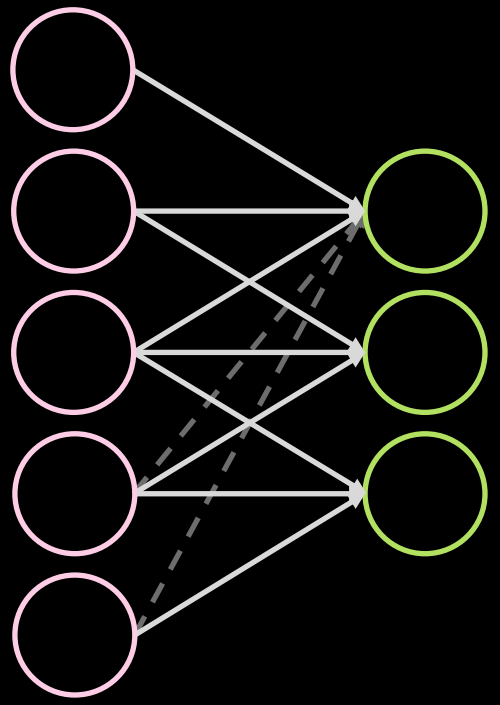
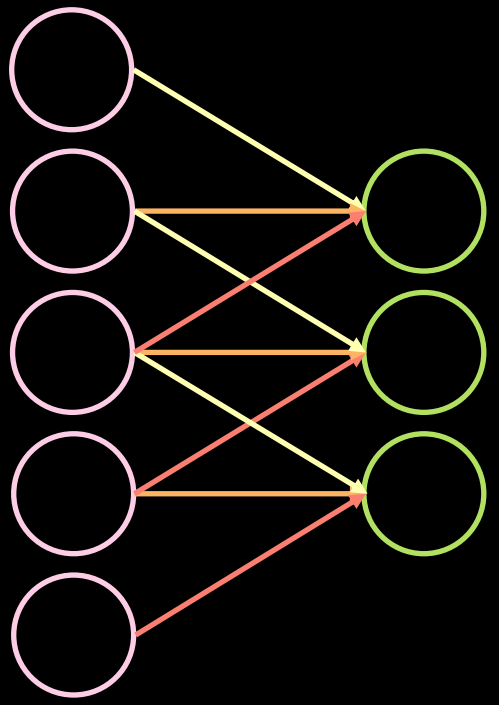
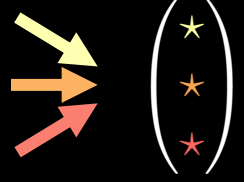
Stationarity ⇒ parameters sharing



