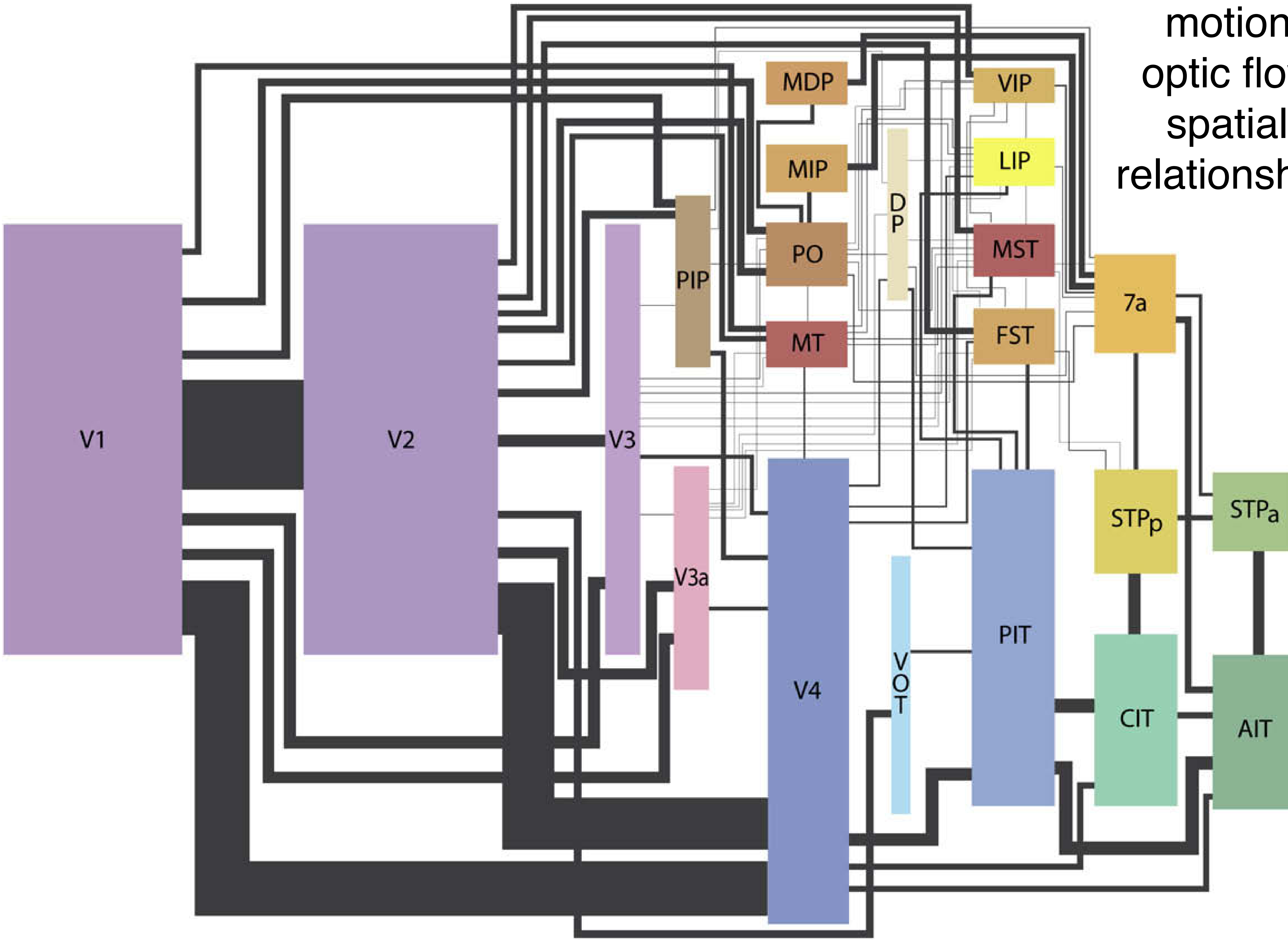
Visual scene analysis

Wallisch & Movshon (2008)

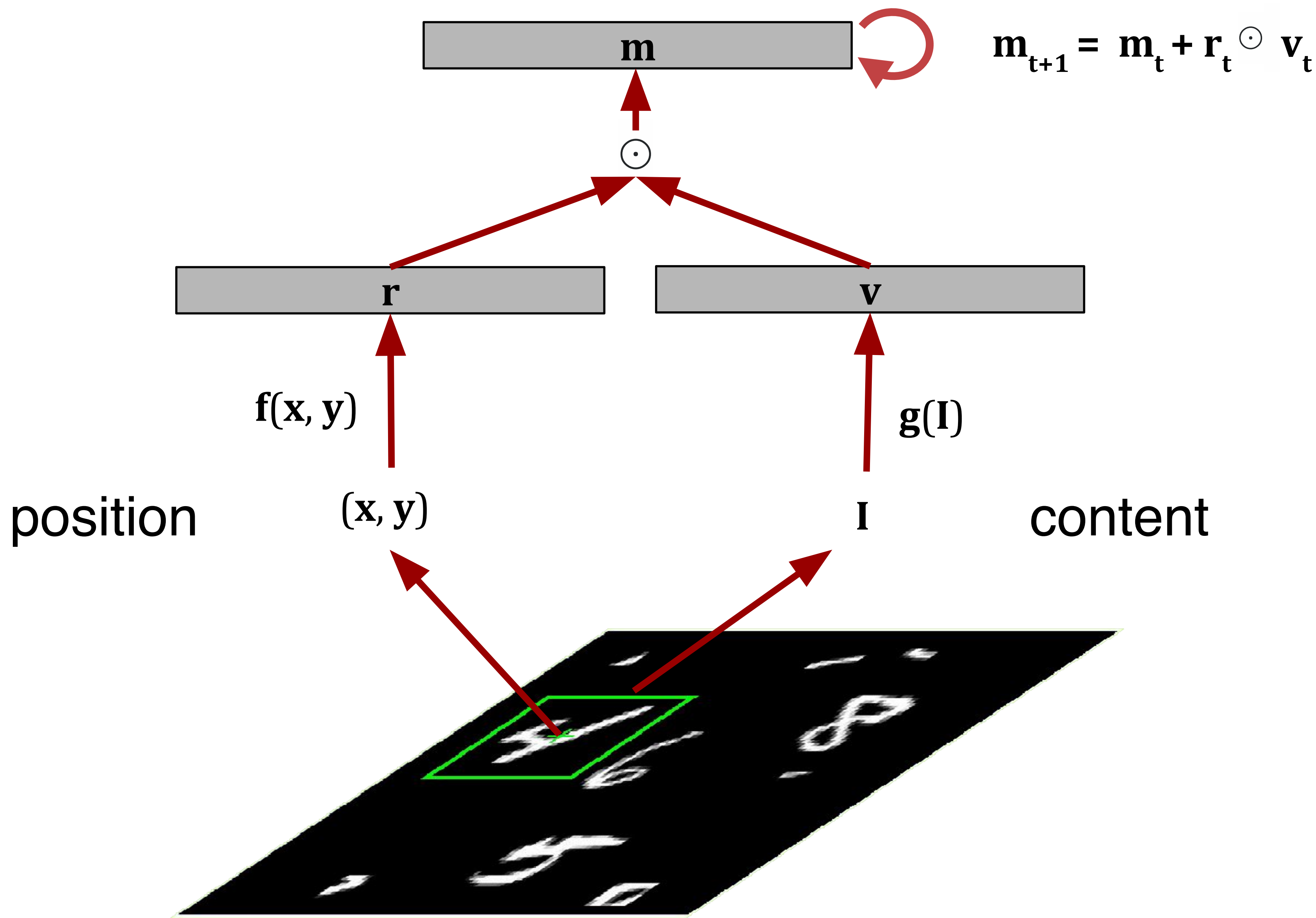
motion,
optic flow,
spatial-
relationships



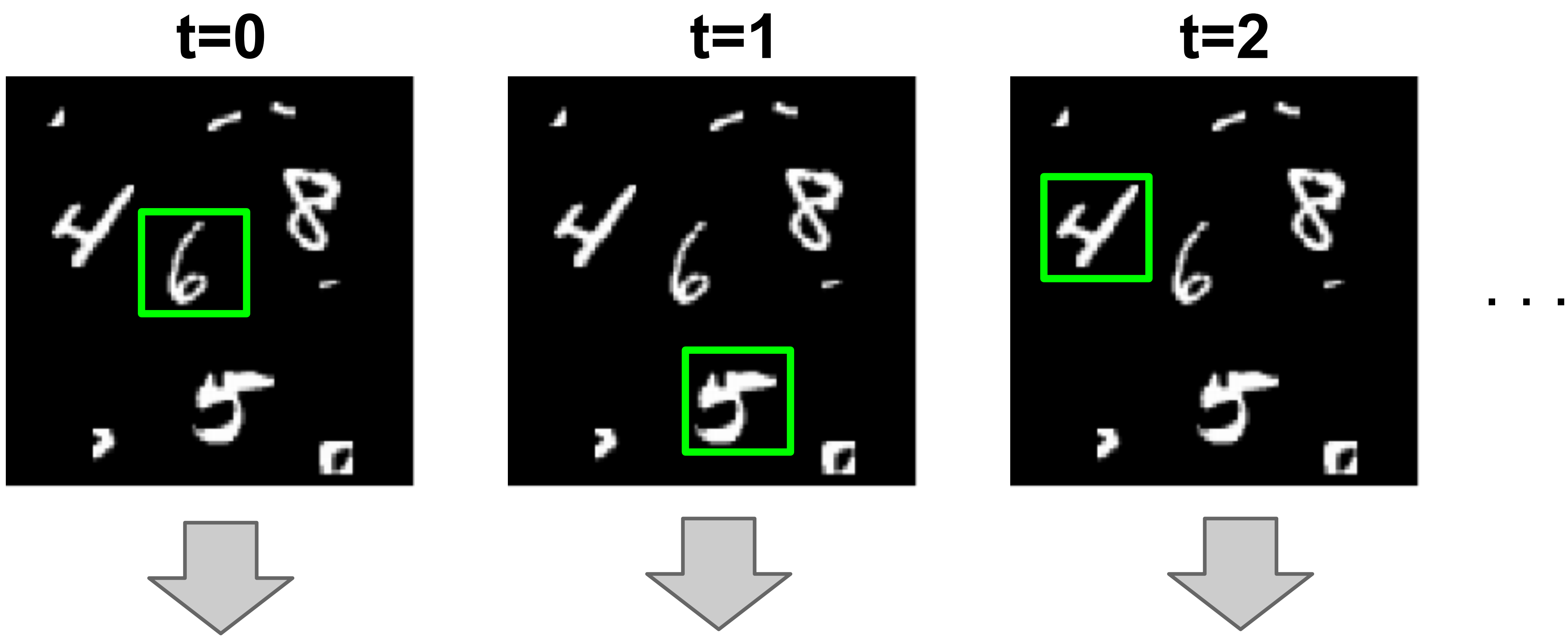
lines, edges surface boundaries ? objects faces

Visual working memory as a superposition of 'what' and 'where' bindings

(Eric Weiss, Ph.D. thesis)



Example encoding



$$\mathbf{m} = \mathbf{v}_6 \odot \mathbf{r}_{t=0} + \mathbf{v}_5 \odot \mathbf{r}_{t=1} + \mathbf{v}_4 \odot \mathbf{r}_{t=2} + \ldots$$

Example queries

Where is the '5'?

$$\text{answer} = \mathbf{v}_5^* \odot \mathbf{m}$$

$$= \mathbf{v}_5^* \odot (\mathbf{v}_6 \odot \mathbf{r}_{t=0} + \mathbf{v}_5 \odot \mathbf{r}_{t=1} + \mathbf{v}_4 \odot \mathbf{r}_{t=2} + \ldots)$$

