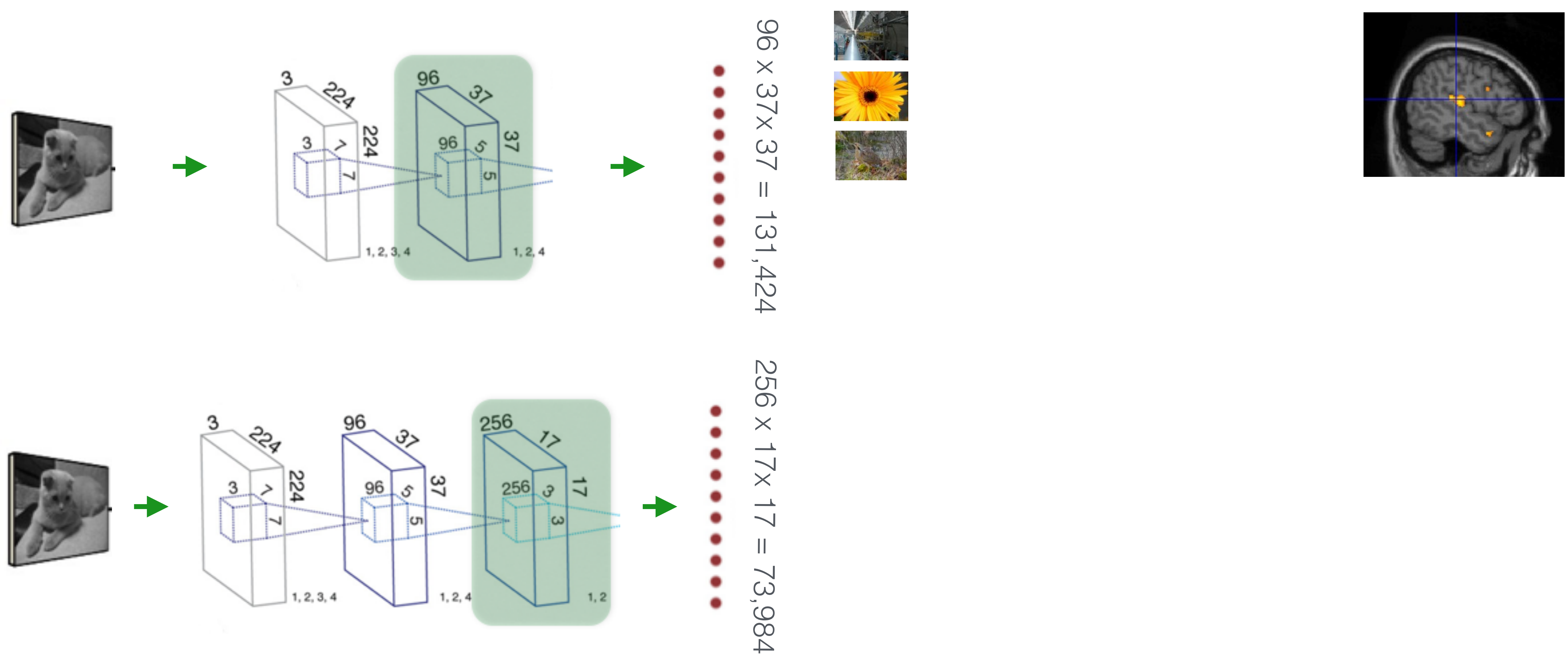
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



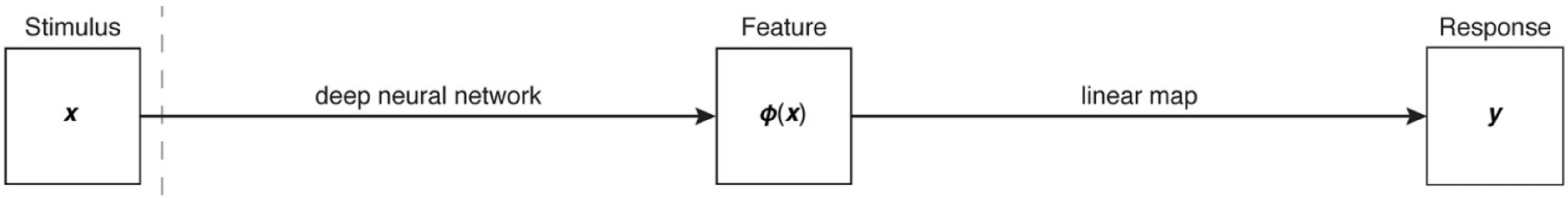
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



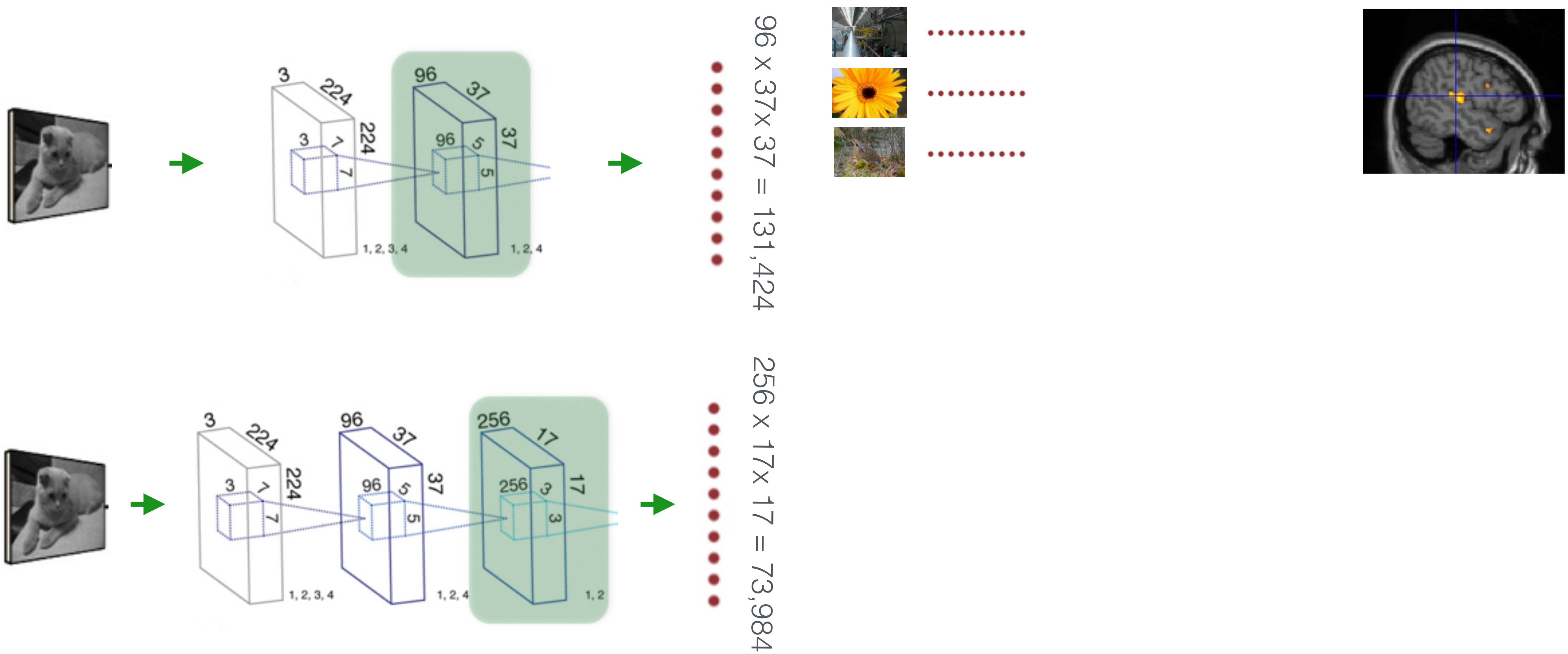
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



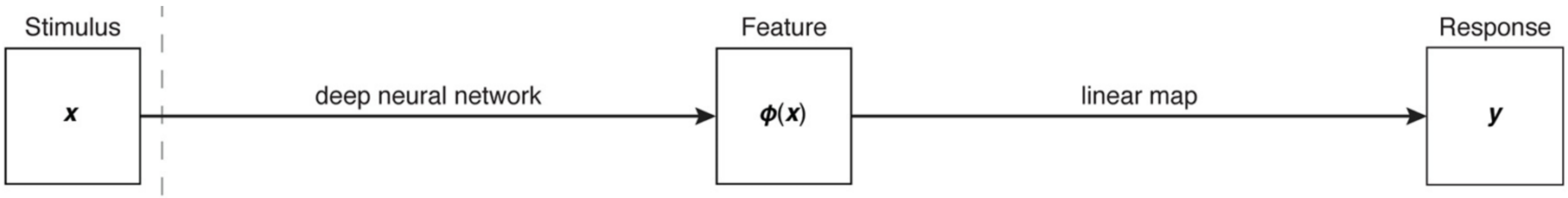
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



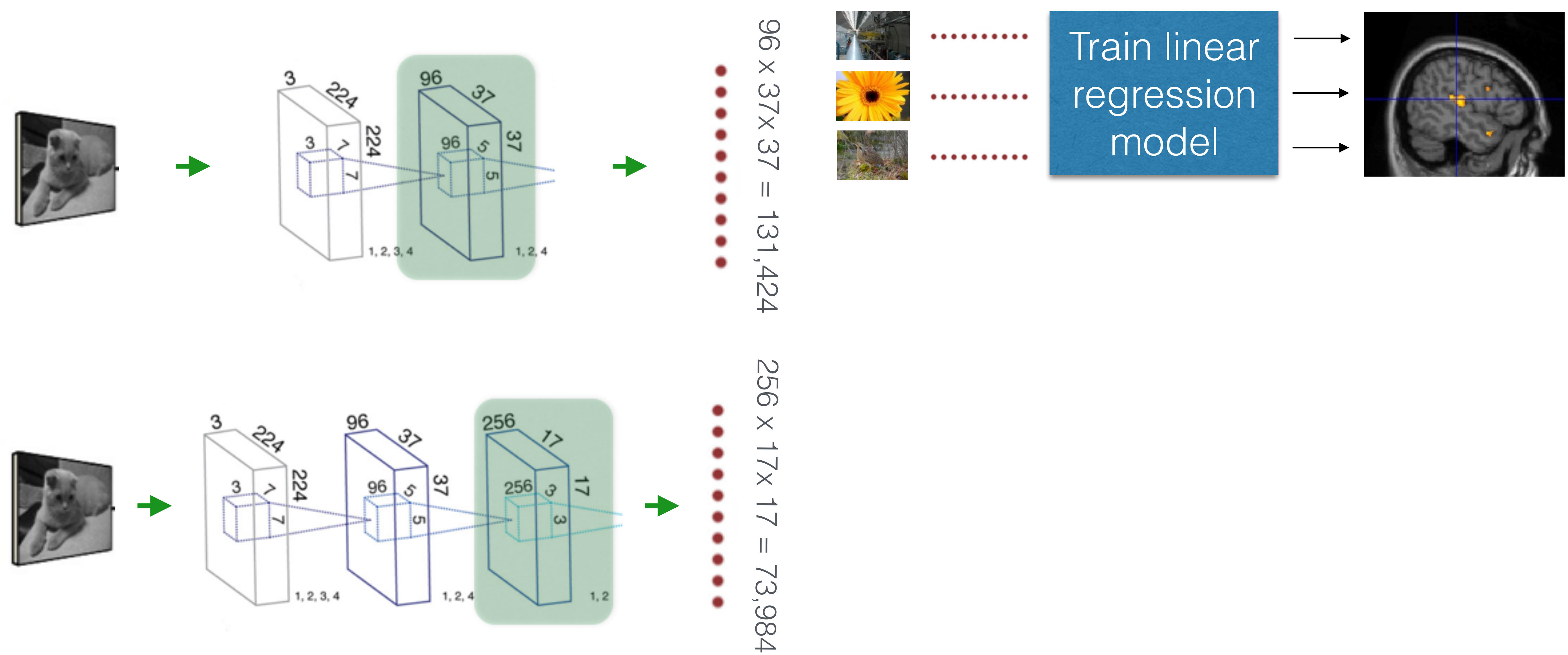
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



