UNIVERSITY OF AMSTERDAM

Group Equivariant Deep Learning

Lecture 1 - Regular group convolutions

Lecture 1.7 - Group convolutions are all you need!

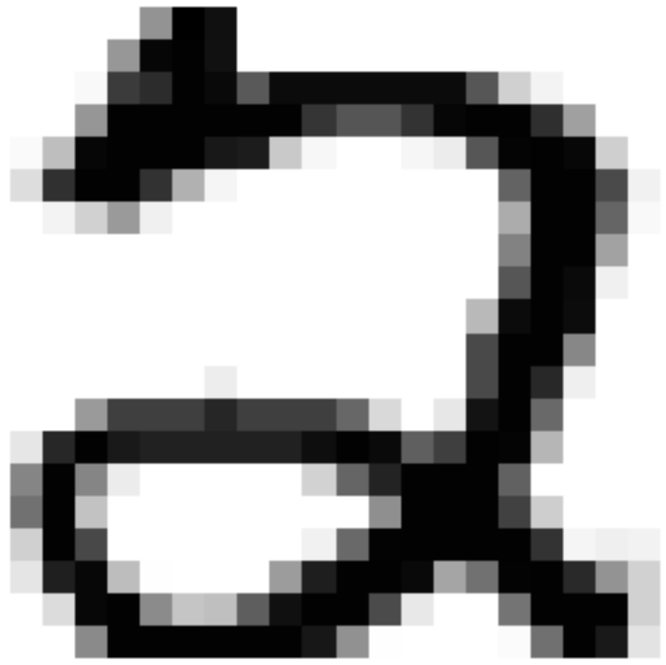
Equivariant linear layers between feature maps are group convolutions

Erik Bekkers, Amsterdam Machine Learning Lab, University of Amsterdam

This mini-course serves as a module with the UvA Master AI course Deep Learning 2 https://uvadl2c.github.io/

Classical artificial neural networks

What's my input? $\underline{x}^0 \in \mathcal{X} = ?$



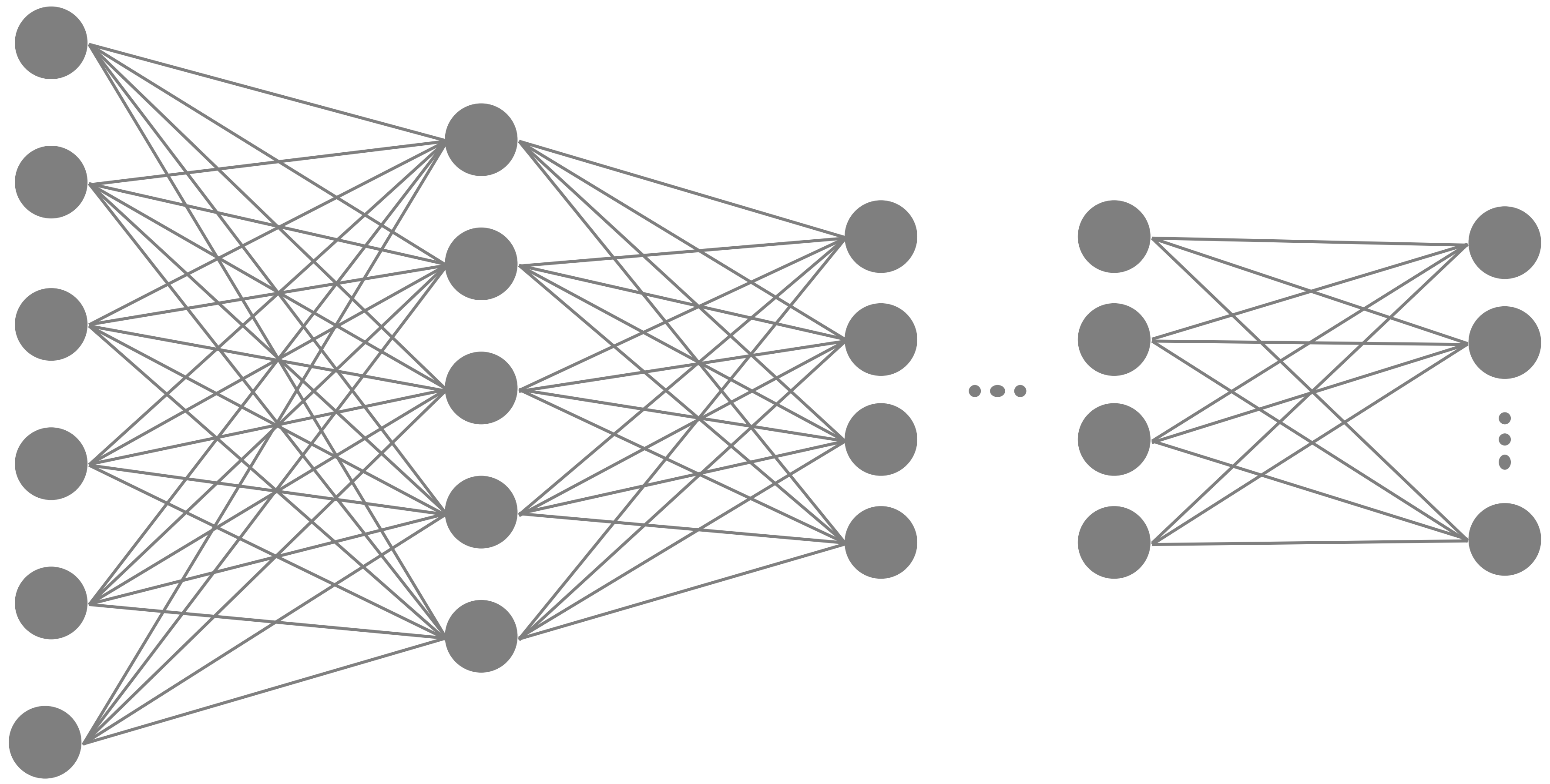
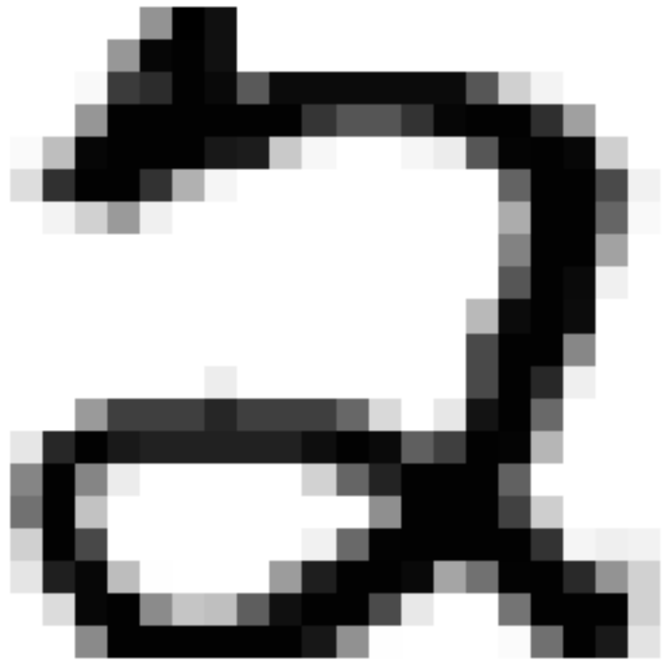


Image analyst: $\underline{x}^0 \in \mathcal{X} = \mathbb{L}_2(\mathbb{R}^2)$

Naive deep learner: $\underline{x}^0 \in \mathcal{X} = \mathbb{R}^{784}$

Classical artificial neural networks

What's my input? $\underline{x}^0 \in \mathscr{X} = ?$



input vector

output probability vector

(e.g. through soft max)

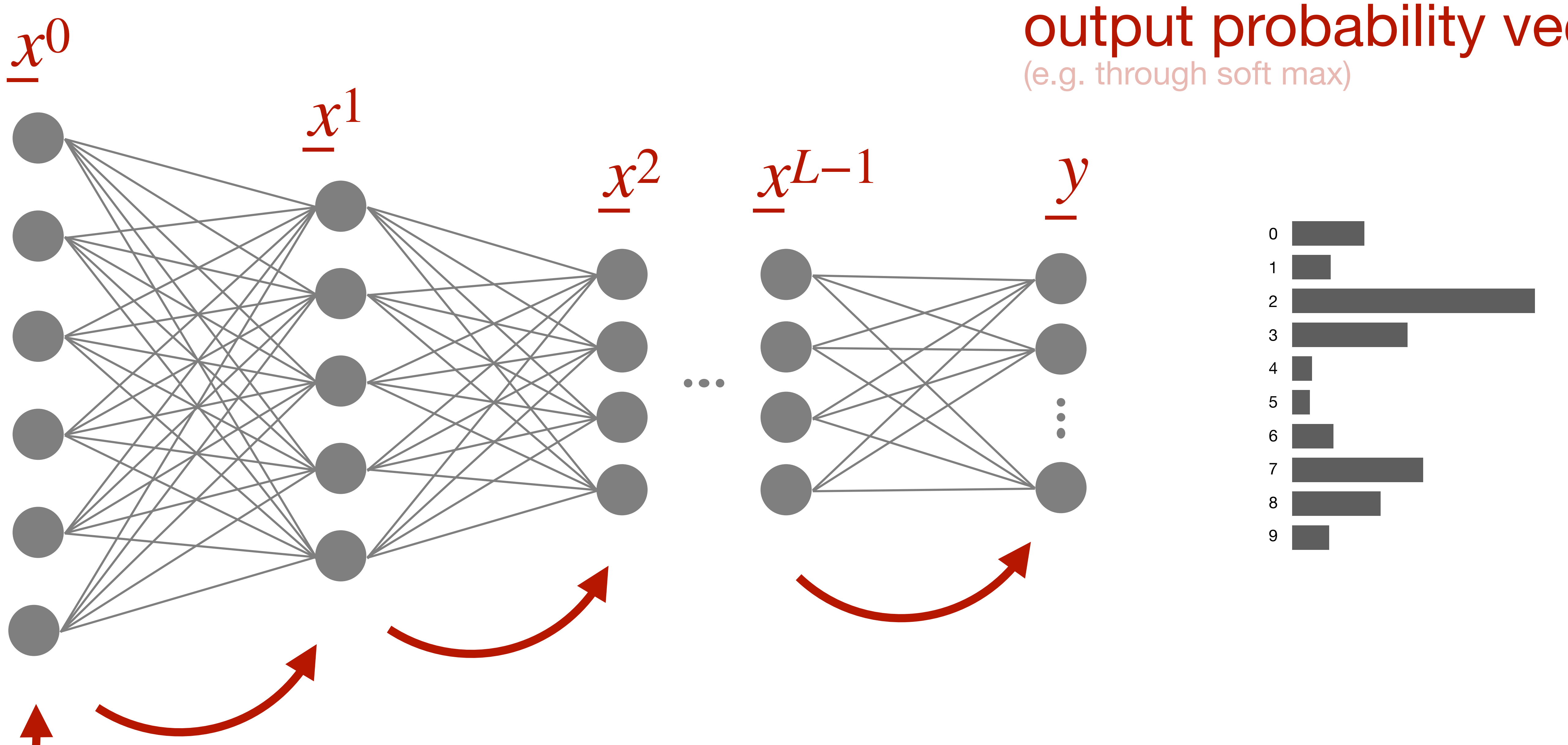
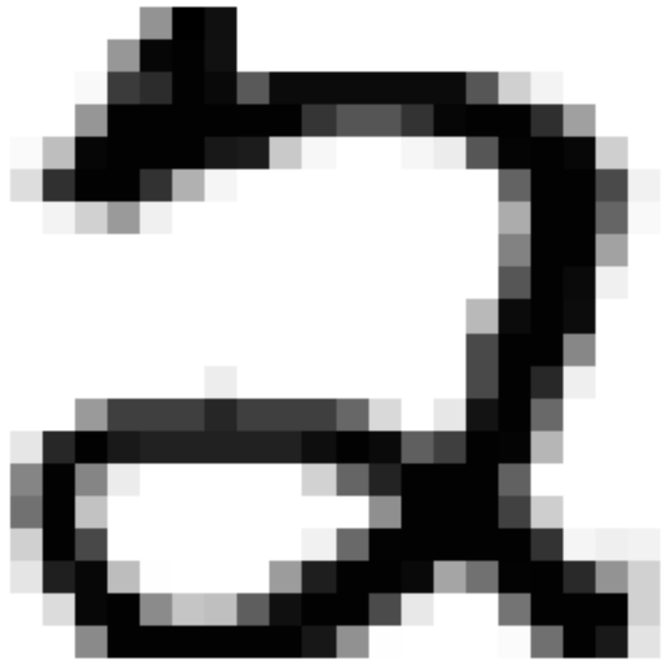


Image analyst: $\underline{x}^0 \in \mathscr{X} = \mathbb{L}_2(\mathbb{R}^2)$

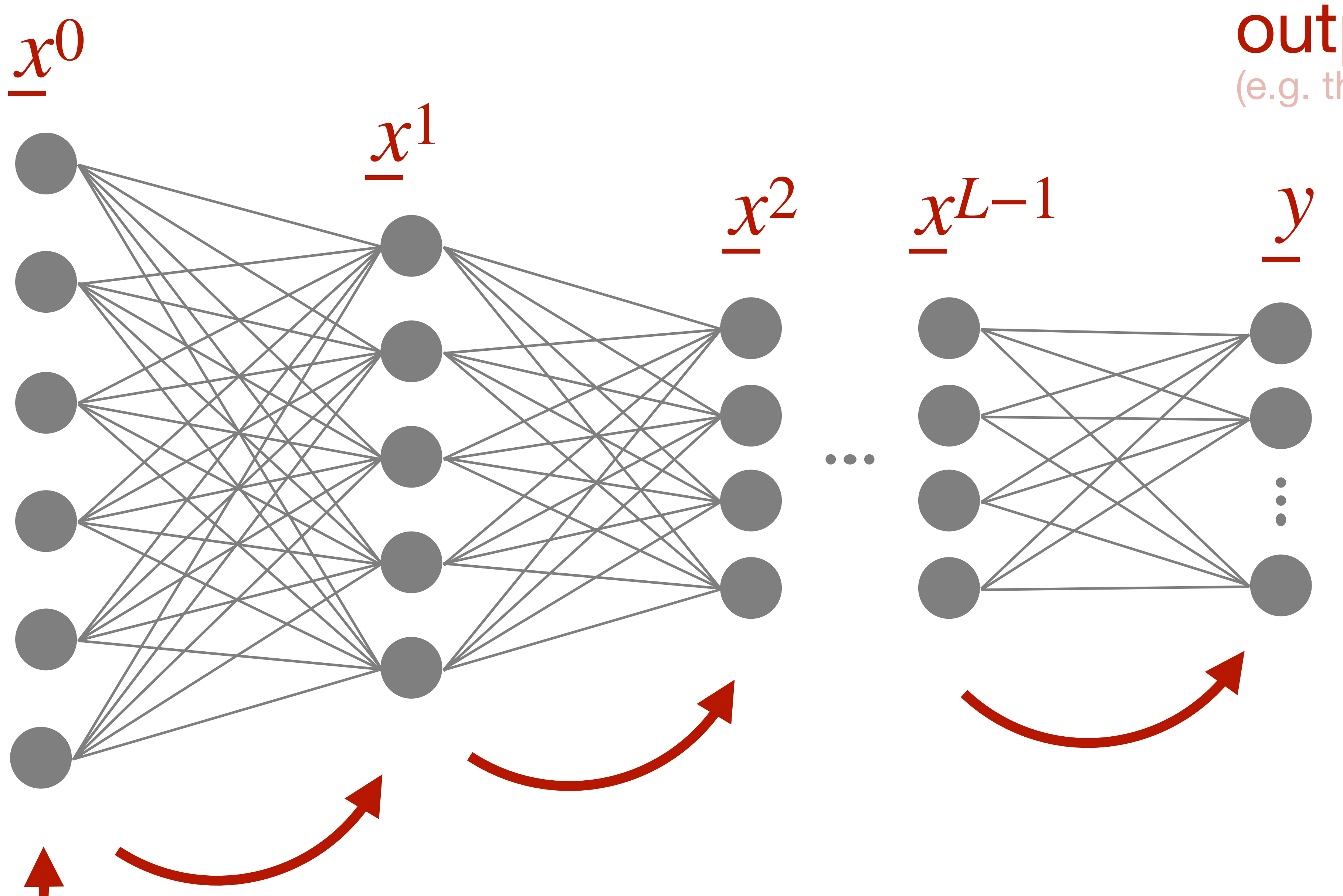
Naive deep learner: $\underline{x}^0 \in \mathscr{X} = \mathbb{R}^{784}$

Classical artificial neural networks

What's my input? $\underline{x}^0 \in \mathcal{X} = ?$



input vector



output probability vector
(e.g. through soft max)

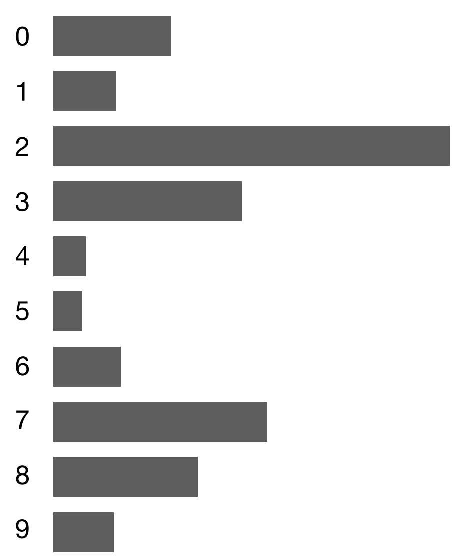


Image analyst: $\underline{x}^0 \in \mathcal{X} = \mathbb{L}_2(\mathbb{R}^2)$

Naive deep learner: $\underline{x}^0 \in \mathcal{X} = \mathbb{R}^{784}$

Iteratively transform the vector via

$$\underline{x}^l = \varphi(K_{\mathbf{w}_l}\underline{x}^{l-1} + \underline{b}^l)$$

Linear map: matrix-vector multiplication with $K_{\mathbf{w}_l} \in \mathbb{R}^{N^l \times N^{l-1}}$

A fully connect layer as convolution on 1D signal

- Way too many degrees of freedom!
- Doest not leverage/preserve structure in data