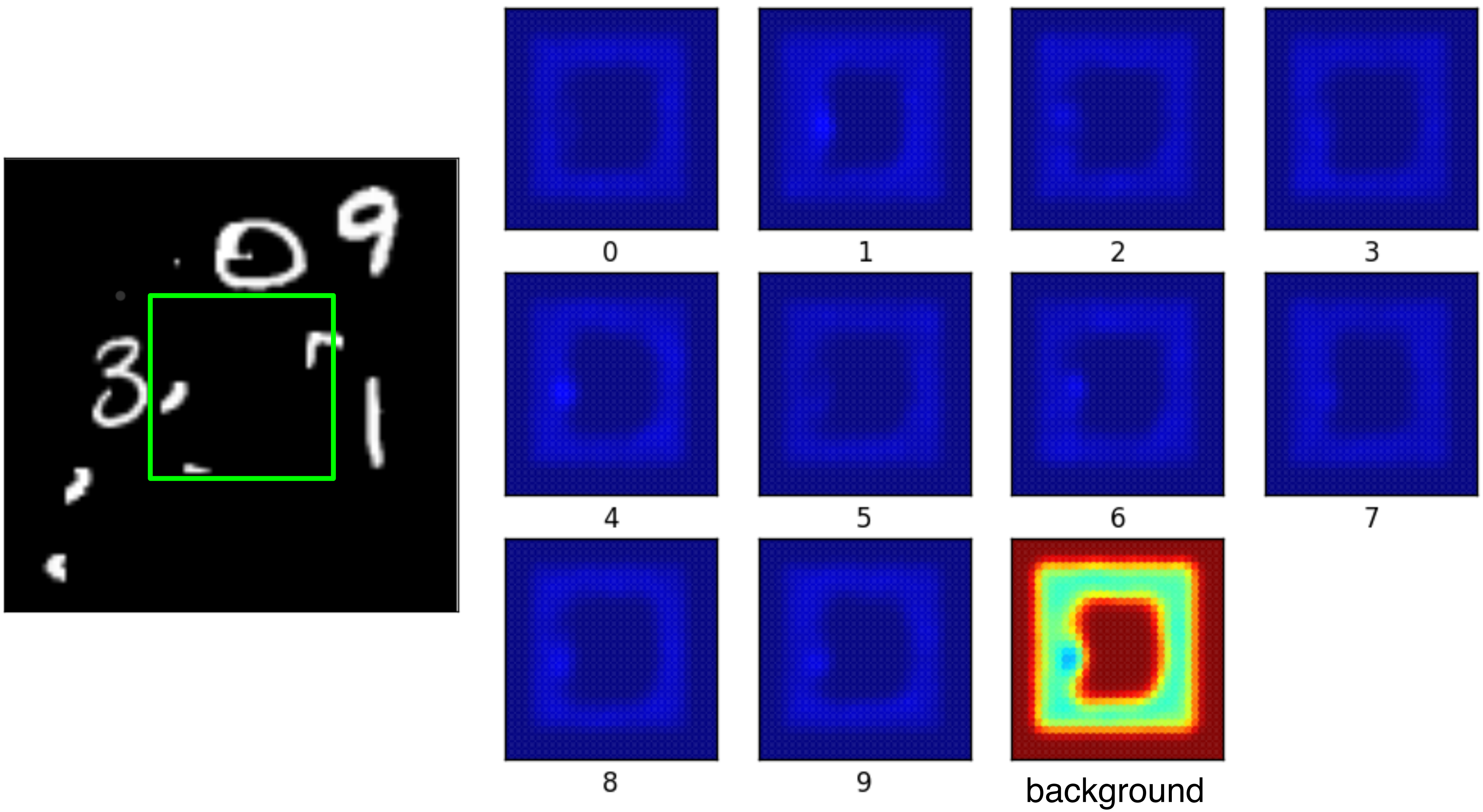
$$\approx \quad \text{noise} \quad + \quad \mathbf{r}_{t=1} \quad + \quad \text{noise} \quad + \ldots$$

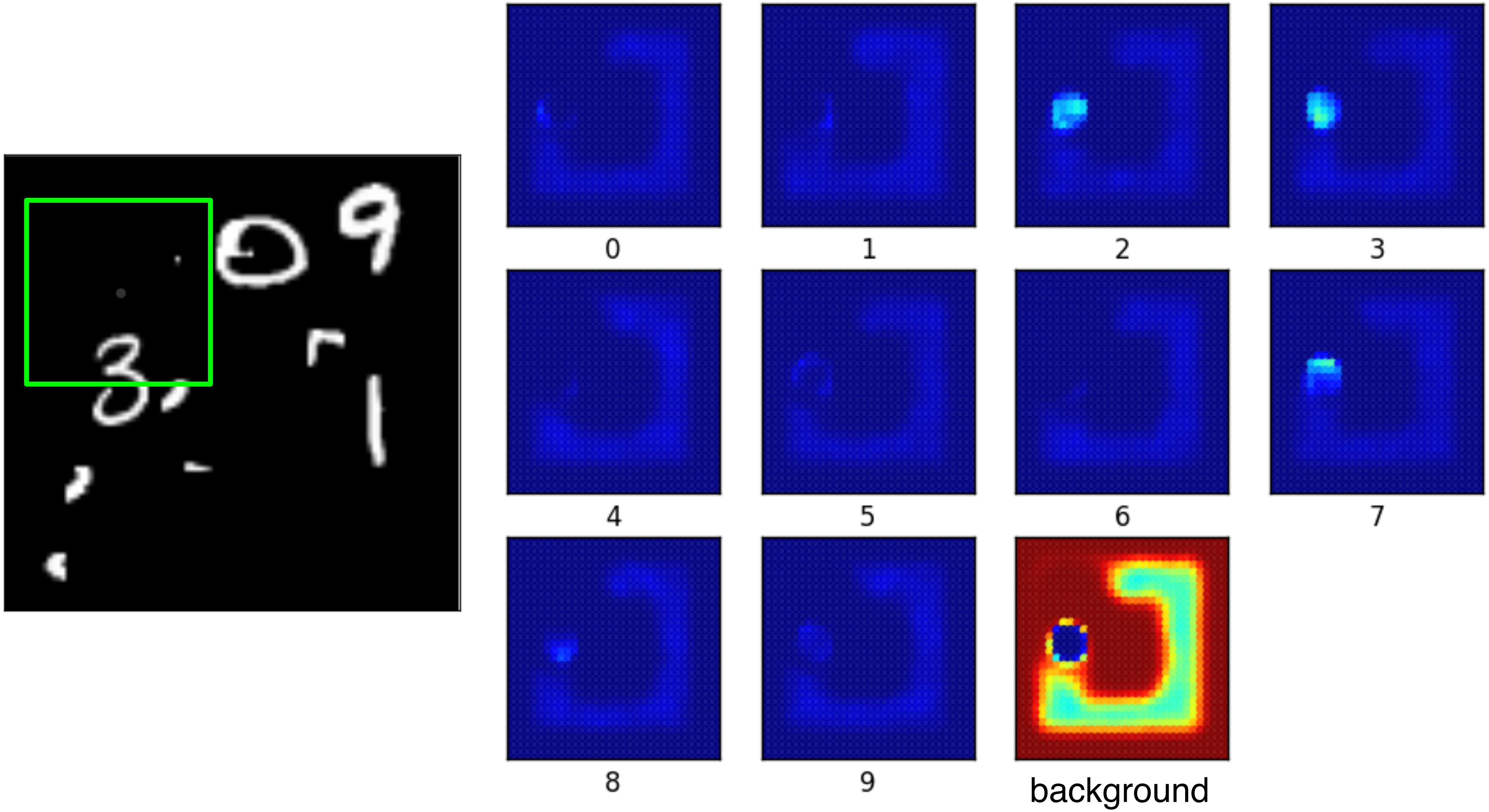
What object is in the center?

$$\text{answer} = \mathbf{r}_{center}^* \odot \mathbf{m}$$

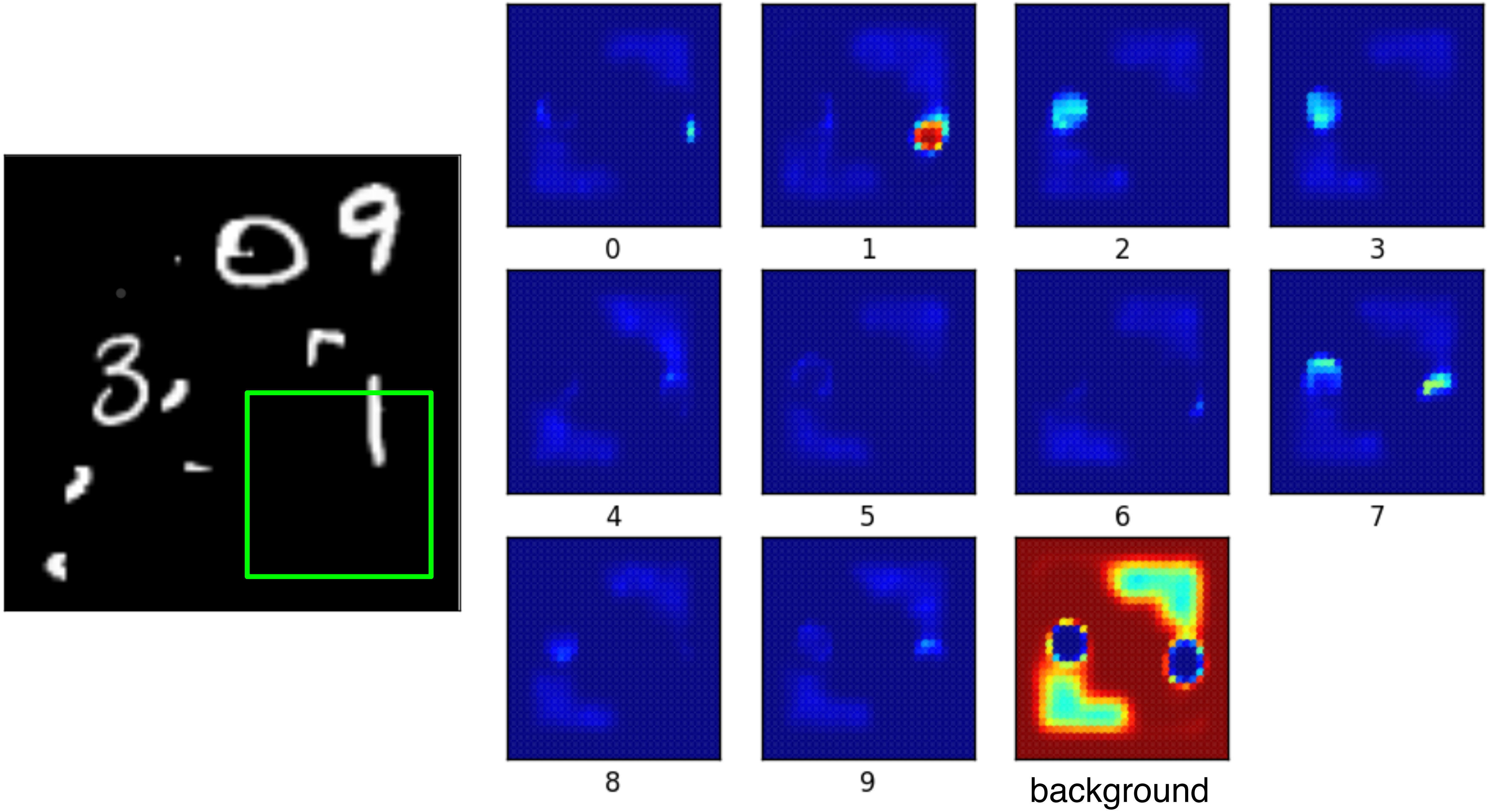
$$= \mathbf{r}_{center}^* \odot (\mathbf{v}_6 \odot \mathbf{r}_{t=0} + \mathbf{v}_5 \odot \mathbf{r}_{t=1} + \mathbf{v}_4 \odot \mathbf{r}_{t=2} + \ldots)$$

