Lecture: Deep Learning and Differential Programming

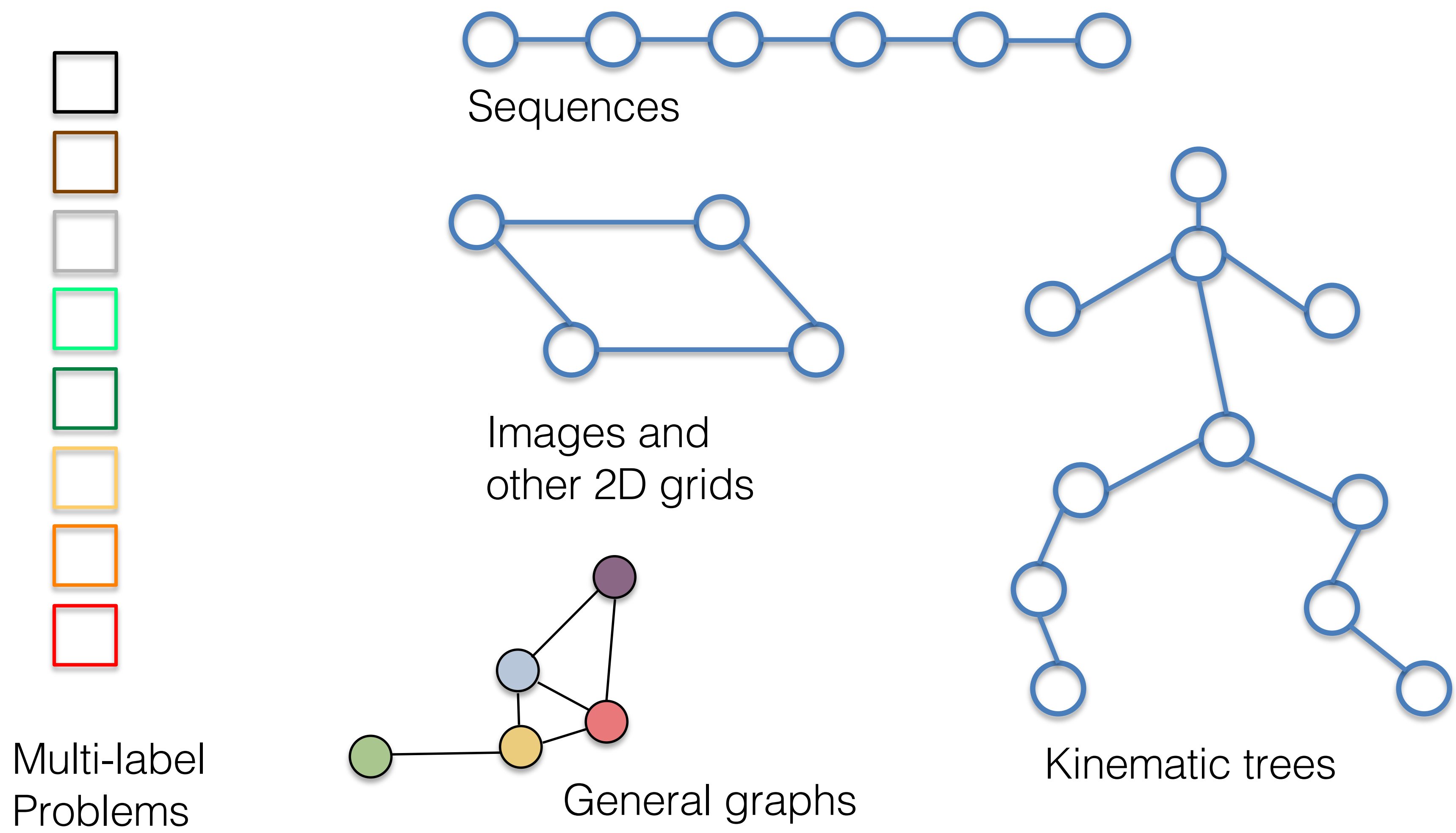
4.3 Graphs and relational reasoning

https://liris.cnrs.fr/christian.wolf/teaching

INSA LYON Christian Wolf

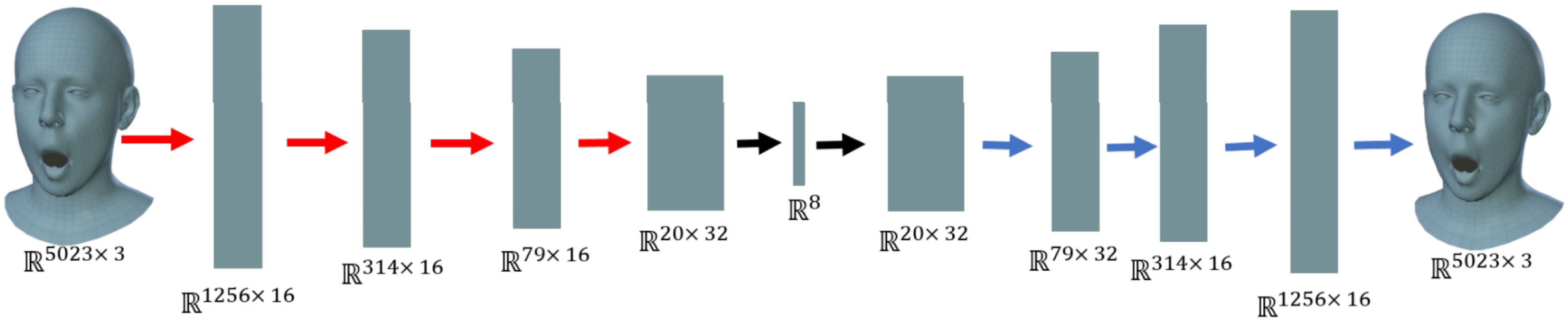
Structured Input and/or structured Output

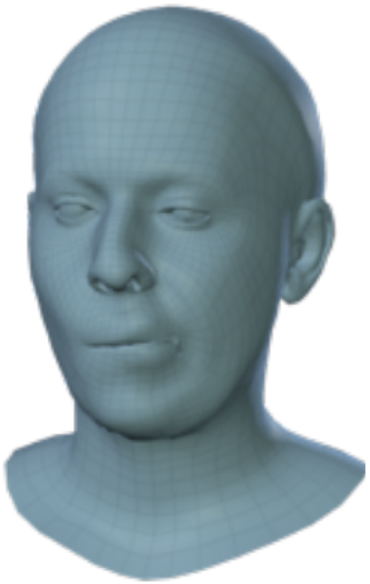
Predicting for multiple inter-dependent variables



Example : meshes

3D triangular meshes are approximations of 3D surfaces / manifolds: a graph with vertices embedded n 3D Euclidean space.