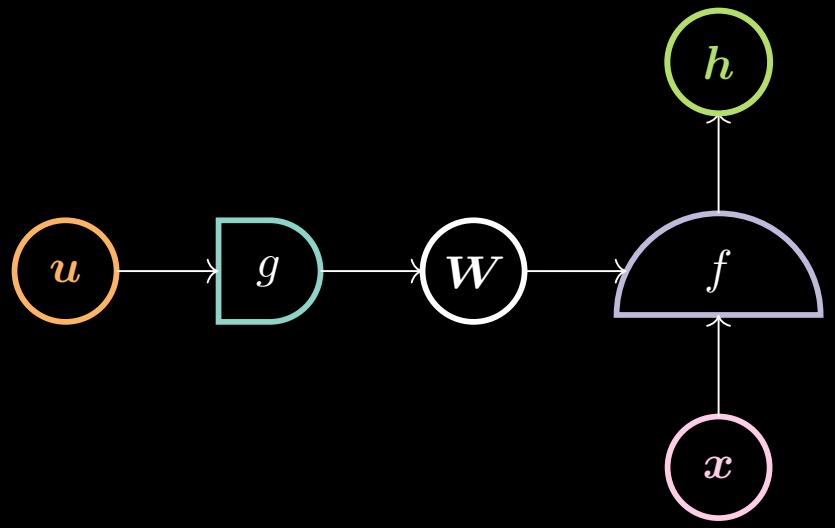


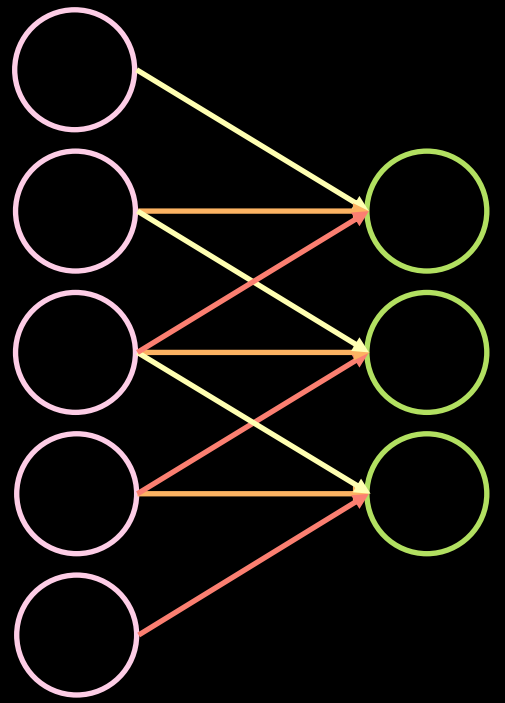
$$\boldsymbol{W} = \begin{bmatrix} & & & 0 & 0 \\ & & & & 0 \\ 0 & & & & \\ 0 & 0 & & & \end{bmatrix}$$



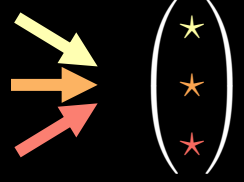
Stationarity ⇒ parameters sharing

$$\boldsymbol{u} = \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix}$$ kernel





$$\boldsymbol{W} = \begin{bmatrix} u_1 & u_2 & u_3 & 0 & 0 \\ 0 & u_1 & u_2 & u_3 & 0 \\ 0 & 0 & u_1 & u_2 & u_3 \end{bmatrix}$$ Töplitz matrix



Convnet benefits

- Sparse connectivity
  - reduced amount of computation
- Weight sharing
  - better generalisation $\Leftarrow \quad \nabla_{u_1} L = \nabla_{w_{11}} L + \nabla_{w_{22}} L + \cdots$
  - not constrained to input size
  - kernel independence ⇒ high parallelisation

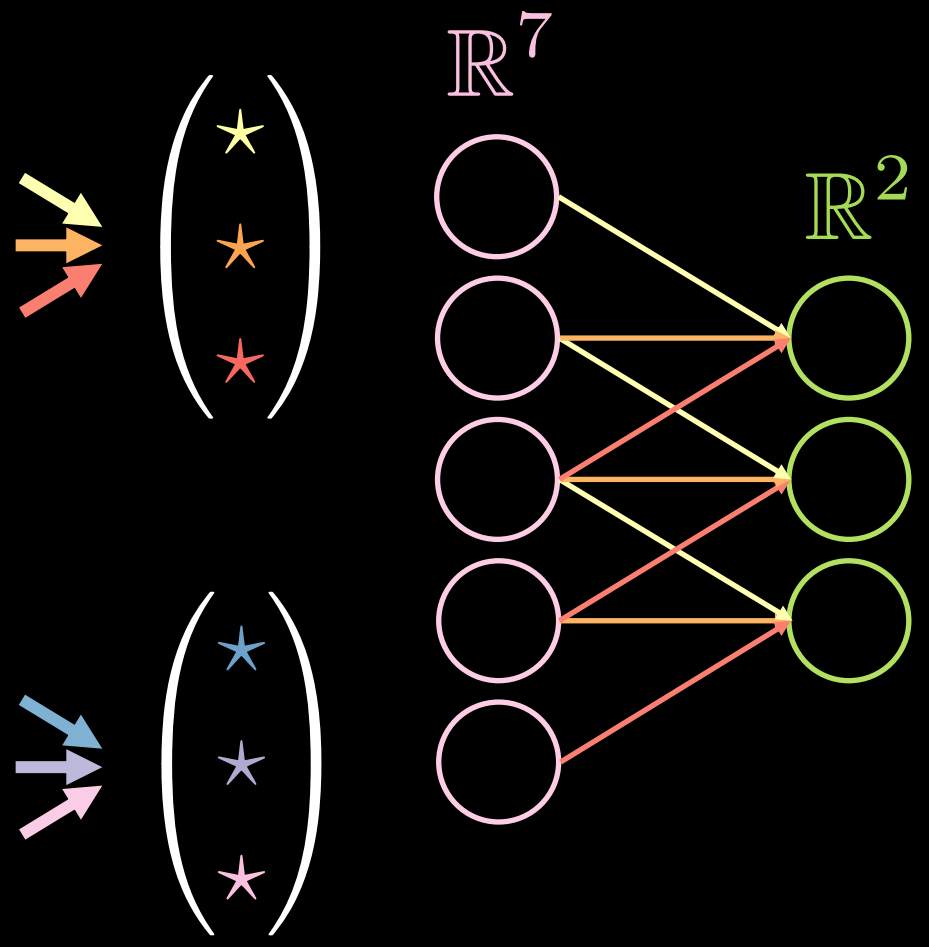
1D data uses 3D kernels-collection!

