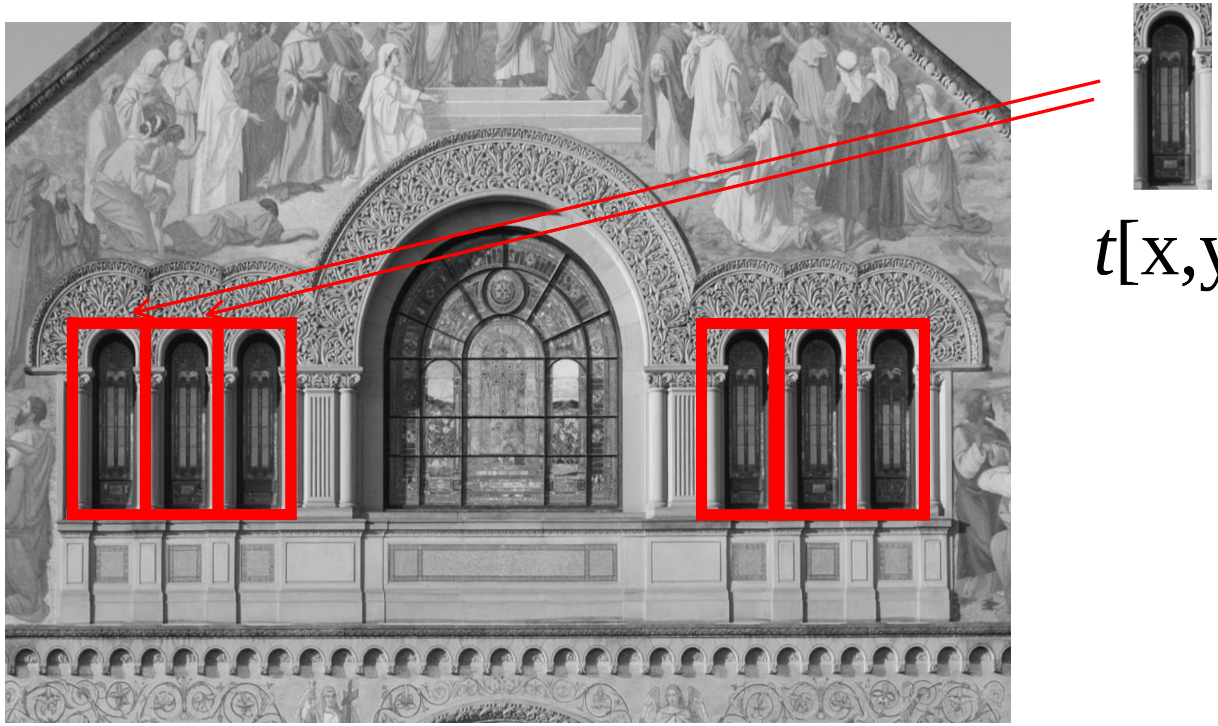
Template matching

- <u>Problem:</u> locate an object, described by a template $t[x,y]$, in the image $s[x,y]$
- Example



$t[x,y]$

$s[x,y]$

Template matching (cont.)

- Search for the best match by minimizing mean-squared error

$$E[p,q] = \sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}\left[s[x,y] - t[x-p,y-q]\right]^2$$

$$= \sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}\left|s[x,y]\right|^2 + \sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}\left|t[x,y]\right|^2 - 2\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}s[x,y]\cdot t[x-p,y-q]$$

- Equivalently, maximize *area correlation*

$$r[p,q] = \sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}s[x,y]\cdot t[x-p,y-q] = s[p,q] * t[-p,-q]$$

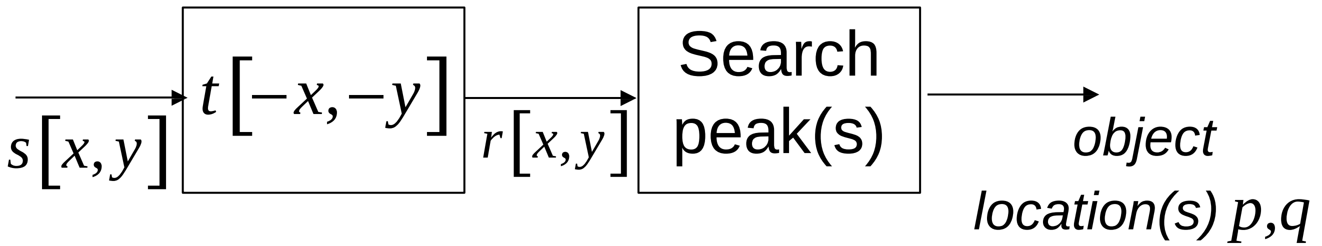
- Area correlation is equivalent to convolution of image $s[x,y]$ with impulse response $t[-x,-y]$

Template matching (cont.)