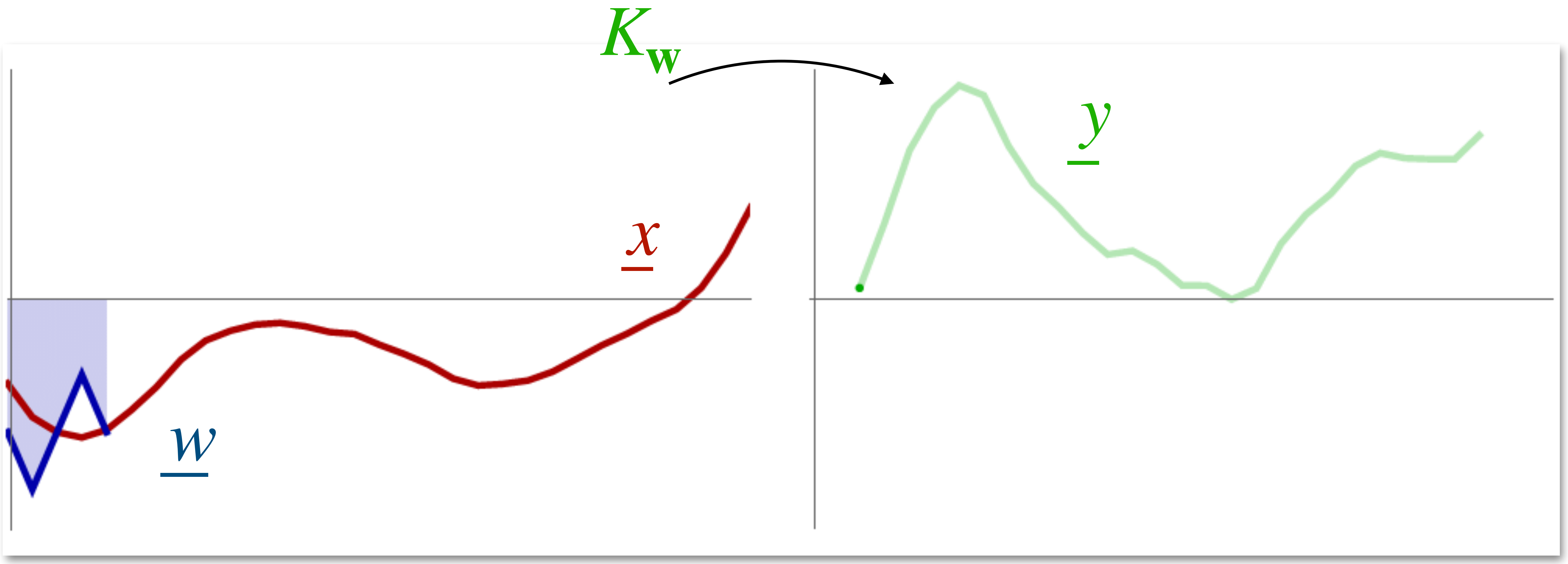
$$\underline{y} = \varphi(\quad K_{\mathbf{w}} \quad \underline{x} + \underline{b} \quad)$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} = \varphi \left(\begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & \cdots \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} & \cdots \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \end{pmatrix} \right)$$

A fully connected layer as convolution on 1D signal

Convolution as linear operator

+ Localized transformations
+ Shift equivariance
+ Sparsification of the linear operator
+ Weightsharing



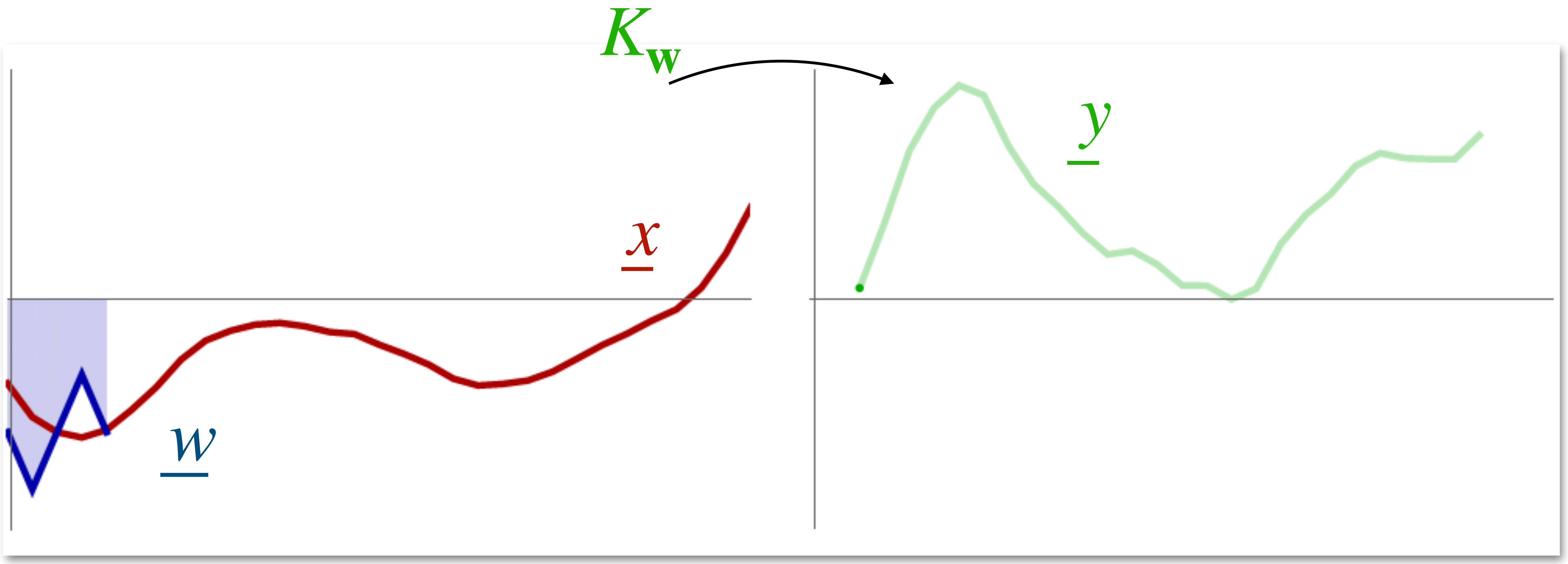
$$\underline{y} = \varphi(K_{\mathbf{w}} \underline{x} + \underline{b})$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} = \varphi \left(\begin{pmatrix} w_1 & w_2 & w_3 & 0 & 0 & \dots \\ 0 & w_1 & w_2 & w_3 & 0 & \dots \\ 0 & 0 & w_1 & w_2 & w_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \end{pmatrix} \right)$$

A fully connected layer as convolution on 1D signal

Convolution as linear operator

+ Localized transformations
+ Shift equivariance
+ Sparsification of the linear operator
+ Weightsharing

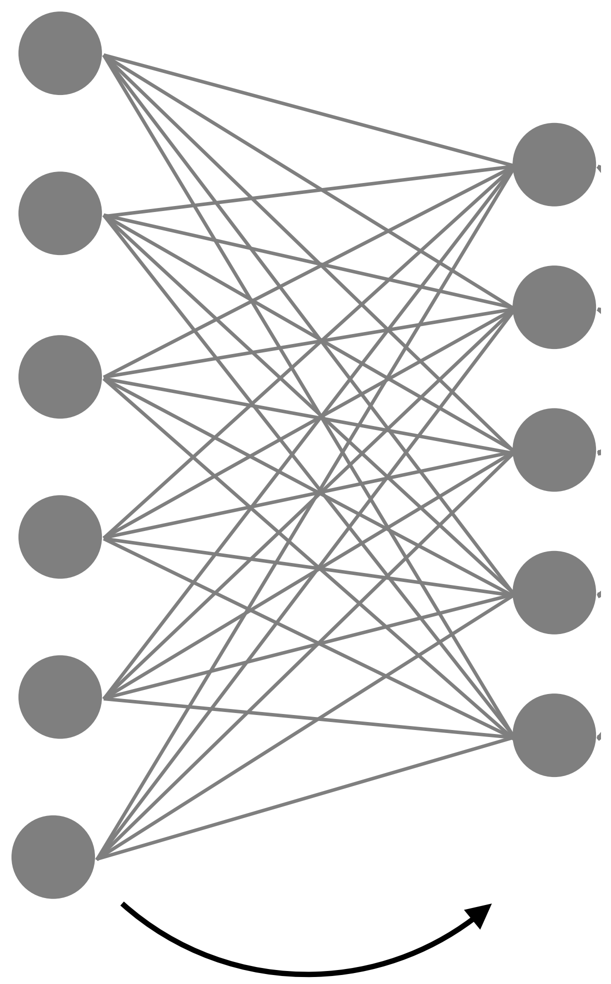


$$\underline{y} = \varphi(K_{\mathbf{w}} \underline{x} + \underline{b})$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} = \varphi \left(\begin{pmatrix} w_1 & w_2 & w_3 & 0 & 0 & \dots \\ 0 & w_1 & w_2 & w_3 & 0 & \dots \\ 0 & 0 & w_1 & w_2 & w_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \end{pmatrix} \right)$$

Classical artificial NNs in the continuous world

Working with vectors $\underline{x} \in \mathcal{X} = \mathbb{R}^{N^x}$



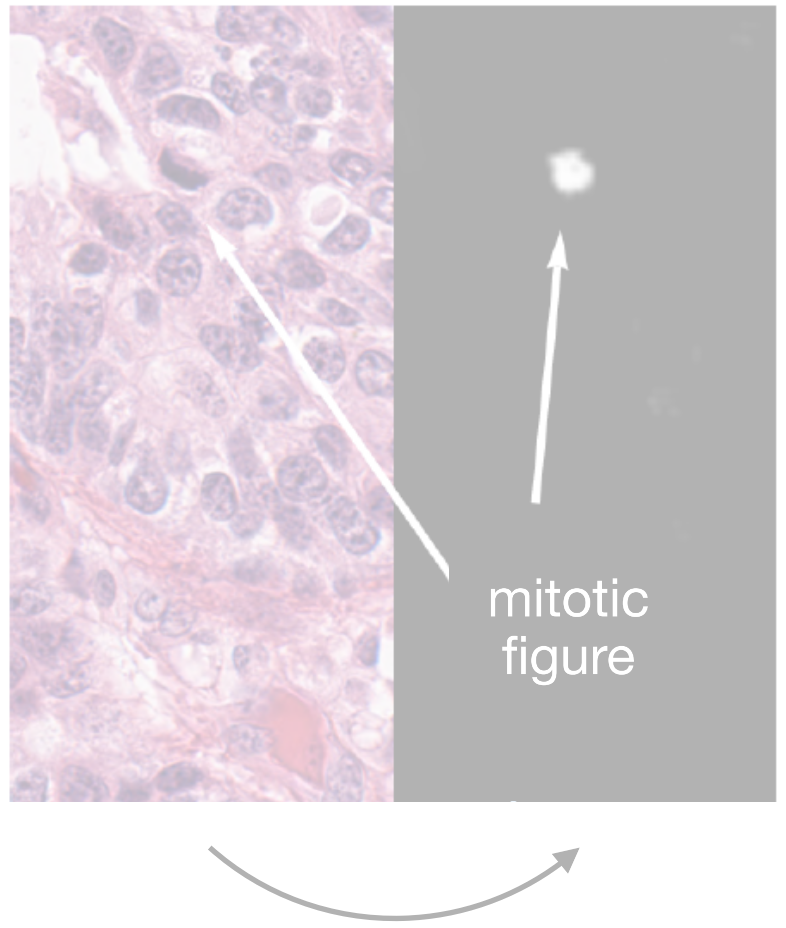
Iteratively transform the **vector** in $\mathbb{R}^{N^x}$ via

$$\underline{y} = \varphi(K\,\underline{x} + \underline{b}^l)$$

Linear map: matrix-vector multiplication with $K \in \mathbb{R}^{N^y \times N^x}$

$$y_j = \sum_i K_{i,j}\, x_i$$

Working with feature maps $f \in \mathcal{X} = \mathbb{L}_2(X)$



Iteratively transform the **feature map** in $\mathbb{L}_2(X)$

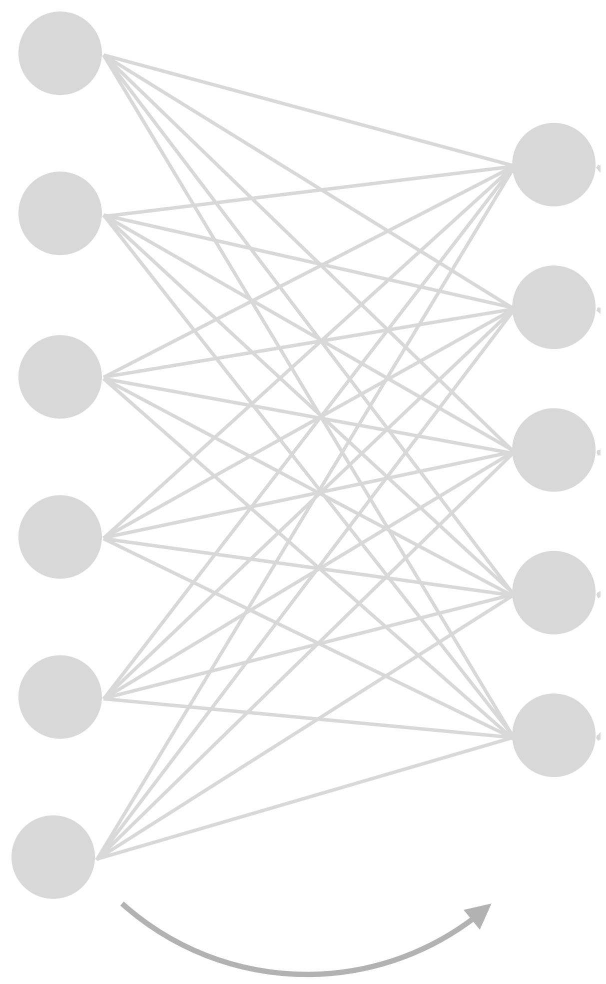
$$f^{out} = \varphi(K\,f^{in} + b^l)$$

Linear map: kernel operator with kernel in $\mathbb{L}_1(Y, X)$

$$(Kf)(y) = \int_X k(y,x) f(x) \mathrm{d}x$$

Classical artificial NNs in the continuous world

Working with vectors $\underline{x} \in \mathscr{X} = \mathbb{R}^{N^x}$



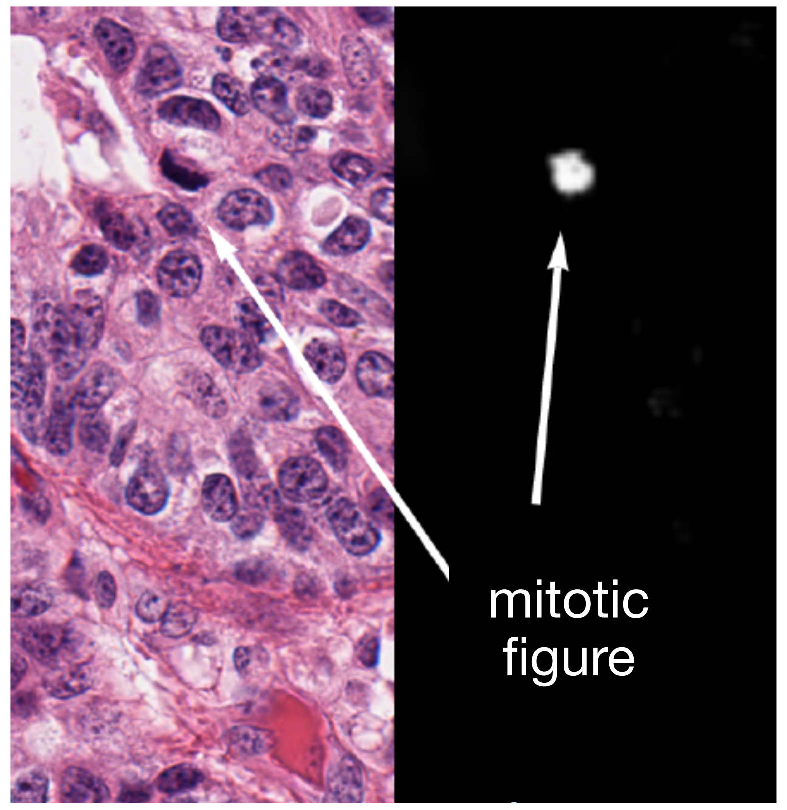
Iteratively transform the **vector** in $\mathbb{R}^{N^x}$ via

$$\underline{y} = \varphi(K\,\underline{x} + \underline{b}^l)$$

Linear map: matrix-vector multiplication with $K \in \mathbb{R}^{N^y \times N^x}$

$$y_j = \sum_i K_{i,j}\, x_i$$

Working with feature maps $f \in \mathscr{X} = \mathbb{L}_2(X)$



Iteratively transform the **feature map** in $\mathbb{L}_2(X)$

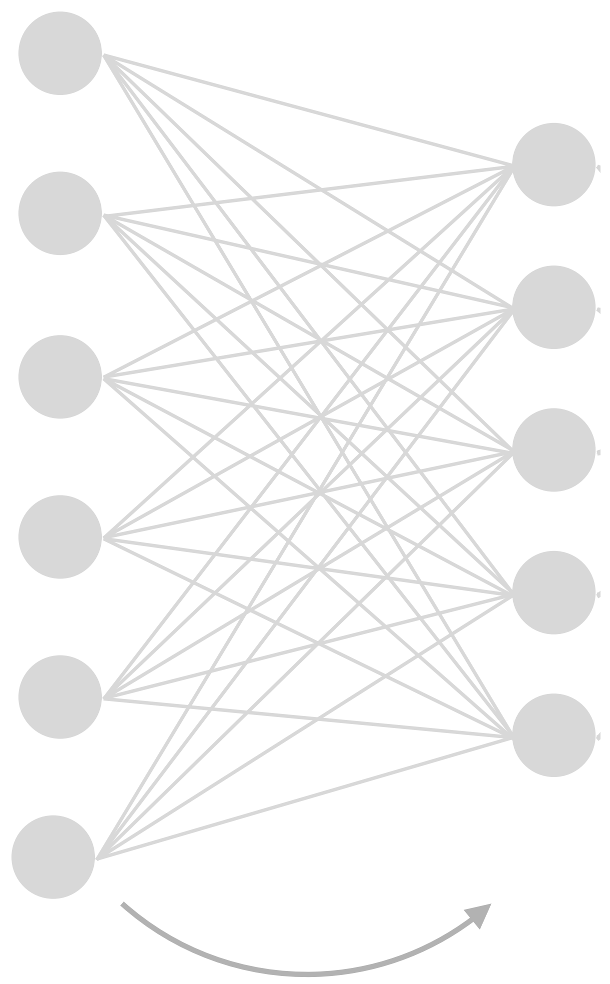
$$f^{out} = \varphi(K f^{in} + b^l)$$

Linear map: kernel operator with kernel in $\mathbb{L}_1(Y, X)$

$$(Kf)(y) = \int_X k(y,x) f(x) \mathrm{d}x$$

Classical artificial NNs in the continuous world

Working with vectors $\underline{x} \in \mathcal{X} = \mathbb{R}^{N^x}$



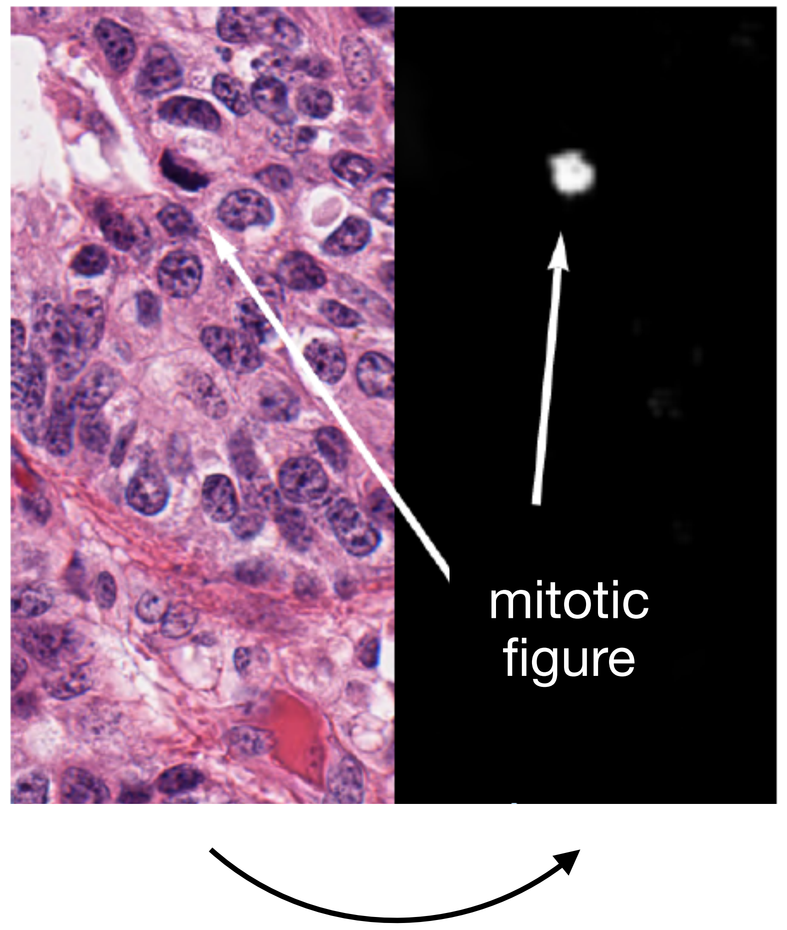
Iteratively transform the **vector** in $\mathbb{R}^{N^x}$ via

We want K to be equivariant!