Kernels – 1D data

kernel size: $2 \times 7 \times 3$

$$\boldsymbol{u} = \begin{pmatrix} (\, \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \boldsymbol{u}_3 \,)_1 \\ (\, \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \boldsymbol{u}_3 \,)_2 \end{pmatrix}$$

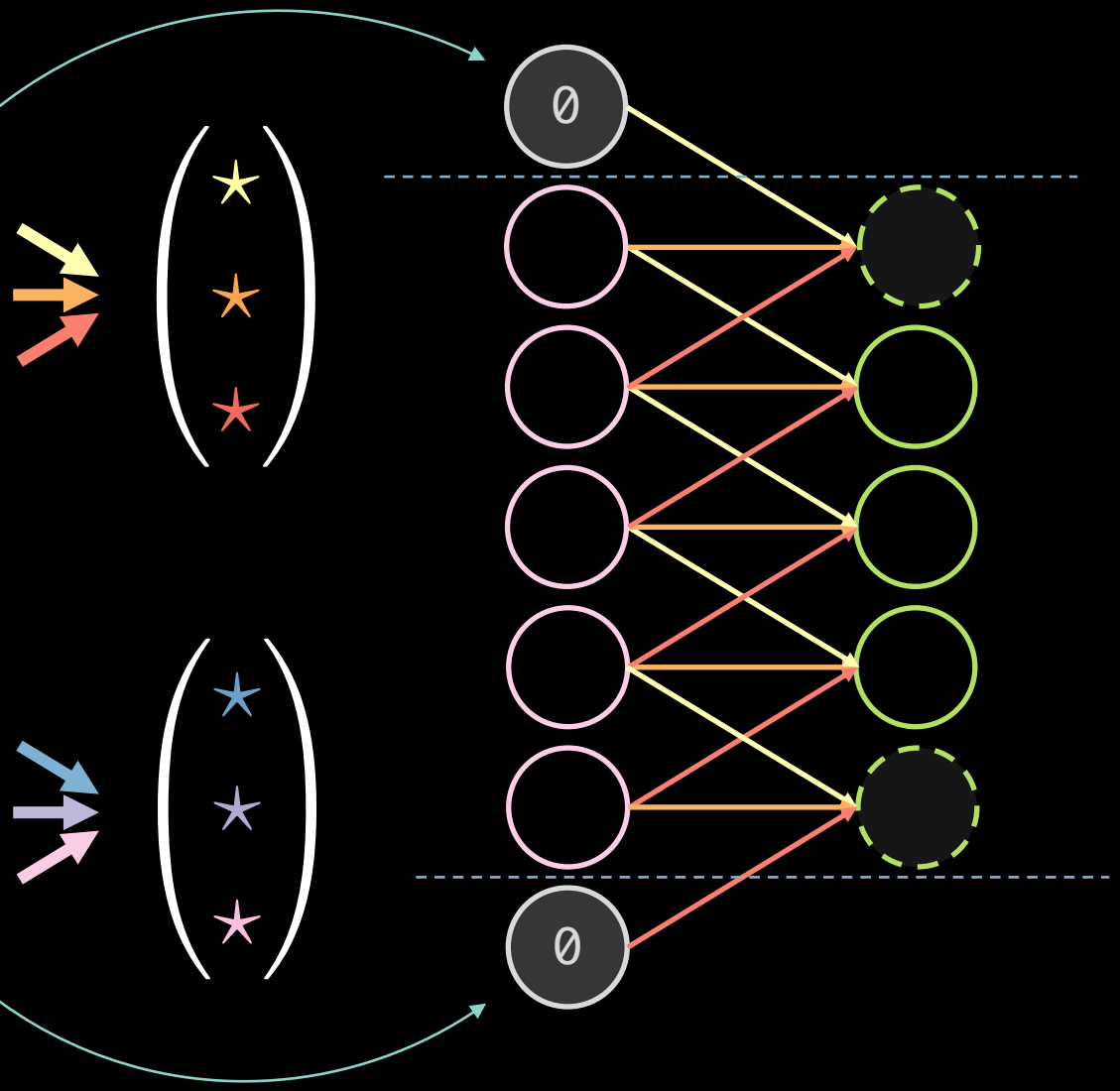
$$\boldsymbol{u}_1 = \begin{pmatrix} u_{11} \\ u_{21} \\ u_{31} \\ \vdots \\ u_{71} \end{pmatrix} \quad \boldsymbol{u}_3 = \begin{pmatrix} u_{13} \\ u_{23} \\ u_{33} \\ \vdots \\ u_{73} \end{pmatrix}$$



