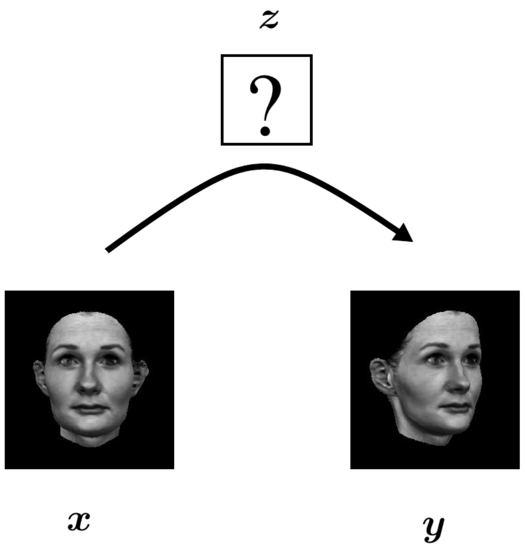
Learning Image Relations with Contrast Association Networks

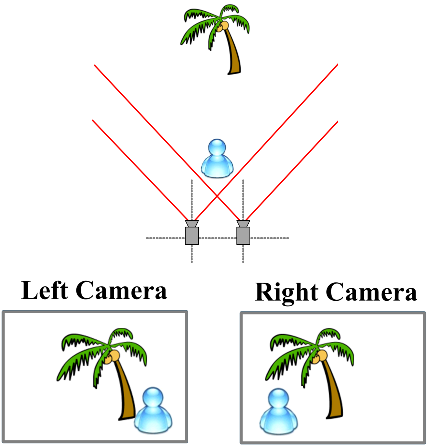
Yao Lu, Zhirong Yang, Juho Kannala, Samuel Kaski

Aalto University

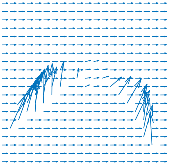
Problem



Stereo Matching



Optical Flow

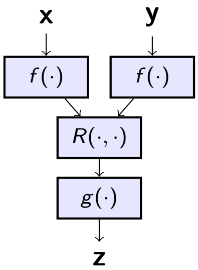


Learning Relations

$$\mathbf{z} = F(\mathbf{x}, \mathbf{y})$$

Learning Relations

$$\mathbf{z} = F(\mathbf{x}, \mathbf{y})$$



$$\begin{aligned} \mathbf{a} &= f(\mathbf{x}), \quad & \mathbf{h} &= R(\mathbf{a}, \mathbf{b}), \\ \mathbf{b} &= f(\mathbf{y}), \quad & \mathbf{z} &= g(\mathbf{h}), \end{aligned} \tag{1}$$

Learning Relations

$$\mathbf{h} = R(\mathbf{a}, \mathbf{b})$$

Concatenation Units

$$\mathbf{h} = [\mathbf{a} \quad \mathbf{b}]$$

Bilinear Units

$$h_k = \sum_{ij} W_{ijk} a_i b_j = \mathbf{a}^T \mathbf{W}_k \mathbf{b},$$

A Toy Example

$$\begin{aligned}
\mathbf{a} &= (1, 2, 3, 4, 5), \\
\mathbf{b}_1 &= (2, 3, 4, 5, 1), \\
\mathbf{b}_2 &= (1, 2, 3, 4, 5), \\
\mathbf{b}_3 &= (5, 1, 2, 3, 4),
\end{aligned}$$

A Toy Example

$$\begin{aligned}
\mathbf{a} &= (1, 2, 3, 4, 5), \\
\mathbf{b}_1 &= (2, 3, 4, 5, 1), \\
\mathbf{b}_2 &= (1, 2, 3, 4, 5), \\
\mathbf{b}_3 &= (5, 1, 2, 3, 4),
\end{aligned}$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?

A Toy Example

$$\begin{aligned}
\mathbf{a} &= (1, 2, 3, 4, 5), \\
\mathbf{b}_1 &= (2, 3, 4, 5, 1), \\
\mathbf{b}_2 &= (1, 2, 3, 4, 5), \\
\mathbf{b}_3 &= (5, 1, 2, 3, 4),
\end{aligned}$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?
Shift $\mathbf{a}$ and compute $\tilde{\mathbf{a}} \cdot \mathbf{b}$!