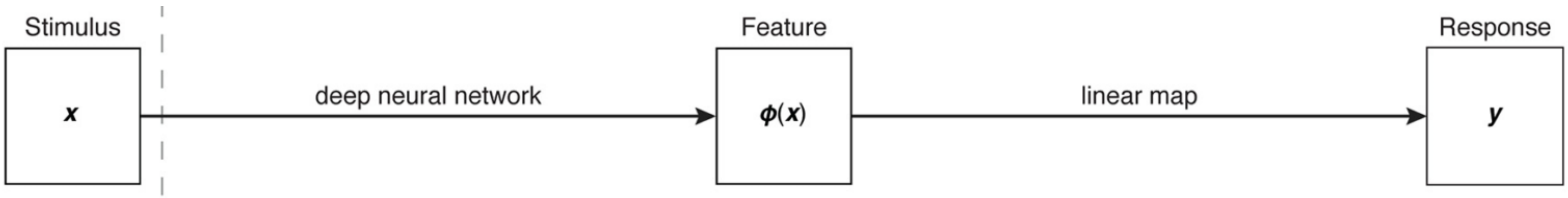
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



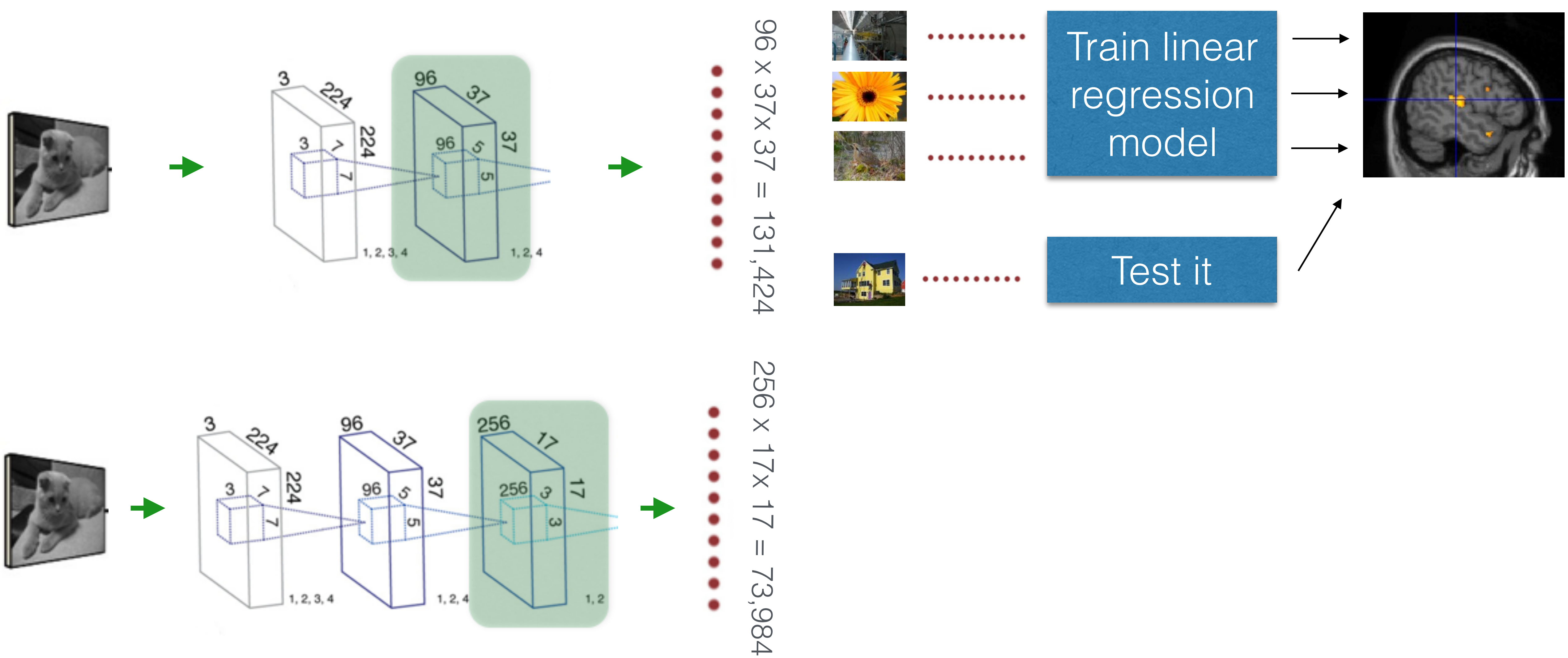
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



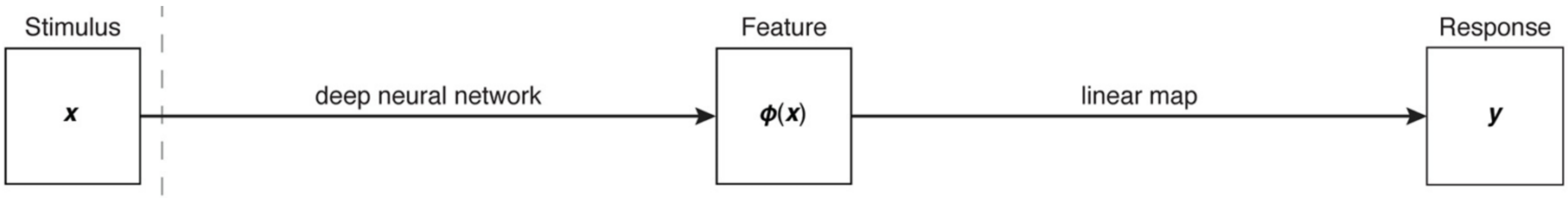
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



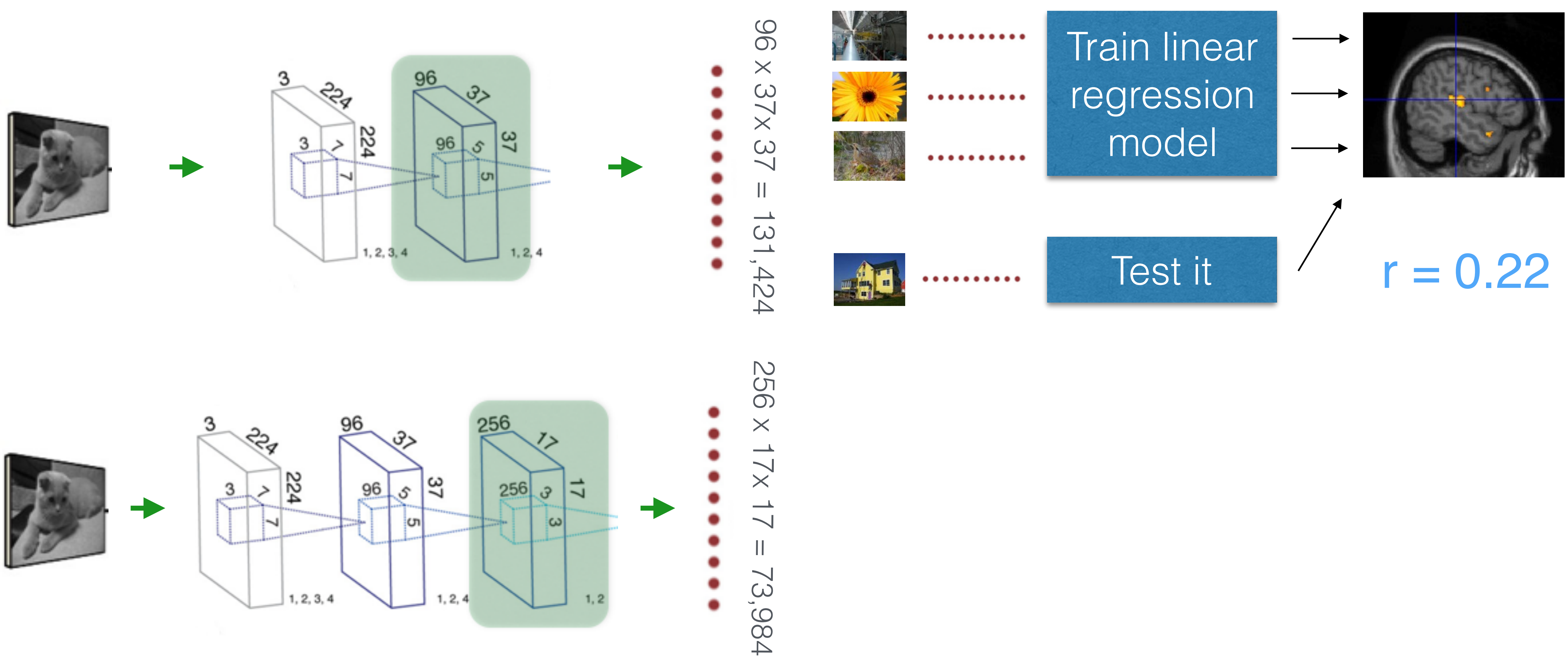
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



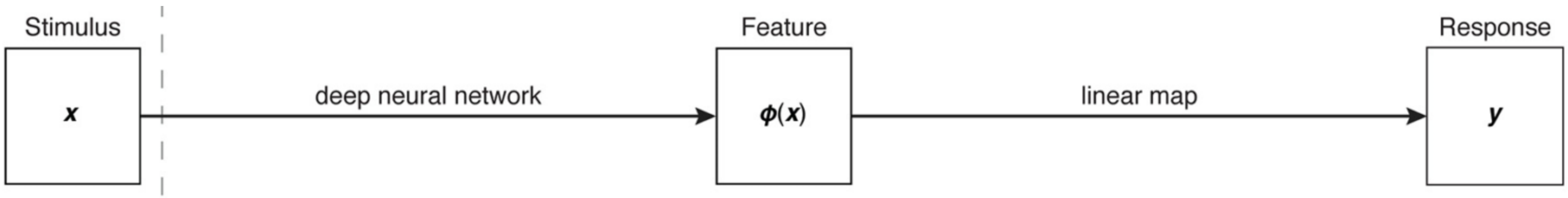
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