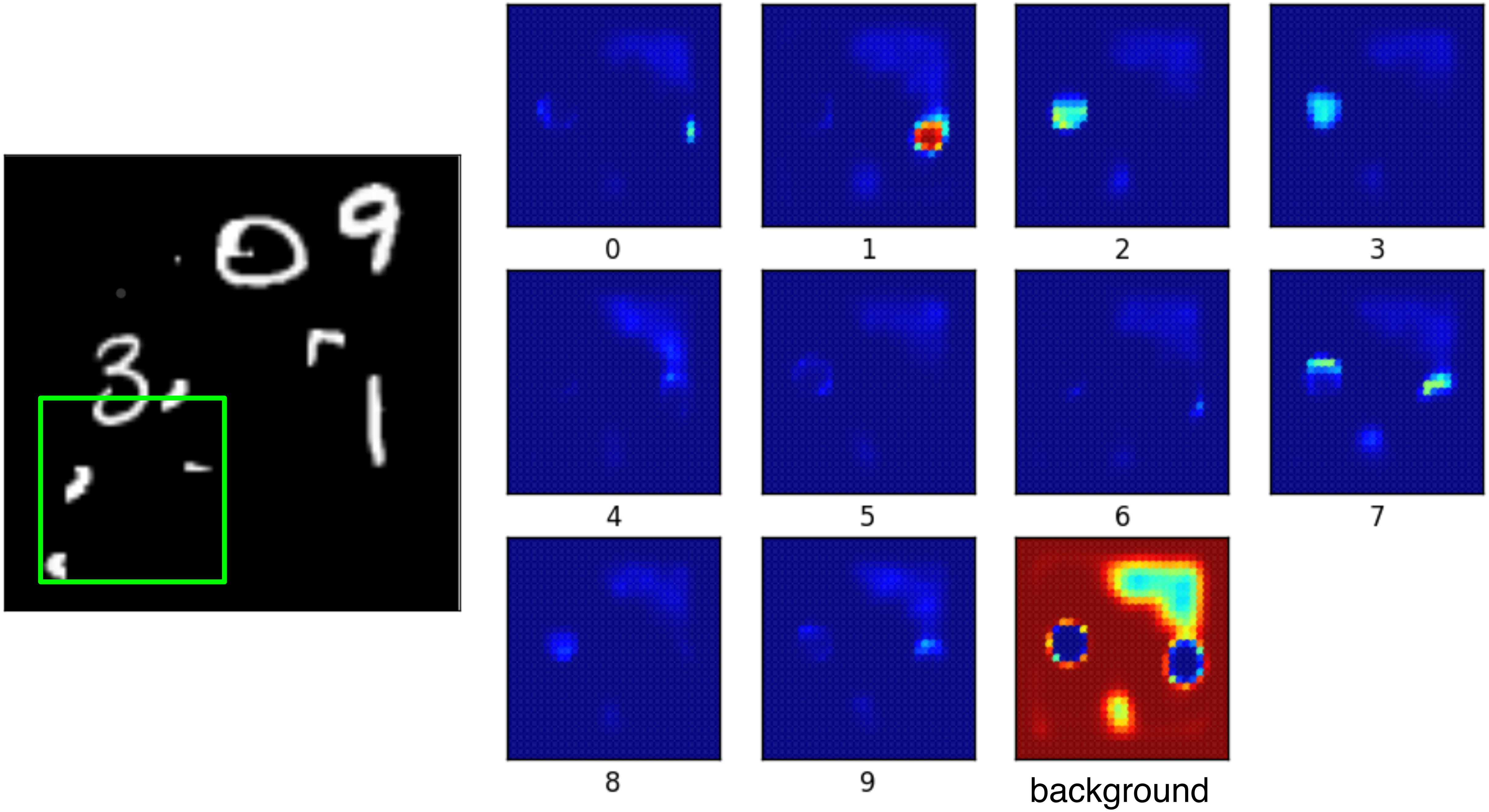
$$\approx \quad \mathbf{v}_6 \quad + \quad \text{noise} \quad + \quad \text{noise} \quad + \ldots$$