- From Cauchy-Schwarz inequality

$$r[p,q] = \sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty} s[x,y]\cdot t[x-p,y-q] \leq \sqrt{\left(\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}\left|s[x,y]\right|^2\right)\left(\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}\left|t[x,y]\right|^2\right)}$$

- Equality, iff $s[x,y] = \alpha \cdot t[x-p, y-q]$ with $\alpha \geq 0$
- Block diagram of template matcher

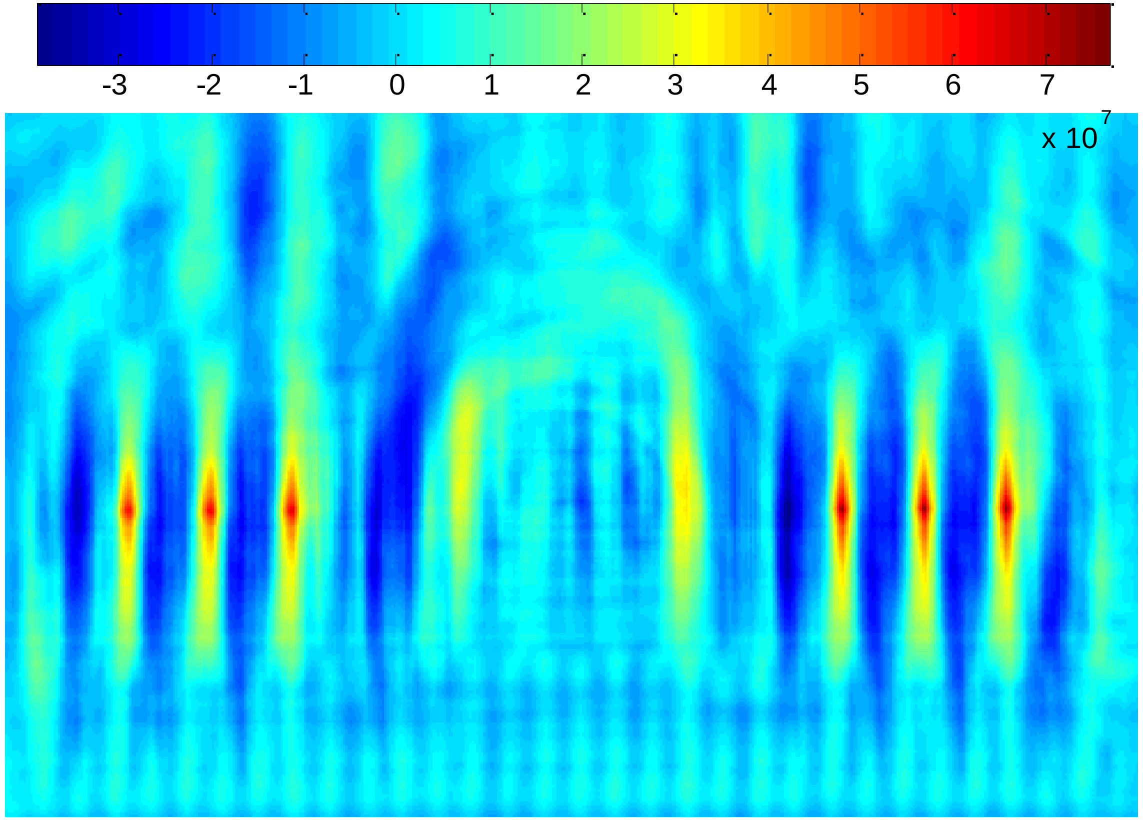


- Remove mean before template matching to avoid bias towards bright image areas

Template matching example

t[x,y]

s[x,y]



r[p,q]

Matched filtering

- Consider signal detection problem

$$s[x,y] \rightarrow \boxed{g[x,y]} \xrightarrow{r[x,y]} \boxed{\text{Search peak(s)}} \xrightarrow{\text{Object location(s) } p,q}$$

- Signal model

shifted template

$$s[x,y] = \underbrace{t[x-p,y-q]}_{\text{shifted template}} + n[x,y]$$

Other objects: "noise" or "clutter"

psd $\Phi_{nn}\left(e^{j\omega_x}, e^{j\omega_y}\right)$

- Problem: design filter $g[x,y]$ to maximize

$$SNR = \frac{\left|r[p,q]\right|^2}{E\left\{\left|n[x,y] * g[x,y]\right|^2\right\}}$$

correct peak

false readings

Vector-matrix formulation

$$r[p,q] = \vec{g}^H \vec{s}$$

covariance

$$R_{nn} = E\left\{\vec{n}\vec{n}^H\right\}$$

$$\vec{s} = \vec{t} + \vec{n}$$

$$SNR = \frac{\left|r[p,q]\right|^2}{E\left\{\left|\vec{g}^H \vec{n}\right|^2\right\}}$$

Matched filtering (cont.)

- Optimum filter has frequency response

$$G\left(e^{j\omega_x}, e^{j\omega_y}\right) = \frac{T^*\left(e^{j\omega_x}, e^{j\omega_y}\right)}{\Phi_{nn}\left(e^{j\omega_x}, e^{j\omega_y}\right)}$$

- Proof:

$$SNR = \frac{\left|r[p,q]\right|^2}{E\left\{\left|n[x,y]*g[x,y]\right|^2\right\}} \approx \frac{\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} G\left(e^{j\omega_x},e^{j\omega_y}\right) T\left(e^{j\omega_x},e^{j\omega_y}\right) d\omega_x d\omega_y\right|^2}{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|G\left(e^{j\omega_x},e^{j\omega_y}\right)\right|^2 \Phi_{nn}\left(e^{j\omega_x},e^{j\omega_y}\right) d\omega_x d\omega_y}$$

$$= \frac{\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[G\Phi_{nn}^{1/2}\right]\left[\Phi_{nn}^{-1/2}T\right] d\omega_x d\omega_y\right|^2}{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|G\right|^2 \Phi_{nn}\, d\omega_x d\omega_y} \le \frac{\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|G\right|^2 \Phi_{nn}\, d\omega_x d\omega_y\right]\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|T\right|^2 \Phi_{nn}^{-1}\, d\omega_x d\omega_y\right]}{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|G\right|^2 \Phi_{nn}\, d\omega_x d\omega_y}$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|T\right|^2 \Phi_{nn}^{-1}\, d\omega_x d\omega_y$$

max. SNR

Cauchy-Schwarz inequality,

with equality, iff $G\Phi_{nn}^{1/2} = \alpha \cdot \left[\Phi_{nn}^{-1/2}T\right]^*$

Vector-matrix formulation

$$\vec{g} = R_{nn}^{-1}\vec{t}$$

Matched filtering (cont.)