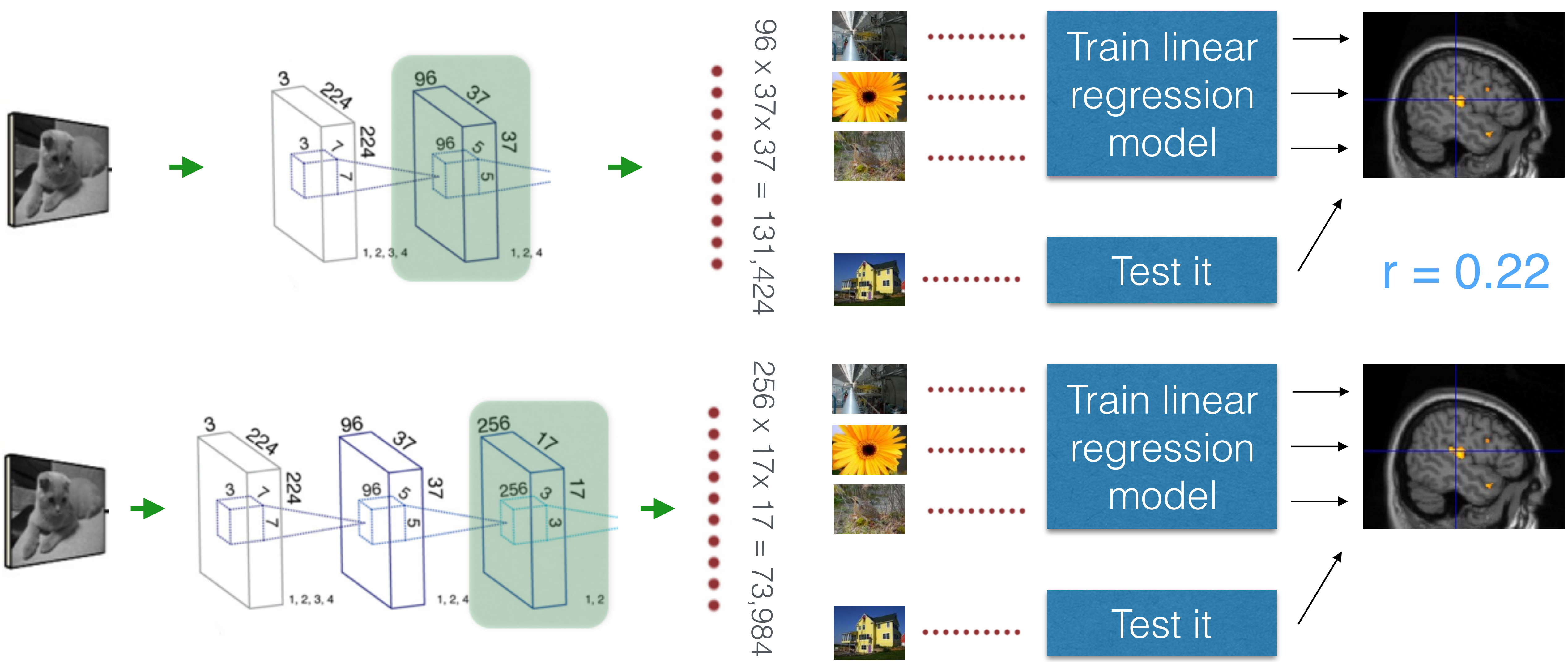
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



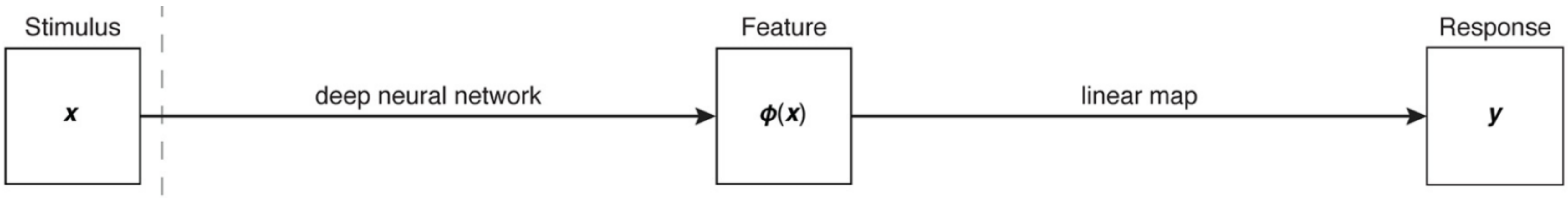
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



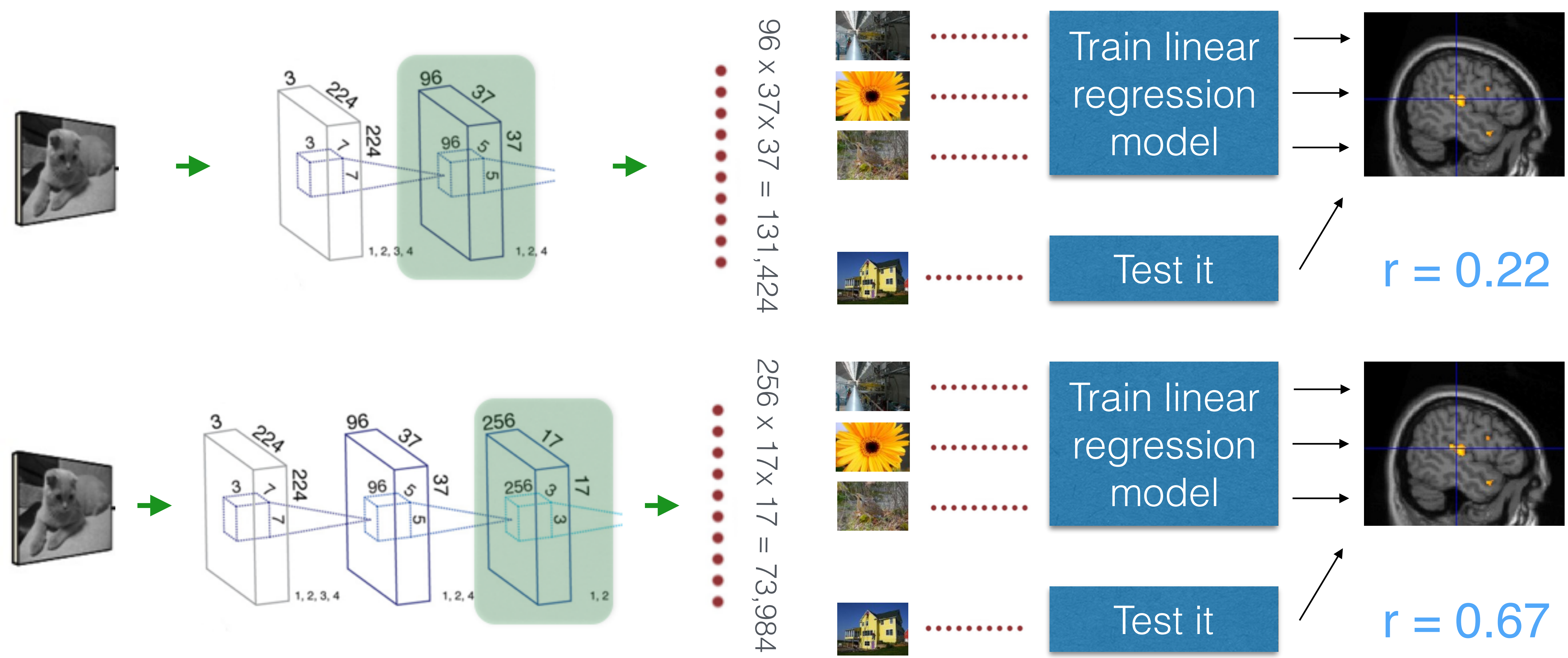
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



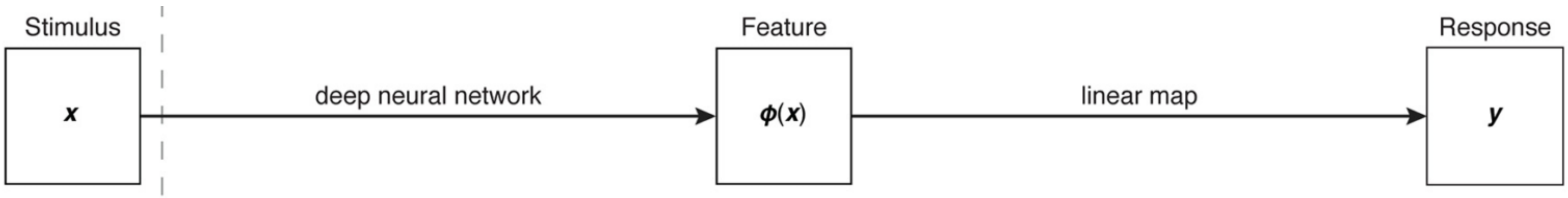
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



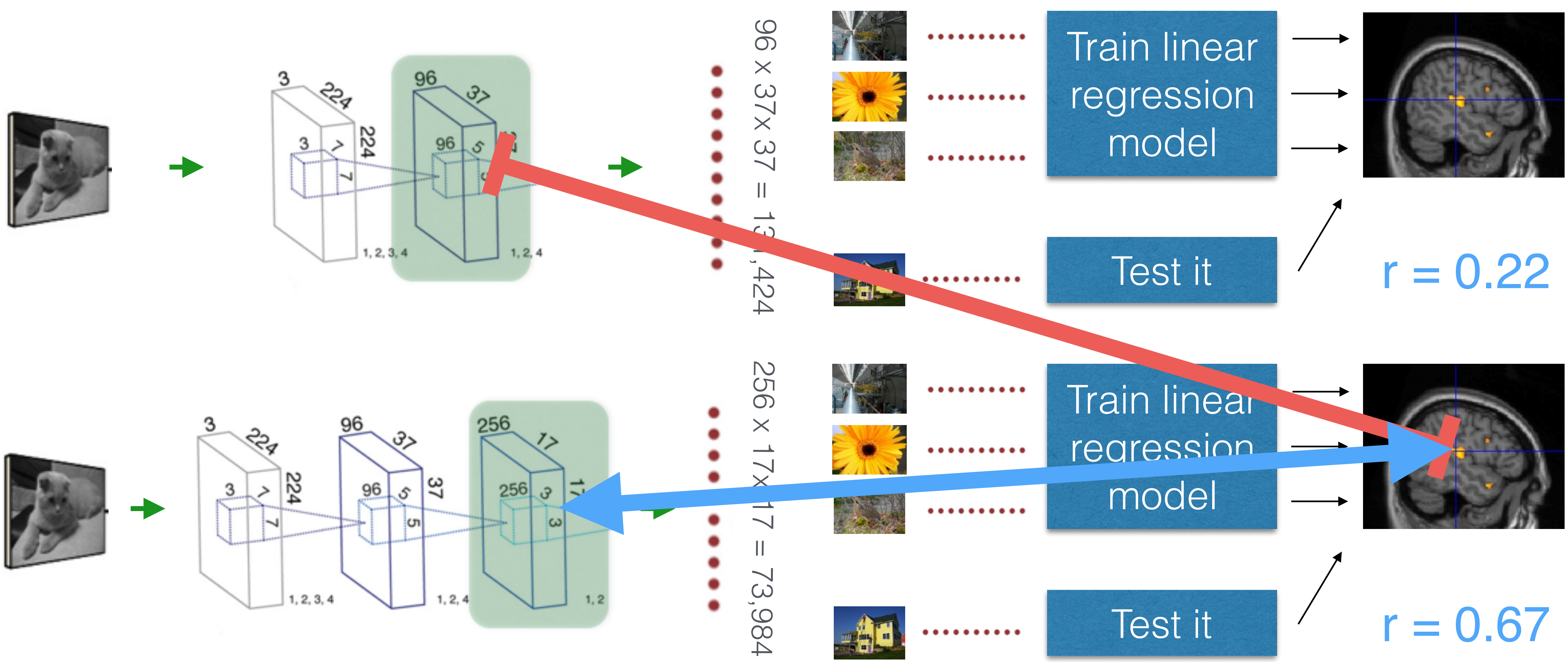
DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).