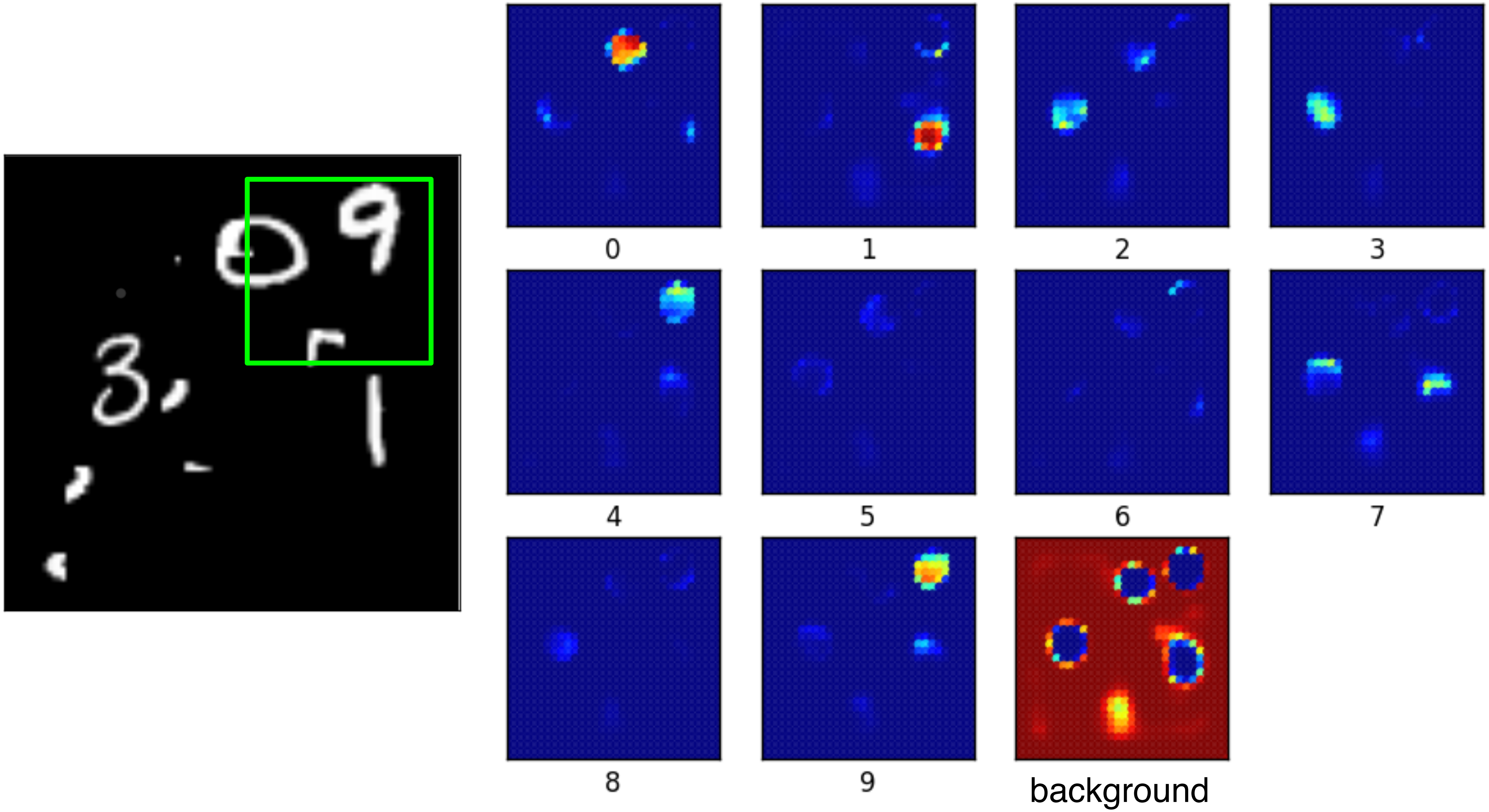

0
1
2
3
4
5
6
7
8
9
background

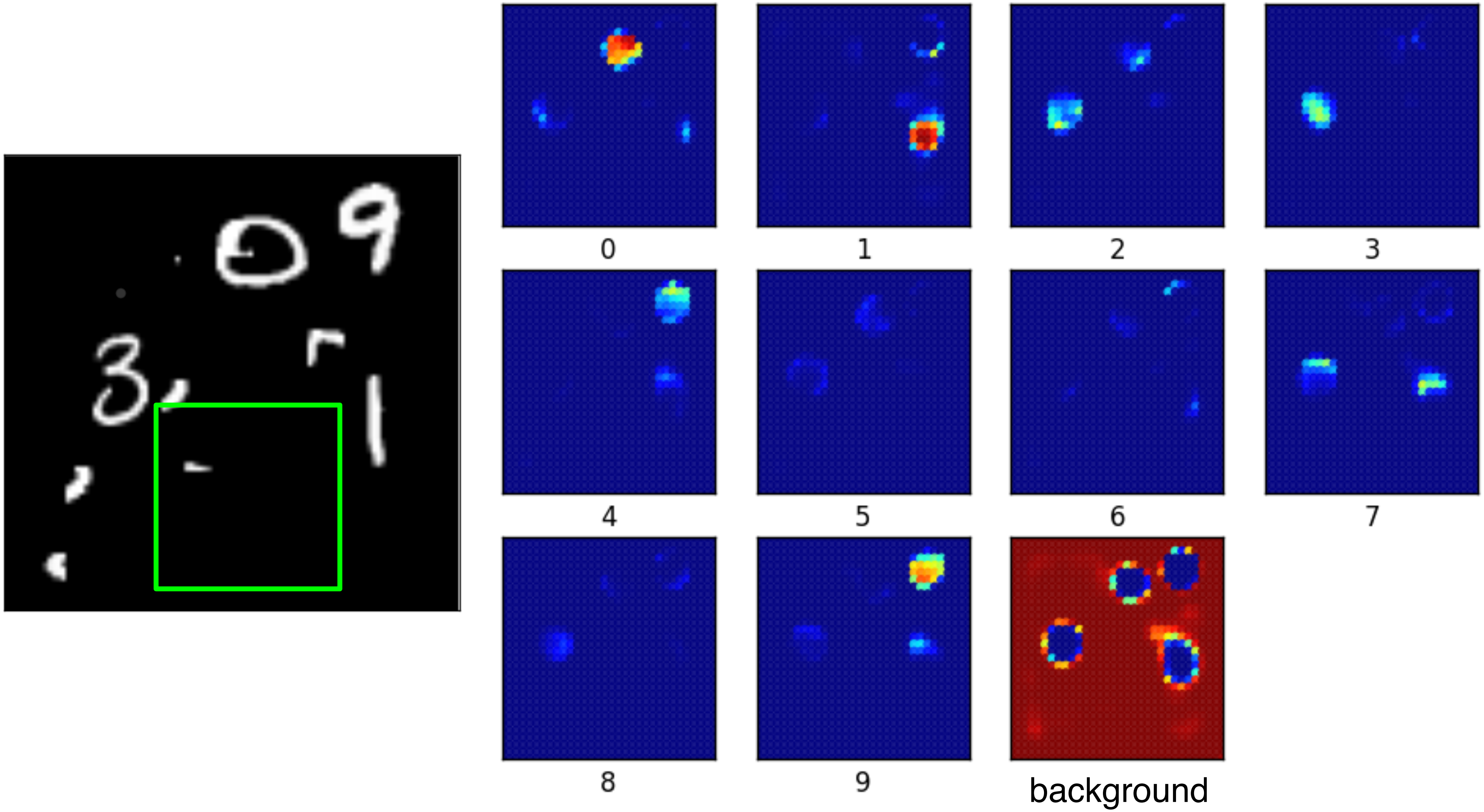


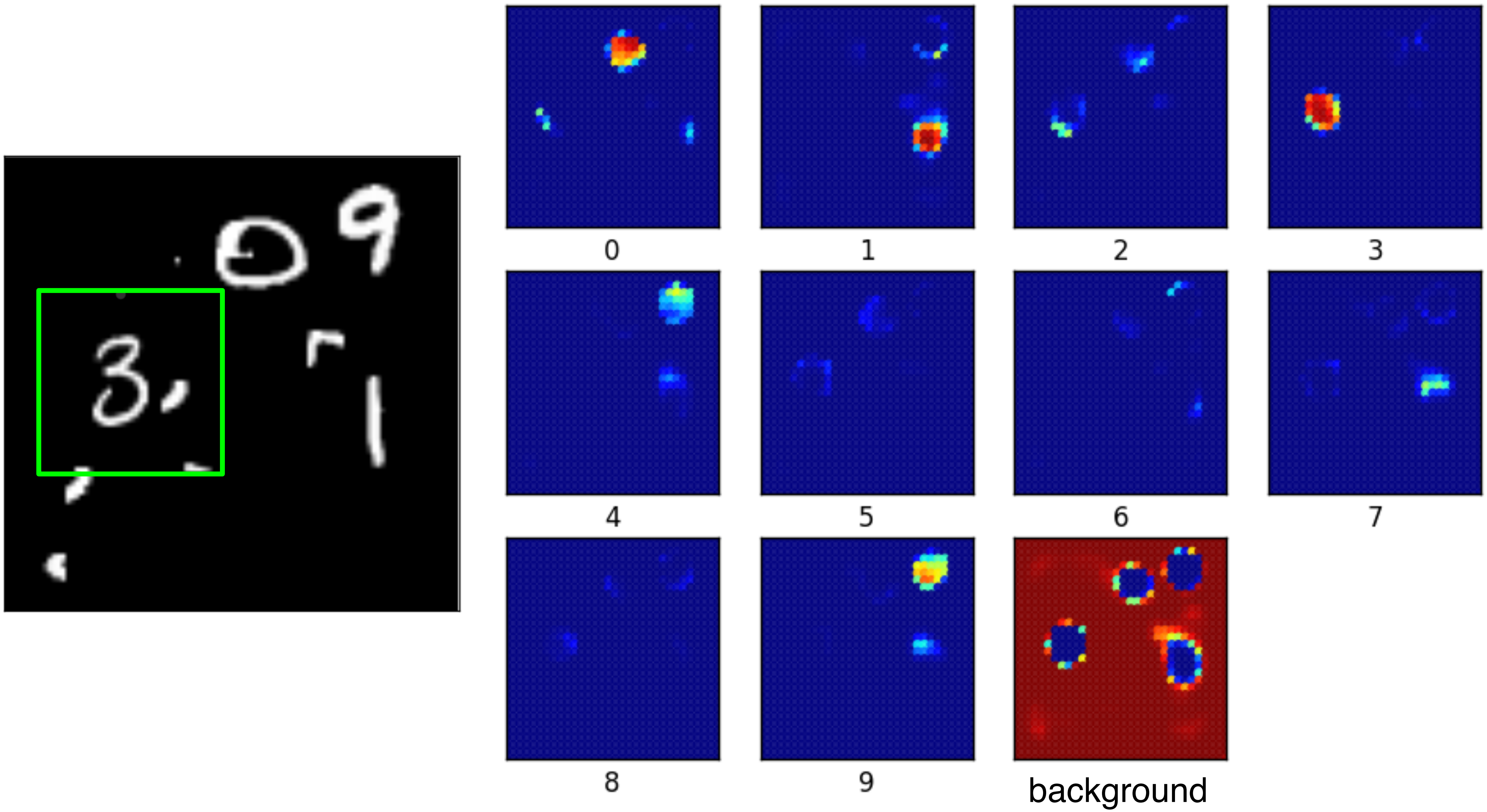
0
1
2
3
4
5
6
7
8
9
background

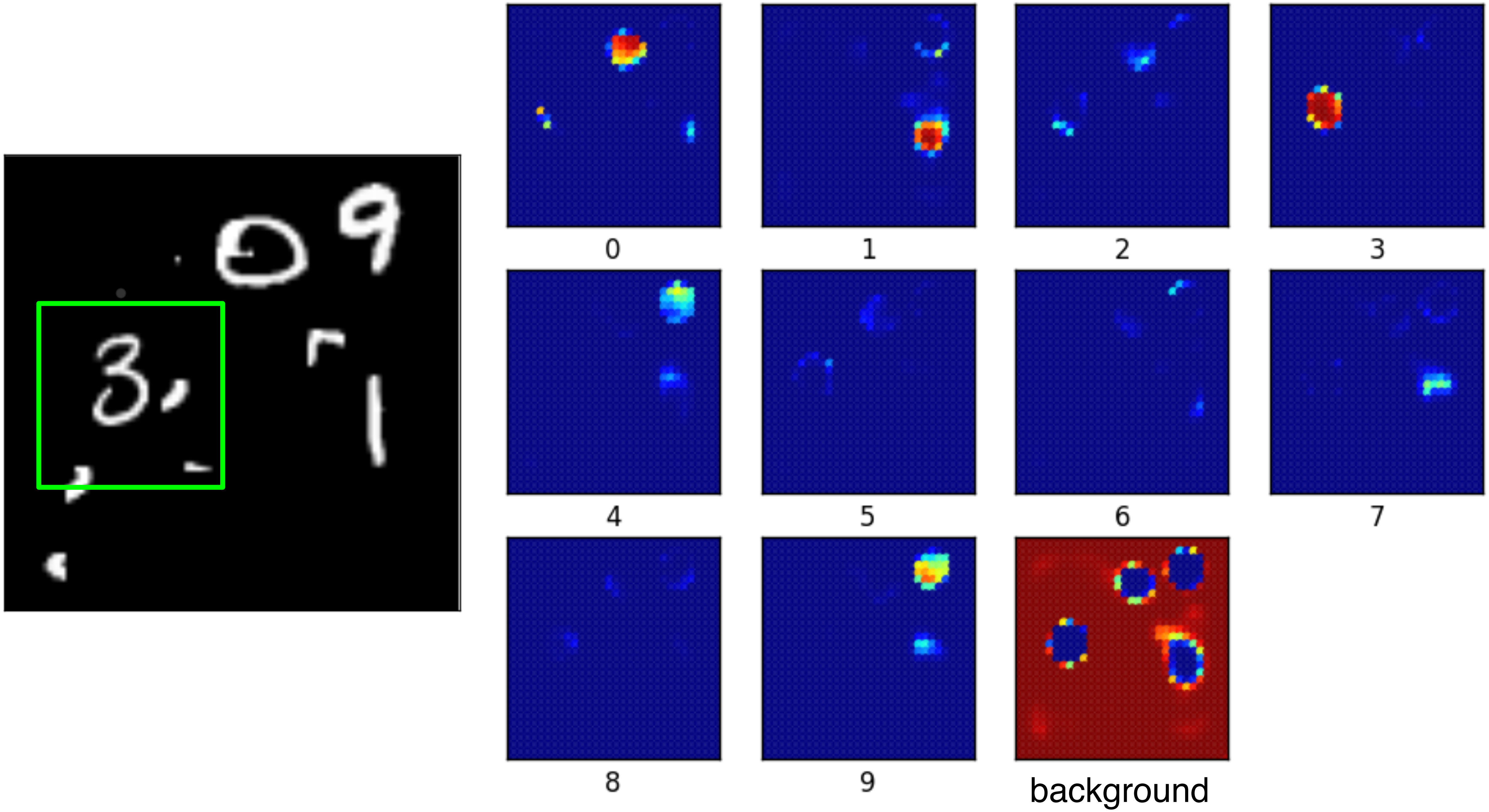

0
1
2
3
4
5
6
7
8
9
background


0
1
2
3
4
5
6
7
8
9
background

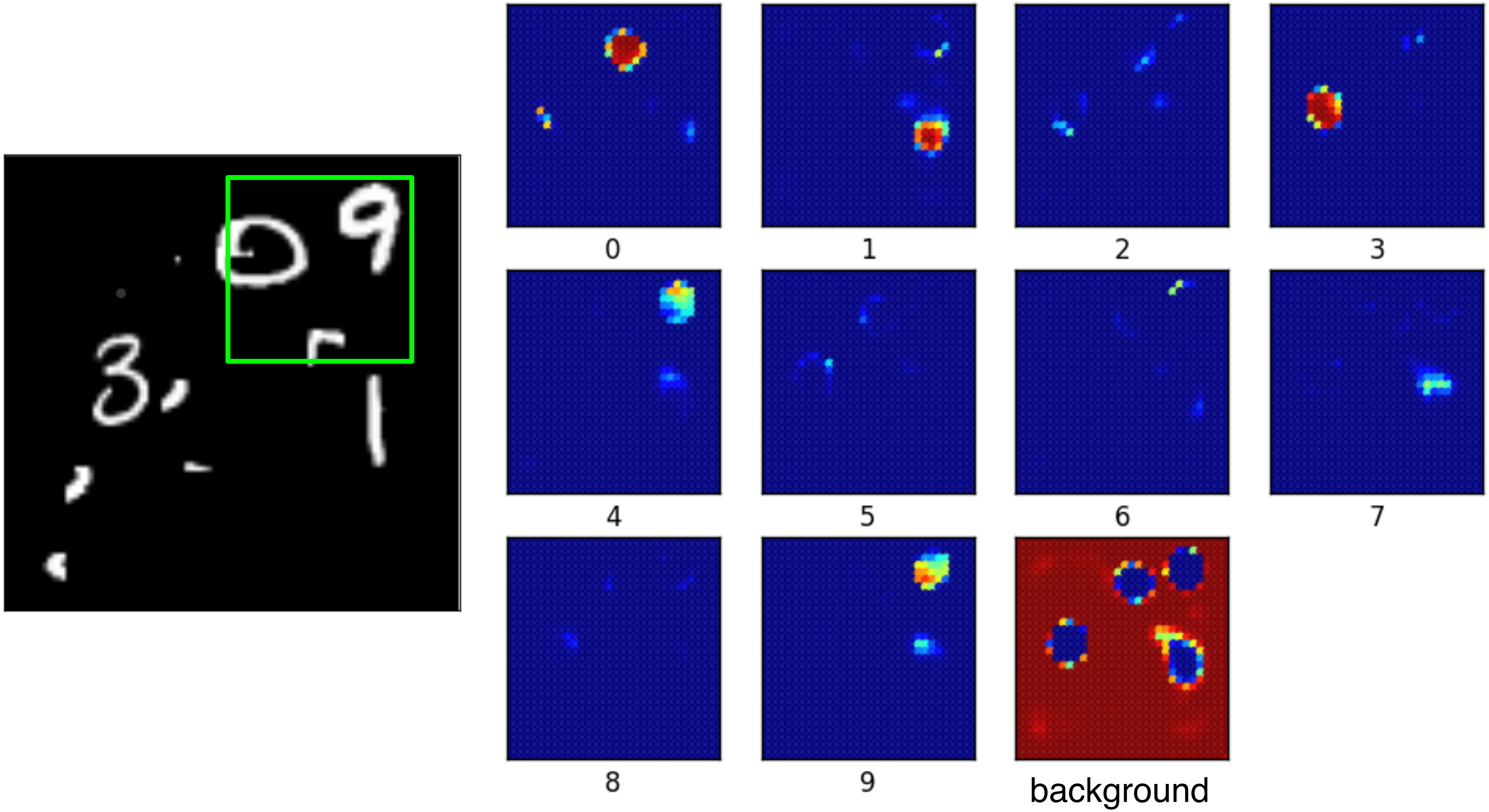

0
1
2
3
4
5
6
7
8
9
background

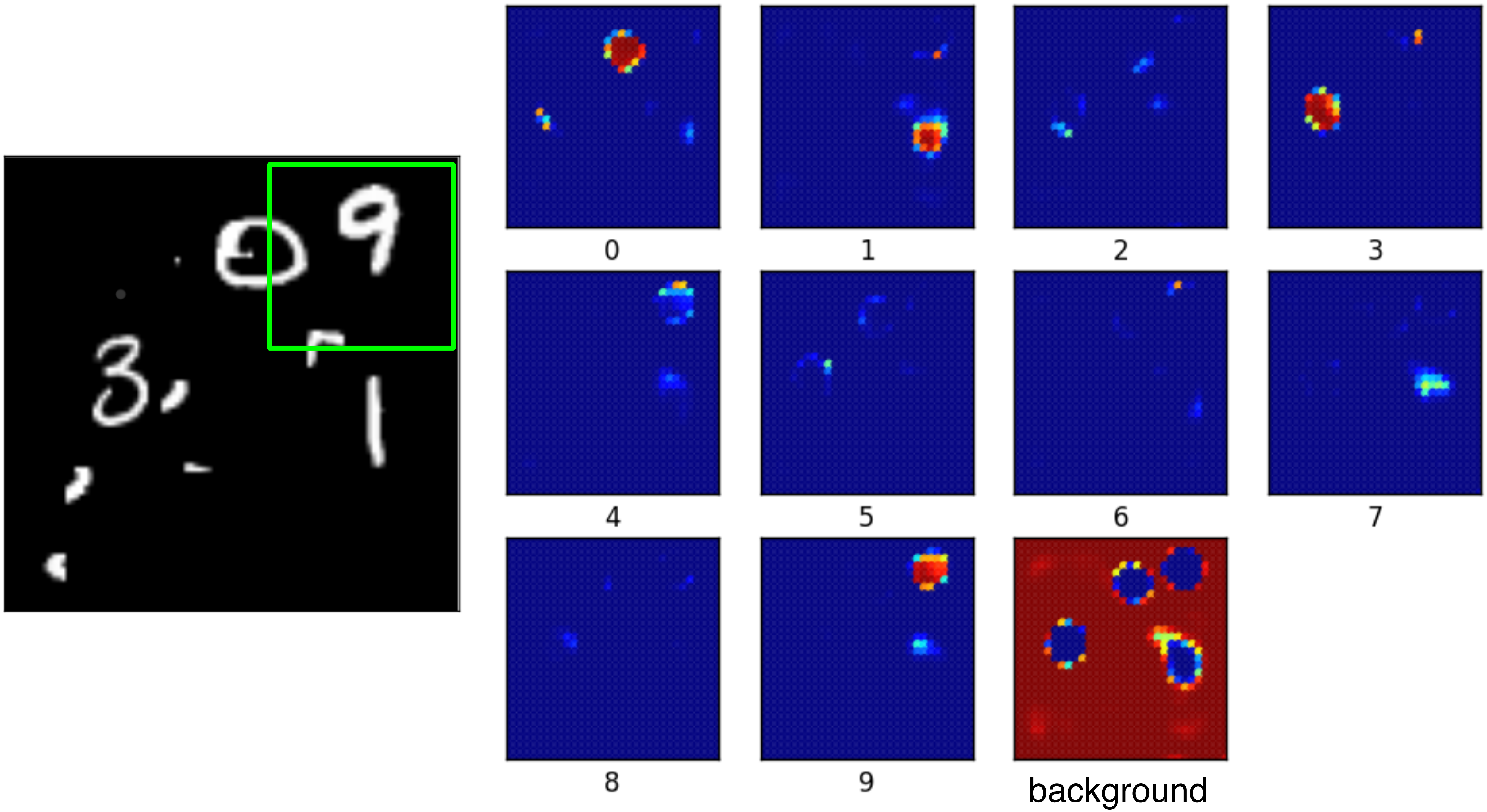

0
1
2
3
4
5
6
7
8
9
background


0
1
2
3
4
5
6
7
8
9
background



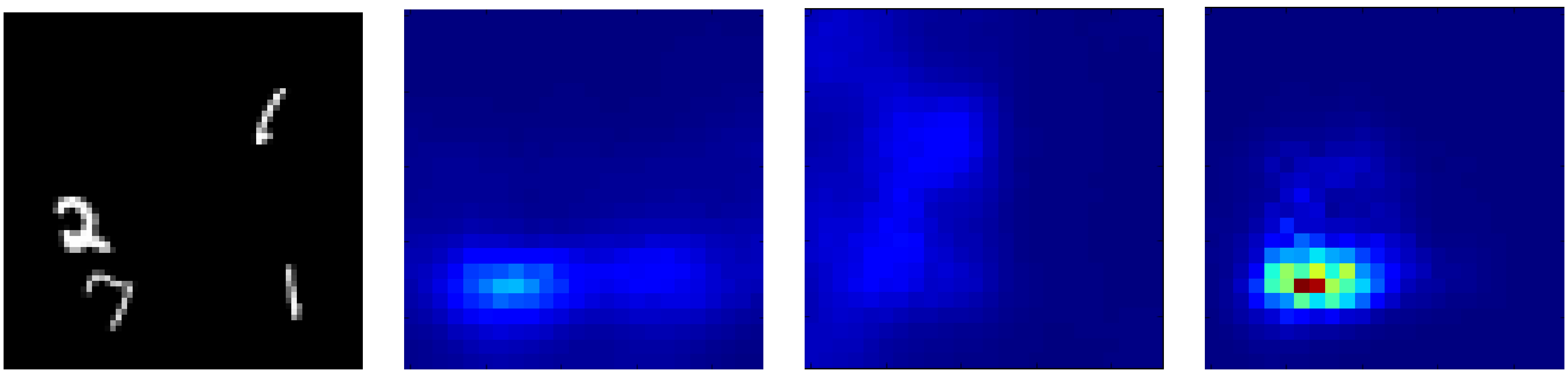
0
1
2
3
4
5
6
7
8
9
background


0
1
2
3
4
5
6
7
8
9
background


0
1
2
3
4
