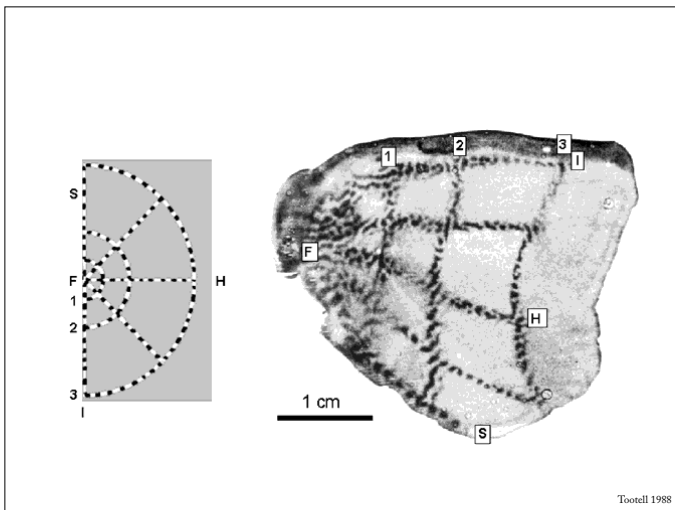


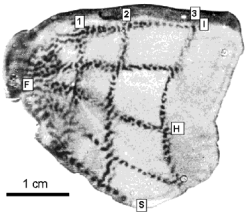


## Retinotopy and Representation

Gabriel Greenberg • Mental Iconicity Seminar  
3.10.26



Tootell 1988

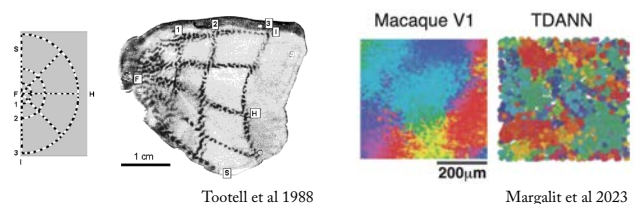


- There appear to be pictures in the brain.
- Can we infer that there are pictures in the brain?
  - **No:** there is no homuncular viewer in the brain.
  - **And no:** we need to show that these brain states are *used* as images.
- But what is it for a brain state to function as an image anyway?
- Today I'll argue that the representations produced by retinotopic areas are iconic; in particular, that they function as picture-like representations.

**What is the computational function of retinotopy?**

### Why are there retinotopic areas?

- Retinotopy is the result of **wiring-optimization** and **functional constraints**. (Covey 1979, Durbin & Mitchison 1990, Kaas 1997, Chklovskii & Koulakov 2004)
  - Compare: **orientation maps** (Margalit et al 2023), **multiple maps** (Doshi & Konkle 2023, Obeid & Konkle 2021)
  - Note: CNN-based methods **presuppose** retinotopy!
- But **all** brain regions are the result of wiring optimization + function.
- So what functional constraints give rise to **retinotopy in particular**?

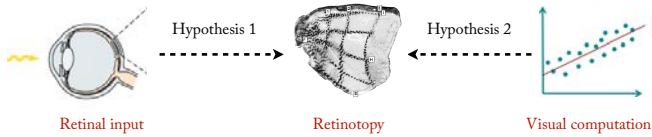


Tootell et al 1988

Margalit et al 2023

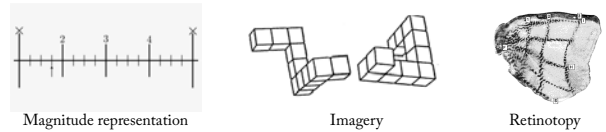
### What are the functional constraints that give rise to retinotopy?