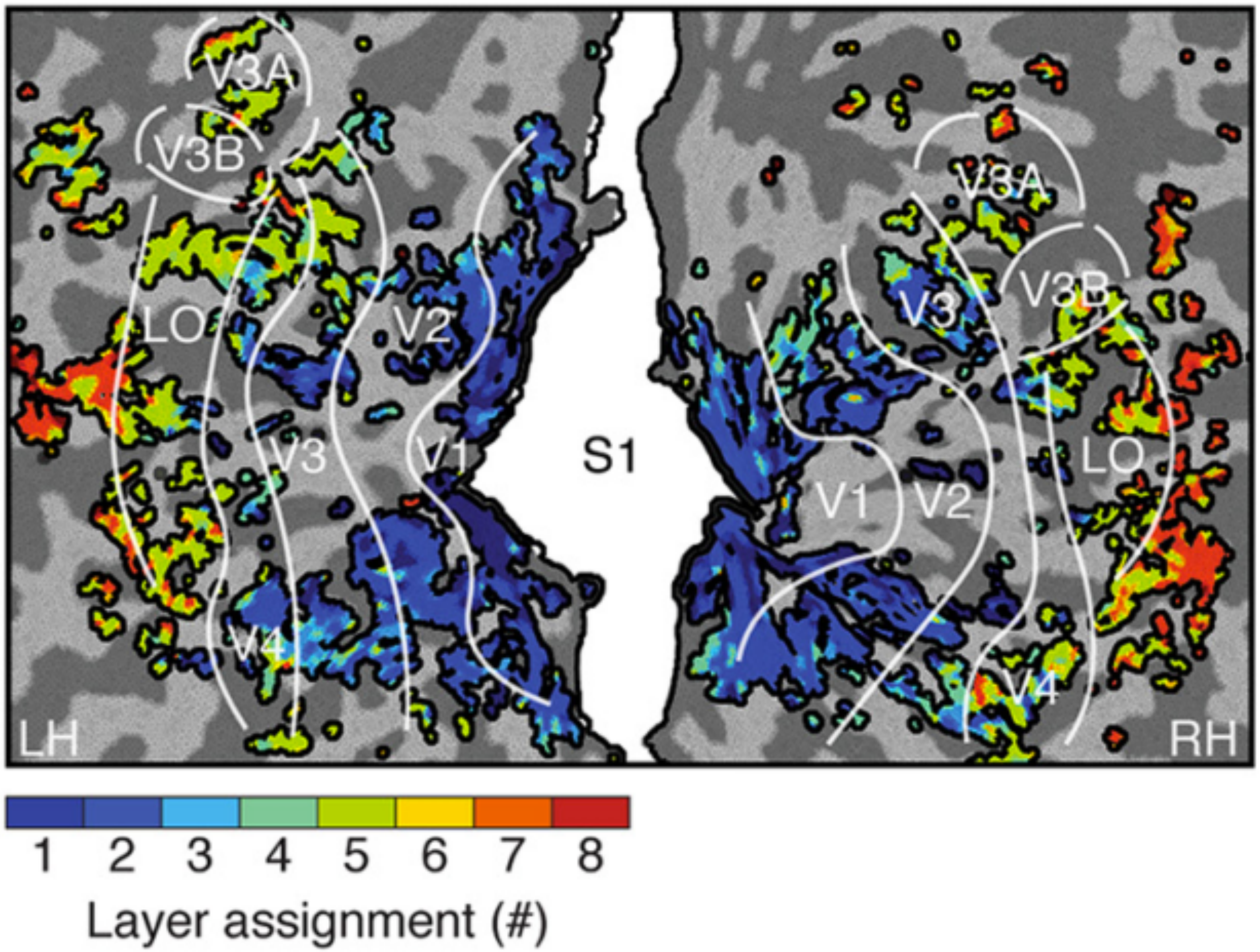
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations



DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural network transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***).

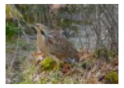


Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream



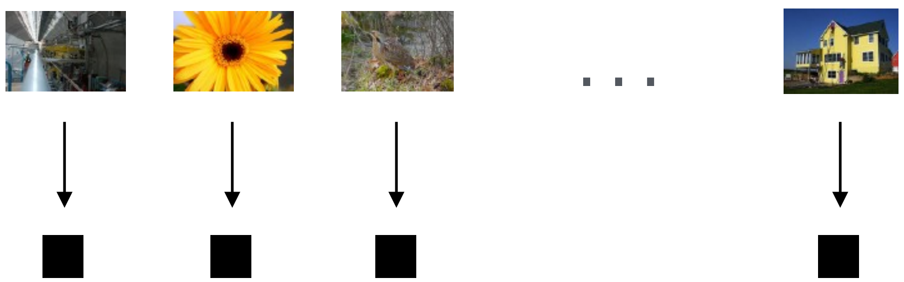
NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set