Linear map: matrix-vector multiplication with $K \in \mathbb{R}^{N^y \times N^x}$

$$y_j = \sum_i K_{i,j} x_i$$

Working with feature maps $f \in \mathcal{X} = \mathbb{L}_2(X)$



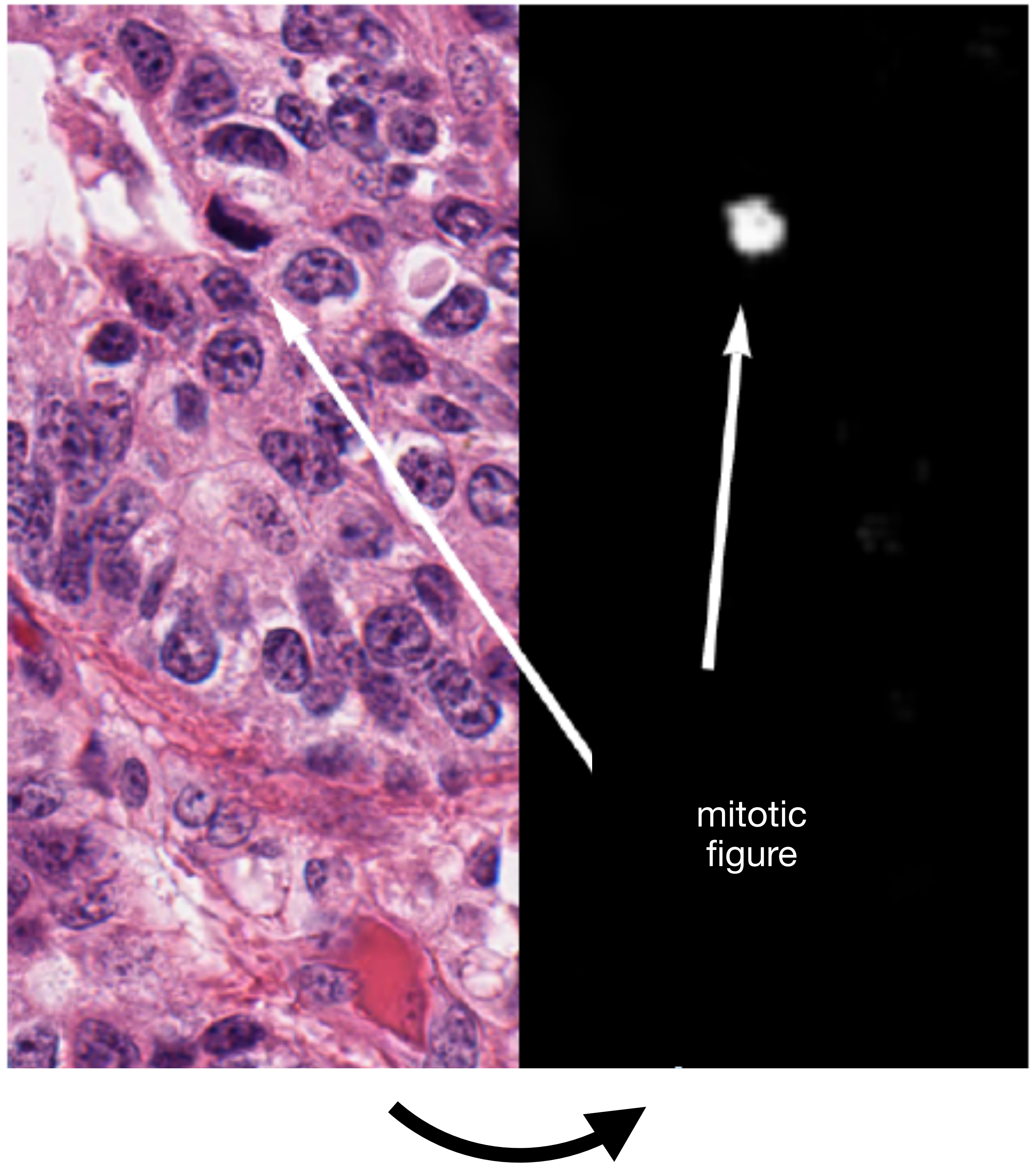
Iteratively transform the **feature map** in $\mathbb{L}_2(X)$

$$f^{out} = \varphi(K f^{in} + b^l)$$

Linear map: kernel operator with kernel in $\mathbb{L}_1(Y, X)$

$$(Kf)(y) = \int_X k(y, x) f(x) \mathrm{d}x$$

Neural Networks for Signal Data



$$\mathscr{K} : \mathbb{L}_2(X)^{N_l} \to \mathbb{L}_2(Y)^{N_{l+1}}$$

Let's build neural networks for signal data via the layers of the form:

$$\underline{f}^{l+1} = \sigma(\mathscr{K}\,\underline{f}^{l} + \mathbf{b}^{l})$$

The linear map has to be an integral transform with a two-argument kernel
(Dunford-Pettis theorem)

$$(\mathscr{K} f)(y) = \int_X \mathbf{k}(y, x)\underline{f}(x)\mathrm{d}x$$

Theorem (G-convs are all you need)

Bekkers ICLR 2020, Thm. 1*

Let $\mathscr{K}: \mathbb{L}_2(X) \rightarrow \mathbb{L}_2(Y)$ map between signals on homogeneous spaces of G.

Let homogeneous space $Y \equiv G/H$ such that $H = \mathrm{Stab}_G(y_0)$ for some chosen origin $y_0 \in Y$ and let $g_y \in G$ such that $\forall_{y \in Y}: y = g_y y_0$.

Then $\mathscr{K}$ is equivariant to group G if and only if:

1. It is a group convolution: $[\mathscr{K} f](y) = \int_X \frac{1}{|g_y|} k(g_y^{-1} x) f(x) \mathrm{d}x$
2. The kernel satisfies a symmetry constraint: $\forall_{h \in H}: \frac{1}{|g_y|} k(hx) = k(x)$

*Work with Remco Duits at TU/e.

See also: Duits 2005 – Thm 25, Cohen, Geiger, Weiler 2018 - Thm 6.1, Kondor, Trivedi 2018 - Thm 1

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform:

$$\mathscr{K}[f](y) = \int_X \tilde{k}(y, x) f(x) \mathrm{d}x$$

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform: $$\mathscr{K}[f](y) = \int_X \tilde{k}(y,x)f(x)\mathrm{d}x$$

$\forall_{g \in G}$ and $\forall_{f \in \mathbb{L}_2(X)}$: $$(\mathscr{K} \circ \mathscr{L}_g^{G \to \mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G \to \mathbb{L}_2(Y)} \circ \mathscr{K})(f)$$

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform:

$$\mathscr{K}[f](y) = \int_X \tilde{k}(y,x)f(x)\mathrm{d}x$$

$\forall_{g \in G}$ and $\forall_{f \in \mathbb{L}_2(X)}$:

$$(\mathscr{K} \circ \mathscr{L}_g^{G \to \mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G \to \mathbb{L}_2(Y)} \circ \mathscr{K})(f)$$

$$\int_X \tilde{k}(y,x)f(g^{-1}x)\mathrm{d}x = \int_X \tilde{k}(g^{-1}y,x)f(x)\mathrm{d}x$$

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform:

$$\mathscr{K}[f](y) = \int_X \tilde{k}(y,x)f(x)\mathrm{d}x$$

$\forall_{g \in G}$ and $\forall_{f \in \mathbb{L}_2(X)}$:

$$(\mathscr{K} \circ \mathscr{L}_g^{G \to \mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G \to \mathbb{L}_2(Y)} \circ \mathscr{K})(f)$$

$$\int_X \tilde{k}(y,x)f(g^{-1}x)\mathrm{d}x = \int_X \tilde{k}(g^{-1}y,x)f(x)\mathrm{d}x$$

(in r.h.s. $x \leftarrow g^{-1}x$)

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform:

$$\mathscr{K}[f](y) = \int_X \tilde{k}(y,x)f(x)\mathrm{d}x$$

$\forall_{g\in G}$ and $\forall_{f\in\mathbb{L}_2(X)}$:

$$(\mathscr{K} \circ \mathscr{L}_g^{G\rightarrow\mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G\rightarrow\mathbb{L}_2(Y)} \circ \mathscr{K})(f)$$

$$\int_X \tilde{k}(y,x)f(g^{-1}x)\mathrm{d}x = \int_X \tilde{k}(g^{-1}y,x)f(x)\mathrm{d}x$$

(in r.h.s. $x \leftarrow g^{-1}x$)

$$\int_X \tilde{k}(y,x)f(g^{-1}x)\mathrm{d}x = \int_X \tilde{k}(g^{-1}y,g^{-1}x)f(g^{-1}x)\mathrm{d}(g^{-1}x)$$

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform: $\mathscr{K}[f](y) = \int_X \tilde{k}(y,x) f(x) \mathrm{d}x$

$\forall_{g \in G}$ and $\forall_{f \in \mathbb{L}_2(X)}$: $(\mathscr{K} \circ \mathscr{L}_g^{G \to \mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G \to \mathbb{L}_2(Y)} \circ \mathscr{K})(f)$

$$\int_X \tilde{k}(y,x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, x) f(x) \mathrm{d}x$$

(in r.h.s. $x \leftarrow g^{-1}x$)

$$\int_X \tilde{k}(y,x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, g^{-1}x) f(g^{-1}x) \mathrm{d}(g^{-1}x)$$

$$\int_X \tilde{k}(y,x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, g^{-1}x) f(g^{-1}x) \frac{1}{|\det g|} \mathrm{d}x \, .$$

Proof (G-convs are all you need)

First of all, $\mathscr{K}$ is an integral transform: $\mathscr{K}[f](y) = \int_X \tilde{k}(y, x) f(x) \mathrm{d}x$

$\forall_{g \in G}$ and $\forall_{f \in \mathbb{L}_2(X)}$:

$$(\mathscr{K} \circ \mathscr{L}_g^{G \to \mathbb{L}_2(X)})(f) = (\mathscr{L}_g^{G \to \mathbb{L}_2(Y)} \circ \mathscr{K})(f)$$

$$\int_X \tilde{k}(y, x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, x) f(x) \mathrm{d}x$$

(in r.h.s. $x \leftarrow g^{-1}x$)

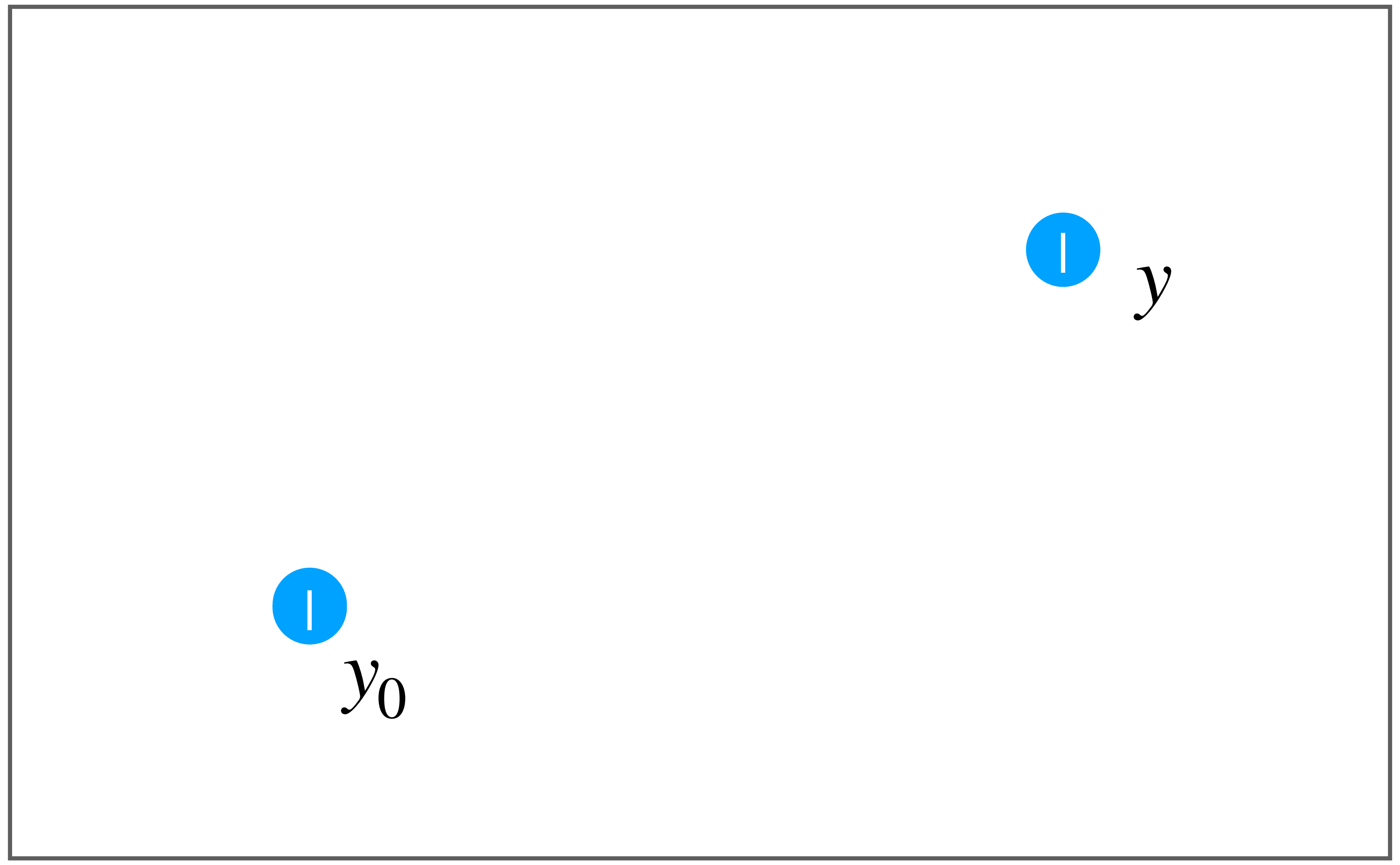
$$\int_X \tilde{k}(y, x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, g^{-1}x) f(g^{-1}x) \mathrm{d}(g^{-1}x)$$

$$\int_X \tilde{k}(y, x) f(g^{-1}x) \mathrm{d}x = \int_X \tilde{k}(g^{-1}y, g^{-1}x) f(g^{-1}x) \frac{1}{|\det g|} \mathrm{d}x\,.$$

Since this should hold or all $f \in \mathbb{L}_2(X)$ we have $\forall_{g \in G}$: $\tilde{k}(y, x) = \frac{1}{|\det g|} \tilde{k}(g^{-1}y, g^{-1}x)$

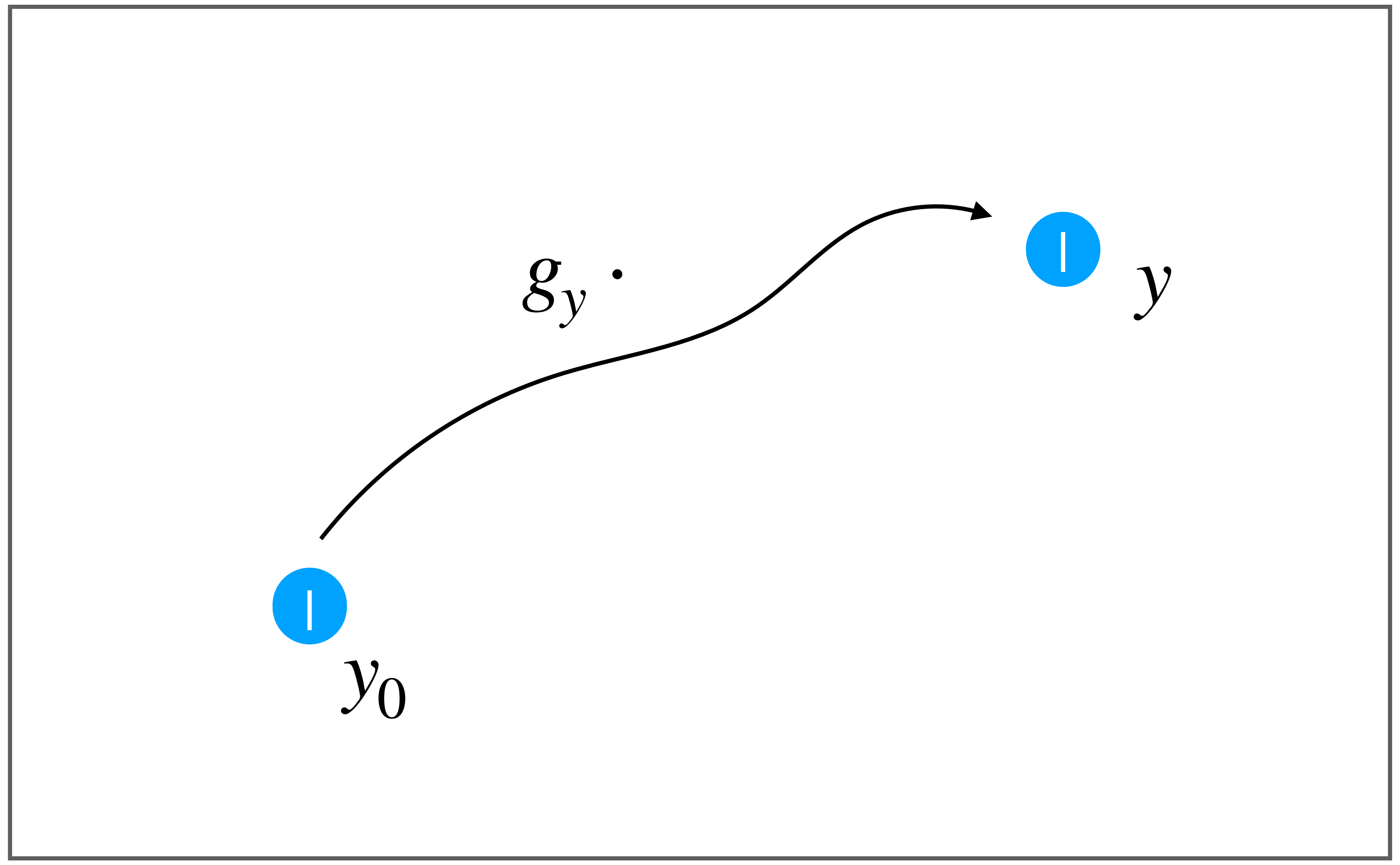
Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