- Optimum filter corresponds to projection on

$$\boxed{\vec{g} = R_{nn}^{-1}\vec{t}}$$

- Proof:

$$SNR = \frac{\left|r[p,q]\right|^2}{E\left\{\left|\vec{g}^H\vec{n}\right|^2\right\}} \approx \frac{\left|\vec{g}^H\vec{t}\right|^2}{\vec{g}^H R_{nn}\vec{g}}$$

$$= \frac{\left|\left(R_{nn}^{1/2}\vec{g}\right)^H\left(R_{nn}^{-1/2}\vec{t}\right)\right|^2}{\left(R_{nn}^{1/2}\vec{g}\right)^H\left(R_{nn}^{1/2}\vec{g}\right)} \le \frac{\left|\left[\left(R_{nn}^{1/2}\vec{g}\right)^H\left(R_{nn}^{1/2}\vec{g}\right)\right]\left[\left(R_{nn}^{-1/2}\vec{t}\right)^H\left(R_{nn}^{-1/2}\vec{t}\right)\right]\right|}{\left(R_{nn}^{1/2}\vec{g}\right)^H\left(R_{nn}^{1/2}\vec{g}\right)}$$

$$= \vec{t}^H R_{nn}^{-1}\vec{t}$$

max. SNR

Cauchy-Schwarz inequality,

with equality, iff $R_{nn}^{1/2}\vec{g} = \alpha \cdot R_{nn}^{-1/2}\vec{t}$

Matched filtering (cont.)

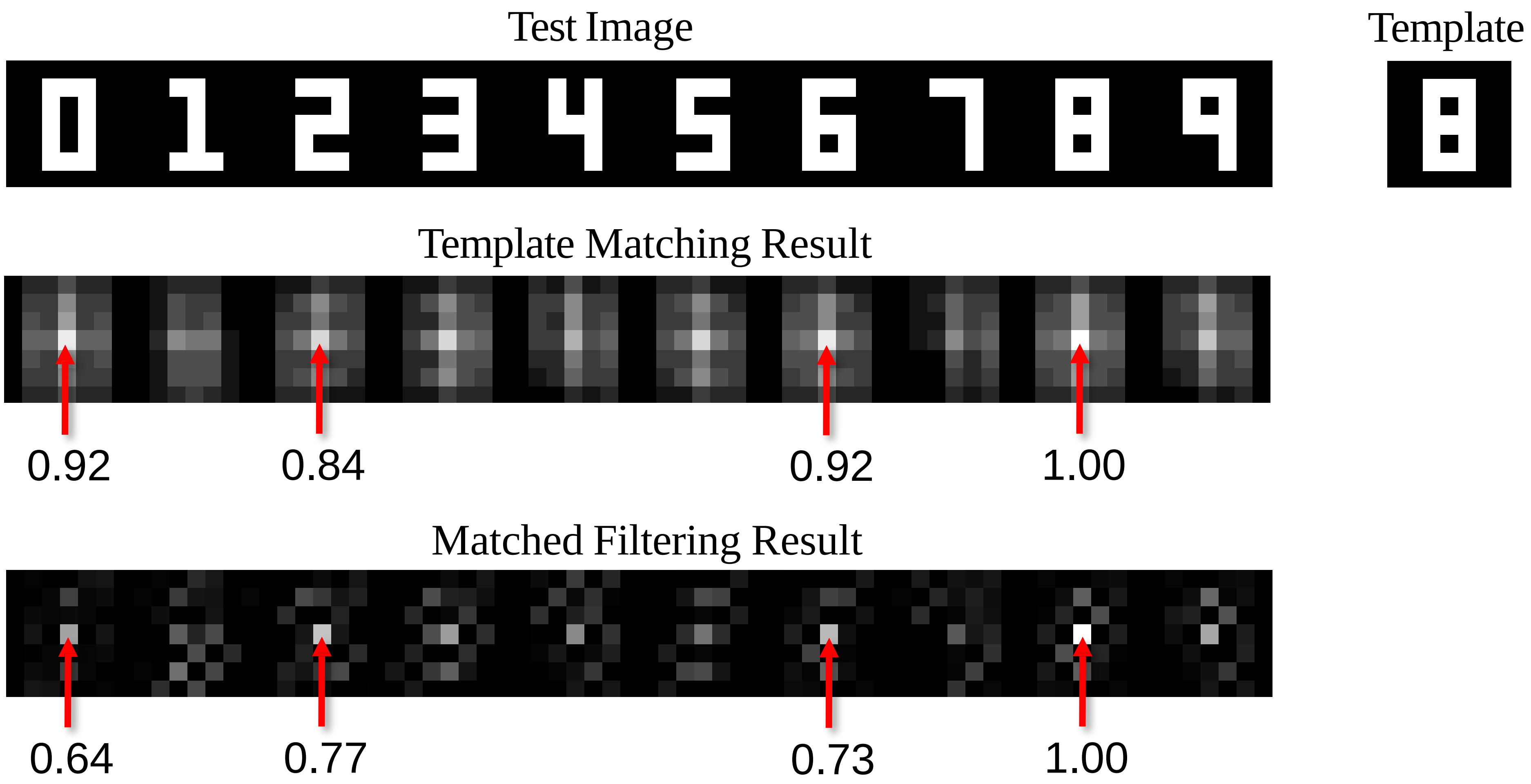
- Optimum detection: prefiltering & template matching