   … 

NEXT COOL THING: CATEGORIES OF FEATURES

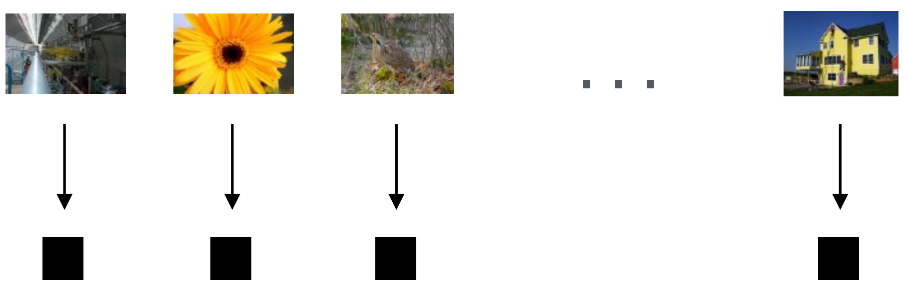
ImageNet validation set



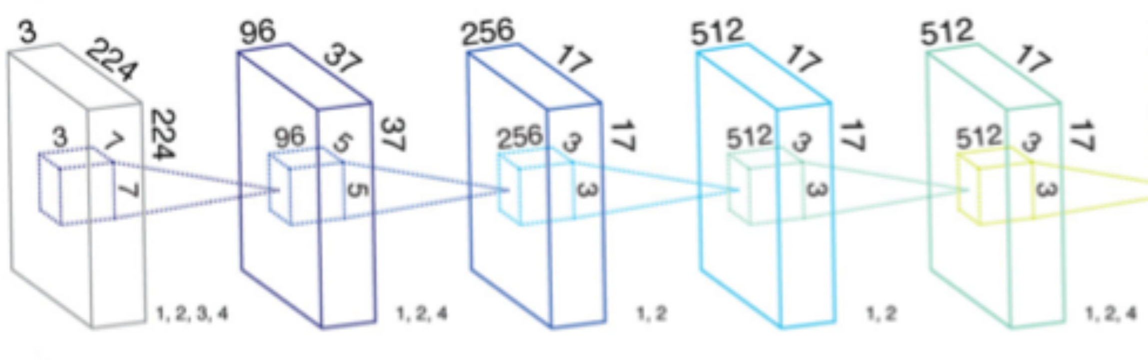
1888

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set

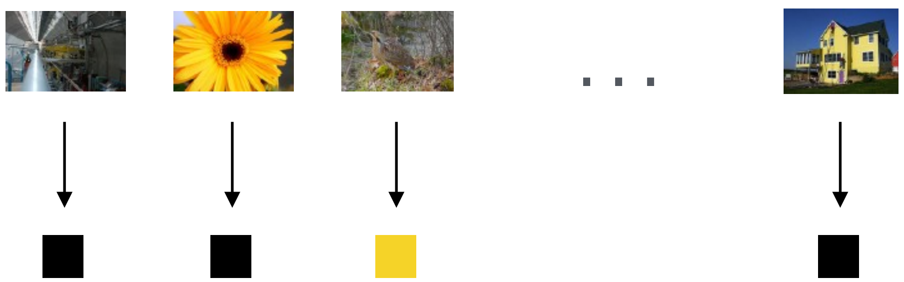


1888

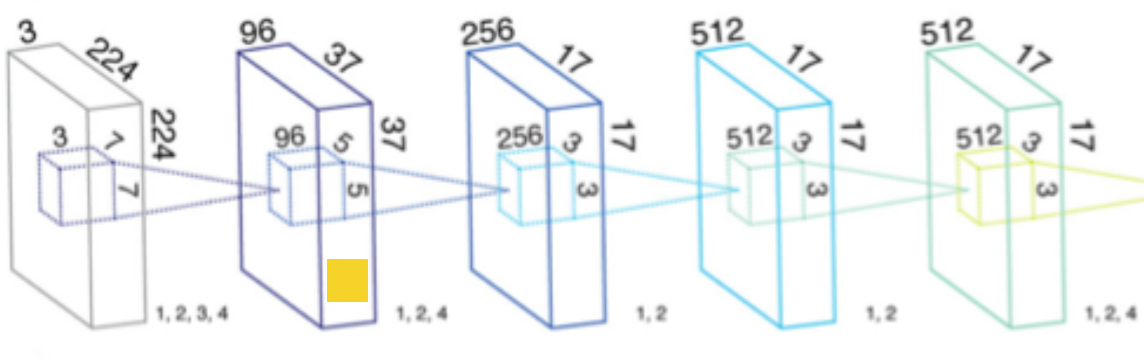


NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set

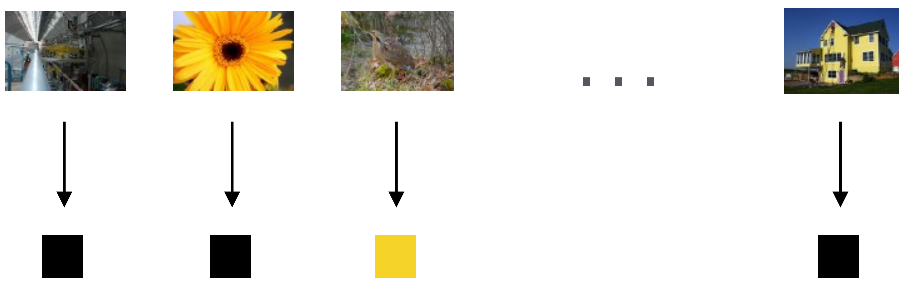


1888

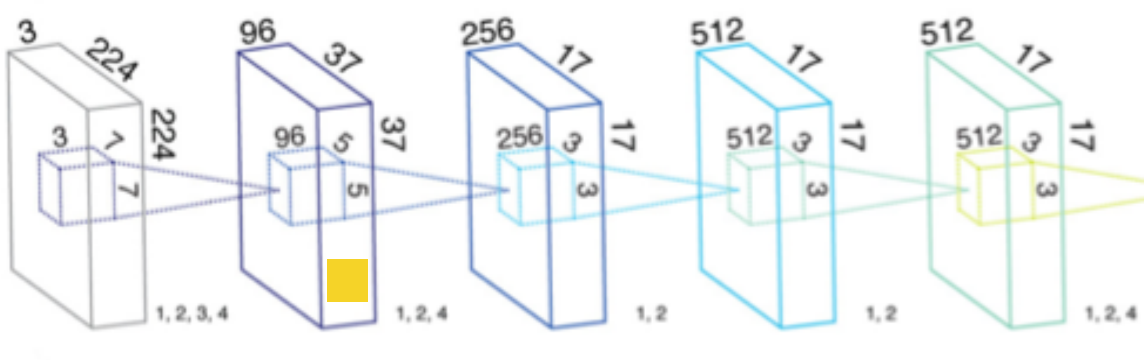


NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



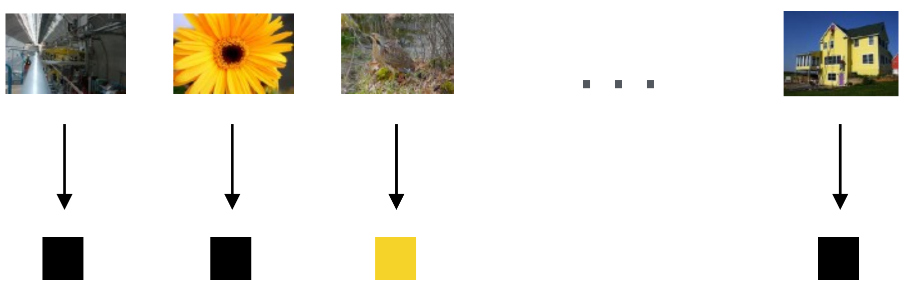
1888



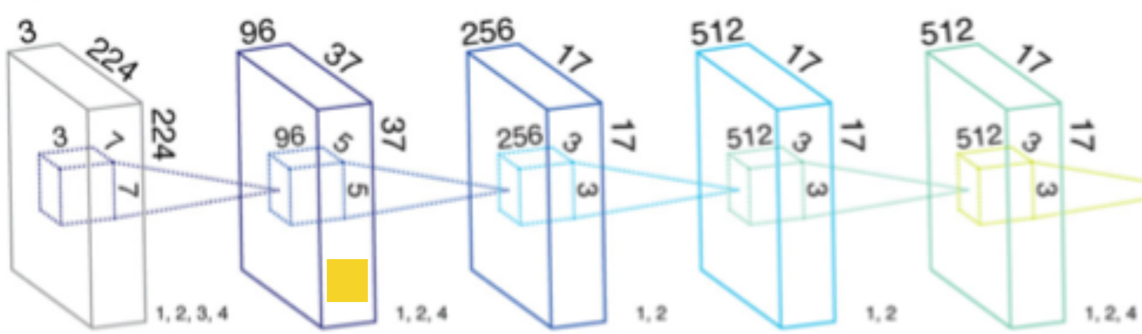
deconvolution

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



1888



deconvolution

human-assigned
to 9 categories

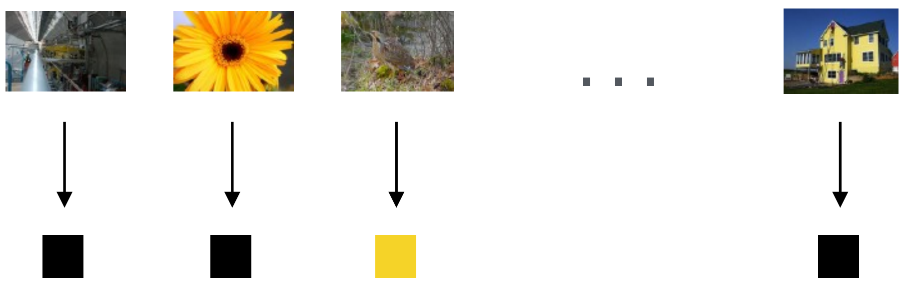
Low
- blob
- contrast
- edge

Mid
- contour
- shape
- texture

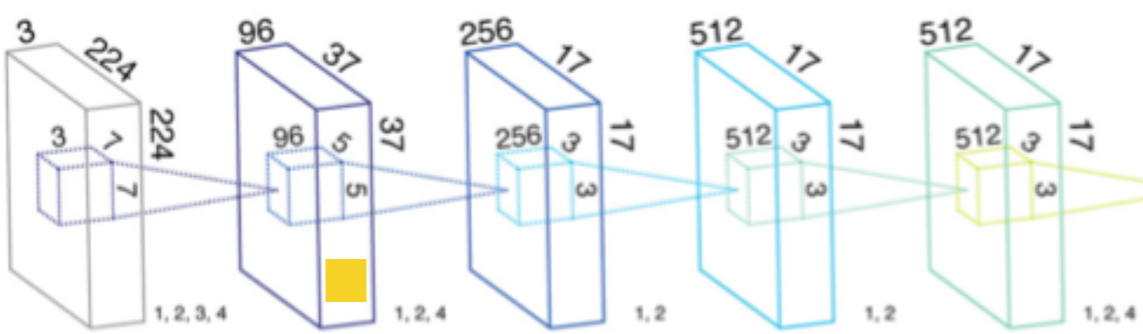
High
- pattern
- object
- object part

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



1888



deconvolution

human-assigned
to 9 categories

Low
- blob
- contrast
- edge

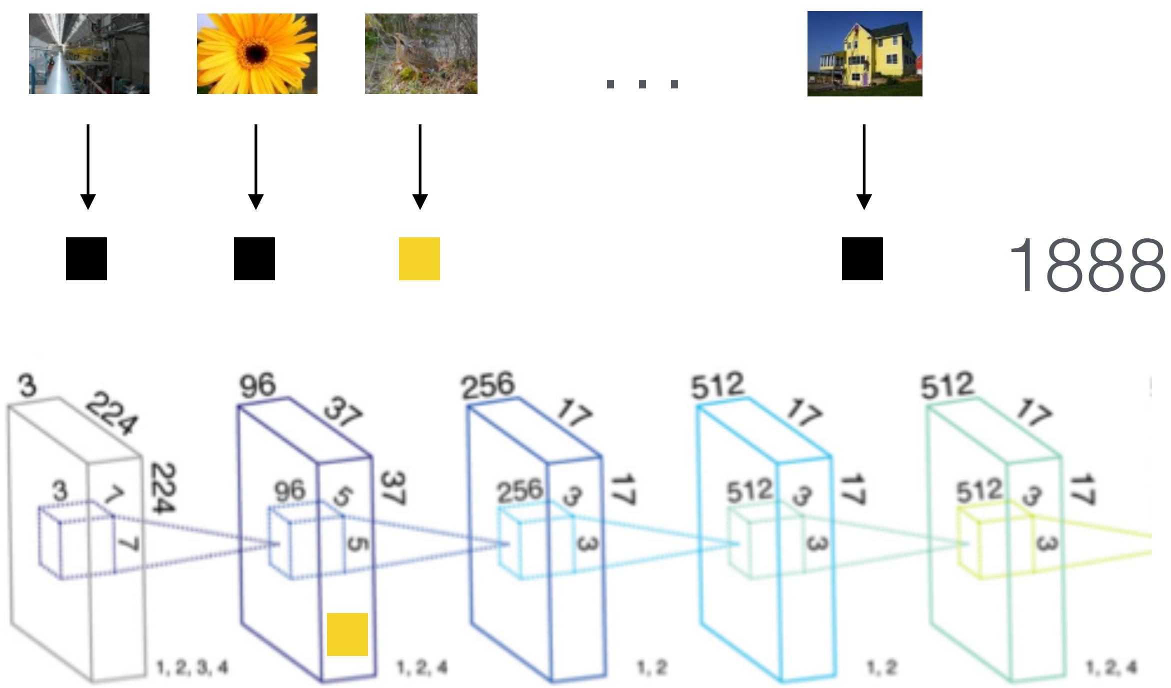
Mid
- contour
- shape
- texture

High
- pattern
- object
- object part

1. Divide 1888 neurons into 9 categories

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



deconvolution

human-assigned
to 9 categories

Low
- blob
- contrast
- edge

Mid
- contour
- shape
- texture

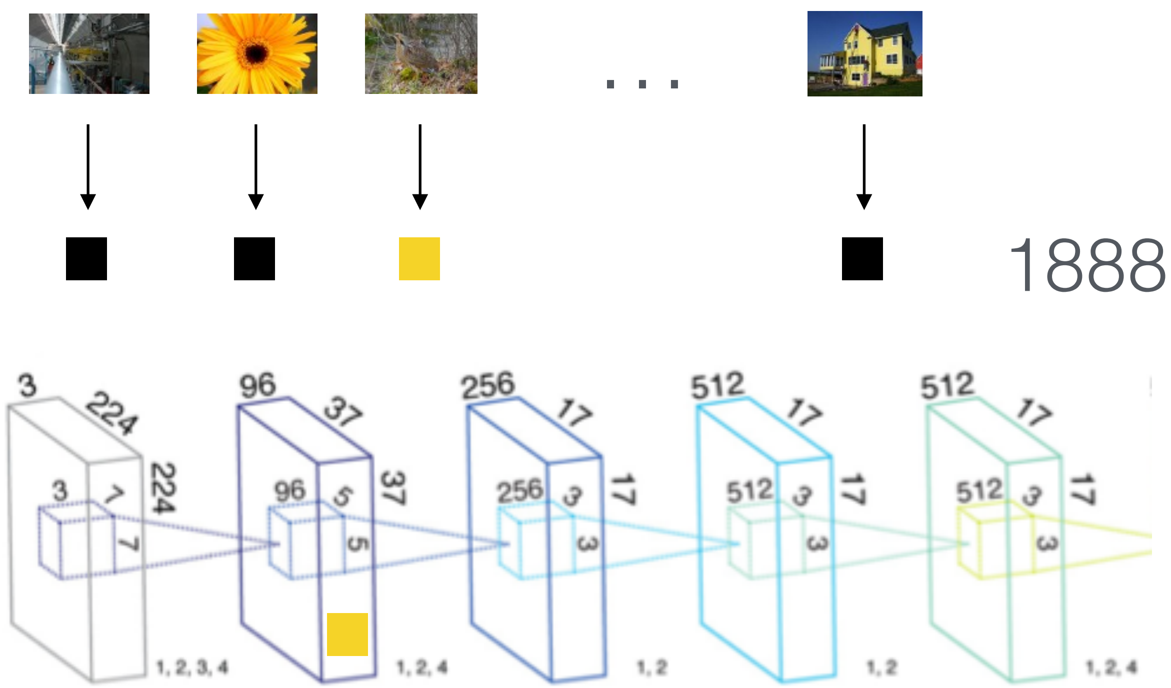
High
- pattern
- object
- object part

1. Divide 1888 neurons into 9 categories

2. Predict activity of each voxel from group-by-group

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



deconvolution

human-assigned
to 9 categories

Low
- blob
- contrast
- edge

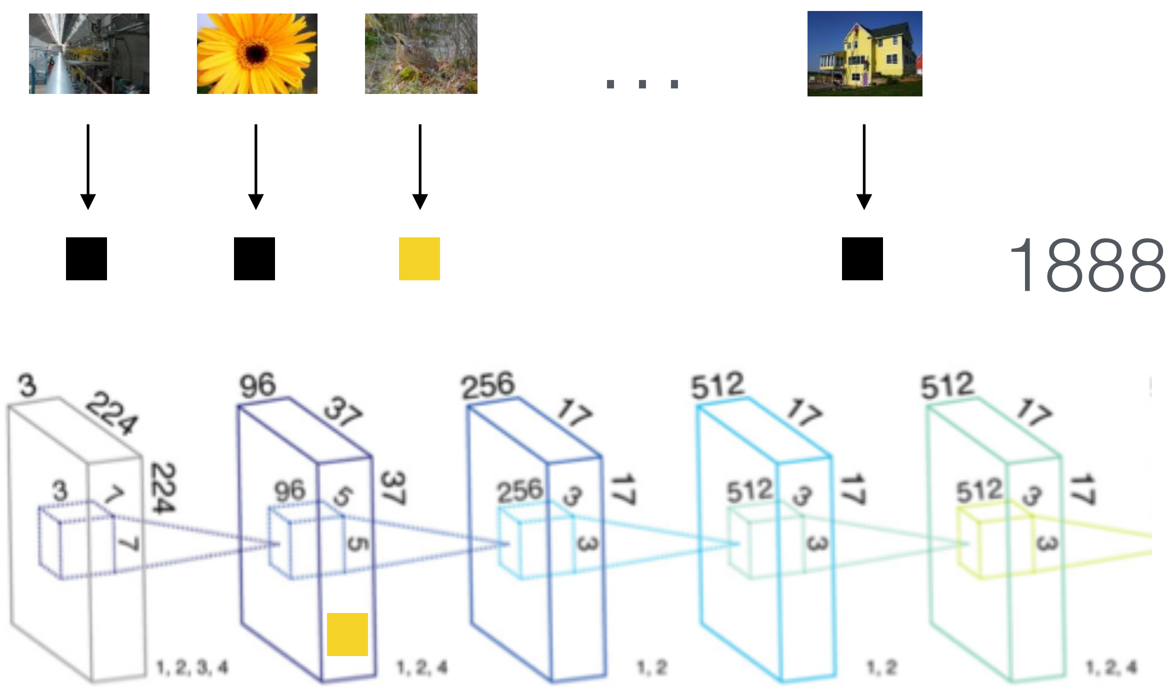
Mid
- contour
- shape
- texture

High
- pattern
- object
- object part

1. Divide 1888 neurons into 9 categories
2. Predict activity of each voxel from group-by-group
3. For each voxel find the group, which best predicts voxel's activity