5
6
7
8
9
background

Spatial reasoning

What is below a ‘2’ and to the left of a ‘1’?



(a) Example image (b) ”below a 2” (c) ”to the left of a 1” (d) Combined

$$\mathbf{a}_1 = f^{-1}(\mathbf{r}_{\mathbf{down}}\,(\mathbf{v}_\mathbf{2}^{*} \odot \mathbf{m}))$$

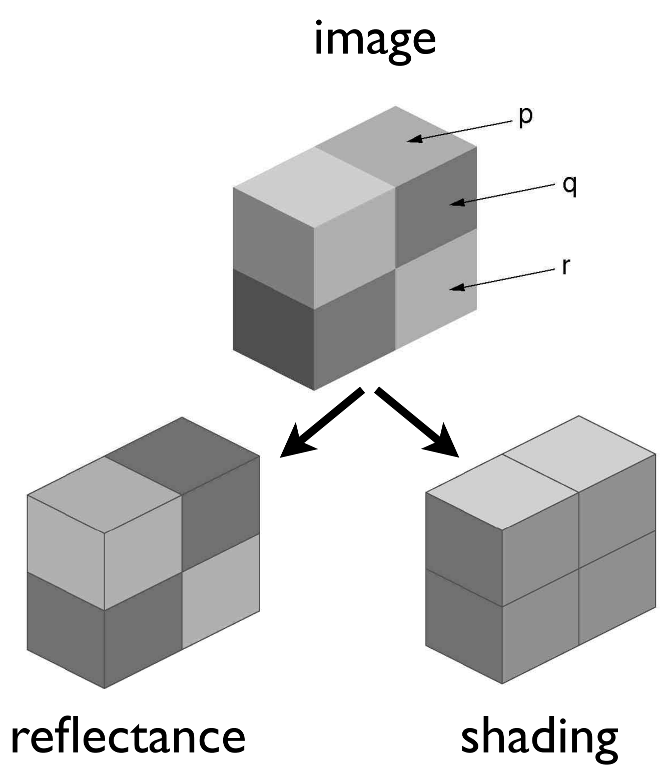
$$\mathbf{a}_2 = f^{-1}(\mathbf{r}_{\mathbf{left}}\,(\mathbf{v}_\mathbf{1}^{*} \odot \mathbf{m}))$$

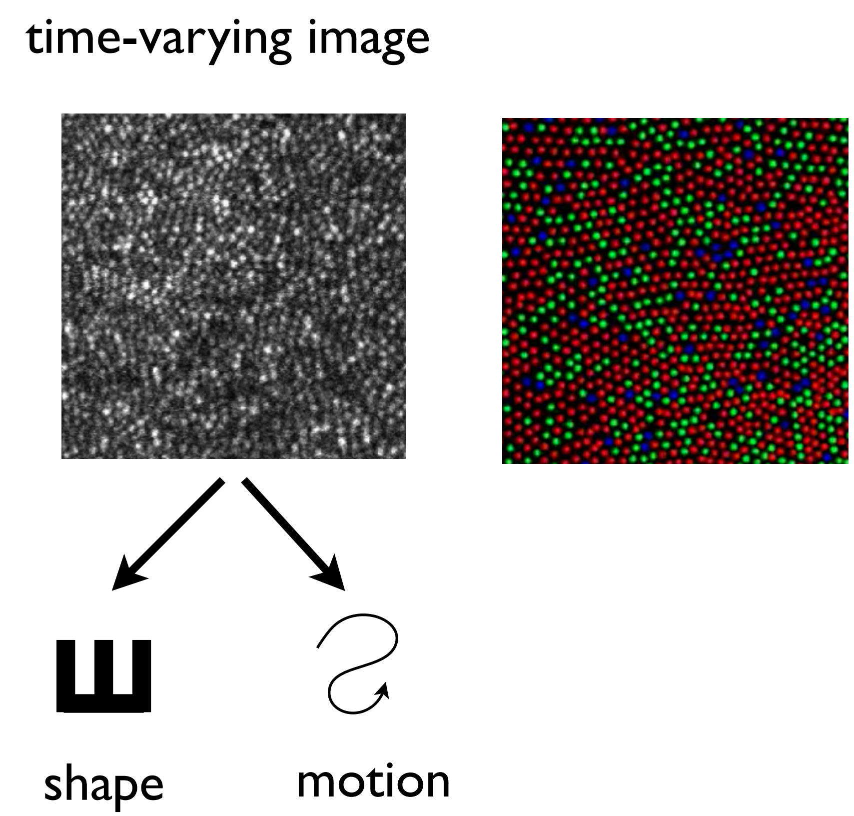
$$\mathbf{a}_1 \odot \mathbf{a}_2$$

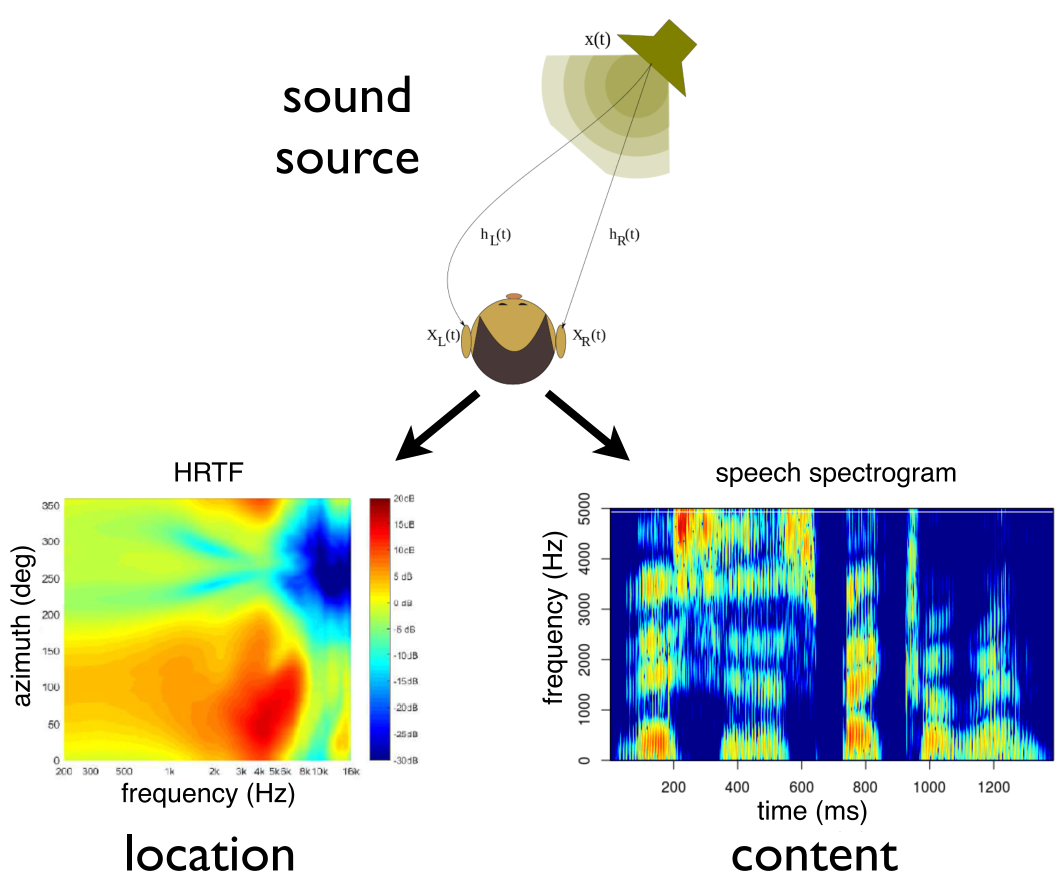
$$\text{answer} = f(\mathbf{a}_1 \odot \mathbf{a}_2) \odot \mathbf{m}$$