$$s[x,y] \longrightarrow \boxed{h[x,y]} \longrightarrow \boxed{t[-x,-y]} \xrightarrow{r[x,y]} \boxed{\text{Search peak(s)}} \longrightarrow \begin{matrix}\textit{object}\\ \textit{location(s)}\\ p,q\end{matrix}$$

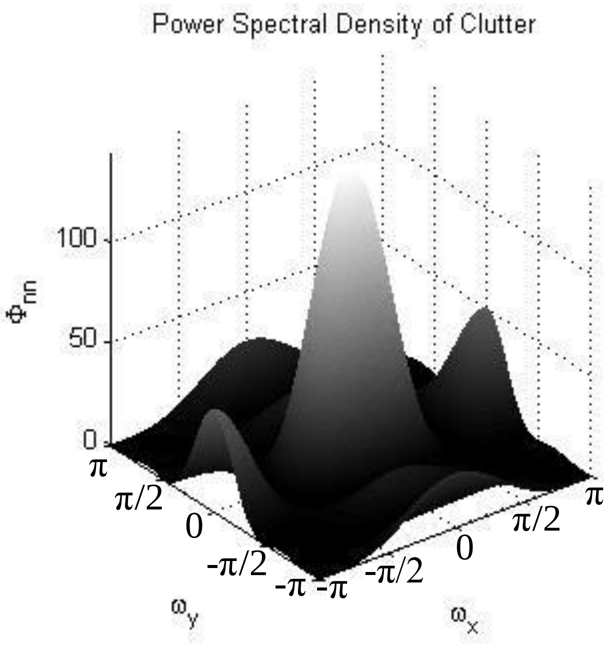
$$h[x,y] = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{\theta_{nn}\left(e^{j\omega_x}, e^{j\omega_y}\right)} e^{j\omega_x x + j\omega_y y} \, d\omega_x d\omega_y$$

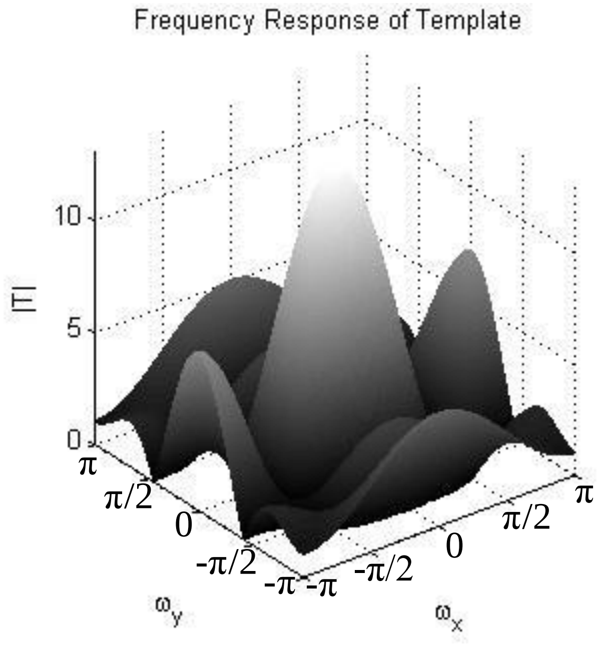
- For white noise $n[x,y]$, no prefiltering $h[x,y]$ required
- Low frequency clutter: highpass prefilter

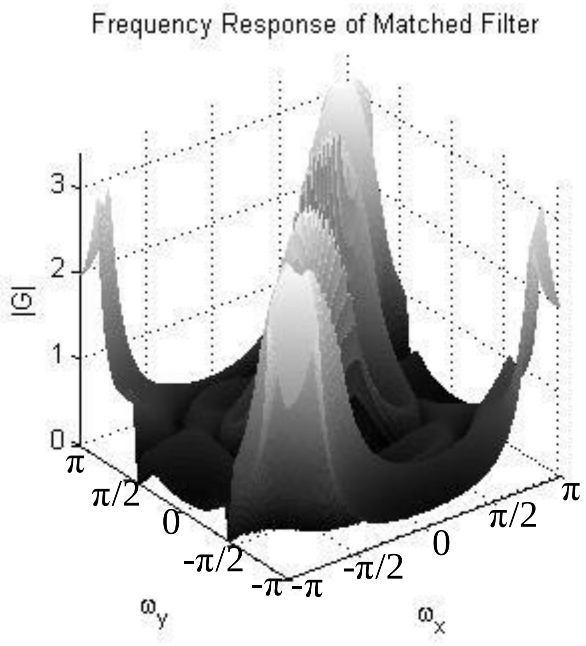
Matched filtering example



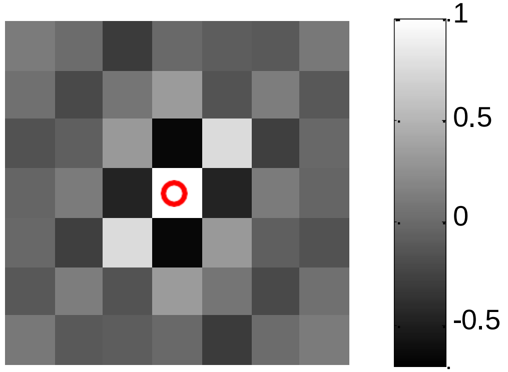
Matched filtering example (cont.)