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



deconvolution

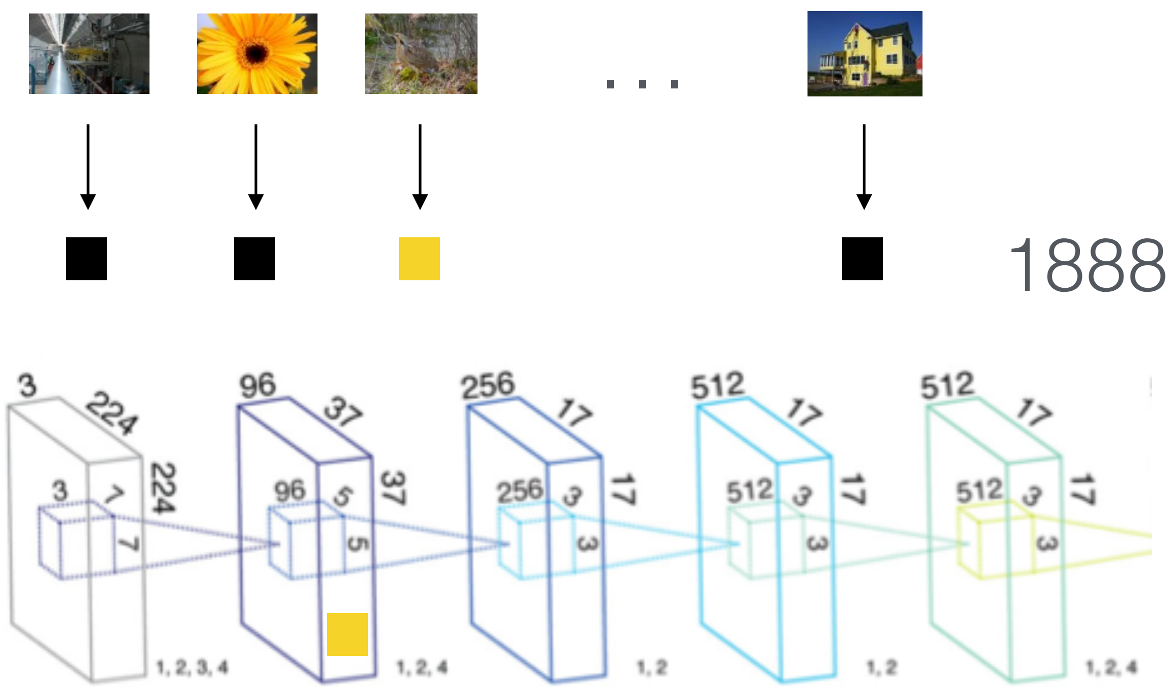
human-assigned
to 9 categories

Low	Mid	High
• blob	• contour	• pattern
• contrast	• shape	• object
• edge	• texture	• object part

1. Divide 1888 neurons into 9 categories
2. Predict activity of each voxel from group-by-group
3. For each voxel find the group, which best predicts voxel's activity
4. Assign each of 1888 DNN neurons to a visual layer: V1, V2, V4, LO

NEXT COOL THING: CATEGORIES OF FEATURES

ImageNet validation set



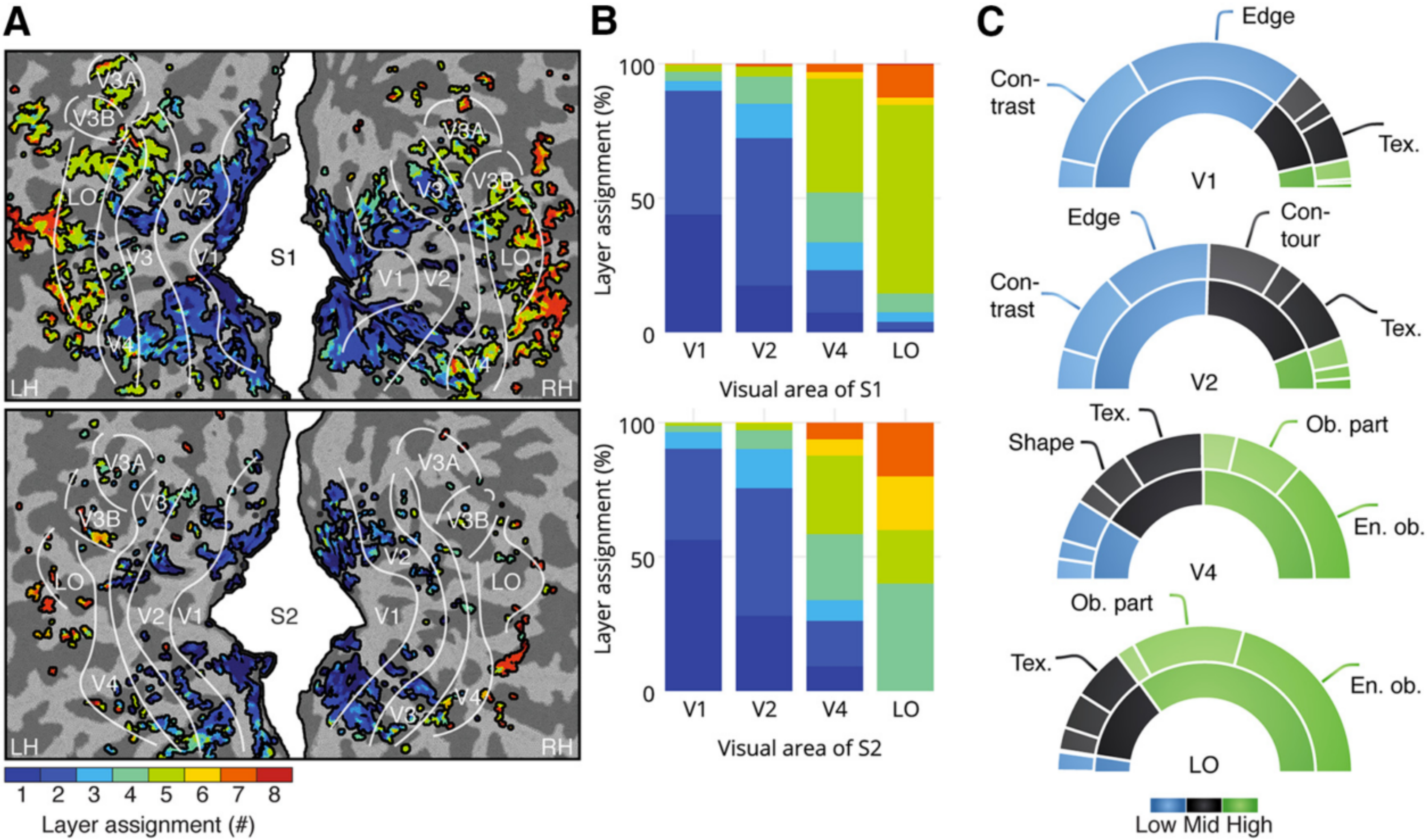
deconvolution

human-assigned
to 9 categories

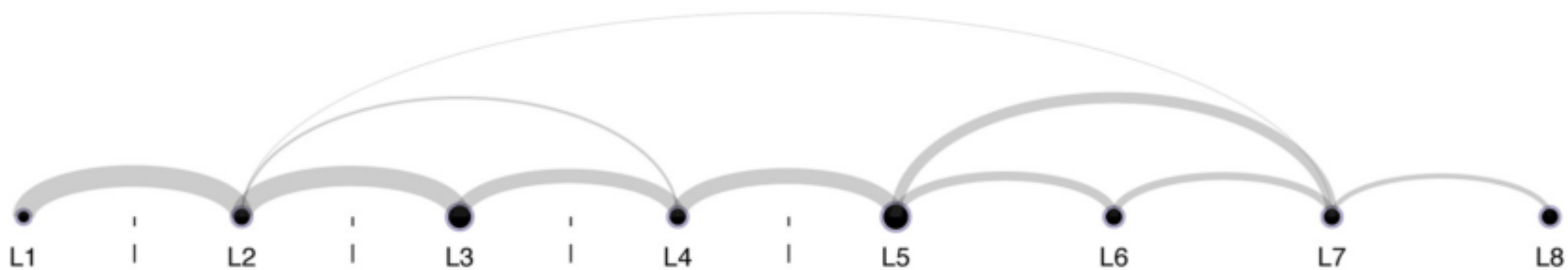
Low	Mid	High
• blob	• contour	• pattern
• contrast	• shape	• object
• edge	• texture	• object part

1. Divide 1888 neurons into 9 categories
2. Predict activity of each voxel from group-by-group
3. For each voxel find the group, which best predicts voxel's activity
4. Assign each of 1888 DNN neurons to a visual layer: V1, V2, V4, LO
5. Map visual layers to categories