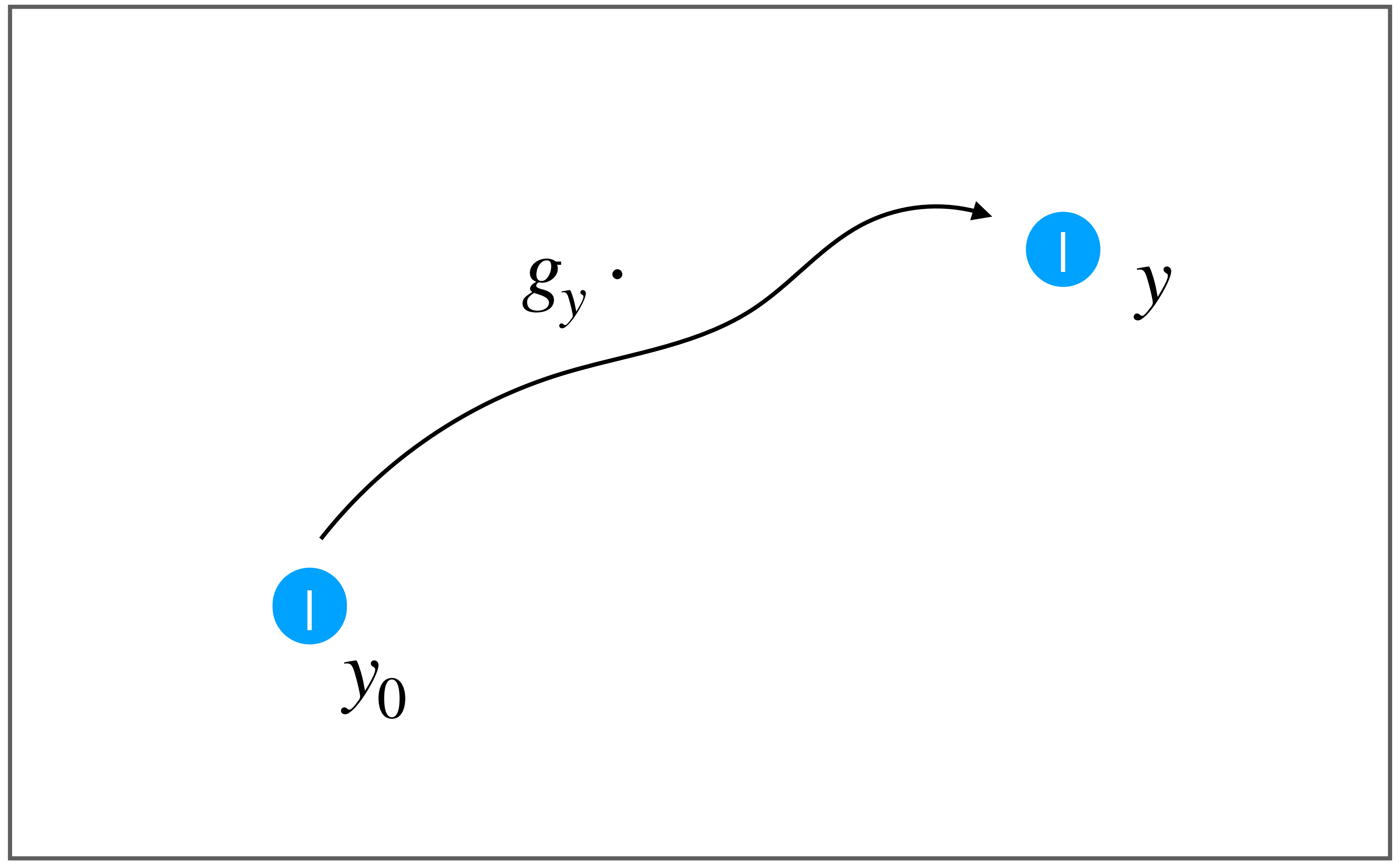
Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Proof (G-convs are all you need)

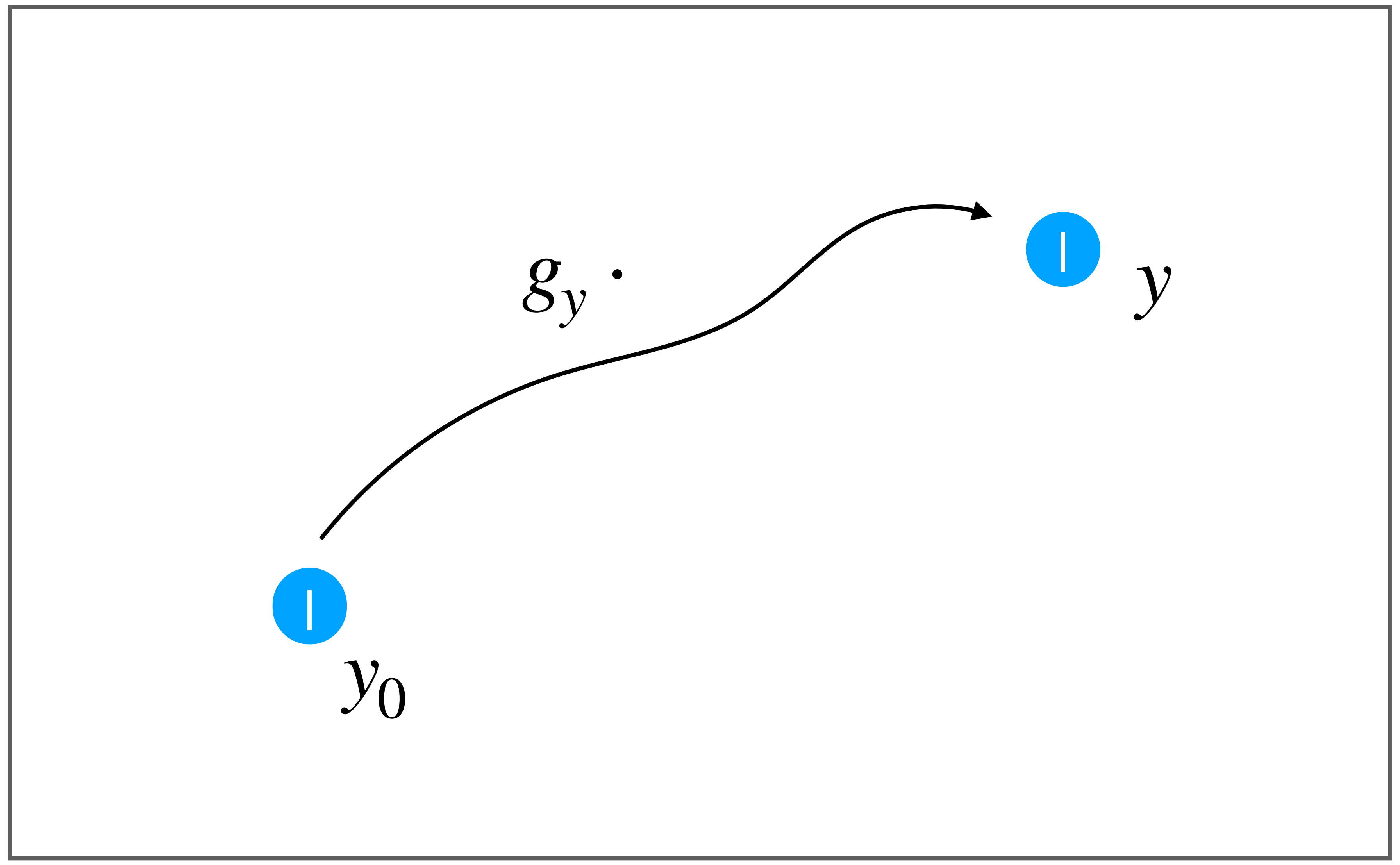
Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\quad \tilde{k}(y, x) = \tilde{k}(g_y y_0, x)$

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



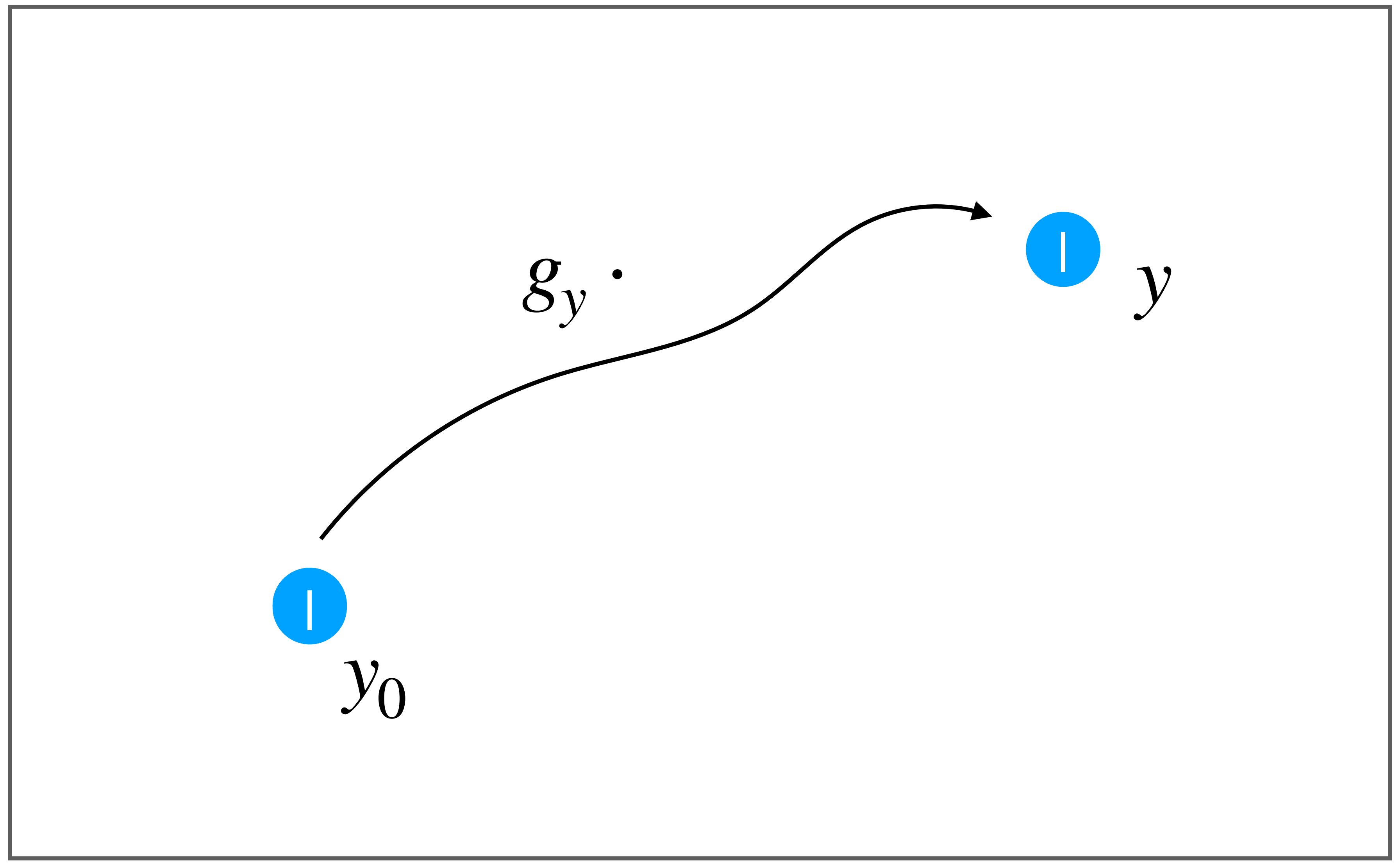
Thus $\tilde{k}(y, x) = \tilde{k}(g_y\, y_0, x)$

(since $\tilde{k}(y, x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\tilde{k}(y,x) = \tilde{k}(g_y\, y_0, x)$

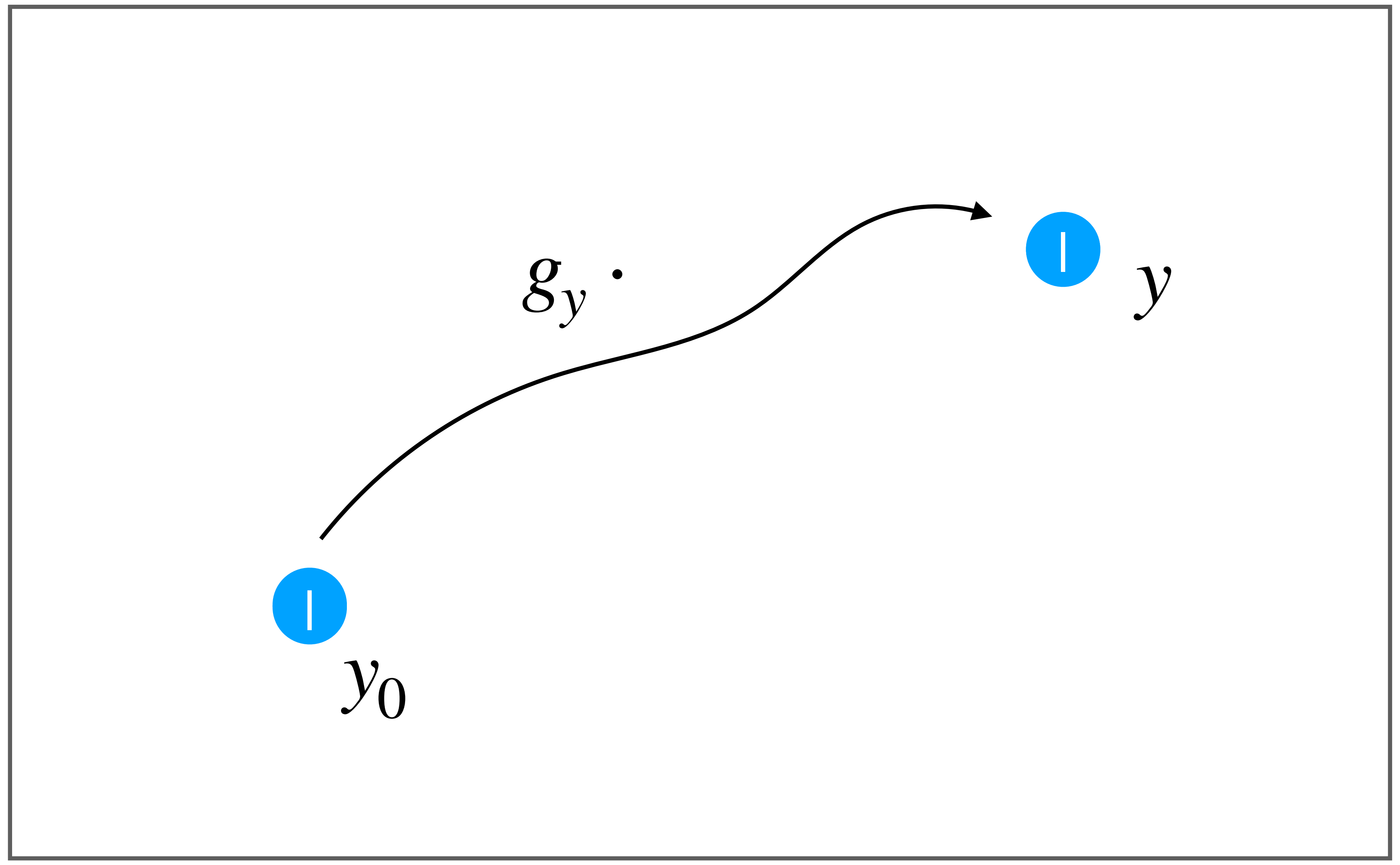
(since $\tilde{k}(y,x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

$$= \frac{1}{|\det g_y|}k(g_y^{-1}\, x)$$

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\tilde{k}(y,x) = \tilde{k}(g_y\, y_0, x)$