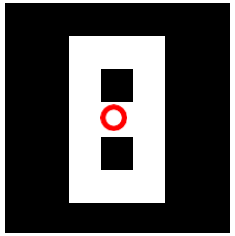




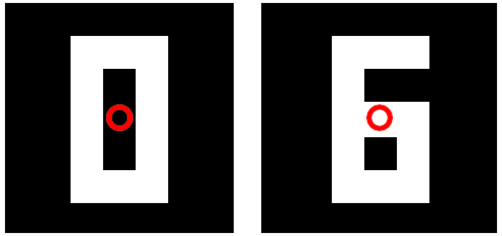
Matched Filter Impulse Response (180^{o} rotated)



Template

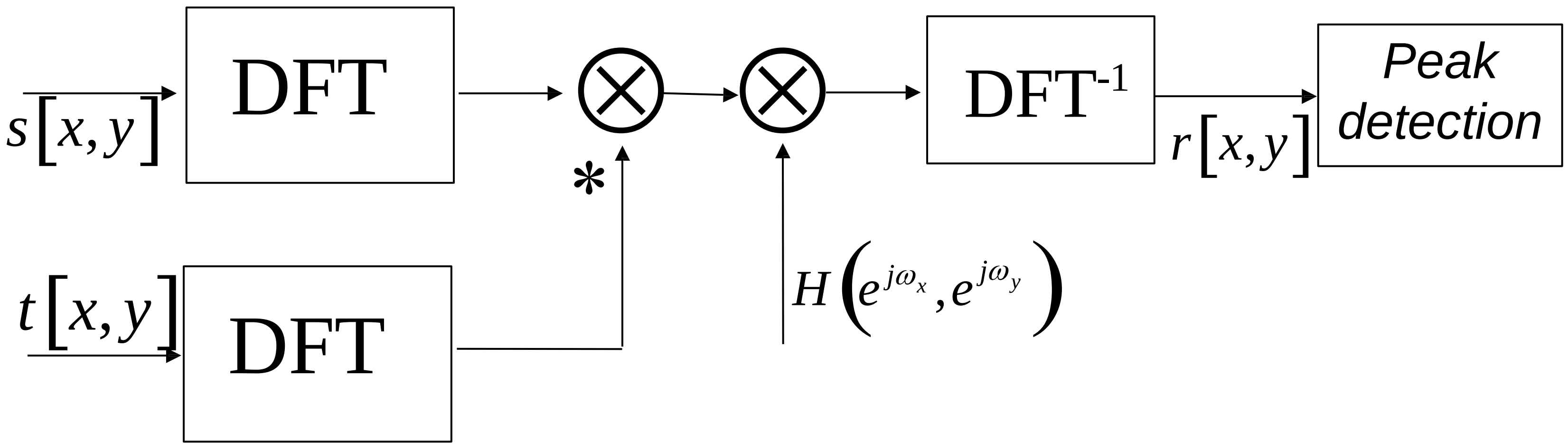


Clutter



Phase correlation

- Efficient implementation employing the Discrete Fourier Transform



- Phase correlation

$$H\left(e^{j\omega_x}, e^{j\omega_y}\right) = \frac{1}{\left|S\left(e^{j\omega_x}, e^{j\omega_y}\right)\right|\left|T\left(e^{j\omega_x}, e^{j\omega_y}\right)\right|}$$