NEXT COOL THING: CATEGORIES OF FEATURES

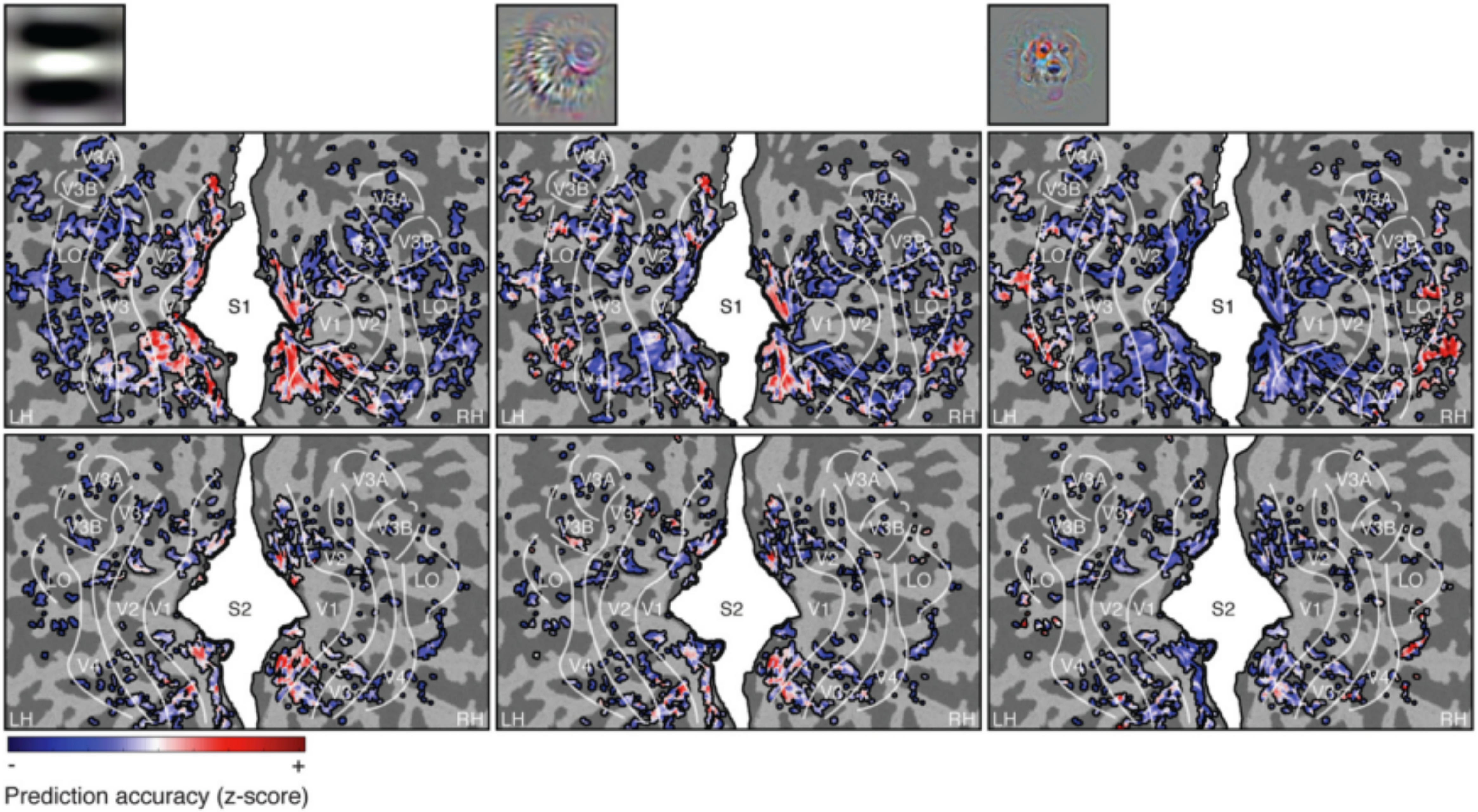


OTHER RESULTS



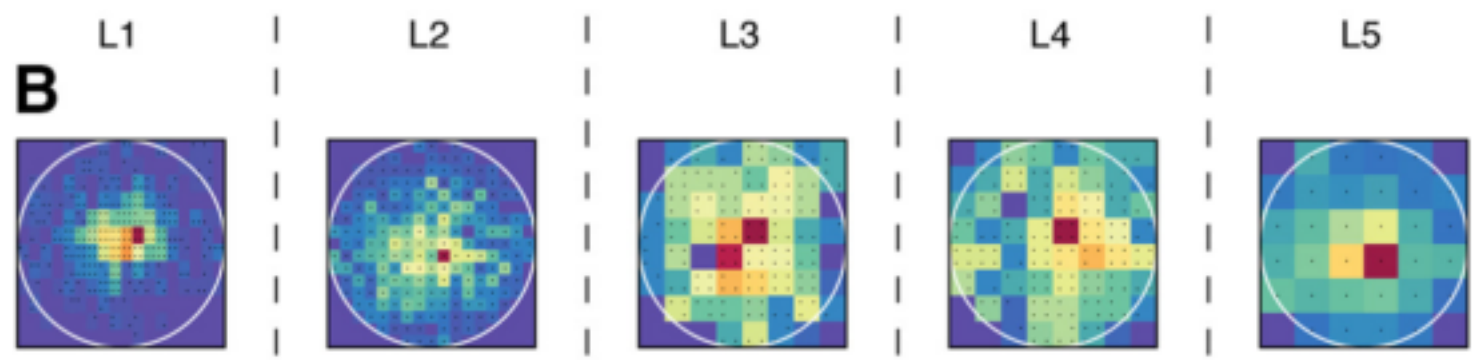
Correlation between predicted responses
between pairs of voxel groups

OTHER RESULTS



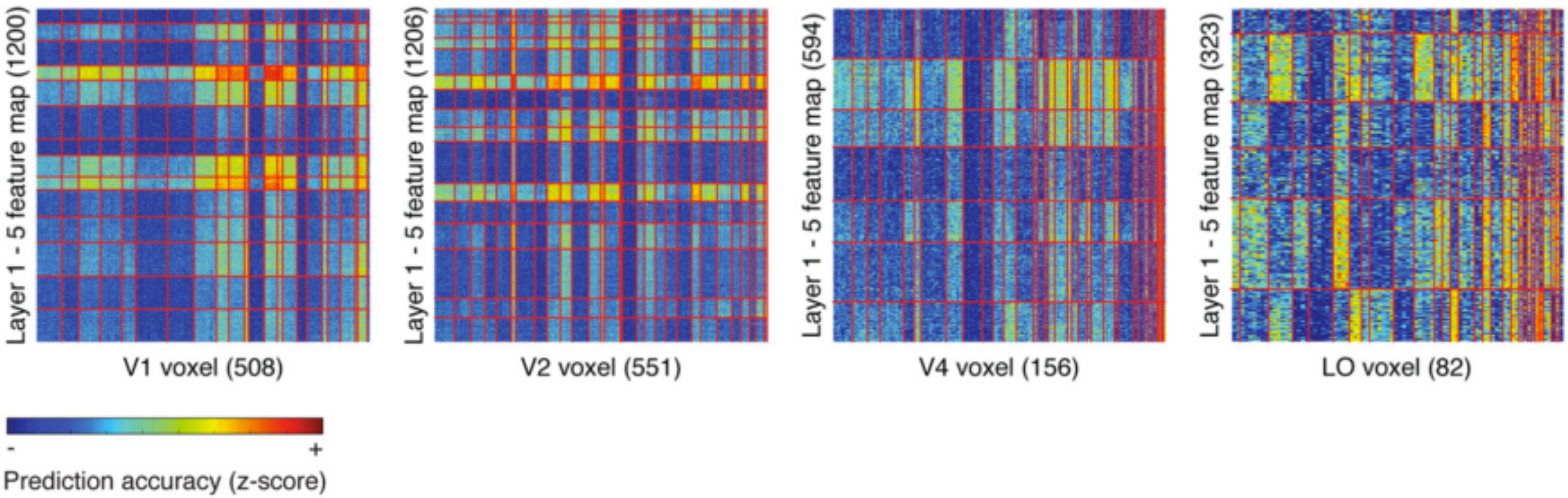
Selectivity of visual areas to feature maps of varying complexity

OTHER RESULTS



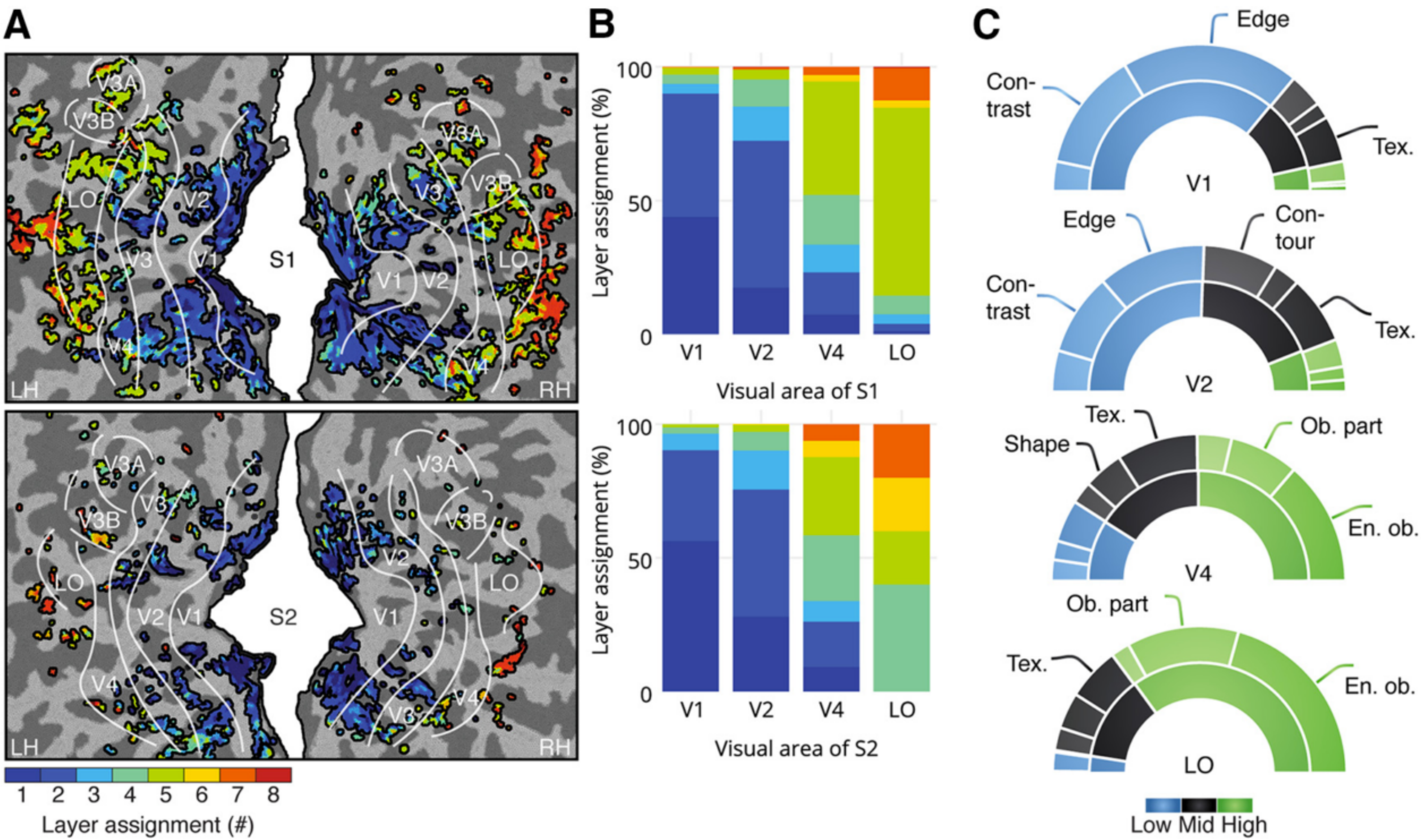
Distribution of the receptive field centers