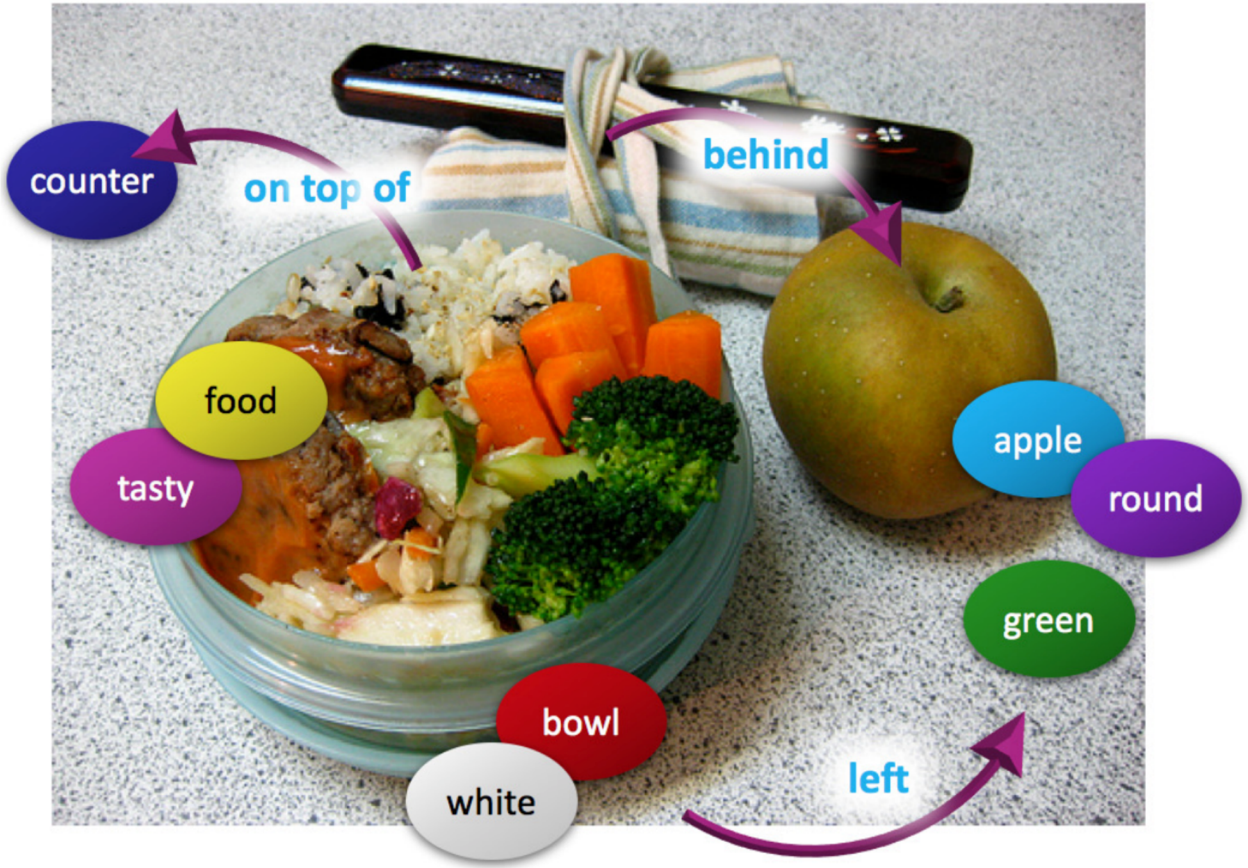


[Ranjan, Bolkart, Sanyal, Black, ECCV 2018]

Example: point clouds

[Figure: Inria Chroma, 2018]

Example : scene graphs



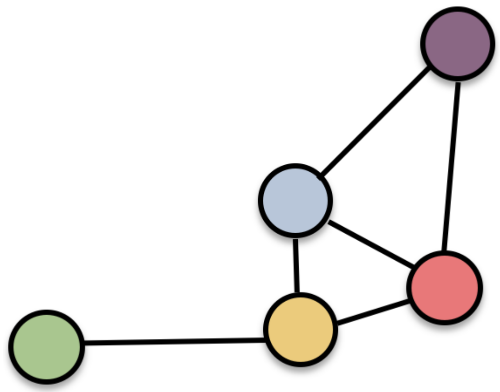
Example image and graph from the GQA dataset

[Hudson, Manning, 2019]

Graphs: definition

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of:

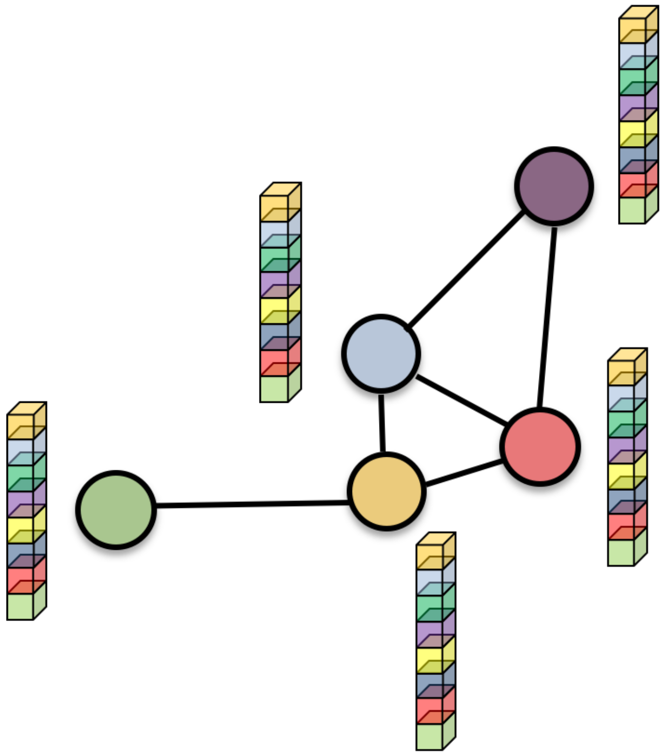
- a set $\mathcal{V}$ of nodes, and
- a set $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ of edges