Padding – 1D data

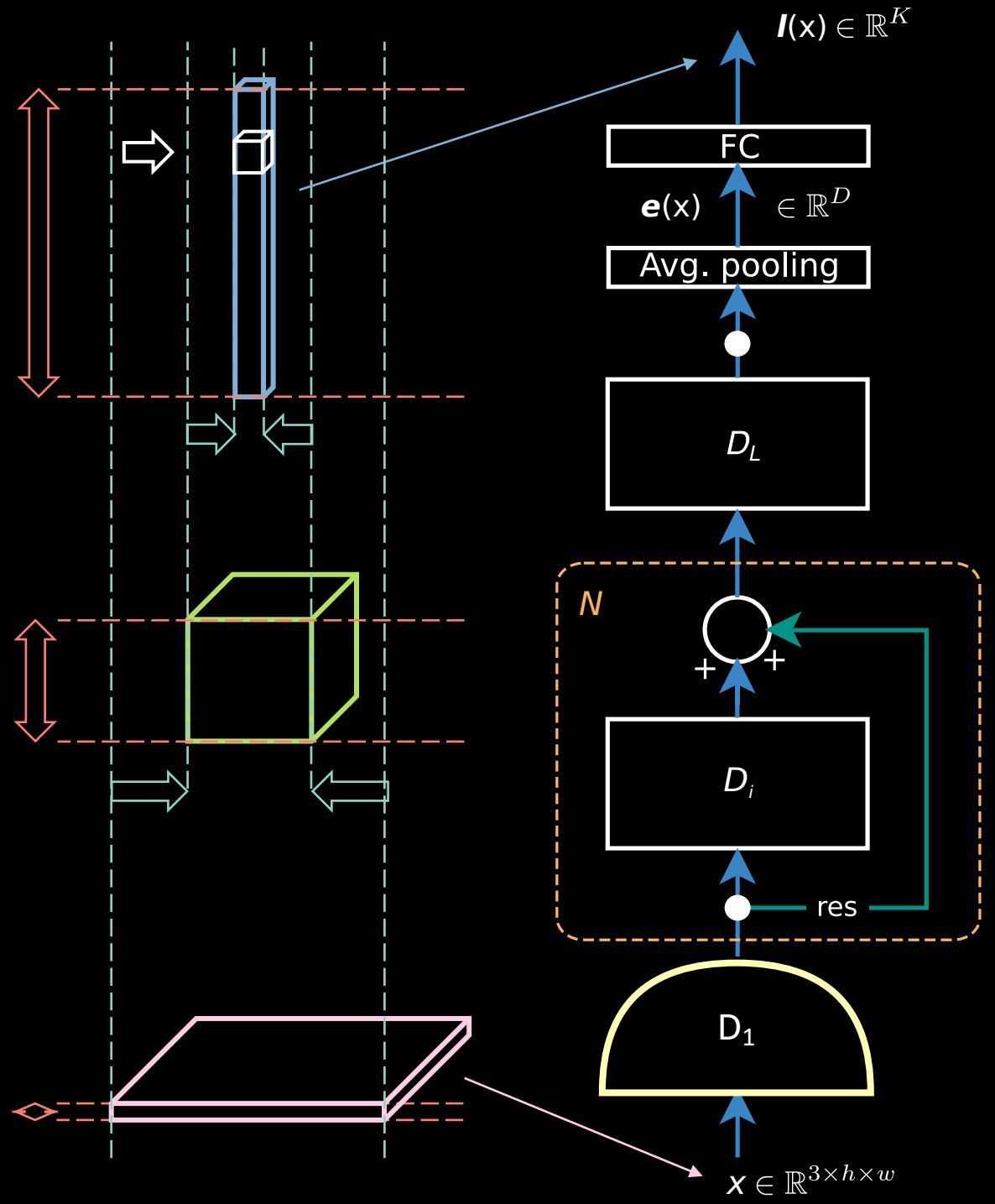
kernel size: $2 \times 7 \times \boxed{3}$