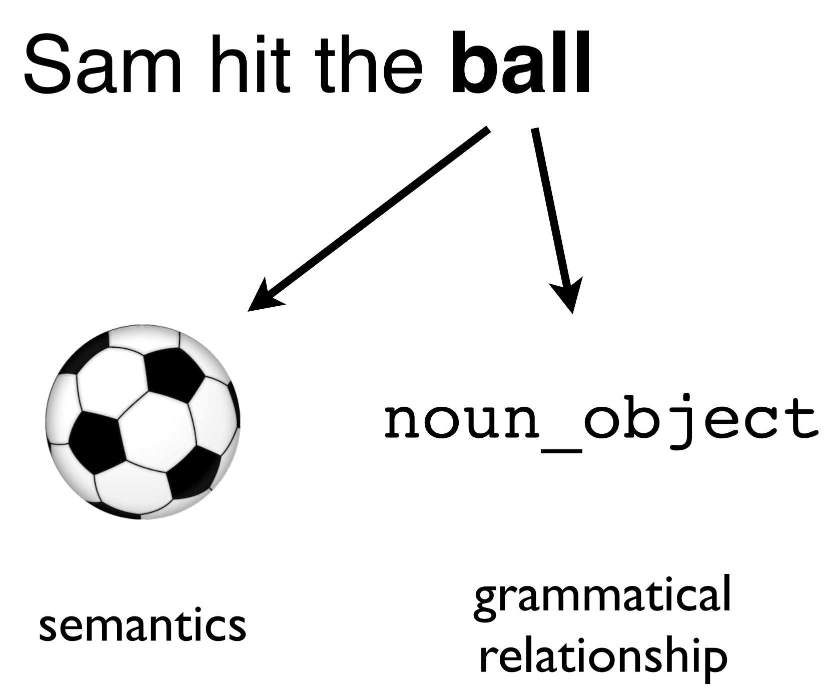
Factorization

Factorization is central to perception and cognition









Resonator Networks for factorizing HD vectors

Let $\mathbf{b} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$

$$\mathbf{x} \in \mathbb{X} := \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$\mathbf{y} \in \mathbb{Y} := \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_n\}$$
$$\mathbf{z} \in \mathbb{Z} := \{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n\}$$

Problem: You are given $\mathbf{b}$, what are $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$?

Solution: Resonate

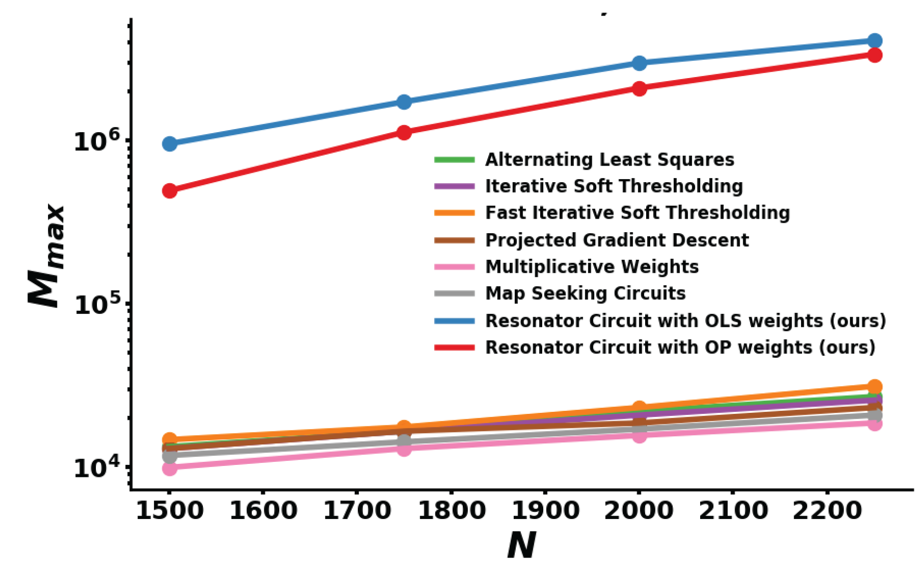
$$\hat{\mathbf{x}}_{t+1} = g\left(\mathbf{X}\mathbf{X}^\top(\mathbf{b} \otimes \hat{\mathbf{y}}_t^{-1} \otimes \hat{\mathbf{z}}_t^{-1})\right) \qquad \mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ | & | & & | \end{bmatrix}$$

$$\hat{\mathbf{y}}_{t+1} = g\left(\mathbf{Y}\mathbf{Y}^\top(\mathbf{b} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{z}}_t^{-1})\right) \qquad \mathbf{Y} = \begin{bmatrix} | & | & & | \\ \mathbf{y}_1 & \mathbf{y}_2 & \ldots & \mathbf{y}_n \\ | & | & & | \end{bmatrix}$$

$$\hat{\mathbf{z}}_{t+1} = g\left(\mathbf{Z}\mathbf{Z}^\top(\mathbf{b} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{y}}_t^{-1})\right) \qquad \mathbf{Z} = \begin{bmatrix} | & | & & | \\ \mathbf{z}_1 & \mathbf{z}_2 & \ldots & \mathbf{z}_n \\ | & | & & | \end{bmatrix}$$

$$g(x) = \mathrm{sgn}(x)$$

Combinatorial capacity exceeds competing methods by *two orders of magnitude*



Three factors, *F* = 3

Frady EP, Kent S, Olshausen BA & Sommer FT (2020) Resonator Networks for factoring distributed representations of data structures. *Neural Computation* (in press) https://arxiv.org/abs/2007.03748

Kent S, Frady EP, Sommer FT & Olshausen BA (2020) Resonator Networks outperform optimization methods at solving high-dimensional vector factorization. *Neural Computation* (in press) https://arxiv.org/abs/1906.11684

Energy function?

$$E = -\mathbf{b} \cdot \overbrace{(\mathbf{x} \bigotimes \mathbf{y} \bigotimes \mathbf{z})}^{(\alpha_1\beta_1\gamma_1\,\mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 \,+\, \ldots \,+\, \alpha_i\beta_j\gamma_k\,\mathbf{x}_i \otimes \mathbf{y}_j \otimes \mathbf{z}_k \,+\, \ldots \,+\, \alpha_n\beta_n\gamma_n\,\mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n)}$$

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i\,\mathbf{x}_i, \quad \mathbf{y} = \sum_{i=1}^{n} \beta_i\,\mathbf{y}_i, \quad \mathbf{z} = \sum_{i=1}^{n} \gamma_i\,\mathbf{z}_i$$

Energy function?

1,000,000 combinations! (n=100)

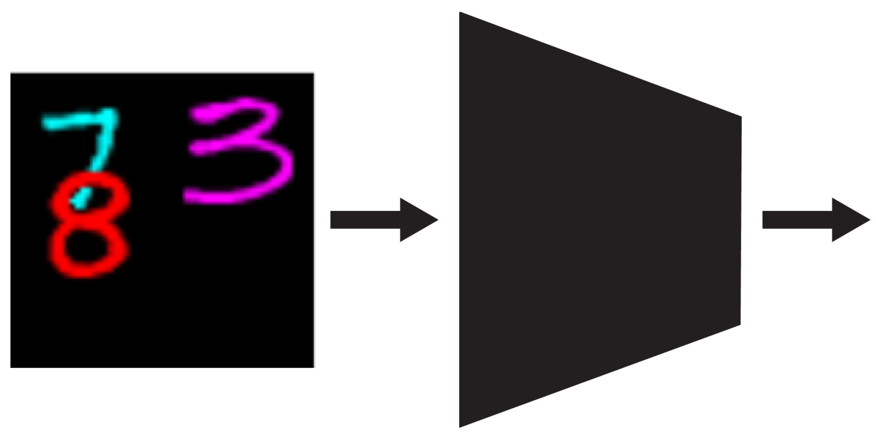
$$(\alpha_1\beta_1\gamma_1\,\mathbf{x}_1\otimes\mathbf{y}_1\otimes\mathbf{z}_1\;+\;\ldots\;+\;\alpha_i\beta_j\gamma_k\,\mathbf{x}_i\otimes\mathbf{y}_j\otimes\mathbf{z}_k\;+\;\ldots\;+\;\alpha_n\beta_n\gamma_n\,\mathbf{x}_n\otimes\mathbf{y}_n\otimes\mathbf{z}_n)$$

$$E = -\mathbf{b}\cdot(\mathbf{x}\otimes\mathbf{y}\otimes\mathbf{z})$$

$$\mathbf{x} = \sum_{i=1}^{n}\alpha_i\,\mathbf{x}_i, \quad \mathbf{y} = \sum_{i=1}^{n}\beta_i\,\mathbf{y}_i, \quad \mathbf{z} = \sum_{i=1}^{n}\gamma_i\,\mathbf{z}_i$$

Visual scene analysis

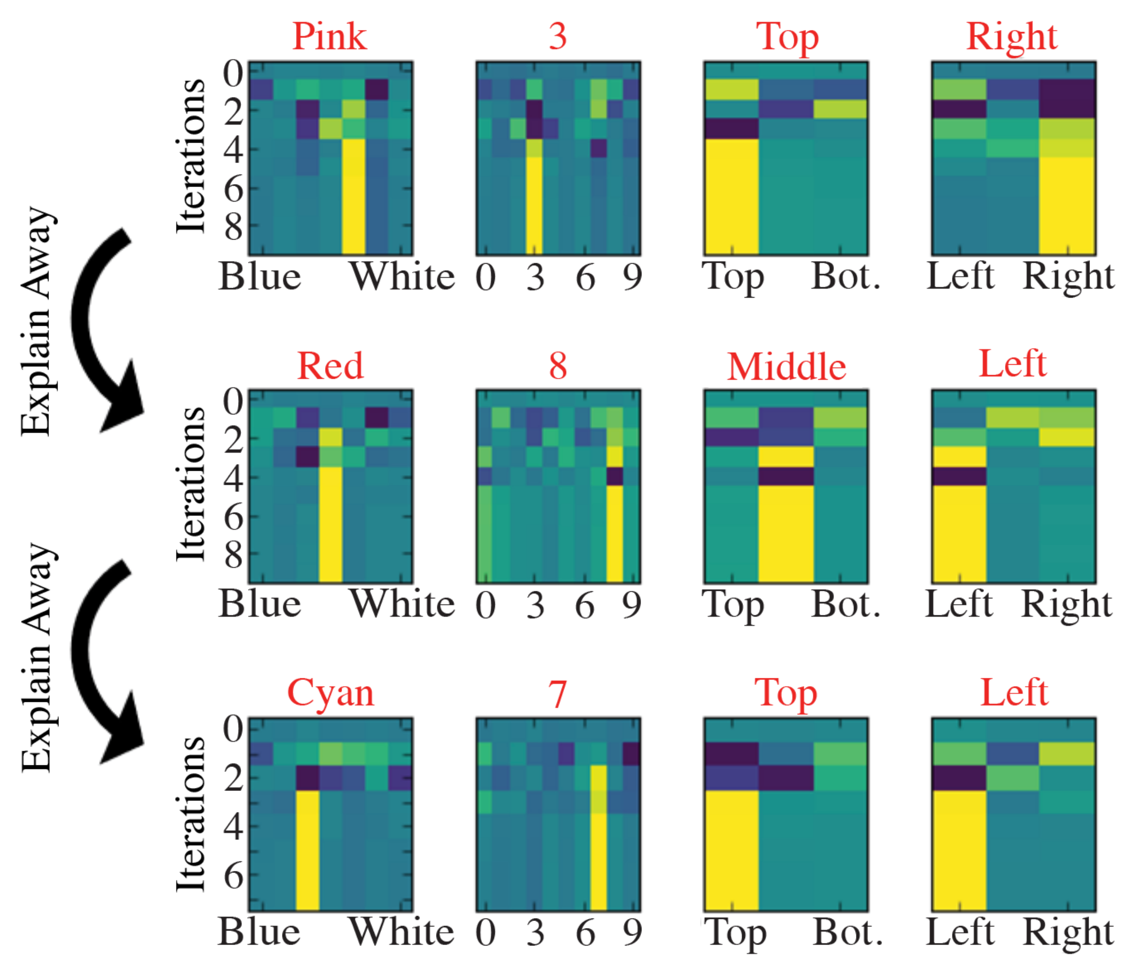
Input Image **Neural Network** **VSA encoding**