- **Hypothesis 1: retina-oriented explanations**
  - The retinotopic layout of V1 is explained in terms of the organization of **input that derives from the retina**.
- **Hypothesis 2: computation-oriented explanations**
  - The retinotopic layout of V1 is explained in terms of the kinds of **computation** made possible by this layout.
- From a **design perspective**, is retinotopy...
  - Something you **tolerate?** (Hypothesis 1)
  - Something you **aim for?** (Hypothesis 2)



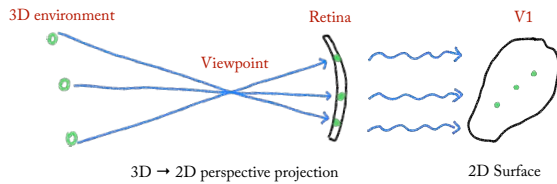
### Retinotopy as pictorial representation

- **Pictorial Hypothesis**
  - Retinotopic areas produce **picture-like representations** of visual space.
  - Picture-like representation is necessary for reliable **convolutional computation**.
- Parallel arguments: Kosslyn 2006, Burge 2022, Beuhler 2025.
- Other **iconic** representations in psychology (Beck 2018, Block 2023):
  - **mental imagery** (Shepard & Metzler 1971, Kosslyn et al 2006)
  - **iconic memory** (Sterling 1960)
  - **echoic buffer** (Neisser 1967)
  - **analogue magnitude** representations (Carey 2009, Beck 2015)
  - **place codes** and **rate codes** (Groh 2001)



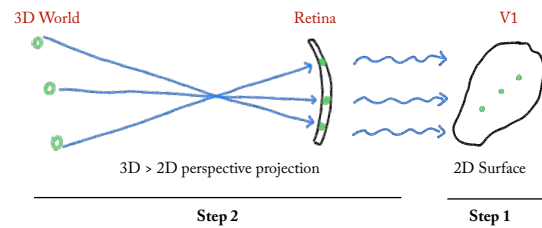
### Picture-like representation

- For a representation P to be **picture-like**:
  - (1) P forms a **2D surface**.
  - (2) P encodes **3D space** via **perspective projection**. (Greenberg 2021)
    - Perspective projection is a family of methods for **3D → 2D transformation** relative to a **viewpoint**. (Willats 1996)
- If retinotopic maps are picture-like, then **2D cortical space** functions as a "place code" for perspectival locations in **3D visual space**.



### The argument

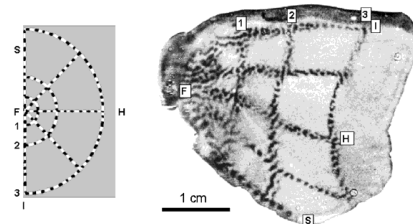
- **Step 1:** retinotopic areas produce **functional 2D representations**.
  - Introduce the concept of **functional space**.
- **Step 2:** functional 2D representations in retinotopic areas are **picture-like**.
  - Argue that **the convolutional model of vision** depends on picture-like representation.



### Functional space in the cortex

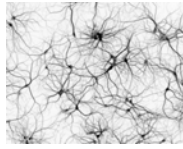
### Retinotopy defined by physical proximity

- Retinotopy is a property of physical space.
- "Neurons whose visual receptive fields lie next to one another in visual space are located next to one another in cortex." (Brewer and Barton 2012.)

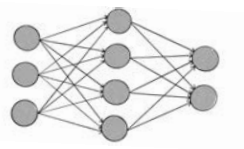


### The problem of spatial sensitivity

- Retinotopy is a property of physical space.
- But physical spatial relations are not (normally) functional for biological neural networks.
- Block's objection:** retinotopy *can't* be functional for brains.
- Kosslyn's reply:** retinotopic areas form "functional spaces" that behave *as if* they formed a space.



Neurons in physical space.



Functional connections.

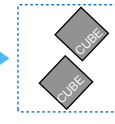
### In search of functional space

- But what is **functional space**?
- To encode a functional space, an area **must to do more than represent space**.
  - Pylyshyn (2003), Rescorla (2009)
  - Criticism of Kosslyn's definition** of functional space (Kosslyn 1980)
- Contrast with representations of **space in the hippocampus**. (Moser et al. 2008)

Using a **spatial representation** to represent space.



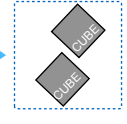
Picture



Content

Using a **non-spatial representation** to represent space.

"There are two grey cubes at angles  $A, A'$ , distances  $D, D'$ , with sizes  $S, S'$ , at orientations  $0, 0'$ ."