zero padding: $(\boxed{3} - 1)/2 = \boxed{1}$



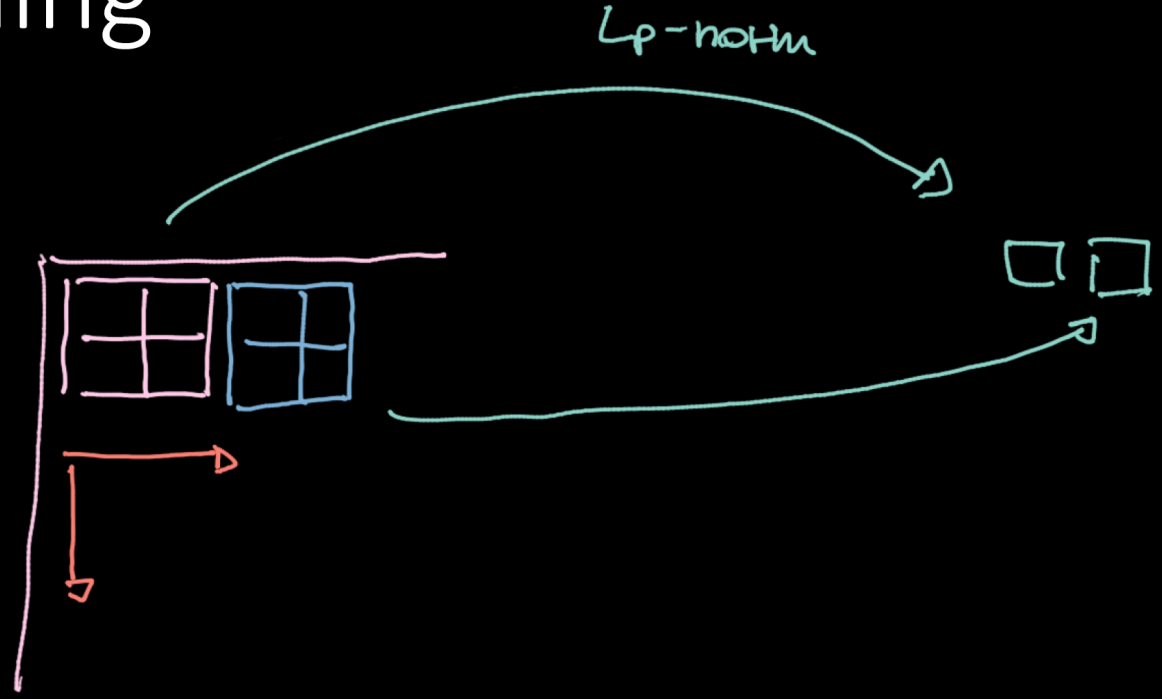
Standard spatial CNN

- Multiple layers
  - Convolution
  - Non-linearity (ReLU and Leaky)
  - Pooling
  - Batch normalisation
- Residual bypass connection



Pooling

$$\|\underline{x}\|_p := \left(\sum_i |x_i|^p \right)^{1/p}$$



$$\|\underline{x}\|_p \to \max(\underline{x}), \; p \to +\infty$$