OTHER RESULTS



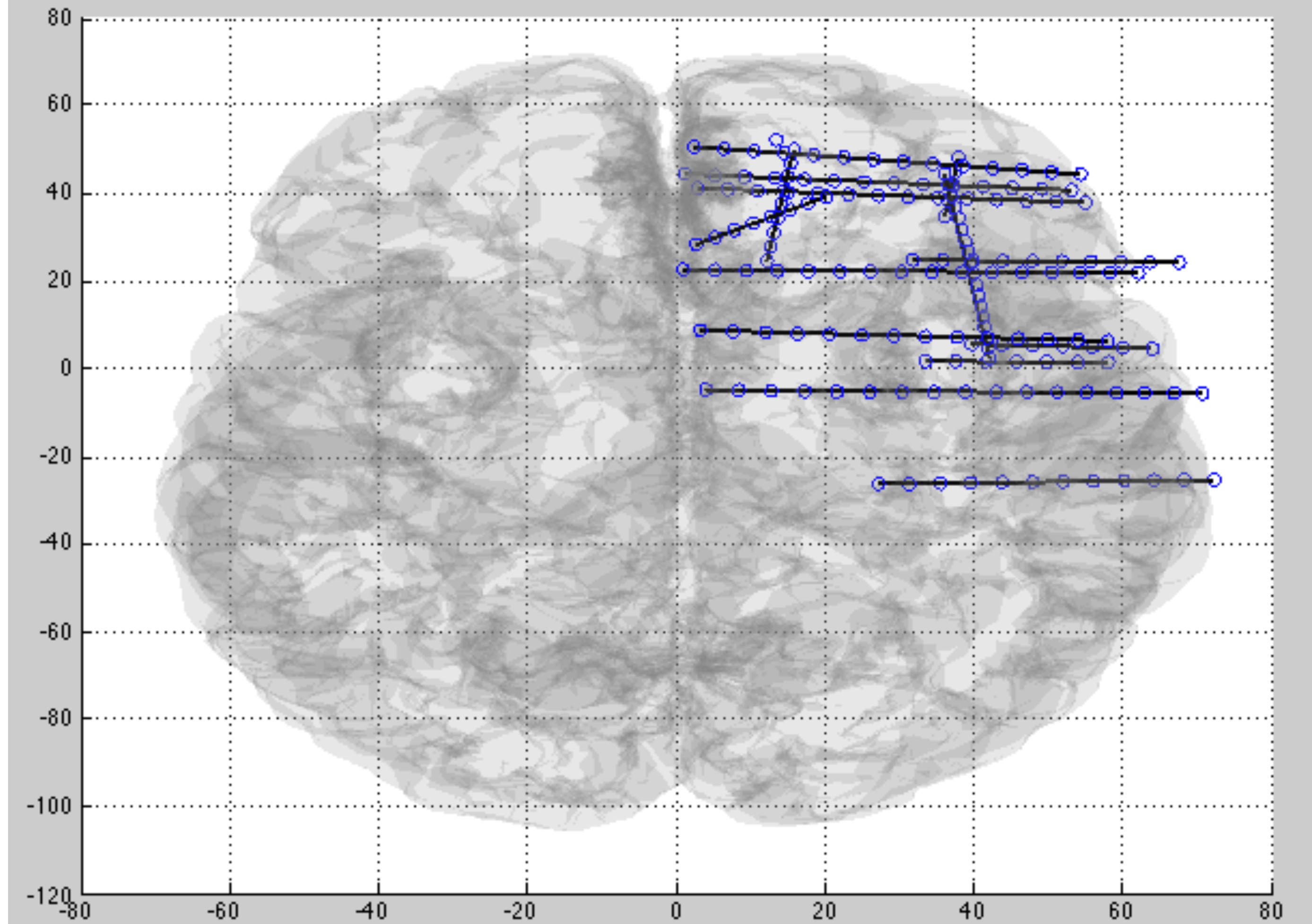
Biclustering of voxels and feature maps

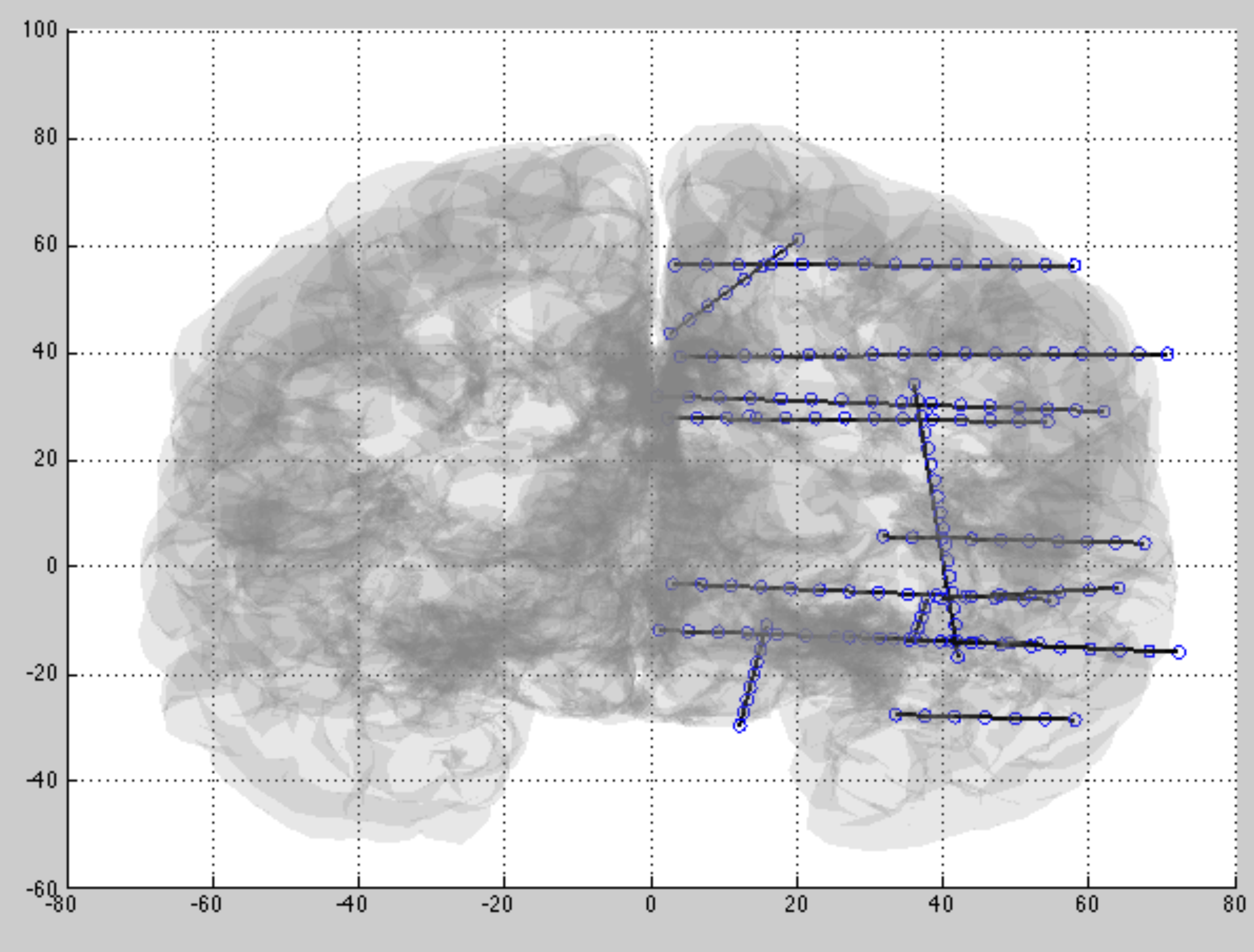
SUMMARY

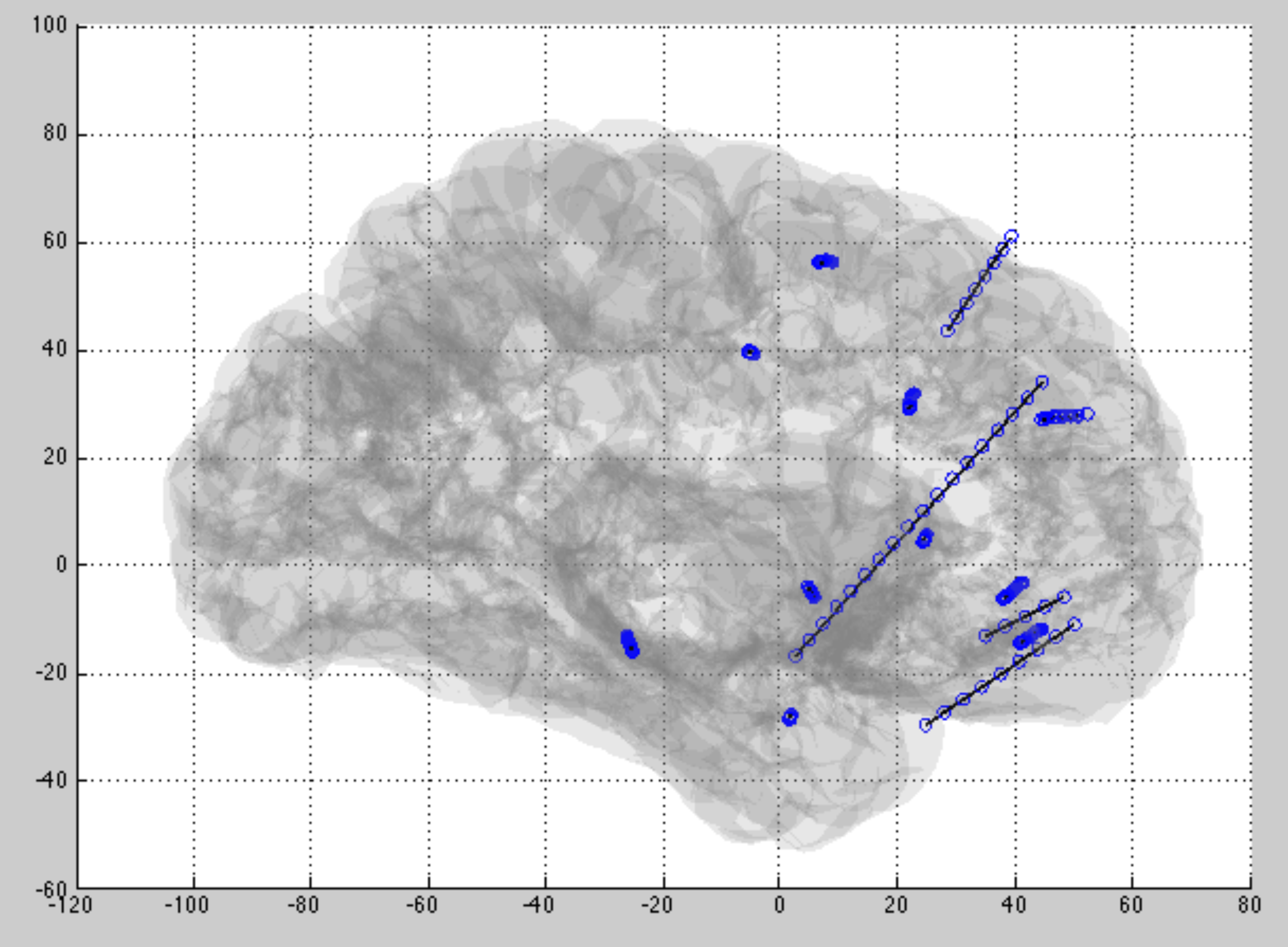

A
V3A
V3B
LO
V2
V3
V1
S1
V4
LH
RH
S2
1 2 3 4 5 6 7 8
Layer assignment (#)
B
Layer assignment (%)
100
50
0
V1 V2 V4 LO
Visual area of S1
Visual area of S2
C
Edge
Con-trast
Tex.
Con-tour
Shape
Ob. part
En. ob.
Low Mid High

An intracranial dataset we have.

How to repeat the result?

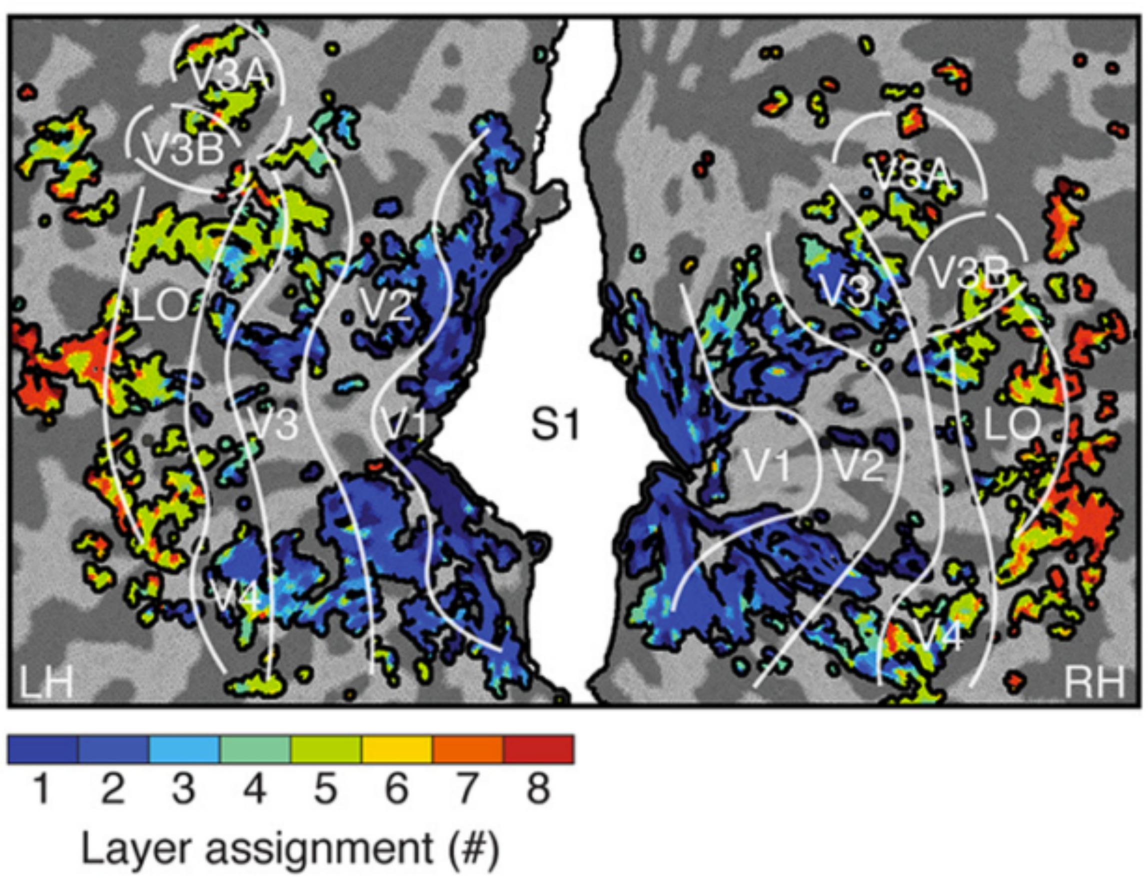






An intracranial dataset we have.

How to repeat the result?



vs.