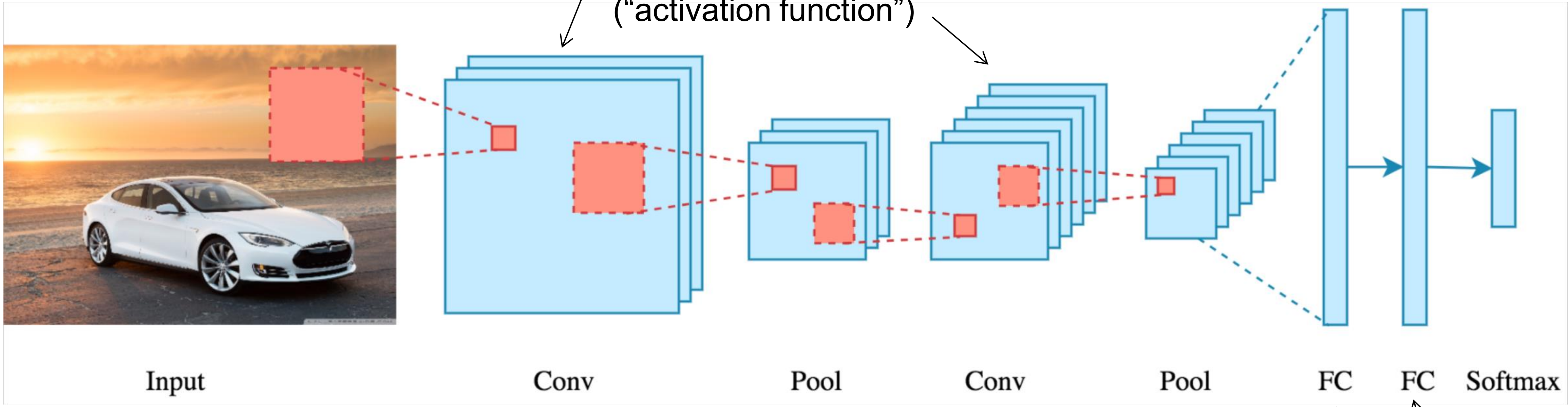
Original image

Magnitude only

Phase only

Convolutional neural networks

MIMO convolution
followed by soft
threshold nonlinearity
("activation function")



Reduce spatial resolution,
e.g. by dilation + subsampling
("Max pooling")

Linear projection
followed by
soft thresholding

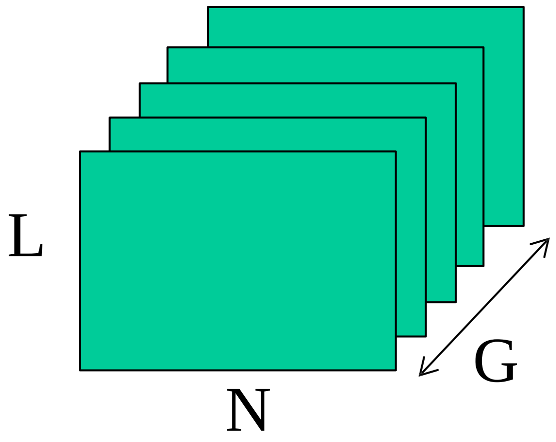
MIMO convolution

- Single-input-single-output: $f\,[x,y]$ and $g\,[x,y]$ are arrays of scalar values

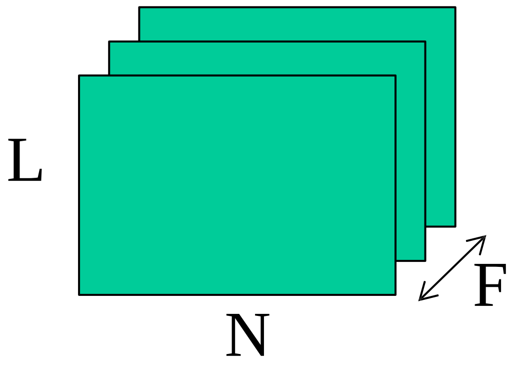
$$g[x,y] = \sum_{x'=0}^{N-1} \sum_{y'=0}^{L-1} f[x',y'] \cdot h[x-x',y-y']$$

- Multiple-input-multiple-output convolution:

$$\boxed{\mathbf{g}[x,y] = \sum_{x'=0}^{N-1} \sum_{y'=0}^{L-1} \mathbf{f}[x',y'] \cdot \mathbf{h}[x-x',y-y']}$$