$$\mathbf{s} = \mathbf{c}_{\texttt{cyan}} \odot \mathbf{d}_7 \odot \mathbf{v}_{\texttt{top}} \odot \mathbf{h}_{\texttt{left}} + \mathbf{c}_{\texttt{pink}} \odot \mathbf{d}_3 \odot \mathbf{v}_{\texttt{top}} \odot \mathbf{h}_{\texttt{right}} + \mathbf{c}_{\texttt{red}} \odot \mathbf{d}_8 \odot \mathbf{v}_{\texttt{middle}} \odot \mathbf{h}_{\texttt{left}}$$

$$\|$$

$$\mathbf{c}_{\texttt{cyan}} \odot \mathbf{d}_7 \odot \mathbf{v}_{\texttt{top}} \odot \mathbf{h}_{\texttt{left}} + \mathbf{c}_{\texttt{pink}} \odot \mathbf{d}_3 \odot \mathbf{v}_{\texttt{top}} \odot \mathbf{h}_{\texttt{right}} + \mathbf{c}_{\texttt{red}} \odot \mathbf{d}_8 \odot \mathbf{v}_{\texttt{middle}} \odot \mathbf{h}_{\texttt{left}}$$



Complex hypervectors

- Useful for representing continuous quantities or attributes such as time, position, angle, color, etc.

- Each element is a complex phasor:

$$\mathbf{z} = \begin{bmatrix} e^{i\,\theta_1} \\ e^{i\,\theta_2} \\ \vdots \\ e^{i\,\theta_N} \end{bmatrix}$$

- Binding shifts phases:

$$\mathbf{z} \otimes \mathbf{x} = \begin{bmatrix} e^{i\,(\theta_1+\xi_1)} \\ e^{i\,(\theta_2+\xi_2)} \\ \vdots \\ e^{i\,(\theta_N+\xi_N)} \end{bmatrix}$$

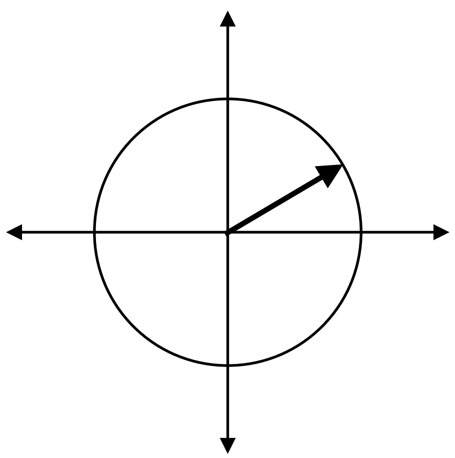
- Permutation (exponentiation) spins phases:

$$\mathbf{z}^t = \begin{bmatrix} e^{i\,\theta_1 t} \\ e^{i\,\theta_2 t} \\ \vdots \\ e^{i\,\theta_N t} \end{bmatrix} \qquad \mathbf{z}^{t_1} \otimes \mathbf{z}^{t_2} = \mathbf{z}^{t_1+t_2}$$

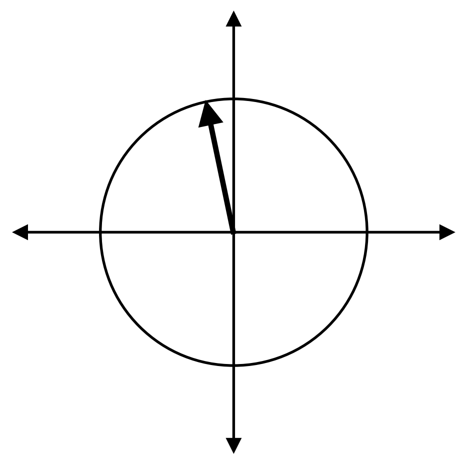
- Superposition:

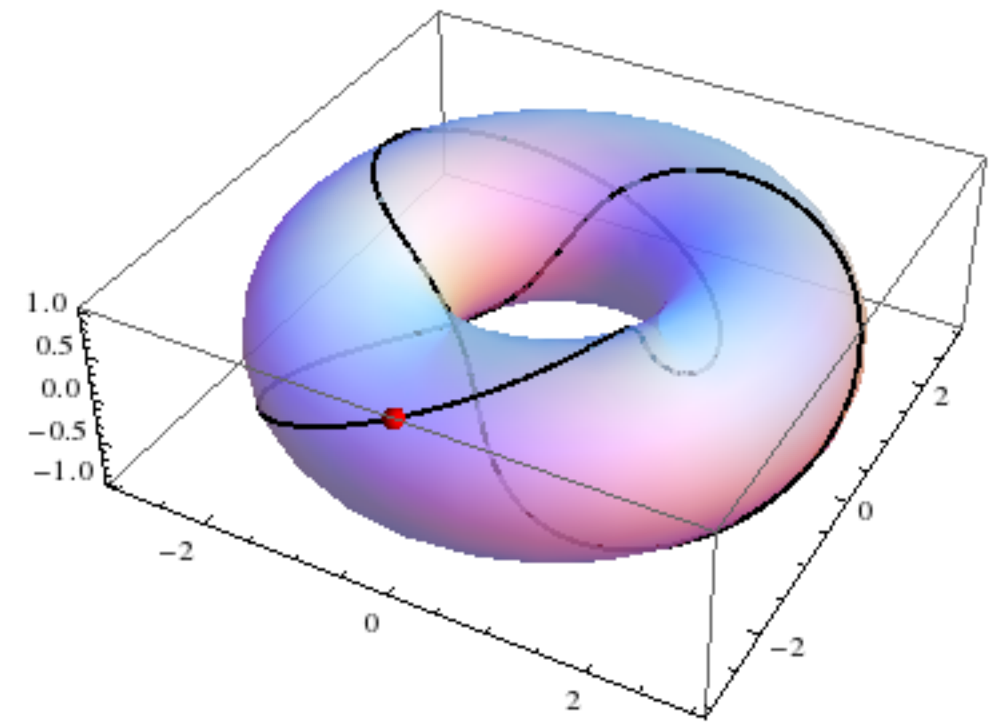
$$\mathbf{z}^{(1)} + \mathbf{z}^{(2)} + \mathbf{z}^{(3)} + \ldots = \begin{bmatrix} e^{i\,\theta_1^{(1)}} + e^{i\,\theta_1^{(2)}} + e^{i\,\theta_1^{(3)}} + \ldots \\ e^{i\,\theta_2^{(1)}} + e^{i\,\theta_2^{(2)}} + e^{i\,\theta_2^{(3)}} + \ldots \\ \vdots \\ e^{i\,\theta_N^{(1)}} + e^{i\,\theta_N^{(2)}} + e^{i\,\theta_N^{(3)}} + \ldots \end{bmatrix}$$

$$e^{i\,\theta_1\,t}$$



$$e^{i\,\theta_2\,t}$$

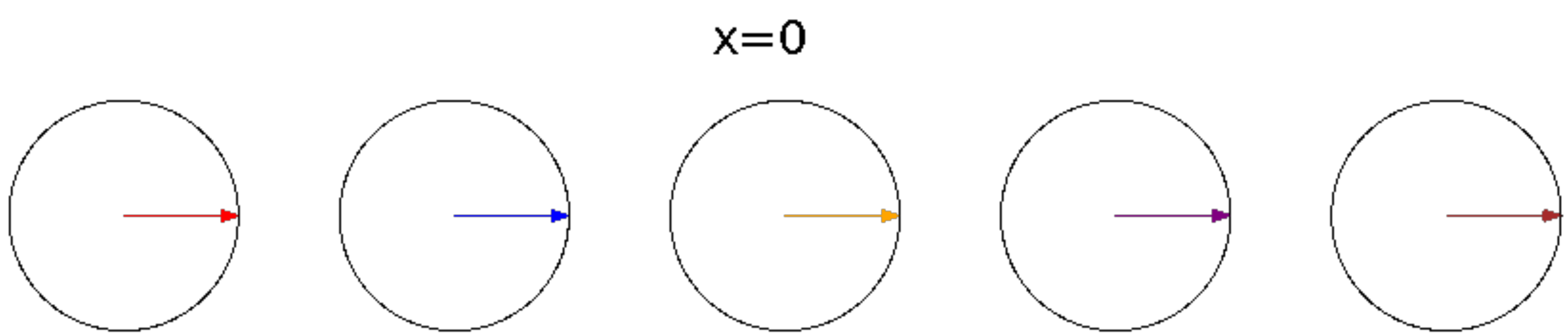




Encoding real numbers via **fractional binding**

Key idea 1: Represent any number x, by binding **z** *x* times with itself:

$$\mathbf{z}(x) = \underbrace{\mathbf{z} \odot \cdots \odot \mathbf{z}}_{\text{x times}} = \mathbf{z}^x$$



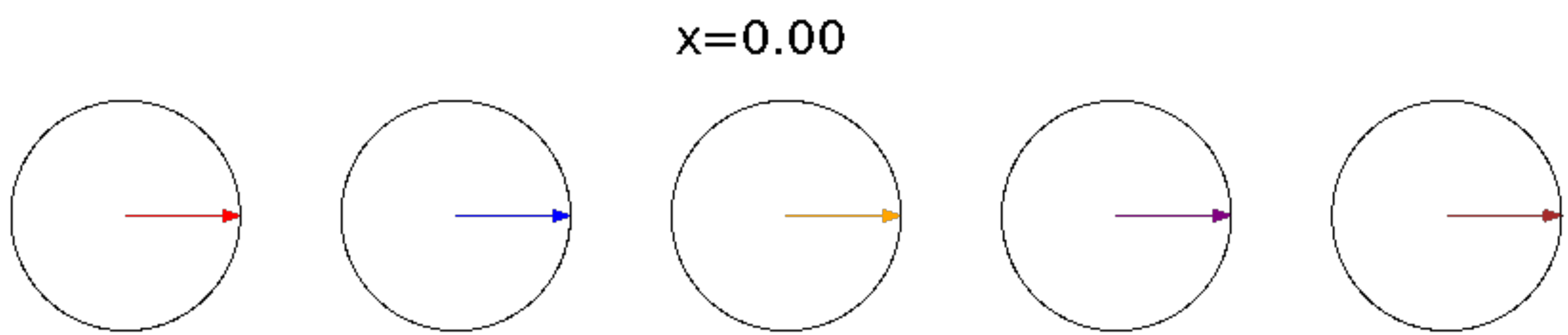
T.A. Plate, "Holographic Recurrent Networks," Advances in Neural Information Processing Systems (NIPS), pp. 34-41, 1992.
T.A. Plate, "Distributed Representations and Nested Compositional Structure," University of Toronto, PhD Thesis, 1994.

Encoding real numbers via **fractional binding**

Key idea 1: Represent any number x, by binding **z** *x* times with itself:

$$\mathbf{z}(x) = \underbrace{\mathbf{z} \odot \cdots \odot \mathbf{z}}_{\text{x times}} = \mathbf{z}^{x}$$

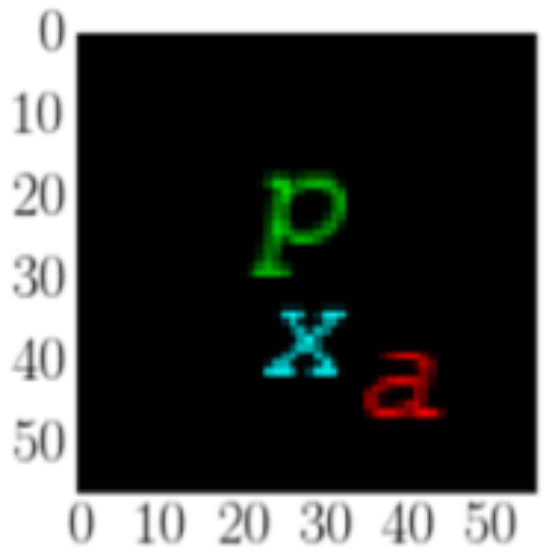
Key idea 2: Extend this definition to support encoding of non-integer *x* values



T.A. Plate, "Holographic Recurrent Networks," Advances in Neural Information Processing Systems (NIPS), pp. 34-41, 1992.
T.A. Plate, "Distributed Representations and Nested Compositional Structure," University of Toronto, PhD Thesis, 1994.