(since $\tilde{k}(y,x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

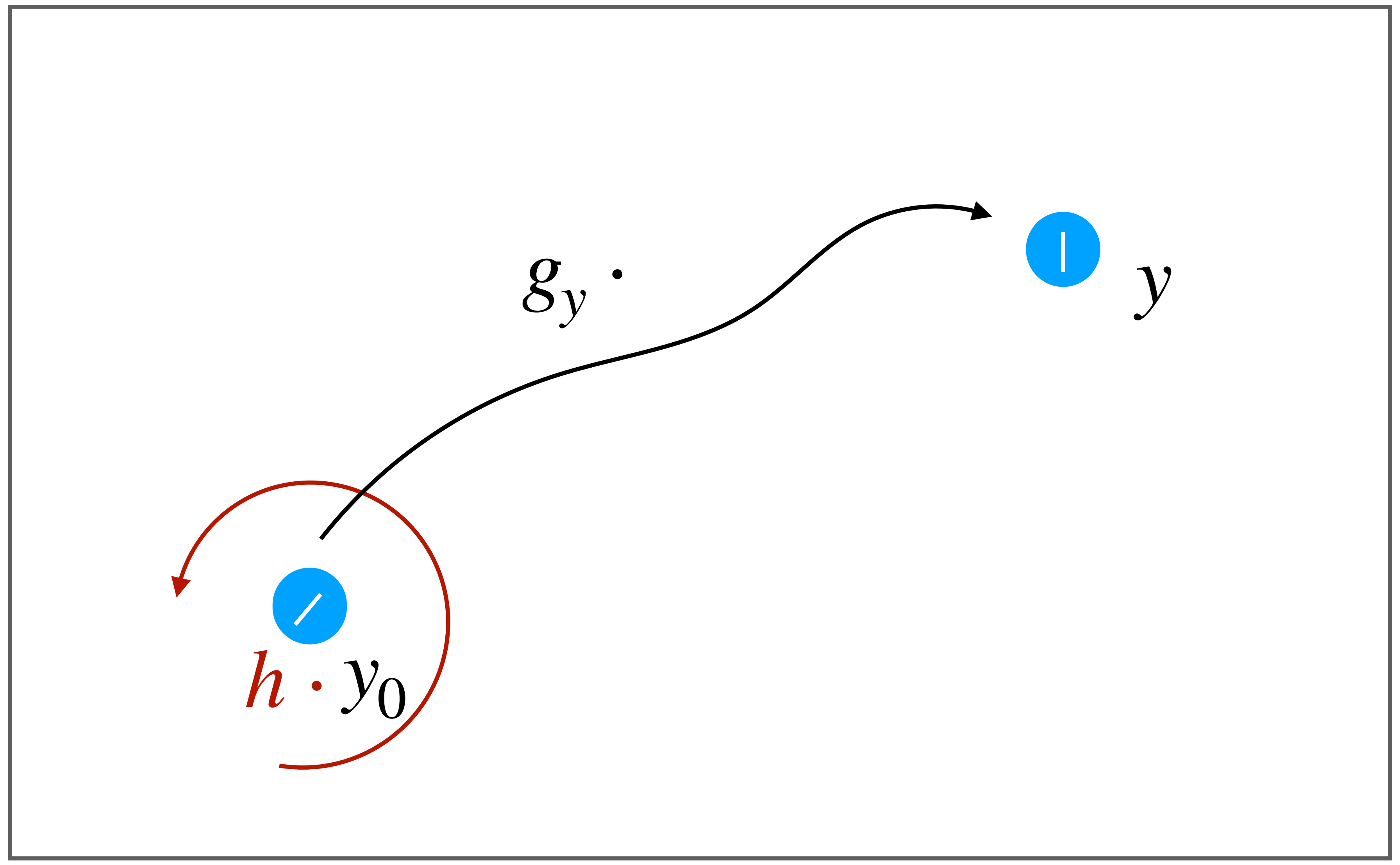
$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

$$= \frac{1}{|\det g_y|}k(g_y^{-1}\, x)$$

If $Y \equiv G/H$ then there could be a whole subgroup $H = \mathrm{Stab}_G\, y_0$ that leaves y_0 invariant

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y, y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\tilde{k}(y, x) = \tilde{k}(g_y\, y_0, x)$

(since $\tilde{k}(y, x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

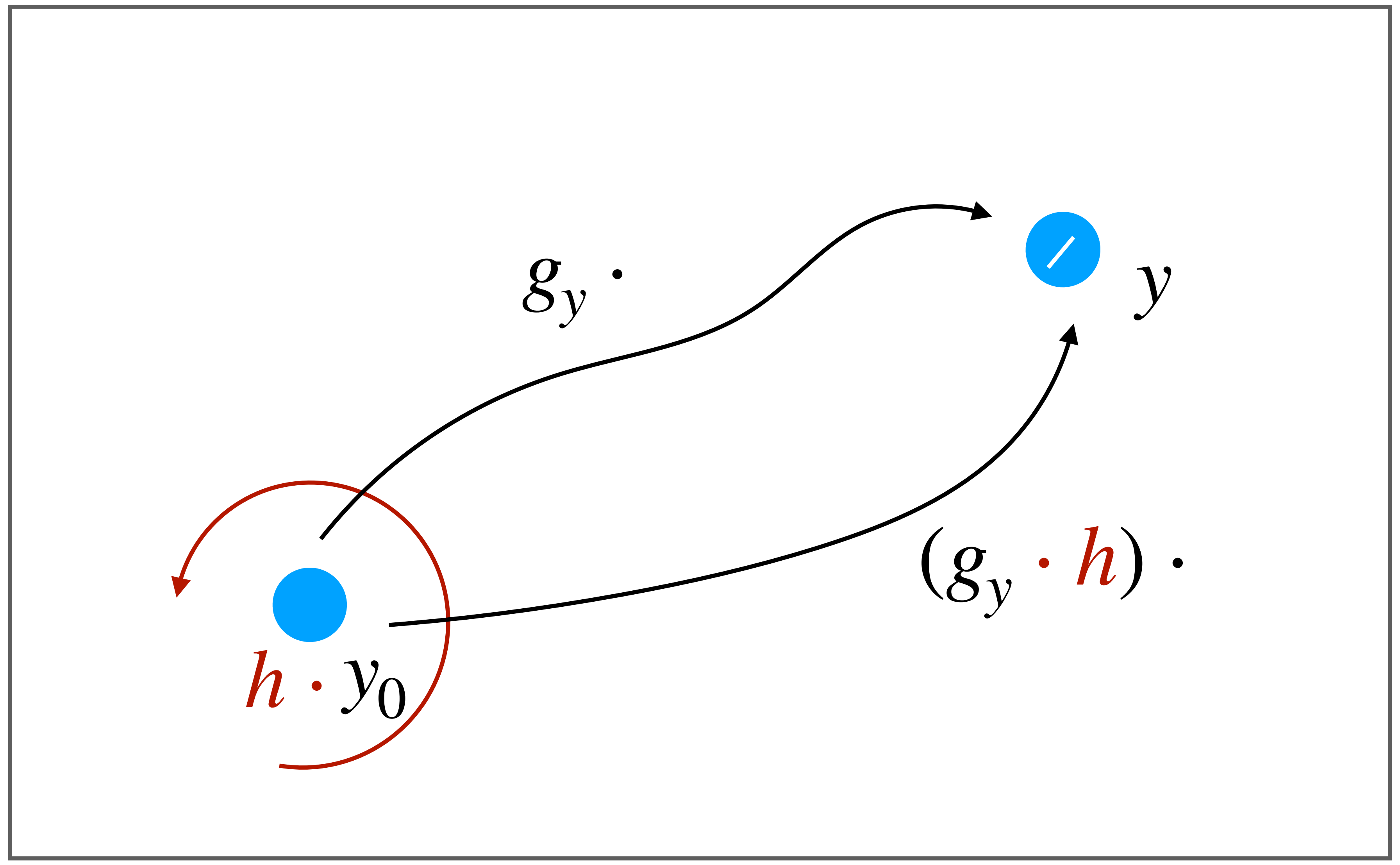
$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

$$= \frac{1}{|\det g_y|}k(g_y^{-1}\, x)$$

If $Y \equiv G/H$ then there could be a whole subgroup $H = \mathrm{Stab}_G\, y_0$ that leaves y_0 invariant

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\quad \tilde{k}(y, x) = \tilde{k}(g_y\, y_0, x)$

(since $\tilde{k}(y,x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

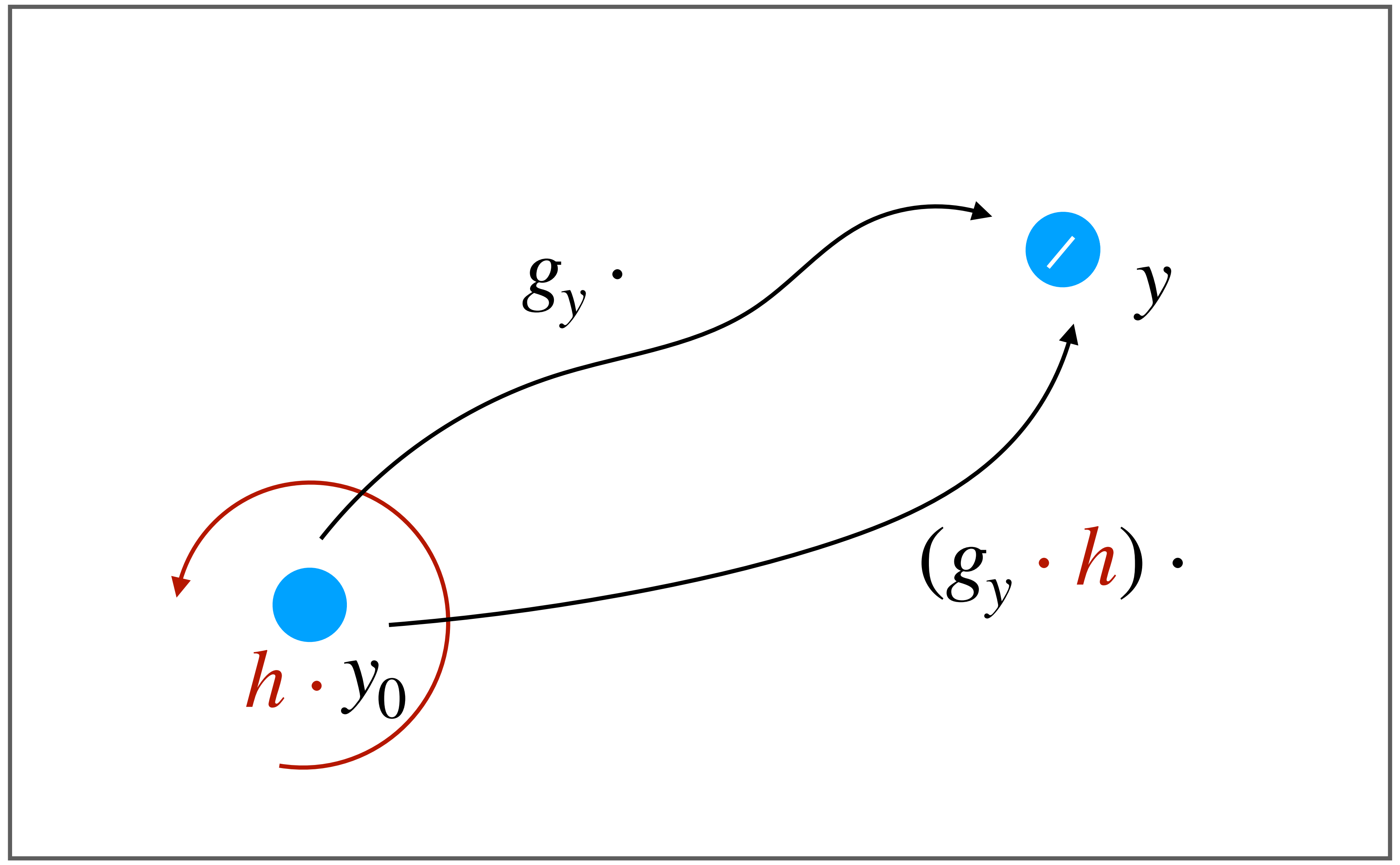
$$= \frac{1}{|\det g_y|}k(g_y^{-1}\, x)$$

If $Y \equiv G/H$ then there could be a whole subgroup $H = \mathrm{Stab}_G\, y_0$ that leaves y_0 invariant

$\forall_{h \in H}: \quad \tilde{k}(h\, y_0, x) = \tilde{k}(y_0, x)$

Proof (G-convs are all you need)

Since G acts transitively on Y we have that $\forall_{y,y_0 \in Y} \exists_{g_y \in G}$ such that $y = g_y y_0$



Thus $\tilde{k}(y,x) = \tilde{k}(g_y\, y_0, x)$

(since $\tilde{k}(y,x) = \frac{1}{|\det g|}\tilde{k}(g^{-1}y, g^{-1}x)$)

$$= \frac{1}{|\det g_y|}\tilde{k}(y_0, g_y^{-1}\, x)$$

$$= \frac{1}{|\det g_y|}k(g_y^{-1}\, x)$$

If $Y \equiv G/H$ then there could be a whole subgroup $H = \mathrm{Stab}_G\, y_0$ that leaves y_0 invariant

$$\forall_{h \in H}: \quad \tilde{k}(h\, y_0, x) = \tilde{k}(y_0, x) \quad \Leftrightarrow \quad k(x) = \frac{1}{|\det h|}k(h^{-1}x)$$

Theorem (G-convs are all you need)

Bekkers ICLR 2020, Thm. 1*

Let $\mathscr{K} : \mathbb{L}_2(X) \rightarrow \mathbb{L}_2(Y)$ map between signals on homogeneous spaces of G.

Let homogeneous space $Y \equiv G/H$ such that $H = \mathrm{Stab}_G(y_0)$ for some chosen origin $y_0 \in Y$ and let $g_y \in G$ such that $\forall_{y \in Y} : y = g_y y_0$.

Then $\mathscr{K}$ is equivariant to group G if and only if:

1. It is a group convolution: $[\mathscr{K}f](y) = \int_X \frac{1}{|g_y|} k(g_y^{-1}x) f(x) \mathrm{d}x$

2. The kernel satisfies a symmetry constraint: $\forall_{h \in H} : \frac{1}{|g_y|} k(hx) = k(x)$

*Work with Remco Duits at TU/e.

See also: Duits 2005 – Thm 25, Cohen, Geiger, Weiler 2018 - Thm 6.1, Kondor, Trivedi 2018 - Thm 1

Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \to \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

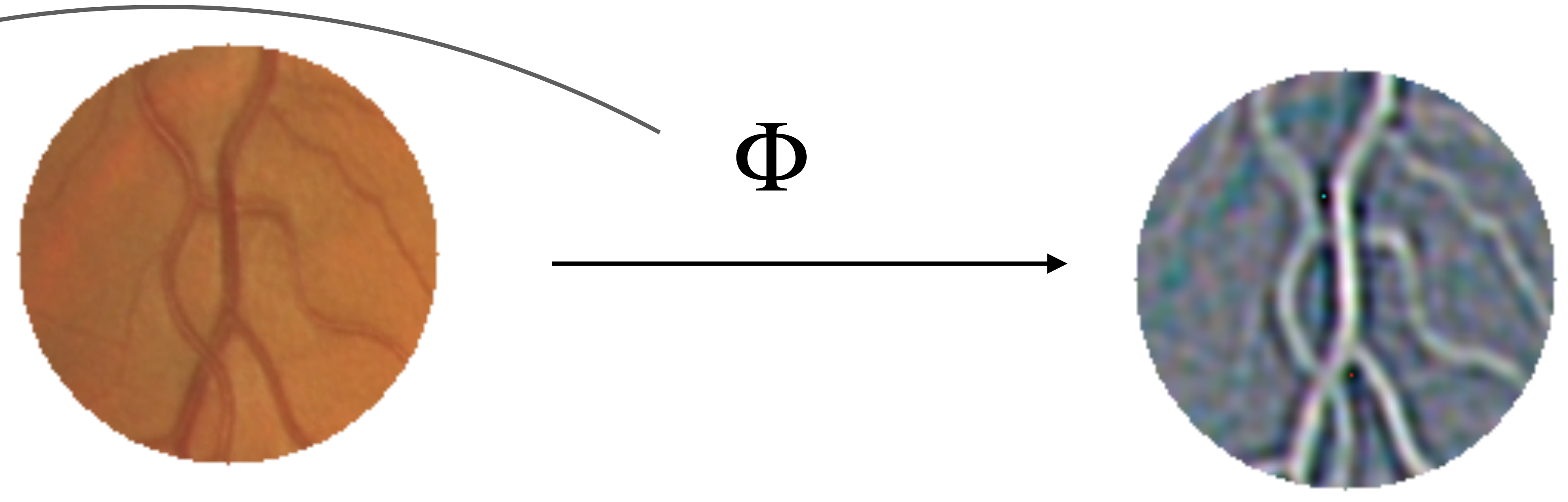
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

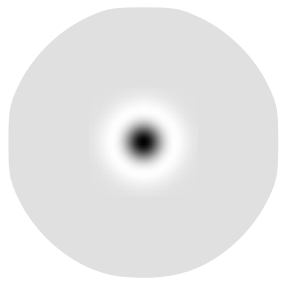
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$



Conv2D(input , )

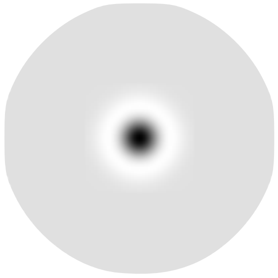
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

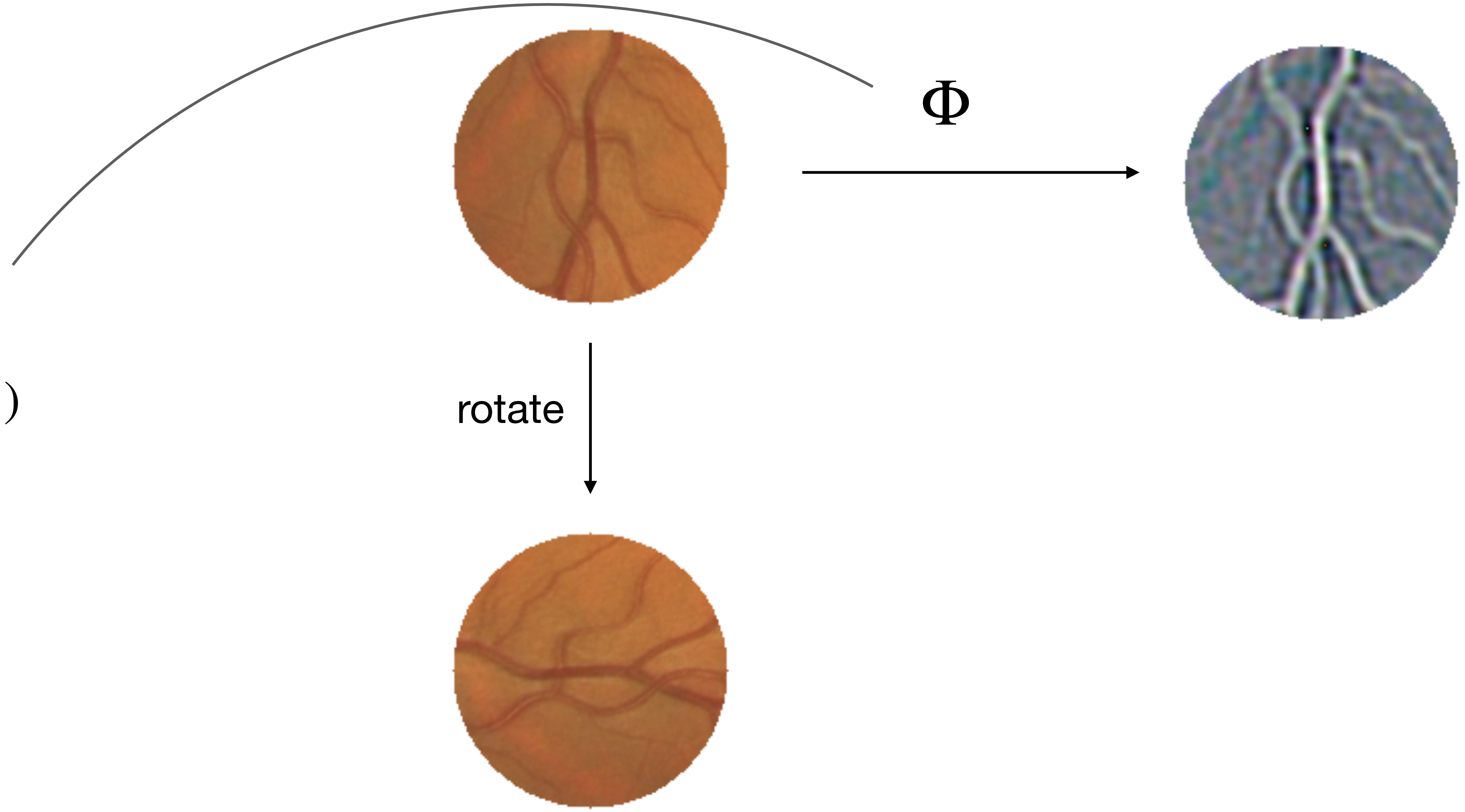
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )



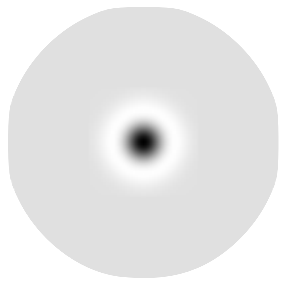
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

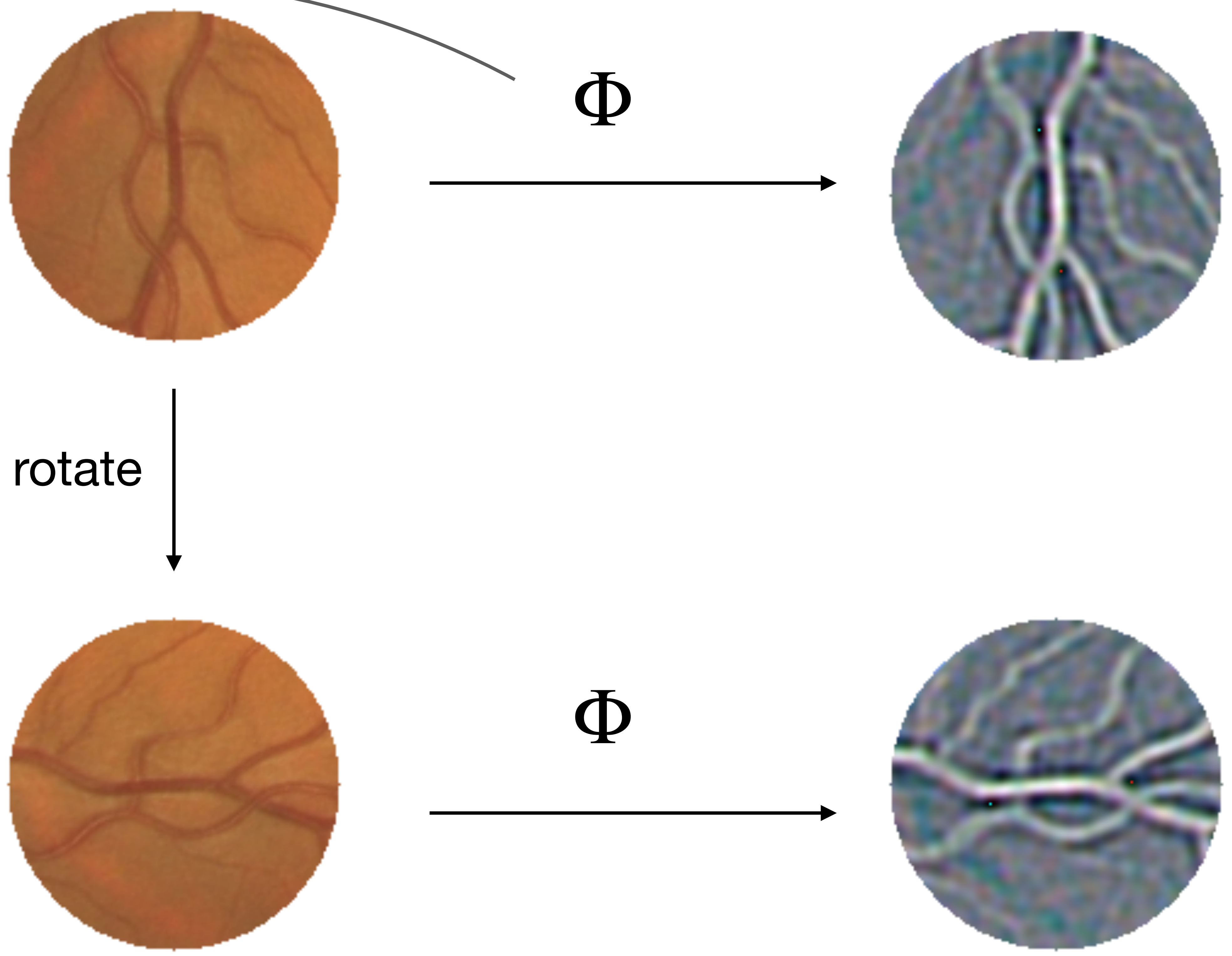
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )



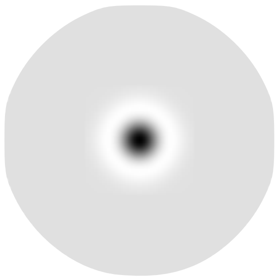
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

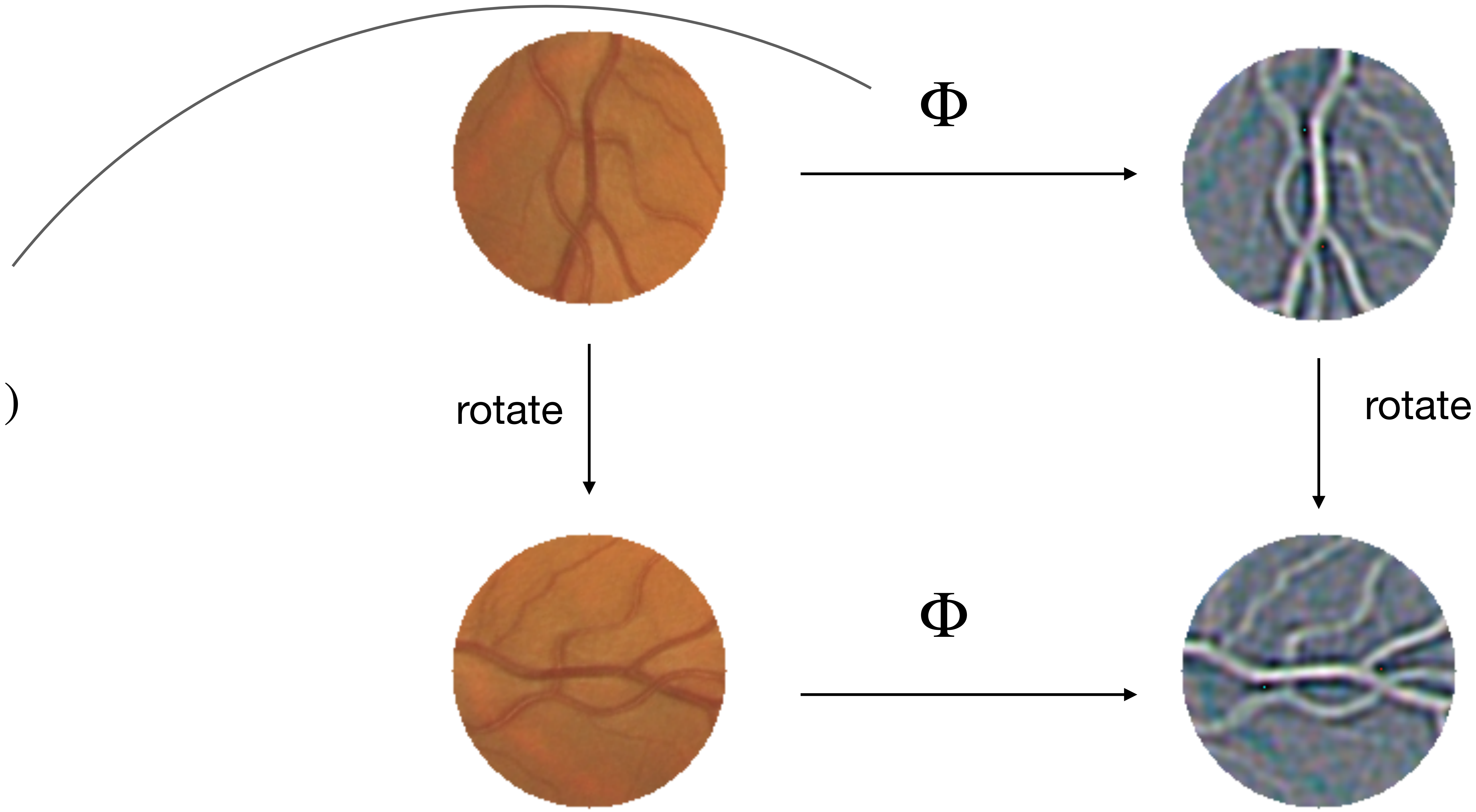
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )



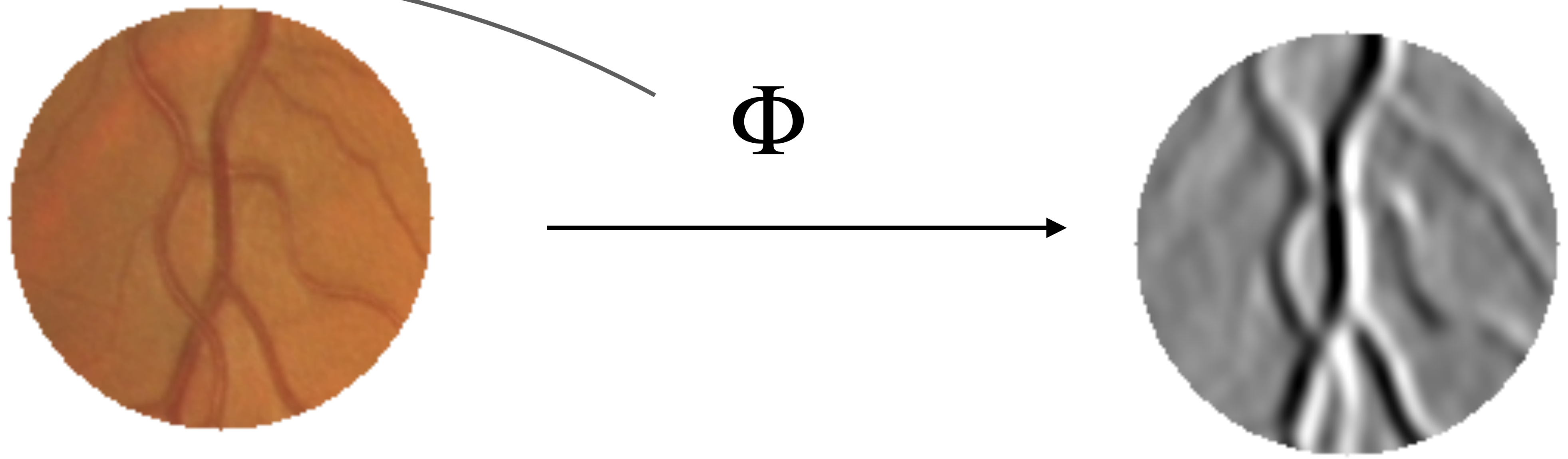
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

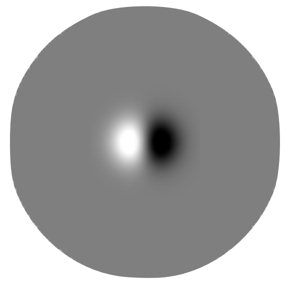
$(X = Y = G/H)$

~~**Isotropic/Constraint convolutions** on spaces of lower dimension than G, $\forall_{h \in H}: k(hx) = k(x)$~~

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$



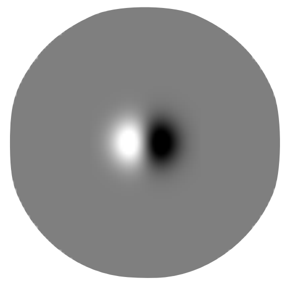
Conv2D(input , )

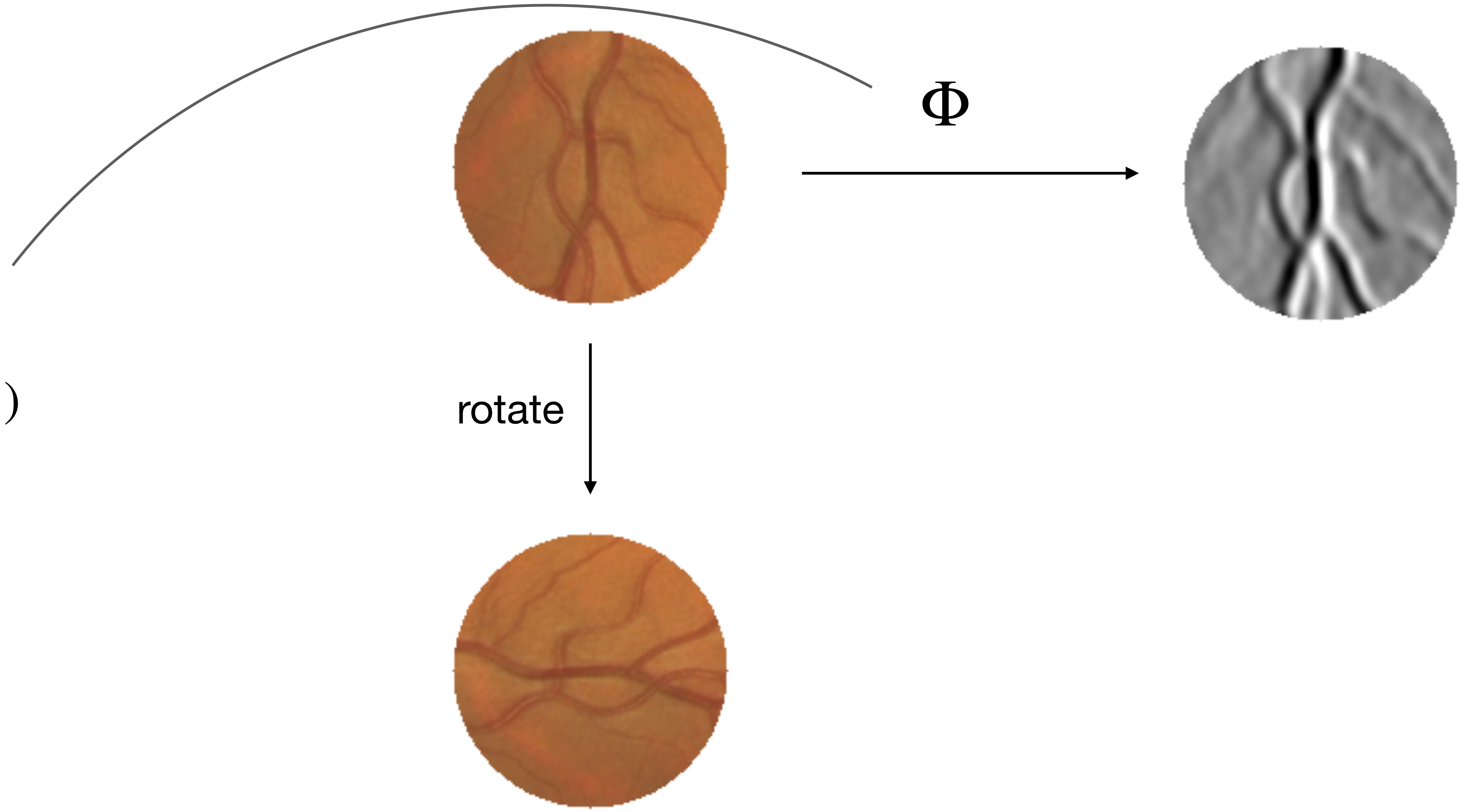
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \to \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN
$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )

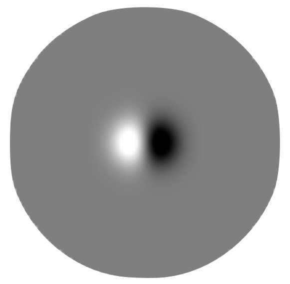


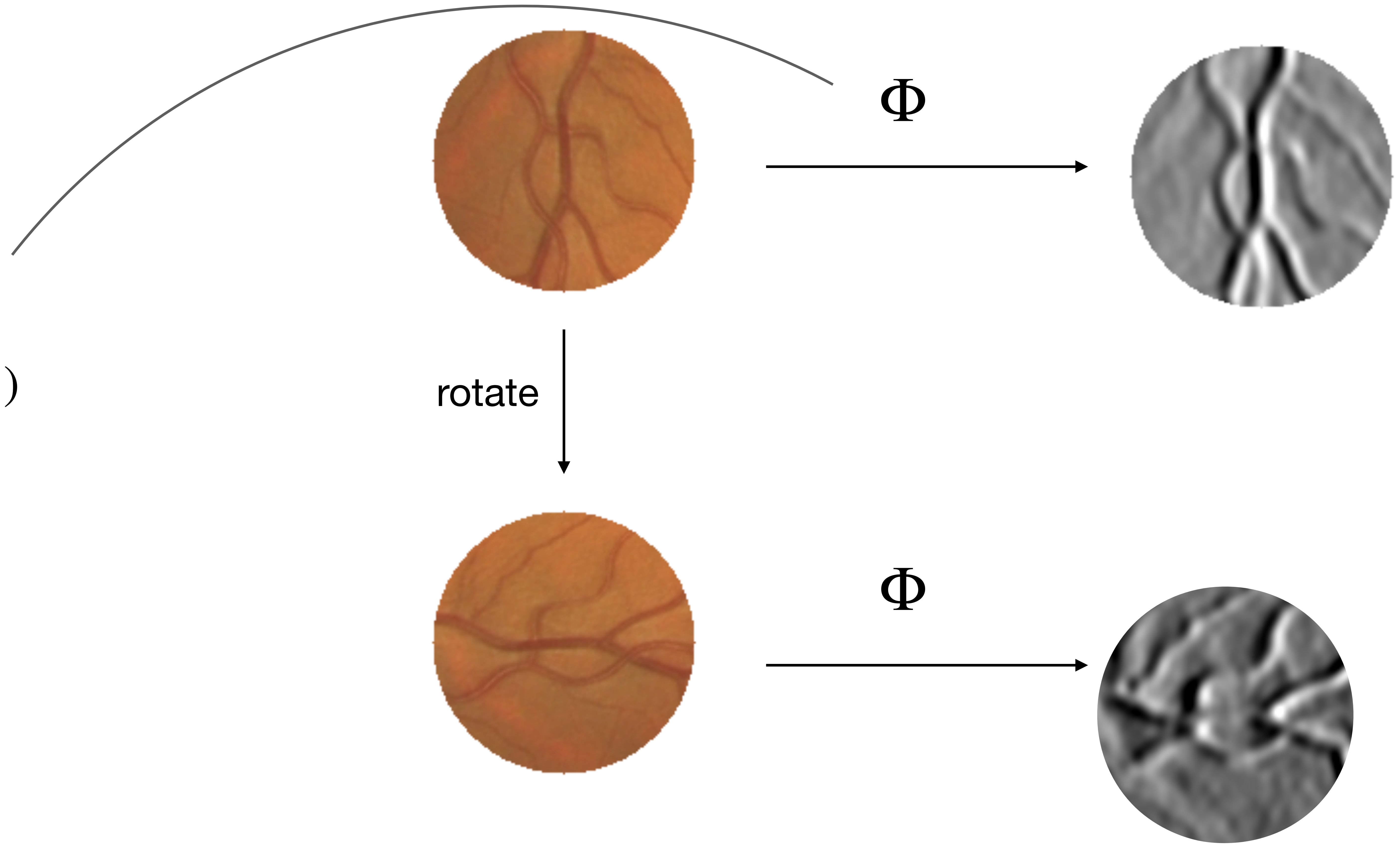
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN
$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )