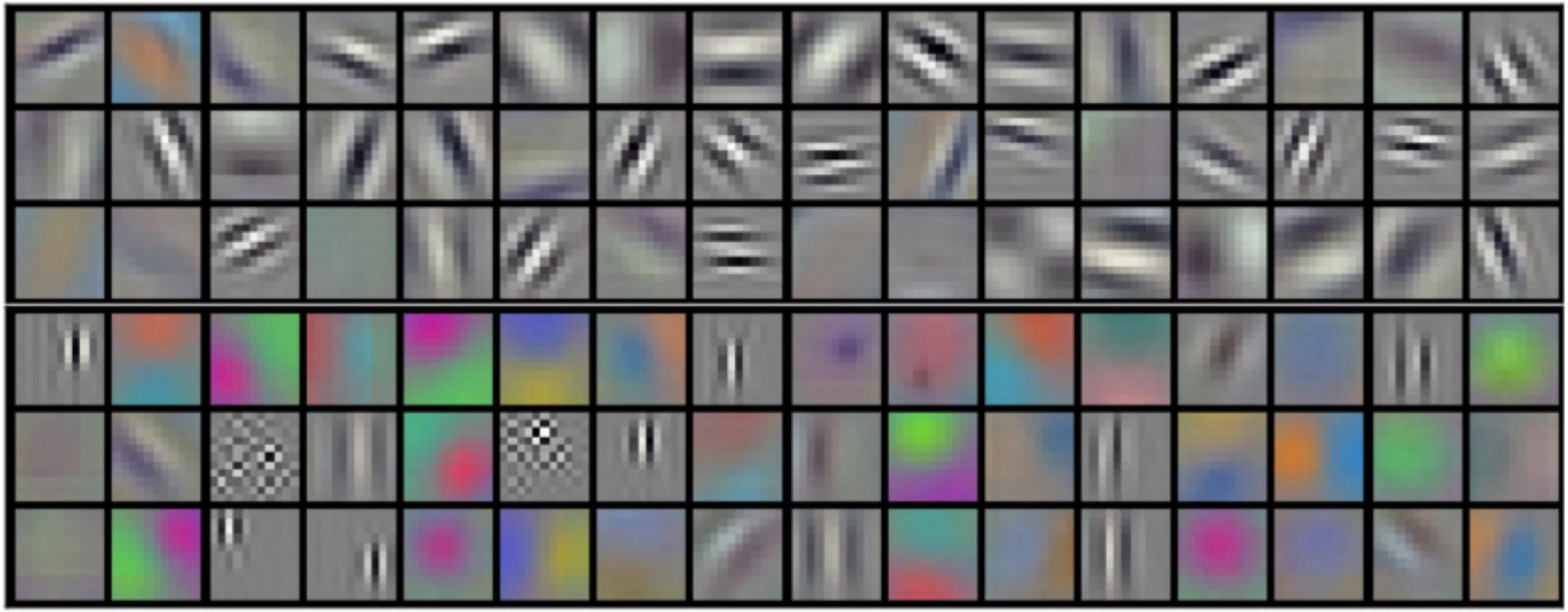


$$\mathbf{g}[x,y] = \begin{pmatrix} g_1[x,y] \\ \vdots \\ g_G[x,y] \end{pmatrix} \qquad \mathbf{f}[x,y] = \begin{pmatrix} f_1[x,y] \\ \vdots \\ f_F[x,y] \end{pmatrix}$$



Example templates of first convolutional layer

AlexNet, F=3, G=96

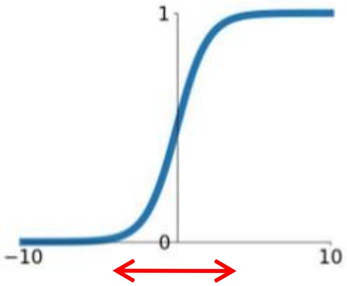


[Krizhevsky et al., 2012]

Activation function

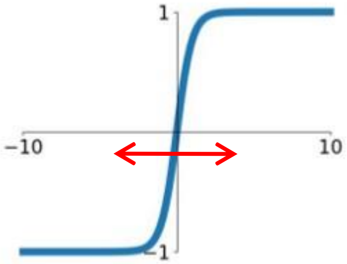
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