Content

### Principle of wiring optimization

- Optimize functionality** and **wiring length** together in circuit construction. (Corey 1979, Chklovskii & Koulikov 2004)
- All else equal: **physical distance** between nodes  $x, y$  is **proportional** to the need for **combining**  $x, y$  values.
- Defeasibly **infer**:
  - From **spatial proximity** in cortex to strong **functional convergence** of outputs.
  - From **retinotopy** to a **topology of informational relations**.

Strong functional convergence.



Shorter spatial distance.

Weak functional convergence.

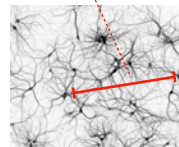


Longer spatial distance.

### Accessibility metric

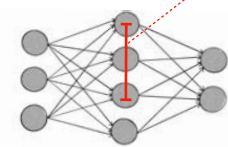
- Definition: an **accessibility metric** measures how informationally accessible each neuron is from the next, relative to a network.
- The accessibility distance between two nodes is proportional to the number of output nodes they both feed into.
- A **functional space** is defined by an accessibility metric.

Physical metric



Neurons in physical space.

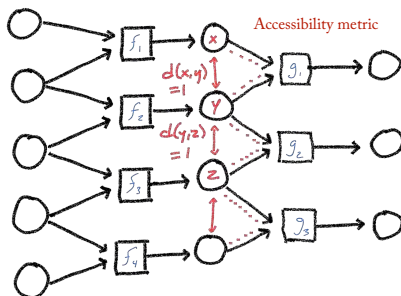
Accessibility metric



Functional connections.

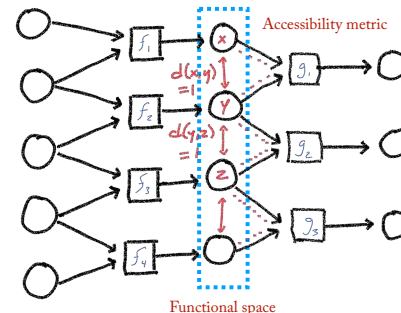
### Accessibility space: an illustration

- Consider a sparse feedforward neural net in which every pair of nodes feeds into a single output node.
- Then  $d(x, y) = 1$  iff  $x$  and  $y$  both feed into the same output node.



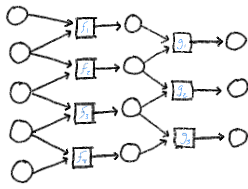
### Accessibility space: an illustration

- Consider a sparse feedforward neural net in which every pair of nodes feeds into a single output node.
- Then  $d(x, y) = 1$  iff  $x$  and  $y$  both feed into the same output node.

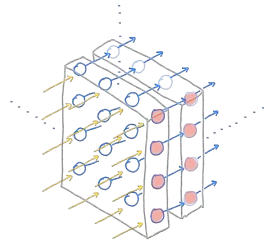


### 2D accessibility metrics

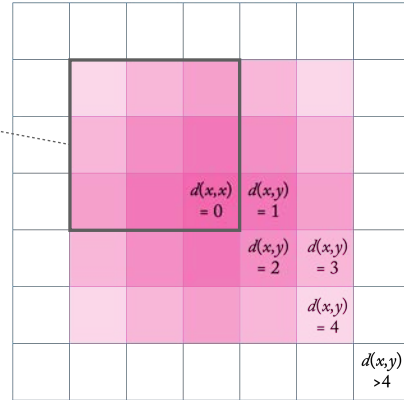
- Any given collection of neurons can form an accessibility metric. But what kind of metric do they form?
- 2D feedforward neural networks with sparse local connections define (+/-) 2D accessibility metrics.



1D sparse feedforward NN.



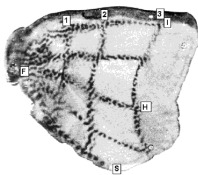
2D sparse feedforward NN.



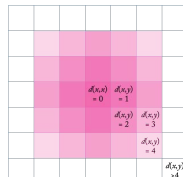
The accessibility metric defined by a convolutional kernel of size 3x3, stride 1 is a **2D city-block metric**.