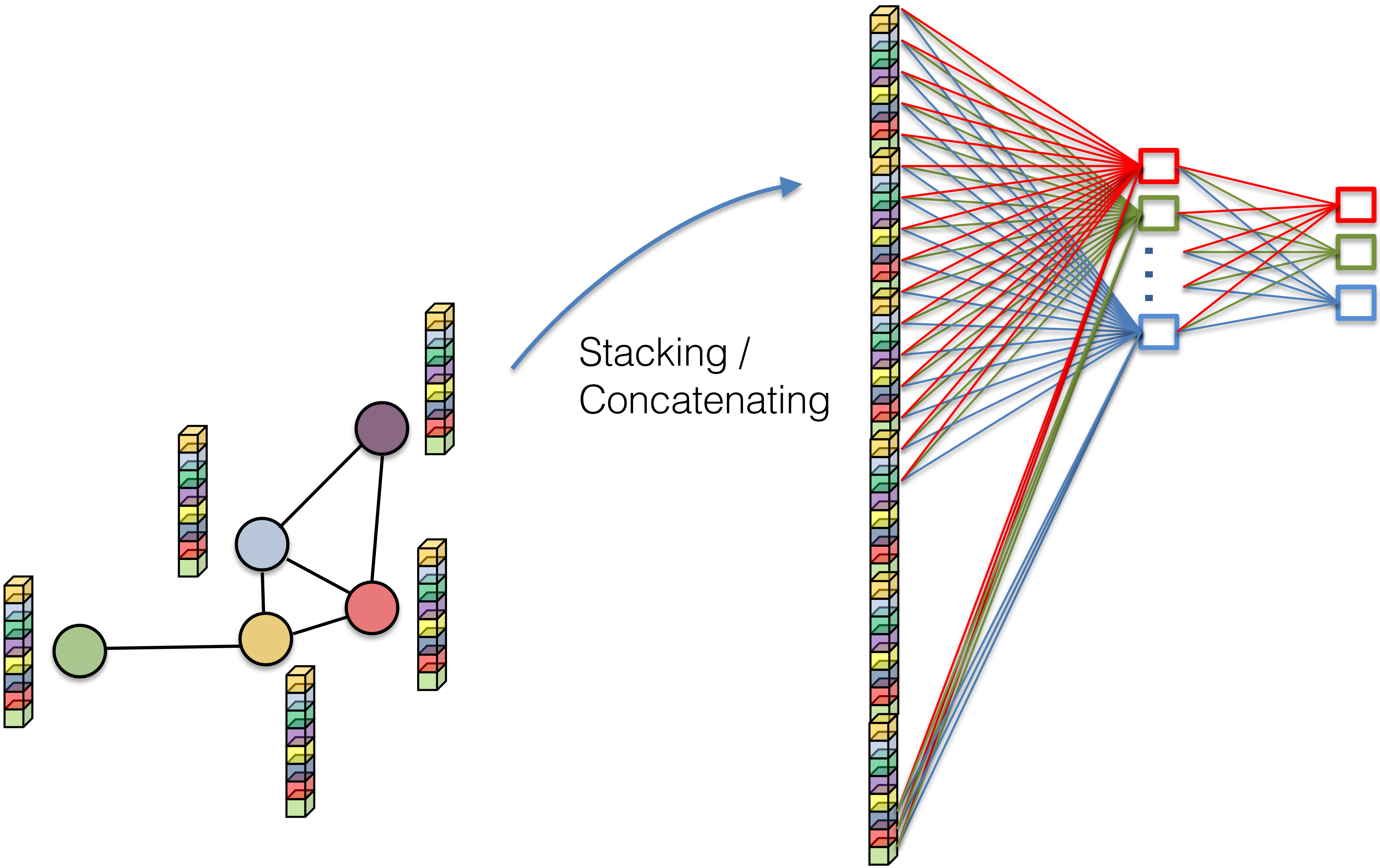
Attributed Graphs

Attributed graphs also have values for each node: $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{X})$:

- a set $\mathcal{V}$ of nodes, and
- a set $\mathcal{E} \in \mathcal{V}\times\mathcal{V}$ of edges
- a set $\mathcal{X} = \{x_0, \ldots, x_N\}$ of values, each value being associated to a node. In our case we find **embeddings** $x_i \in \mathbb{R}^d$ for each node.



Trivial solutions (don't do this at home)



Relational Reasoning

$$g(x_1, x_2, \ldots, x_N) = \max(h(x_1), h(x_2), \ldots, h(x_N))$$

"PointNet" [Qi, Su, Mo, Guibas, CVPR 2017]