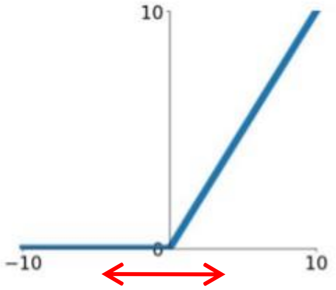
tanh

$\tanh(x)$



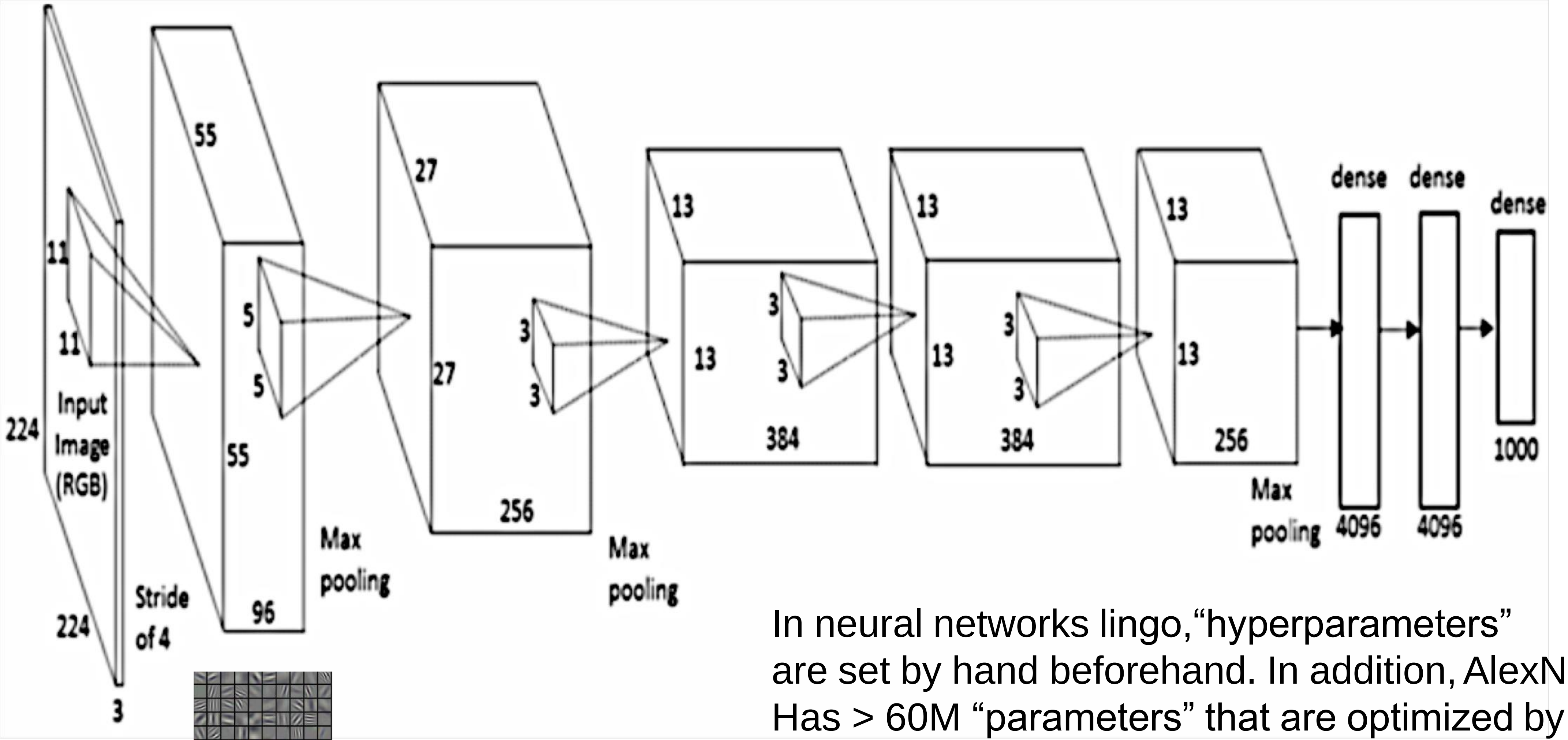
ReLU

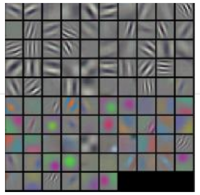
$\max(0, x)$



- Sigmoid and tanh traditionally used
- ReLU (rectified linear unit) simpler and improves convergence of training
- Trained bias is added before activation function to set the best threshold

AlexNet hyperparameters

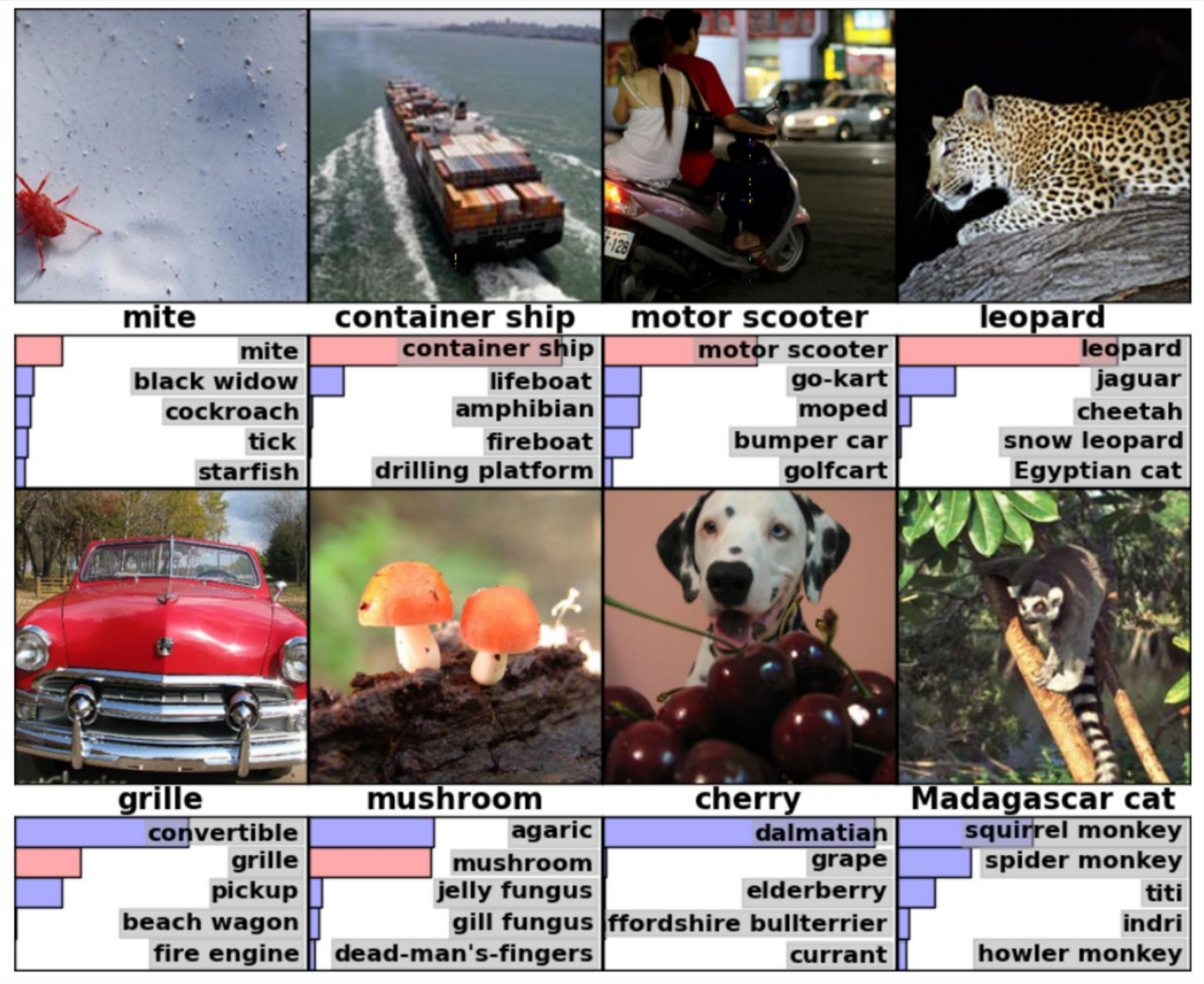




In neural networks lingo,“hyperparameters” are set by hand beforehand. In addition, AlexNet Has > 60M “parameters” that are optimized by supervised learning.

[Krizhevsky et al., 2012]

AlexNet Image Classification Results



[Krizhevsky et al., 2012]

AlexNet Image-based Retrieval Results

Query

Most similar images in database

[Krizhevsky et al., 2012]