### Functional retinotopy

- Physical retinotopy:** neurons that are nearby in 2D cortical space have receptive fields that are nearby in 3D visual space.
- Functional retinotopy:** neurons that are nearby in 2D functional space have receptive fields that are nearby in 3D visual space.
- By wiring optimization:** infer from **physical** retinotopy to **functional** retinotopy.
- Hypothesis:** picture-like representations are encoded by functional 2D surfaces in retinotopic areas.



From retinotopy...

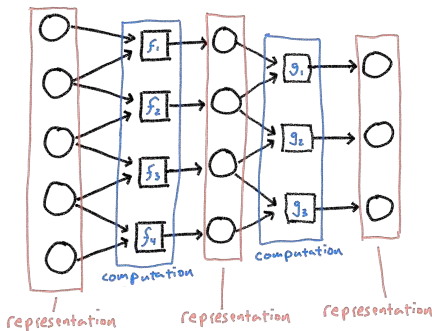


... infer 2D functional space.

## Picture-like representations in the cortex

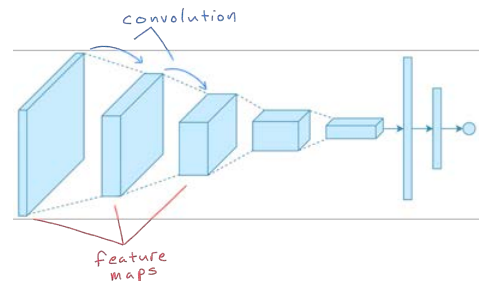
### Local vs global visual computation

- Local visual computation** minimizes the role of 2D layout.
  - E.g. Simple and complex cells (Hubel and Wiesel 1968)
- Global visual computations** provides a role for 2D representational layout.
  - Visual field maps as global representations (Wandell et al. 2007)



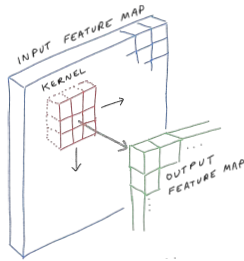
### Local vs global visual computation

- Local visual computation** minimizes the role of 2D layout.
  - E.g. Simple and complex cells (Hubel and Wiesel 1968)
- Global visual computations** provides a role for 2D representational layout.
  - Visual field maps as global representations (Wandell et al. 2007)
- Convolution models** of vision as global visual computations.

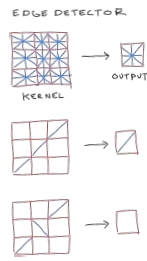


### What is convolution?

- Convolution is an algorithm that
  - reads an **input feature map** and writes an **output feature map**
  - by applying a single **spatially local** non-linear transformations **uniformly** across the input map
  - to generate output values **at the same location** in the output map. (Prince 2023)



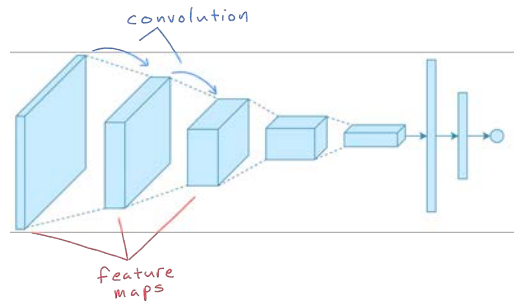
Convolution as a spatial algorithm.



A kernel trained for edge detection.

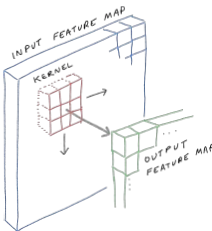
### What is convolution?

- A **convolutional neural network (CNN)** iterates convolution over a hierarchy of feature maps (Fukushima 1980, LeCun et al. 1989, Krizhevsky et al. 2013)

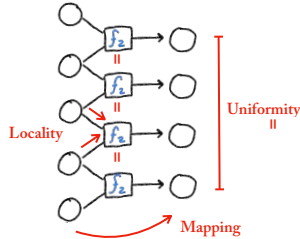


### The role of space in convolutional algorithms

- Convolution makes essential use of the 2D metric of its inputs and outputs.
- Convolution applies non-linear transformations to
  - spatially local** values
  - uniformly** across the input map
  - to generate values **at the same location** in the output map

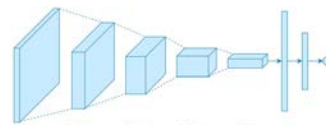


- Locality:** kernels only take inputs from nearby nodes.
- Uniformity:** the same kernel is applied uniformly across space.
- Mapping:** kernels map input values to the "same location" on the output map.