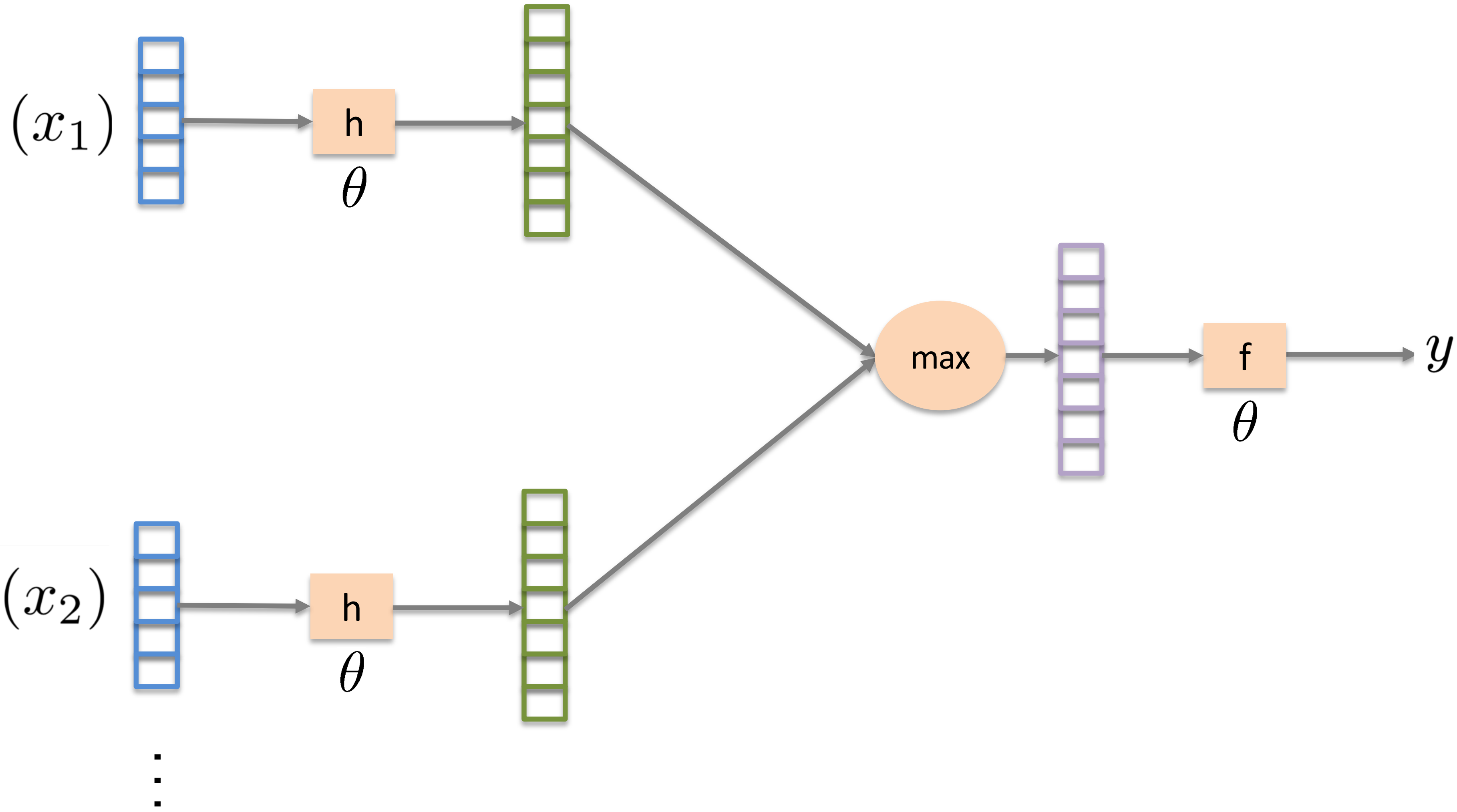
Defined over points of a point cloud

$$g(x_1, x_2, \ldots, x_N) = \sum_{i,j} h(x_i, x_j)$$

"Relational Reasoning" [Santoro et al., NIPS 2017]