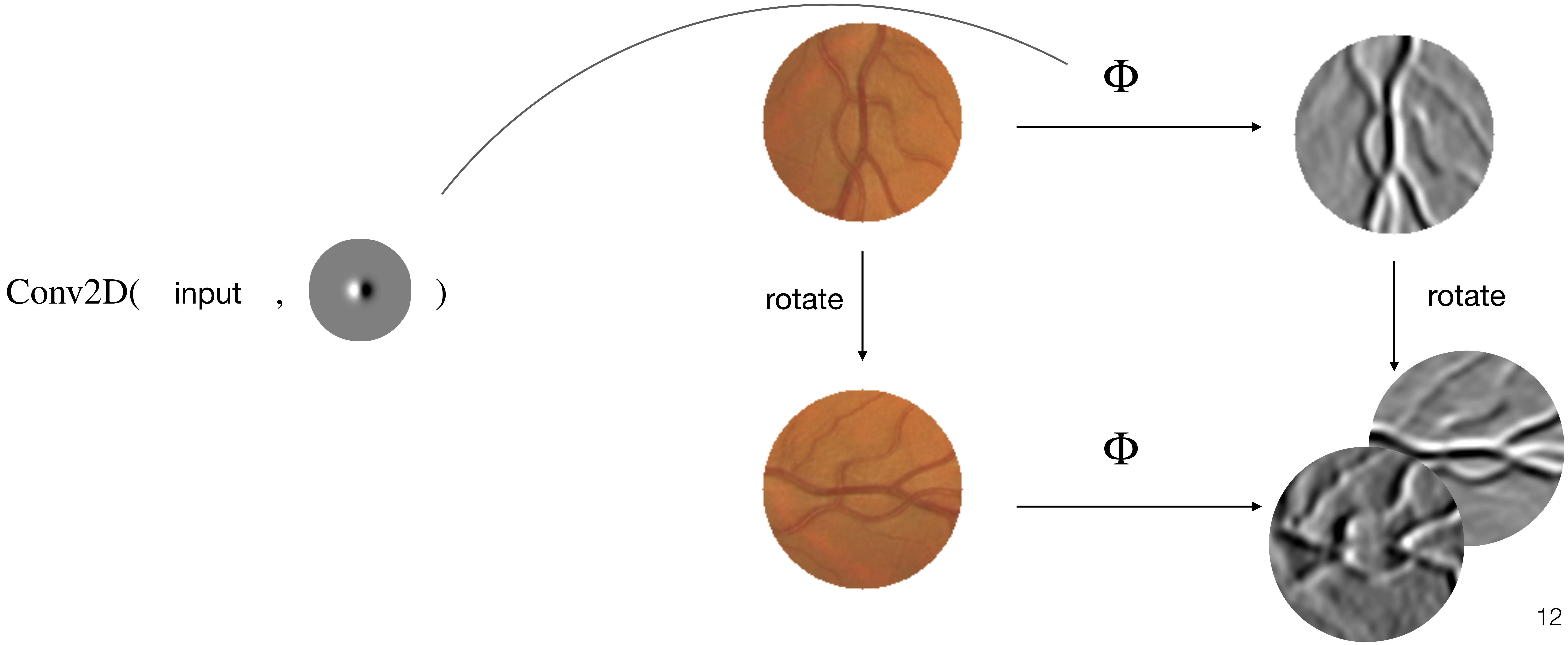


Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN
$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$



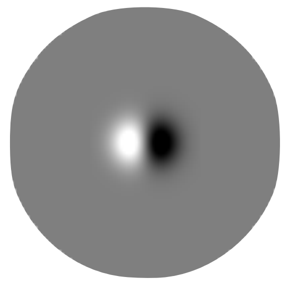
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

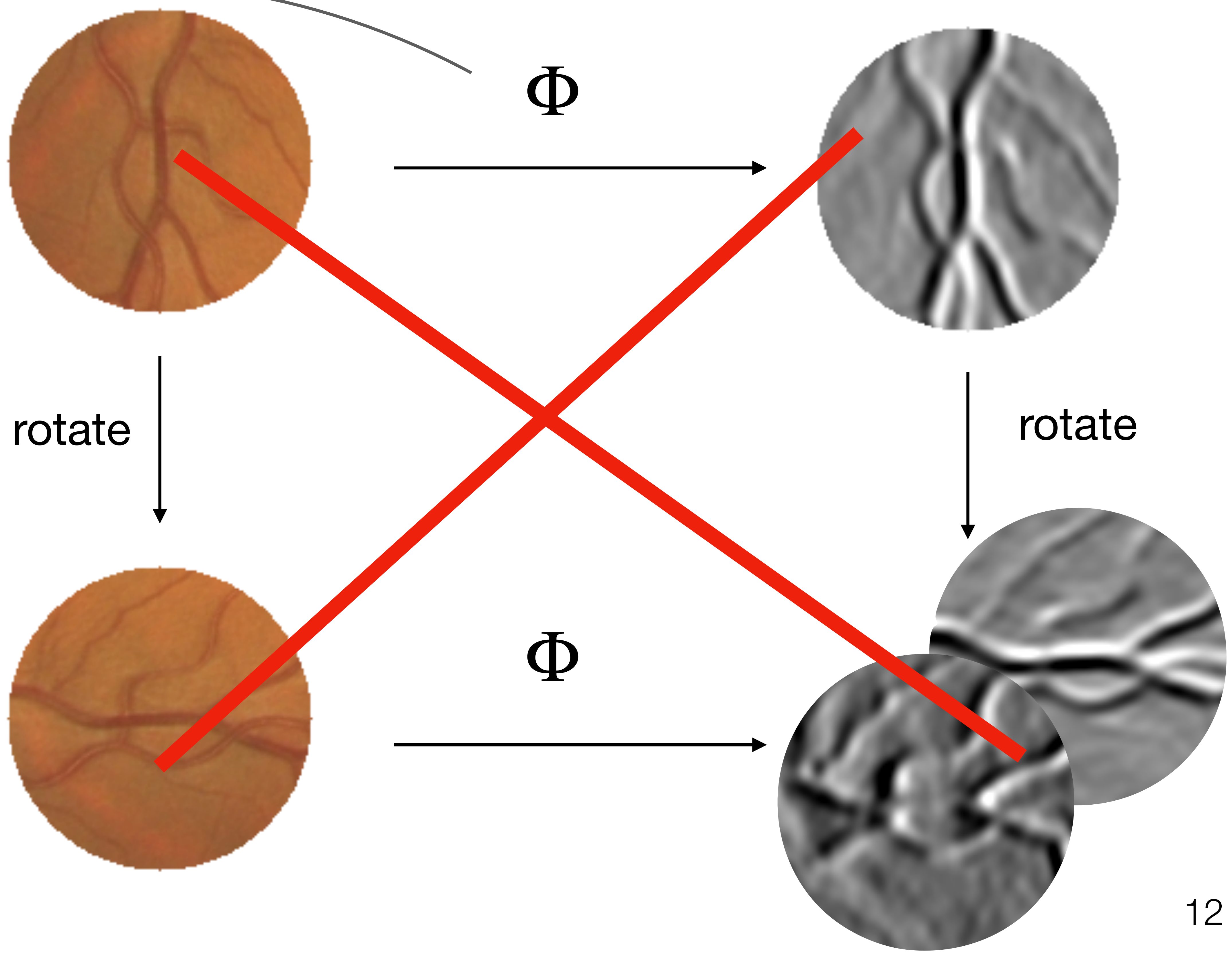
$(X = Y = G/H)$

~~**Isotropic/Constraint convolutions** on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$~~

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )



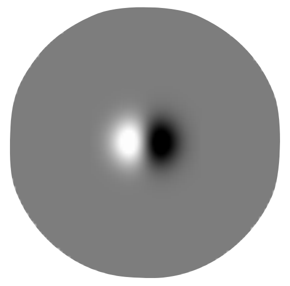
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

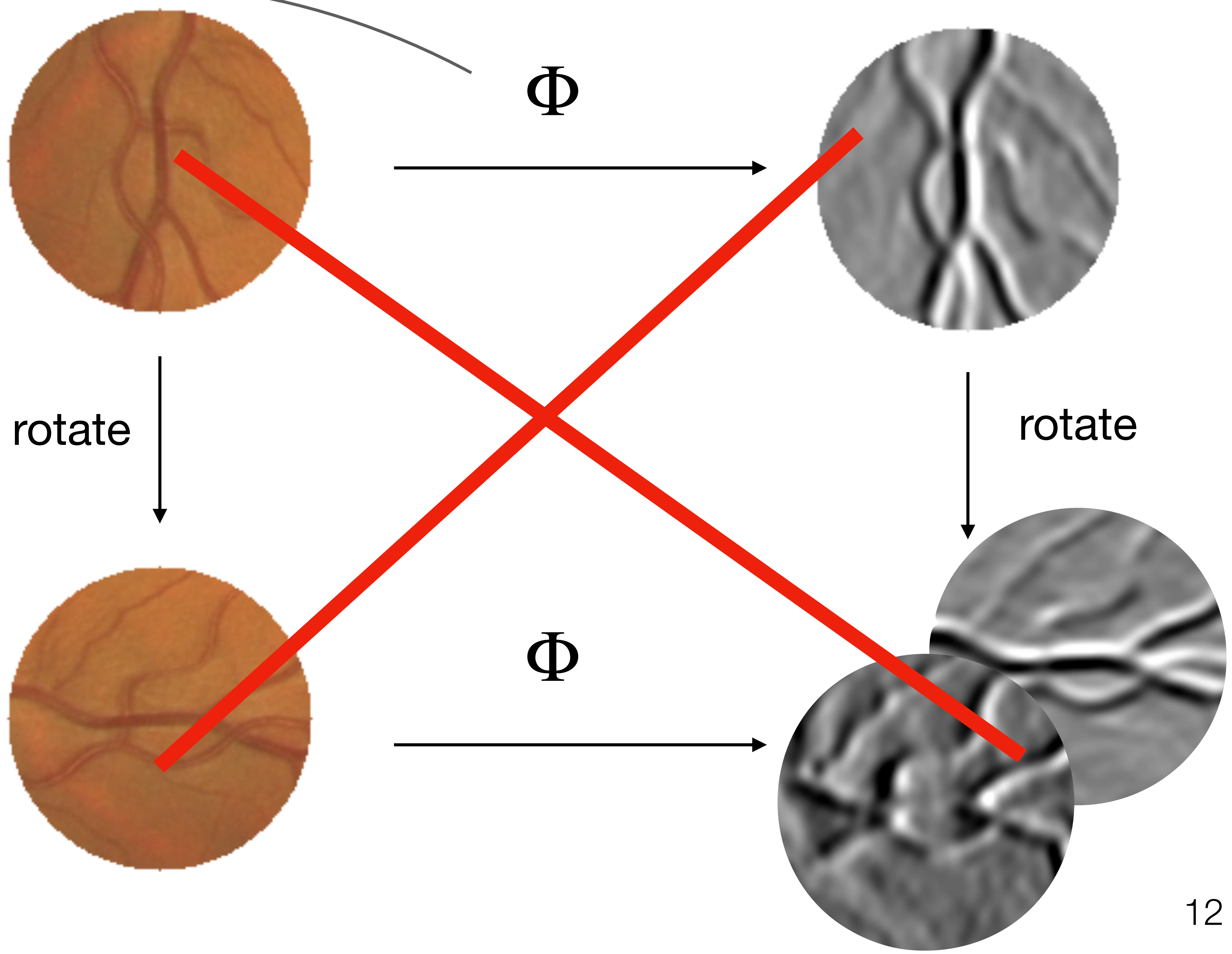
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

Example 2D CNN

$X = \mathbb{R}^2 \equiv SE(2)/SO(2)$

Conv2D(input , )



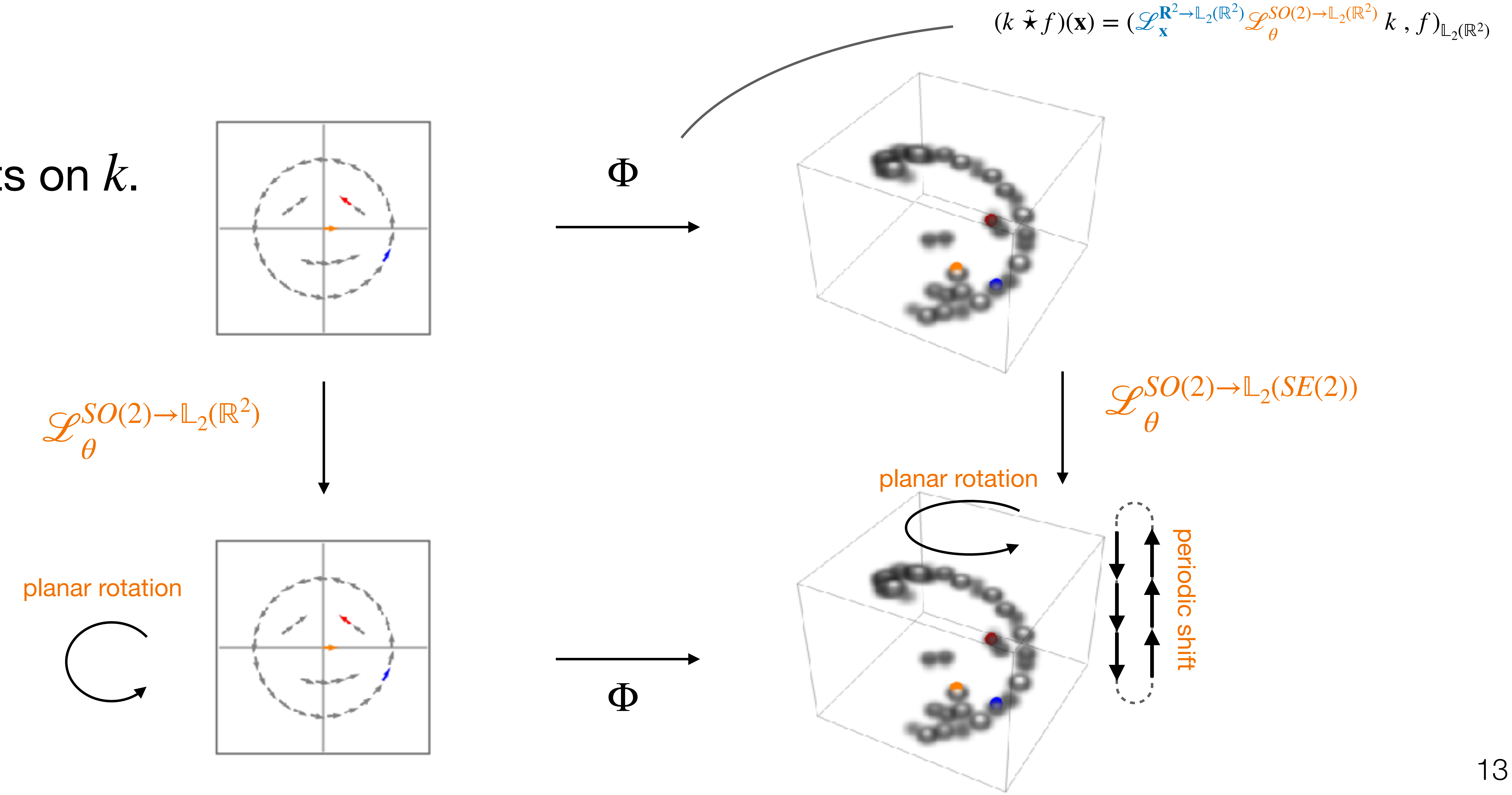
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \to \mathbb{L}_2(SE(2))$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

$(X = G/H, Y = G)$

Lifting convolution. No constraints on k.



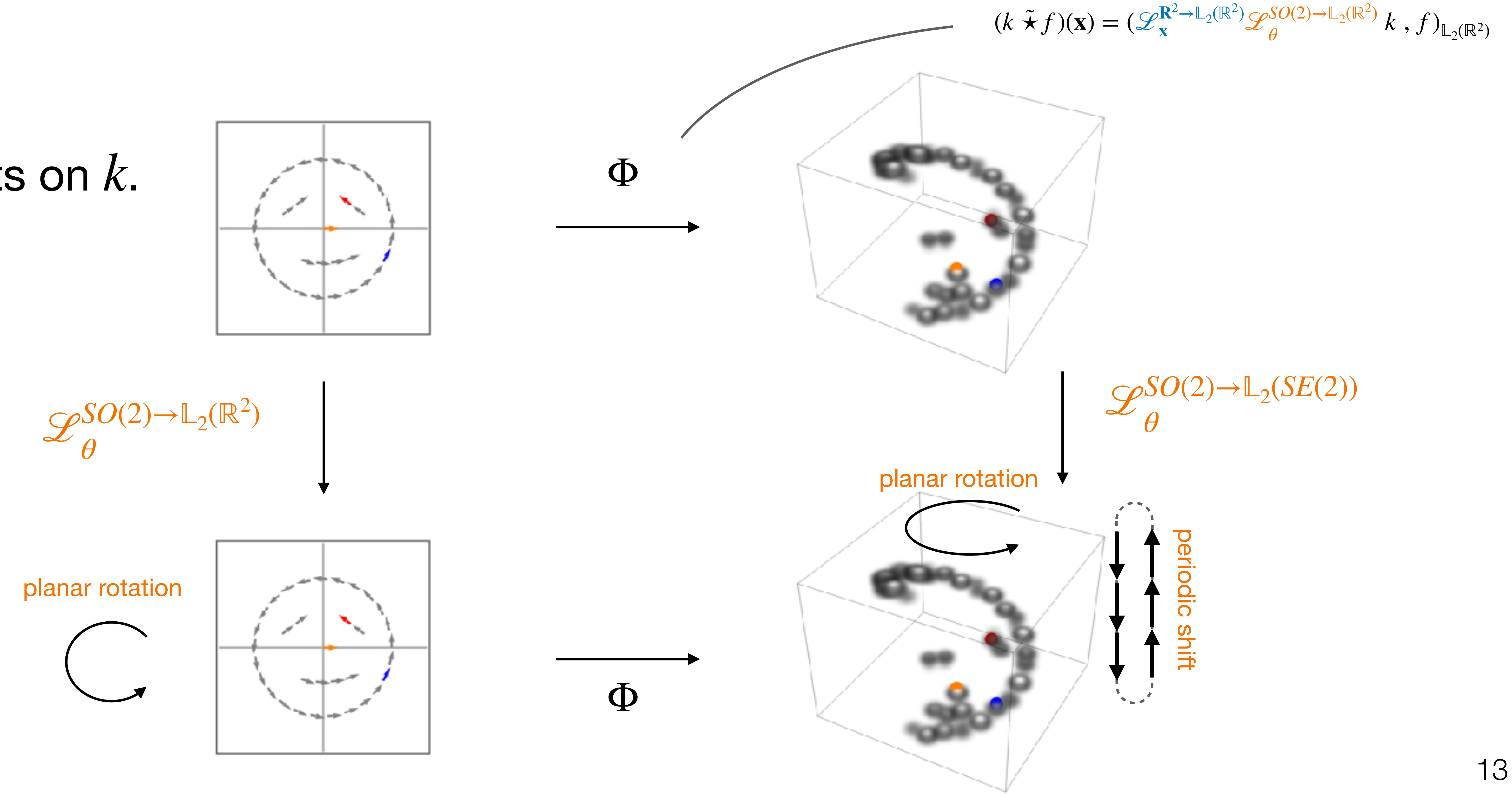
Types of layers $K : \mathbb{L}_2(\mathbb{R}^2) \to \mathbb{L}_2(SE(2))$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

$(X = G/H, Y = G)$

Lifting convolution. No constraints on k.