A Toy Example

$$\begin{aligned}
\mathbf{a} &= (1, 2, 3, 4, 5), \\
\mathbf{b}_1 &= (2, 3, 4, 5, 1), \\
\mathbf{b}_2 &= (1, 2, 3, 4, 5), \\
\mathbf{b}_3 &= (5, 1, 2, 3, 4),
\end{aligned}$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?
Shift $\mathbf{a}$ and compute $\tilde{\mathbf{a}} \cdot \mathbf{b}$!

$$h_k = \sum_{ij} W_{ijk} a_i b_j = \mathbf{a}^T \mathbf{W}_k \mathbf{b}$$

A Toy Example

$$
\begin{aligned}
\mathbf{a} &= (1,2,3,4,5),\\
\mathbf{b}_1 &= (2,3,4,5,1),\\
\mathbf{b}_2 &= (1,2,3,4,5),\\
\mathbf{b}_3 &= (5,1,2,3,4),
\end{aligned}
$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?
Shift $\mathbf{a}$ and compute $\tilde{\mathbf{a}} \cdot \mathbf{b}$!

$$
h_k = \sum_{ij} W_{ijk} a_i b_j = \mathbf{a}^T \mathbf{W}_k \mathbf{b}
$$

$$
\underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{W}_1}, \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{W}_2}, \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{W}_3} \tag{2}
$$

A Toy Example

$$\begin{aligned} \mathbf{a} &= (1,2,3,4,5), \\ \mathbf{b}_1 &= (2,3,4,5,6), \\ \mathbf{b}_2 &= (1,2,3,4,5), \\ \mathbf{b}_3 &= (0,1,2,3,4), \end{aligned}$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?

A Toy Example

$$\begin{aligned} \mathbf{a} &= (1,2,3,4,5), \\ \mathbf{b}_1 &= (2,3,4,5,6), \\ \mathbf{b}_2 &= (1,2,3,4,5), \\ \mathbf{b}_3 &= (0,1,2,3,4), \end{aligned}$$

Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?

~~Shift $\mathbf{a}$ and compute $\tilde{\mathbf{a}} \cdot \mathbf{b}$!~~

A Toy Example

$$
\begin{aligned}
\mathbf{a} &= (1,2,3,4,5),\\
\mathbf{b}_1 &= (2,3,4,5,6),\\
\mathbf{b}_2 &= (1,2,3,4,5),\\
\mathbf{b}_3 &= (0,1,2,3,4),
\end{aligned}
$$

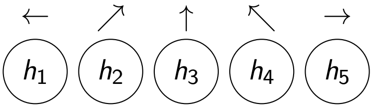
Unknown $\mathbf{b}$, which ones in $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is $\mathbf{b}$?
~~Shift $\mathbf{a}$ and compute $\tilde{\mathbf{a}} \cdot \mathbf{b}$!~~

$$
h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0
$$

$$
\underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{W}_1}, \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{W}_2}, \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{W}_3} \qquad (3)
$$

Contrast Association Units

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$



Contrast Association Units

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

$$\begin{array}{ccccc} \leftarrow & \nearrow & \uparrow & \nwarrow & \rightarrow \\ h_1 & h_2 & h_3 & h_4 & h_5 \end{array}$$

Competition

$$h'_k = \frac{e^{-h_k}}{\sum_i e^{-h_i}}$$

Low-rank Approximation

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

Low-rank Approximation

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

1. $\mathbf{W}_k = \mathbf{u}_k \mathbf{v}_k^T$
2. $\mathbf{h}^* = \frac{1}{2}\Big[(\mathbf{V1}) \circ \mathbf{U}(\mathbf{a})^2 + (\mathbf{U1}) \circ \mathbf{V}(\mathbf{b})^2\Big] - (\mathbf{Ua}) \circ (\mathbf{Vb})$
3. Pooling over $\mathbf{h}^*$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient
- Parametrization $w(s) = \frac{1}{1+\exp(-s)}$ or $w(s) = \log(1 + \exp(s))$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient
- Parametrization $w(s) = \frac{1}{1+\exp(-s)}$ or $w(s) = \log(1 + \exp(s))$
- Multiplicative update

$$\frac{\partial E}{\partial \mathbf{W}} = \nabla^+ - \nabla^-$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient
- Parametrization $w(s) = \frac{1}{1+\exp(-s)}$ or $w(s) = \log(1 + \exp(s))$
- Multiplicative update

$$\frac{\partial E}{\partial \mathbf{W}} = \nabla^+ - \nabla^-$$

$$\nabla^+ = \frac{1}{2}\left(\mathrm{abs}\left(\frac{\partial E}{\partial \mathbf{W}}\right) + \frac{\partial E}{\partial \mathbf{W}}\right) + \epsilon$$

$$\nabla^- = \frac{1}{2}\left(\mathrm{abs}\left(\frac{\partial E}{\partial \mathbf{W}}\right) - \frac{\partial E}{\partial \mathbf{W}}\right) + \epsilon$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient
- Parametrization $w(s) = \frac{1}{1+\exp(-s)}$ or $w(s) = \log(1 + \exp(s))$
- Multiplicative update

$$\frac{\partial E}{\partial \mathbf{W}} = \nabla^+ - \nabla^-$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