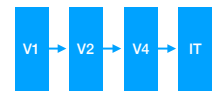


### Convolutional models of visual processing

- Convolutional models** of ventral stream vision:
  - Explicitly **inspired by V1** architecture. (Fukushima 1980)
  - Achieve **human-level categorization** accuracy. (Krizhevsky et al 2019)
  - Recreate **size and abstraction hierarchy**. (Zeiler and Fergus 2013)
  - Predict **contour maps** organization. (Margalit et al 2023)
  - Activity in early/middle/late stages of CNNs **predict neural activity** in early/middle/late stages of ventral pathway. (Yamins and DiCarlo 2016, Zhuang et al. 2021)
- Still **far** from complete. (Bowers 2023)



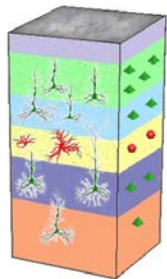
Hierarchical convolutional neural net



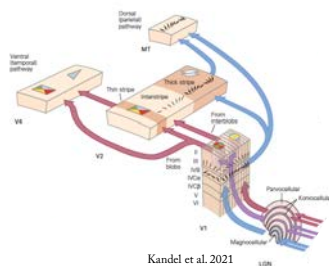
Ventral pathway

### Convolutional models of visual processing

- Taking the models **literally**: layer-to-layer computations are convolutions.
  - Intra-area:** layer 4 → layer 2/3. **Inter-area:** V1 → V2.



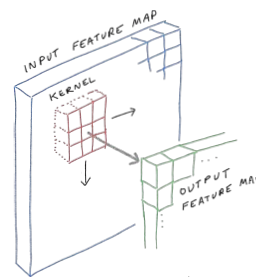
Intra-area computations



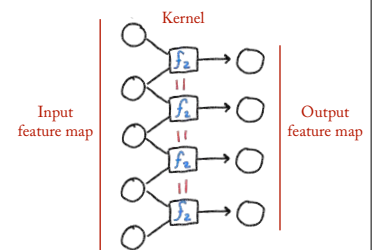
Inter-area computations

### Convolutional models of visual processing

- Taking the models **literally**: layer-to-layer computations are convolutions.
  - Intra-area:** layer 4 → layer 2/3. **Inter-area:** V1 → V2.
  - Serial to parallel** implementation.



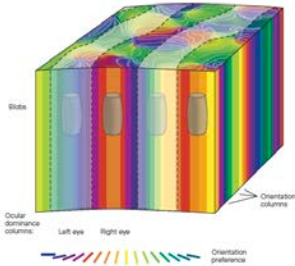
Serial read/write implementation



Parallel circuit implementation

### Convolutional models of visual processing

- Taking the models **literally**: layer-to-layer computations are convolutions.
  - **Intra-area**: layer 4 → layer 2/3. **Inter-area**: V1 → V2.
  - **Serial to parallel** implementation.
  - The **kernel** = the **contour pinwheel (hypercolumn)**.



Kandel et al. 2021

### Convolutional models of visual processing

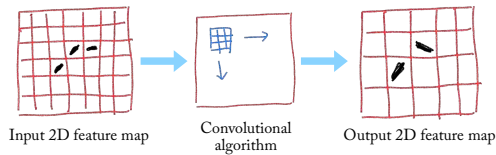
- Taking the models **literally**: layer-to-layer computations are convolutions.
  - **Intra-area**: layer 4 → layer 2/3. **Inter-area**: V1 → V2.
  - **Serial to parallel** implementation.
  - The **kernel** = the **contour pinwheel (hypercolumn)**.
  - The **tiling pattern** of the kernel = **mosaic** of contour pinwheels. (Presslof & Cowan 2003)



Bosking et al. 1997

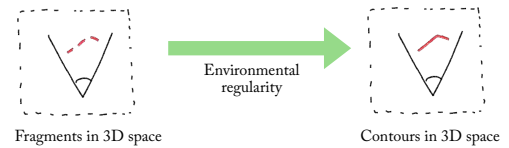
### Why does convolution “work”?