Types of layers $K : \mathbb{L}_2(SE(2)) \rightarrow \mathbb{L}_2(SE(2))$

$(X = Y = G/H)$

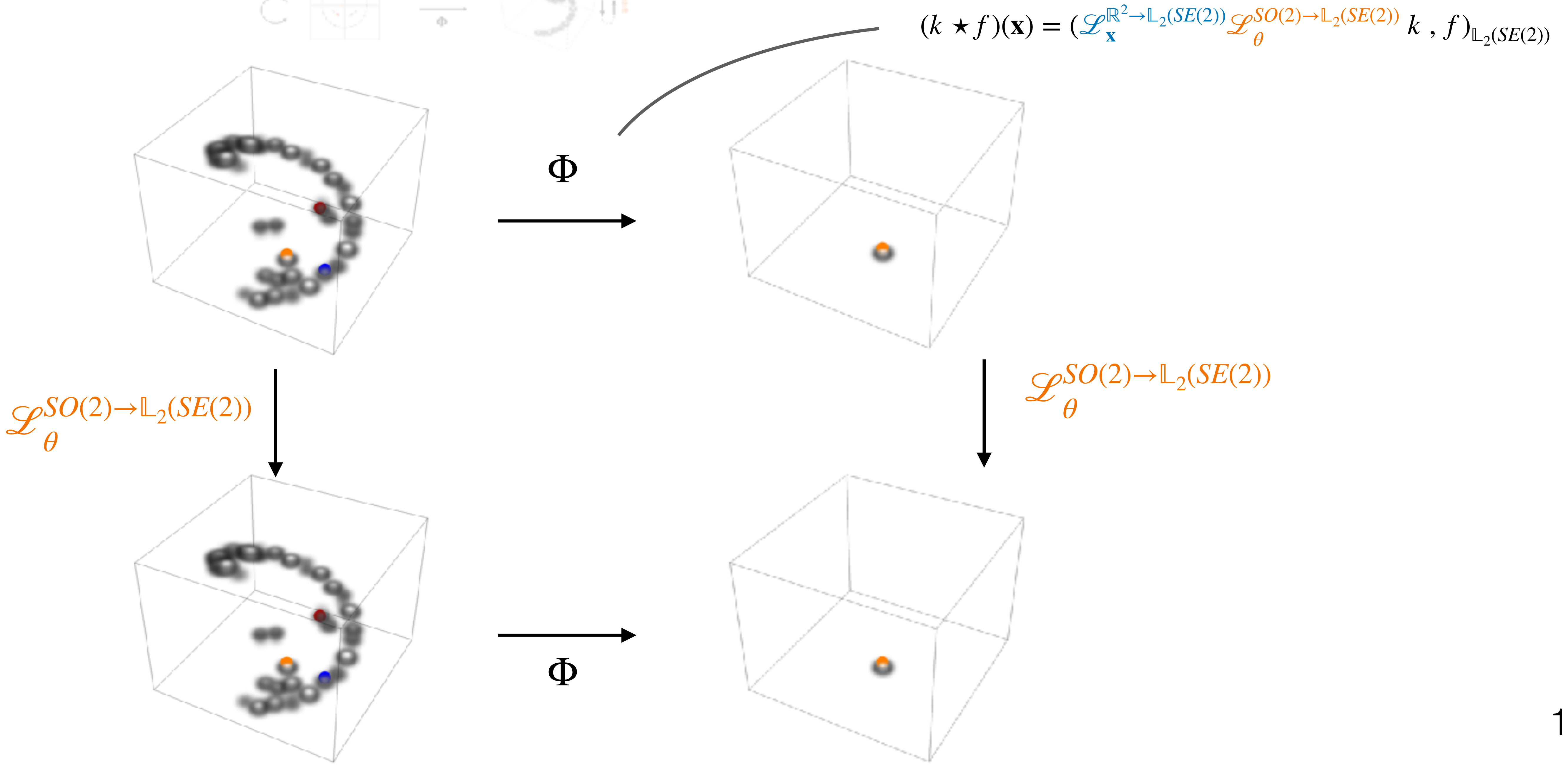
Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

$(X = G/H, Y = G)$

Lifting convolution. No constraints on k.

$(X = Y = G)$

Group convolutions. No constraints on k.

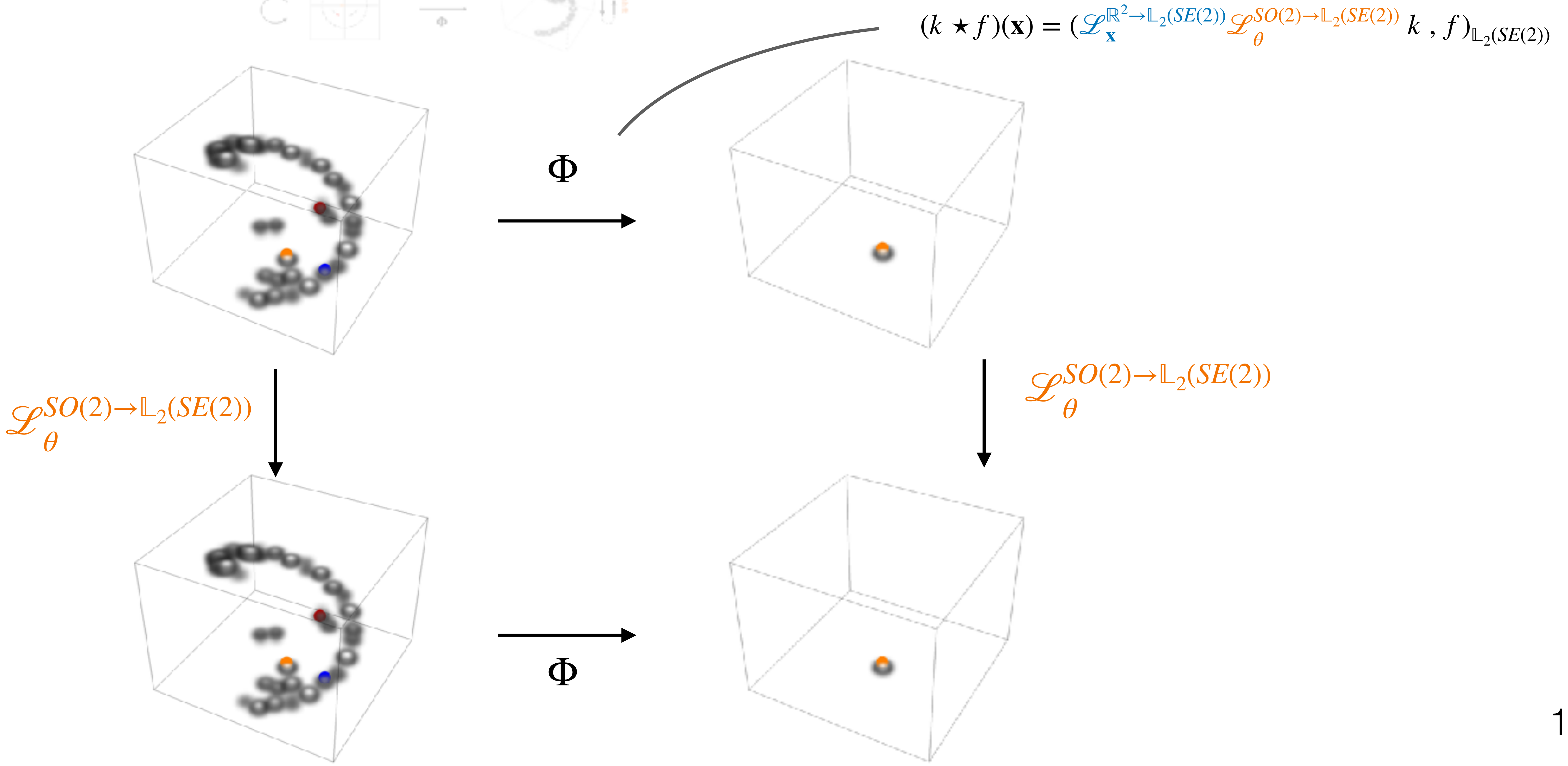


Types of layers $K : \mathbb{L}_2(SE(2)) \rightarrow \mathbb{L}_2(SE(2))$

$(X = Y = G/H)$
Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

$(X = G/H, Y = G)$
Lifting convolution. No constraints on k.

$(X = Y = G)$
Group convolutions. No constraints on k.



Types of layers $K : \mathbb{L}_2(SE(2)) \rightarrow \mathbb{L}_2(\mathbb{R}^2)$

$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$

$(X = G/H, Y = G)$

Lifting convolution. No constraints on k.

$(X = Y = G)$

Group convolutions. No constraints on k.

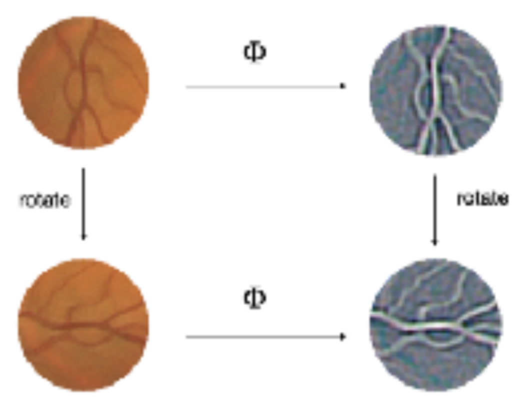
$(X = G, Y = G/H)$

Projection layer. Mean pooling over H.

Types of layers

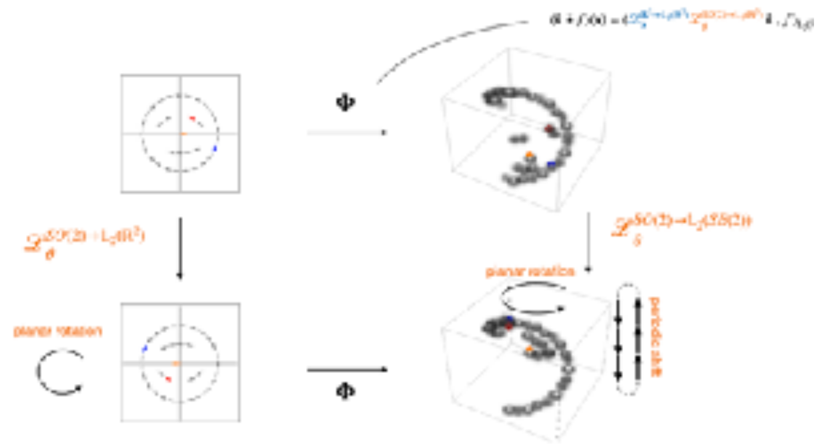
$(X = Y = G/H)$

Isotropic/Constraint convolutions on spaces of lower dimension than G, $\forall_{h \in H} : k(hx) = k(x)$



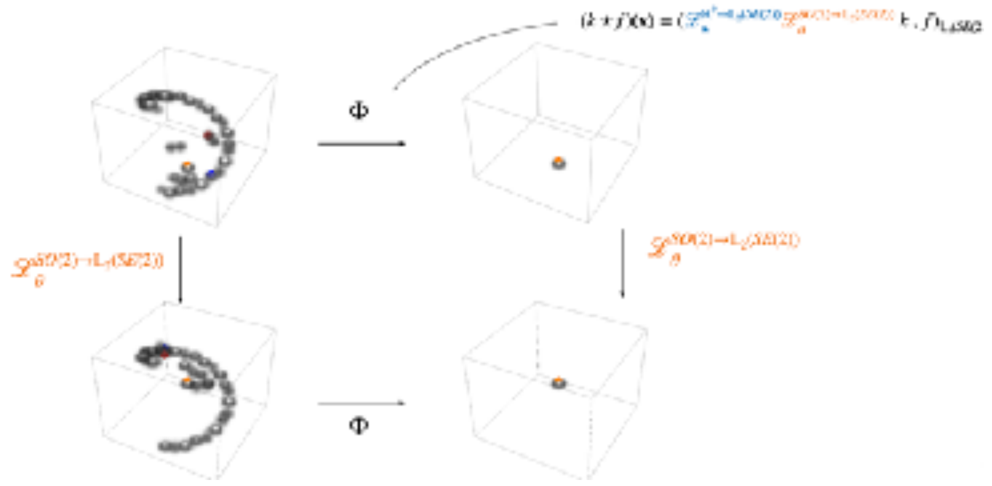
$(X = G/H, Y = G)$

Lifting convolution. No constraints on k.



$(X = Y = G)$

Group convolutions. No constraints on k.

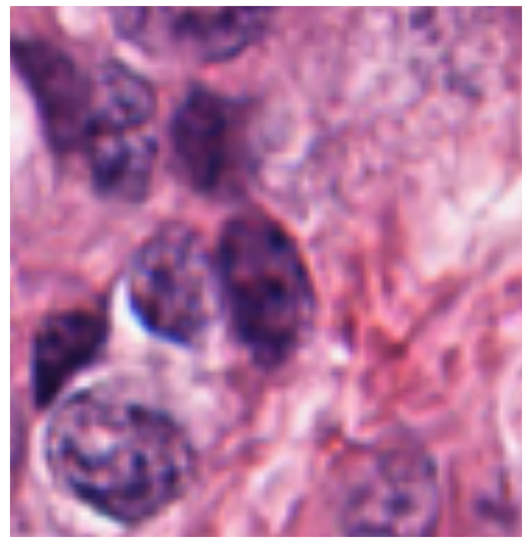


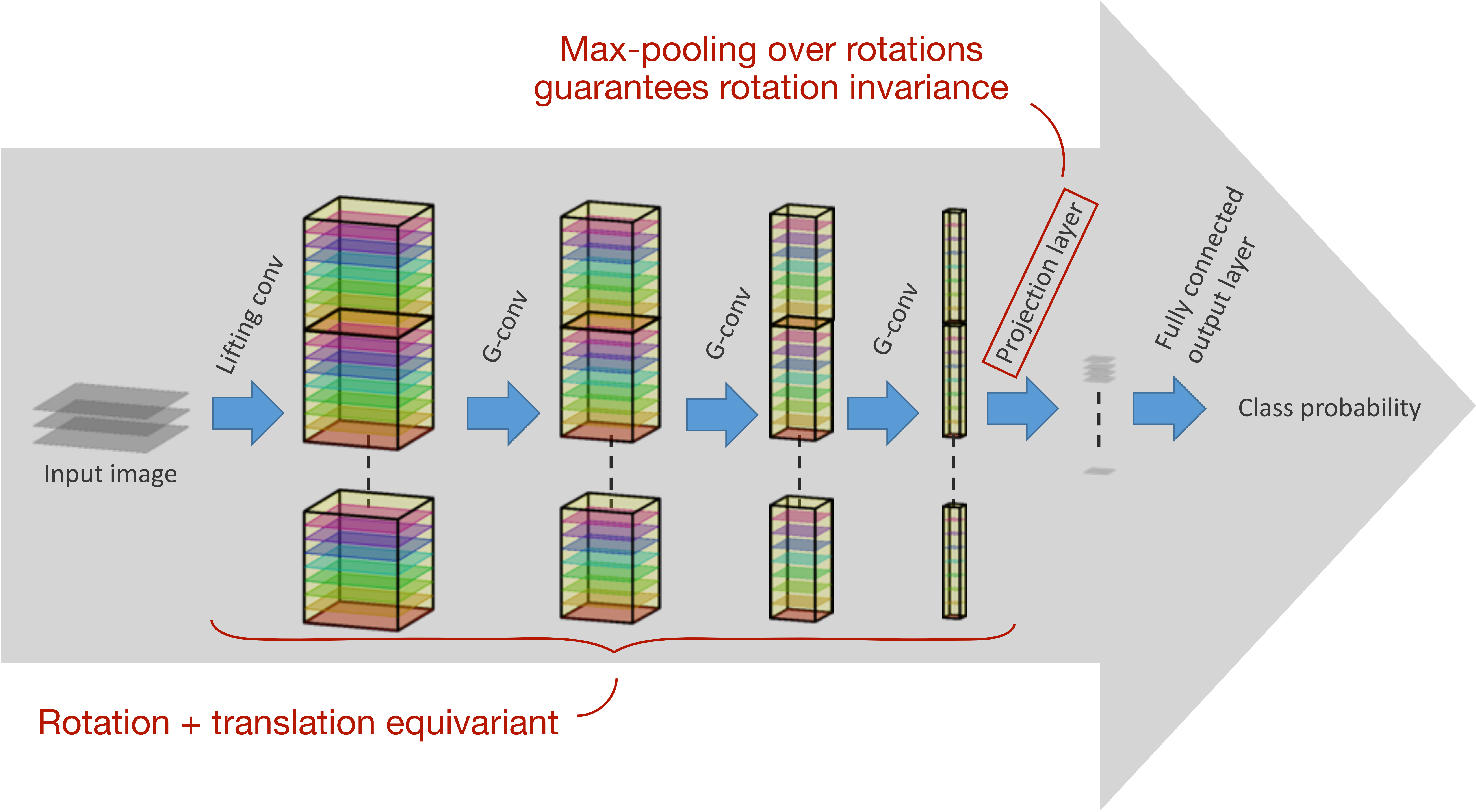
$(X = G, Y = G/H)$

Projection layer. Mean pooling over H.

The most expressive group equivariant architectures are obtained by lifting the feature maps to the group

General group equivariant architecture





“normal” (0)
vs
“mitotic” (1)

Conclusion

If you want to build **equivariant neural networks**

Conclusion

If you want to build **equivariant neural networks**

Group convolutions are all you need!