- Projected gradient
- Parametrization $w(s) = \frac{1}{1+\exp(-s)}$ or $w(s) = \log(1 + \exp(s))$
- Multiplicative update

$$\frac{\partial E}{\partial \mathbf{W}} = \nabla^{+} - \nabla^{-}$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \left(\frac{\nabla^{-}}{\nabla^{+}}\right)^{\eta}$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \left(\frac{\nabla^-}{\nabla^+} \right)^\eta$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \left(\frac{\nabla^-}{\nabla^+}\right)^\eta$$

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{W}} &= (3, -2) \\ &= (3 + \epsilon, \epsilon) - (\epsilon, 2 + \epsilon) \end{aligned}$$

Optimization

$$h_k = \sum_{ij} W_{ijk}(a_i - b_j)^2, \quad W_{ijk} \geq 0$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \left(\frac{\nabla^-}{\nabla^+}\right)^\eta$$

$$\begin{aligned}\frac{\partial E}{\partial \mathbf{W}} &= (3, -2) \\ &= (3+\epsilon, \epsilon) - (\epsilon, 2+\epsilon)\end{aligned}$$

$$\frac{\nabla^-}{\nabla^+} = (\frac{\epsilon}{3+\epsilon}, \frac{2+\epsilon}{\epsilon})$$

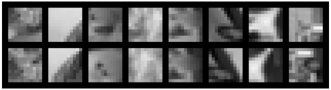
Very sparse solution!

Experiments

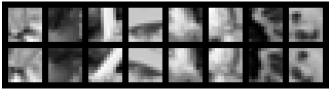
$\mathbf{p}' = \mathbf{H}\mathbf{p}$ with homography matrix

Experiments

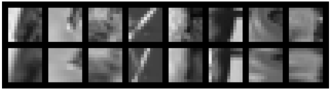
$\mathbf{p}' = \mathbf{H}\mathbf{p}$ with homography matrix



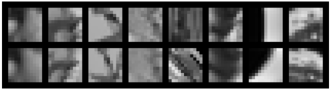
(f) Translation



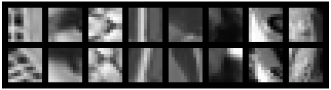
(g) Rotation



(h) Scaling



(i) Affine



(j) Projective

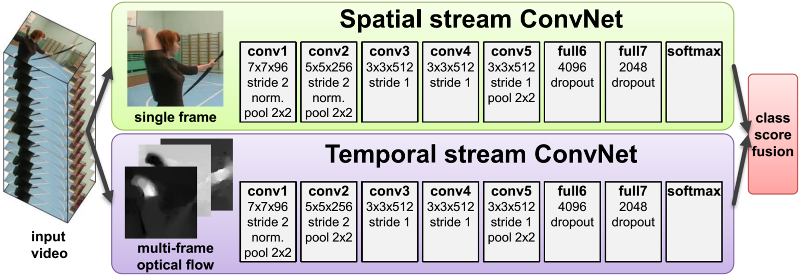
CTN	BLN	CAN
Concat	Bilinear*, 1200	CAU*, 1200
Linear, 1200	Sum-Pooling, 300	Sum-Pooling, 300
PReLU	l_2 Norm	Softmin
Linear, 300	Linear, 100	Linear, 100
PReLU	PReLU	PReLU
Linear, 100	Linear, 100	Linear, 100
PReLU	PReLU	PReLU
Linear, 100	Linear, $\dim(\mathbf{z})$	Linear, $\dim(\mathbf{z})$
PReLU		
Linear, $\dim(\mathbf{z})$		

Task	CTN	BLN	CAN
Translation	0.773	1.893	**0.049**
Rotation	9.854	5.925	**3.518**
Scaling	0.018	0.025	**0.017**
Affine	0.014	0.020	**0.010**
Projective	**0.030**	0.032	**0.030**

Conclusion

- CNN is a hierarchical template matcher for representing appearance
- Special purpose neurons for representing relations

Puzzle



Why we still need hand-crafted features for motion?