- Why is convolution a **reliable algorithm** for extracting visual features?
- For an algorithm to be **reliable**, there must be...
  - (1) A **environmental regularity** that it is tracking.
  - (2) A **systematic relation** between **inputs/outputs** of the algorithm and **inputs/outputs** of the environmental regularity. (“**Semantics**”)

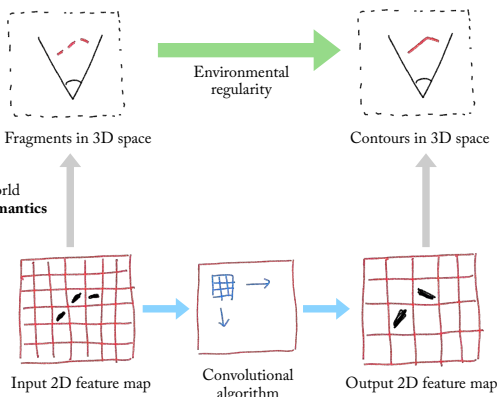


### Why does convolution “work”?

- Why is convolution a **reliable algorithm** for extracting visual features?
- For an algorithm to be **reliable**, there must be...
  - (1) A **environmental regularity** that it is tracking.
  - (2) A **systematic relation** between **inputs/outputs** of the algorithm and **inputs/outputs** of the environmental regularity. (“**Semantics**”)



Brain-world  
Relations: **semantics**



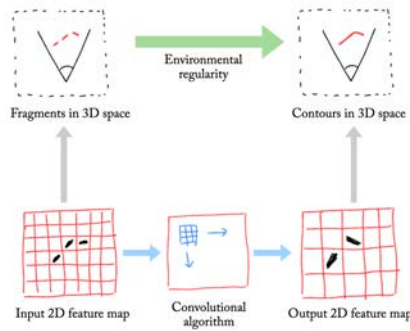
### Why does convolution “work”?

- Why is convolution a **reliable algorithm** for extracting visual features?
- In **vision**, the environmental regularities often relate
  - **spatially local features** in 3D-space
  - to more **complex features at the same location** in 3D space.
- Examples:
  - **Fragments > contours** (Palmer 1977, Lande 2023)
  - Face parts > face (Tsao & Livingston 2008)



### Why does convolution “work”?

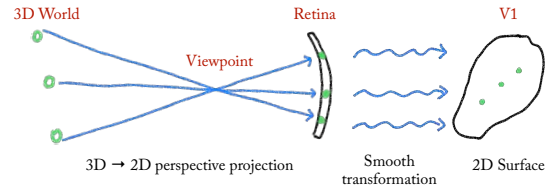
- In vision, the environmental regularities often relate:
  - spatially local features in 3D-space
  - to more complex features at the same location in 3D space.
- Visual convolution maps:
  - spatially local values in 2D input feature maps
  - to values at the same location in a 2D output feature maps.



- For convolution to be reliable:
  - There must be a **systematic mapping** between **2D locations** on the feature map locations and **3D locations** in the environment.