Factorization of shape, color and position

(Paxon Frady)



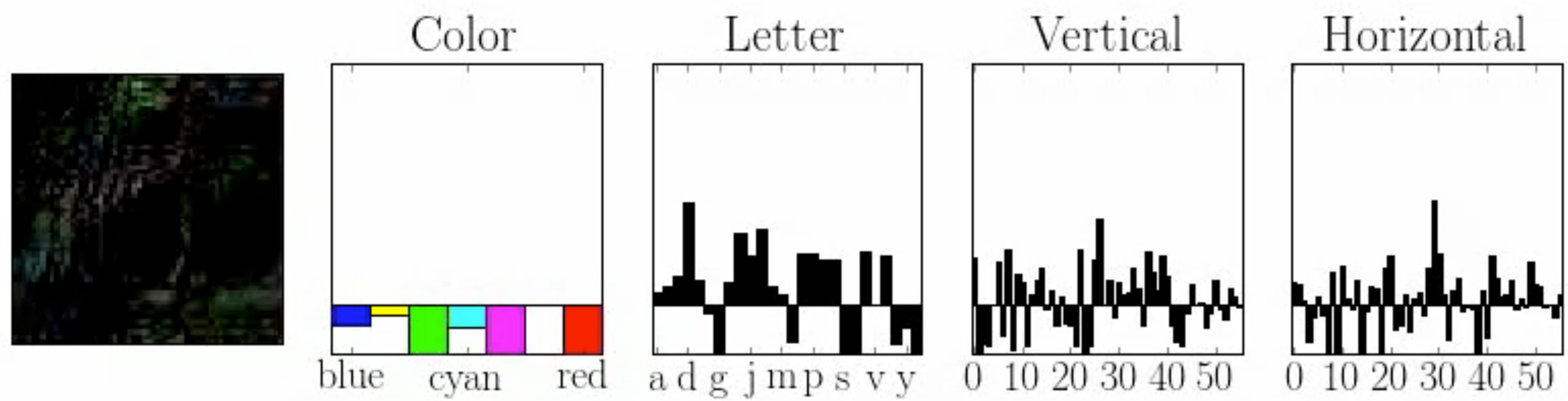
$$\mathbf{u}^{x_i} = \text{horizontal position } x_i$$
$$\mathbf{v}^{y_i} = \text{vertical position } y_j$$
$$\mathbf{w}_c = \text{color channel } c$$

$$\mathbf{s} = \sum_{i,j,c} I(x_i, y_j, c)\, \mathbf{u}^{x_i}\, \mathbf{v}^{y_j}\, \mathbf{w}_c$$

Given $\mathbf{s}$, find $\mathbf{x}$, $\mathbf{y}$, $\mathbf{c}$ and $\mathbf{p}$ via resonator:

$$\hat{\mathbf{x}}_{t+1} = g(\mathbf{X}\mathbf{X}^\top (\mathbf{s} \otimes \hat{\mathbf{y}}_t^{-1} \otimes \hat{\mathbf{c}}_t^{-1} \otimes \hat{\mathbf{p}}_t^{-1})) \quad \text{horizontal position}$$
$$\hat{\mathbf{y}}_{t+1} = g(\mathbf{Y}\mathbf{Y}^\top (\mathbf{s} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{c}}_t^{-1} \otimes \hat{\mathbf{p}}_t^{-1})) \quad \text{vertical position}$$
$$\hat{\mathbf{c}}_{t+1} = g(\mathbf{C}\mathbf{C}^\top (\mathbf{s} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{y}}_t^{-1} \otimes \hat{\mathbf{p}}_t^{-1})) \quad \text{color}$$
$$\hat{\mathbf{p}}_{t+1} = g(\mathbf{P}\mathbf{P}^\top (\mathbf{s} \otimes \hat{\mathbf{x}}_t^{-1} \otimes \hat{\mathbf{y}}_t^{-1} \otimes \hat{\mathbf{c}}_t^{-1})) \quad \text{pattern}$$

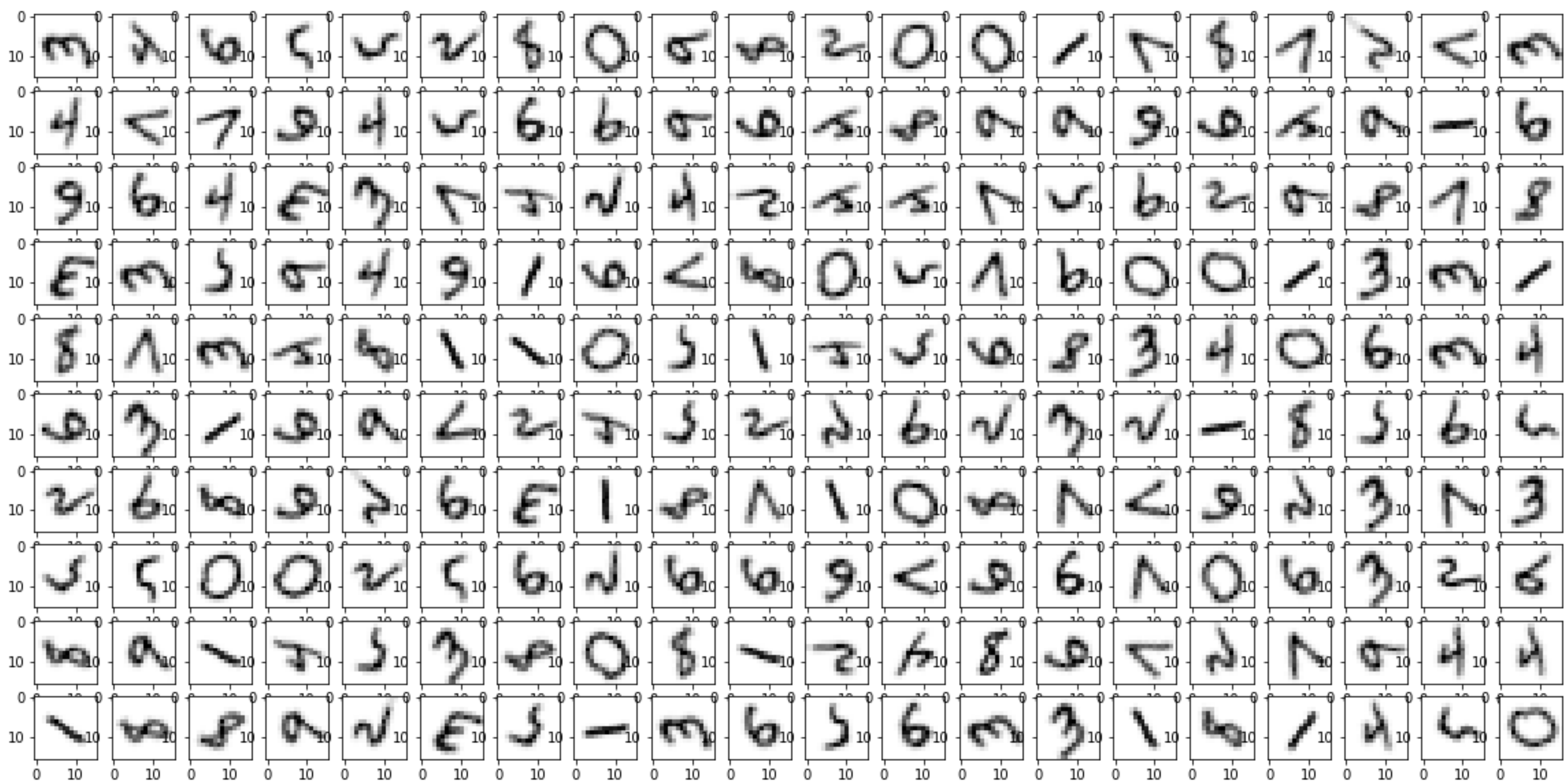
Visual scene analysis via factorization of HD vectors
(Paxon Frady)



Learning to separate shape and transformations via Lie group operators and sparse coding

(Ho Yin Chau, Yubei Chen, Frank Qiu)

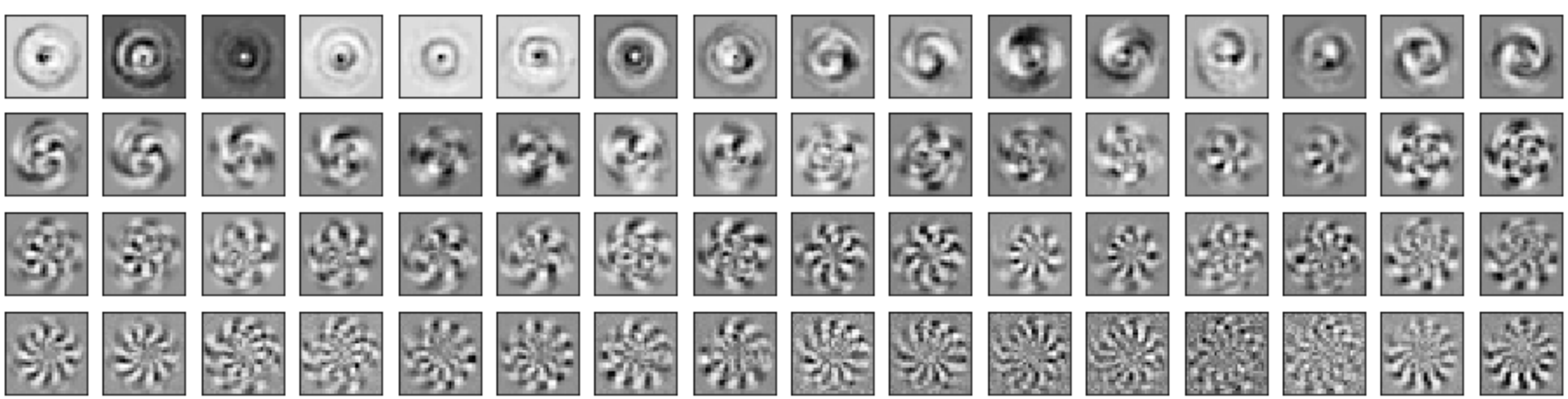
$$\mathbf{I} = \mathbf{T}(s)\,\mathbf{\Phi}\,\alpha + \epsilon$$
$$= e^{\mathbf{A}s}\,\mathbf{\Phi}\,\alpha + \epsilon$$



$$\begin{aligned}\mathbf{T}(s) &= e^{\mathbf{A}s} \\ &= \mathbf{W}\, e^{\mathbf{\Sigma} s}\, \mathbf{W}^T \;=\; \mathbf{W}\,\mathbf{R}(s)\,\mathbf{W}^T\end{aligned}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 0 & -\omega_1 & & & \\ \omega_1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & -\omega_{D/2} \\ & & & \omega_{D/2} & 0 \end{bmatrix} \qquad \mathbf{R}(s) = \begin{bmatrix} \cos(\omega_1 s) & -\sin(\omega_1 s) & & & \\ \sin(\omega_1 s) & \cos(\omega_1 s) & & & \\ & & \ddots & & \\ & & & \cos\left(\omega_{D/2} s\right) & -\sin\left(\omega_{D/2} s\right) \\ & & & \sin\left(\omega_{D/2} s\right) & \cos\left(\omega_{D/2} s\right) \end{bmatrix}$$

$$\mathbf{I} = \mathbf{W}\,\mathbf{R}(s)\,\mathbf{W}^T\,\mathbf{\Phi}\,\alpha + \epsilon$$



Results