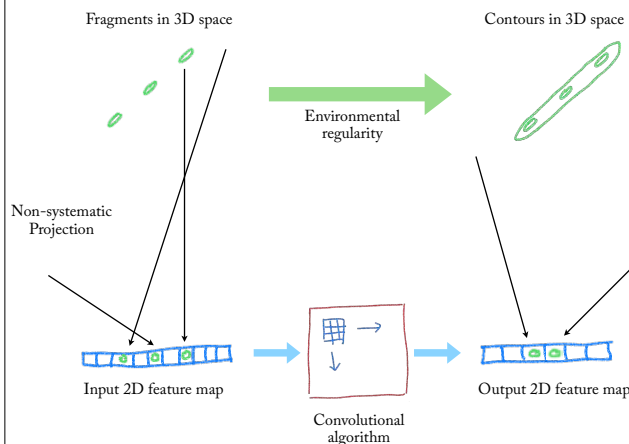
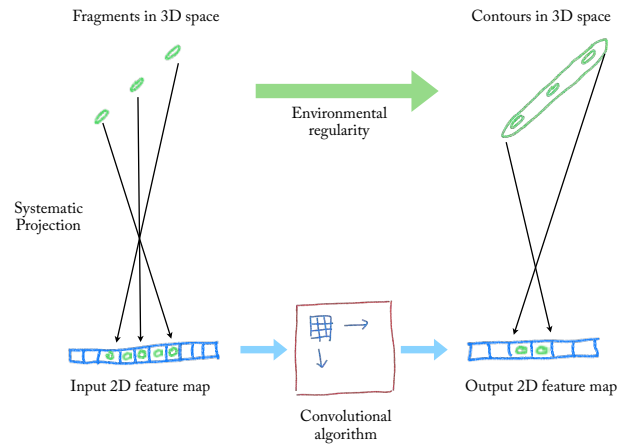
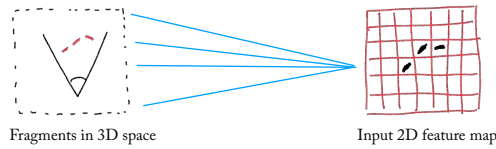
### Projection semantics

- Systematic 3D-2D relations are provided by
  - perspective projection
  - smooth transformation
- This is a **projection semantics** for retinotopic representations (a theory of representation-environment relations).
  - Computer vision models trained on photographs build in projection semantics from the ground.
- Extending **semantics** to vision (Lande 2023).



### Projection semantics and convolution

- Projection semantics** makes **reliable convolution** possible.
- For many features F, G related by environmental regularities:
  - If Fs **systematically project** to visual field map V, then a convolution of V can **reliably extract** representations of G's.
- For many features F, G related by environmental regularities:
  - If Fs **do not** systematically project to visual field map V, then a convolution of V **cannot** reliably extract representations of G's.



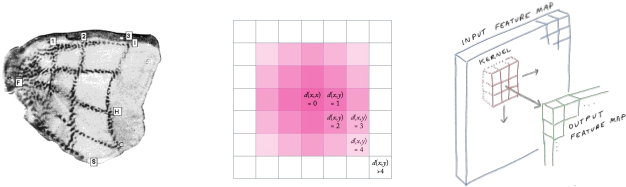
Data A					2x2 kernel		Data A						
								A	2	3	90	.	B
								3	4	65	.	C	4
				45				5	45	.	D	5	4
								102	.	E	6	3	59
								90	51	87	62	59	.
								.	F	5	3	62	.
								G	4	3	87	.	H
								3	3	51	.		

**Picture-like** encoding of visual space supports convolution.

**Language-like** encoding of visual space does not support convolution.

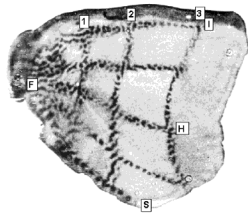
## Summary

- Retinotopy implies the existence of **functional 2D visual field maps**.
- If layer-to-layer visual computations are **convolutional**, then 2D visual field maps must stand in functional relations of perspective projection to 3D visual space.
- **Pictorial hypothesis**: there are computationally functional picture-like representations in retinotopic areas.



## Pictures in the brain?

- **Retinotopic representations differ from pictures** because one derives its content by being viewed, while the other does not.
- But (arguably) the essence of pictorial representation is a semantics that encodes 3D visual content in 2D signs **via perspective projection from a viewpoint**.
- This turns out to be an **extremely efficient** way of storing 3D information, and 2D surfaces are relatively easy to come by.
- Retinotopic areas store the same kind of 3D information in 2D surfaces, via the same kinds of sign-to-world relations.
- So retinotopy really does mean that **there are pictures in the brain**.



Thank you!

For invaluable discussion, thanks to Dave Barker-Plummer, Jacob Beck, Ned Block, Rosa Cao, Susan Carey, Judy Fan, Aaron Hertzmann, Kevin Lande, Elvie Lin, Eric Mandelbaum, Jake Quilty-Dunn, Gabe Rabin, Dan Yamins, and audiences at NYU and Stanford.