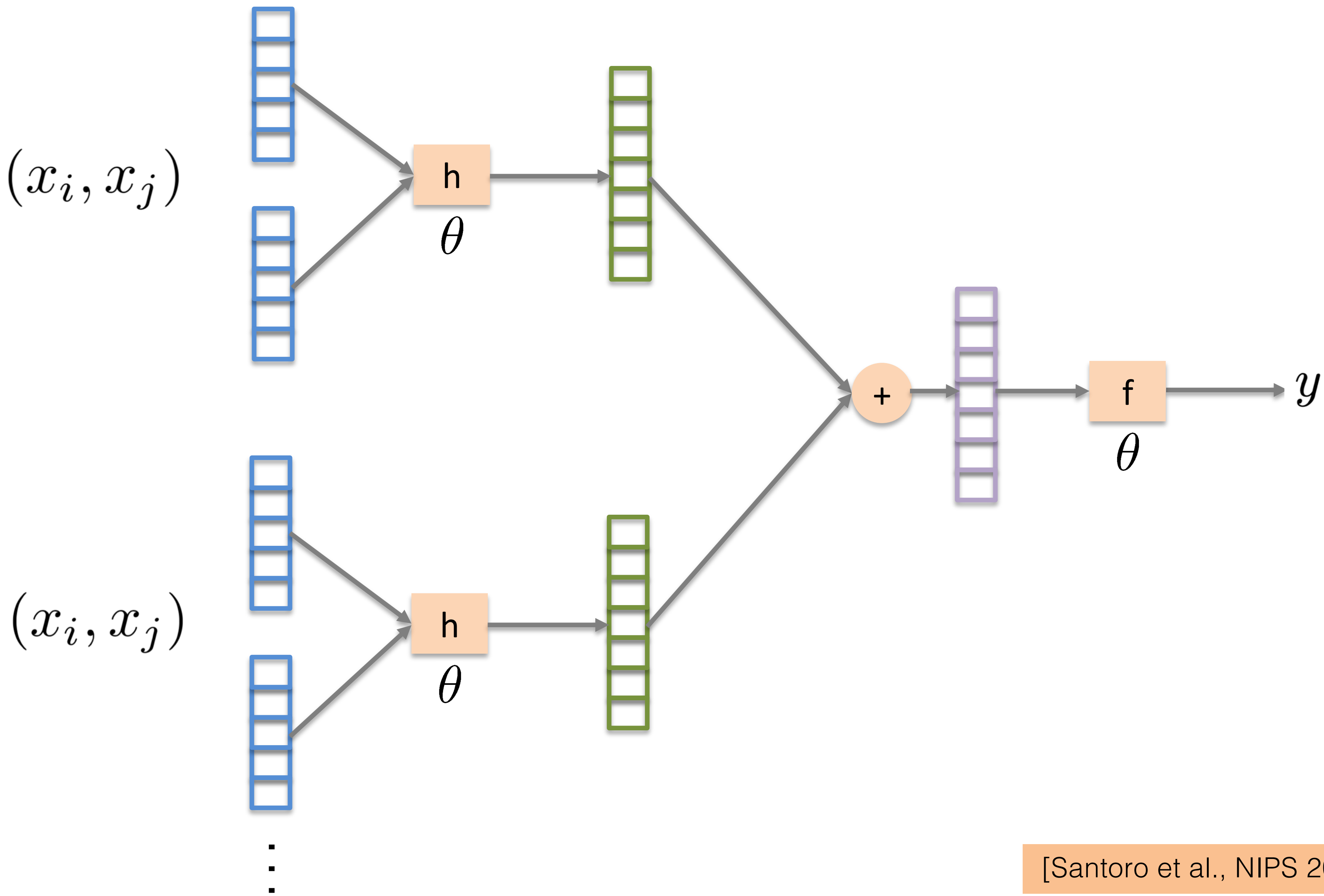
Defined over feature map cells

PointNet



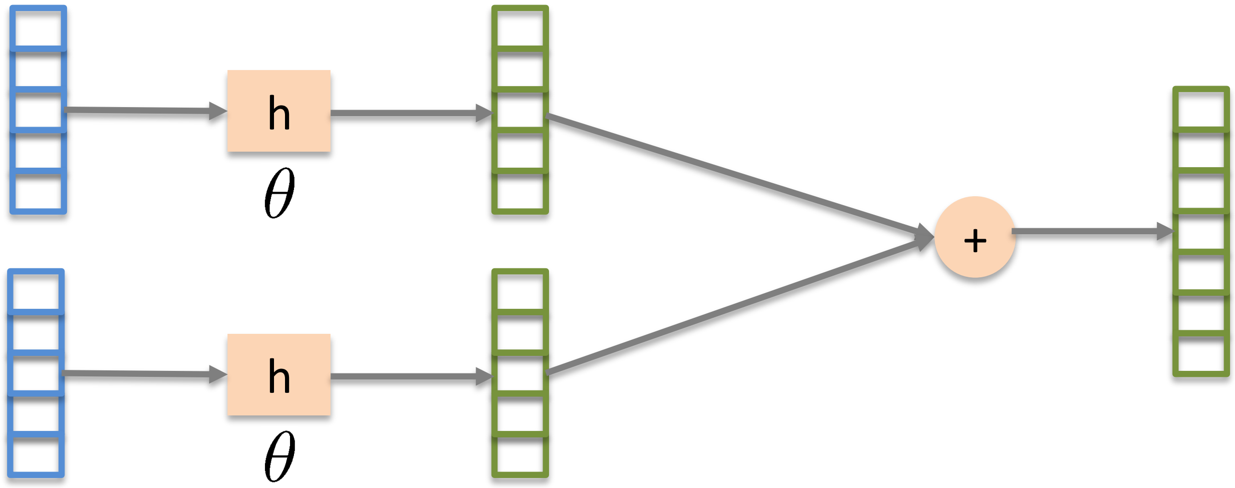
[Qi, Su, Mo, Guibas, CVPR 2017]

Relational Reasoning: pairwise terms



[Santoro et al., NIPS 2017]

Relational Reasoning: deep learning

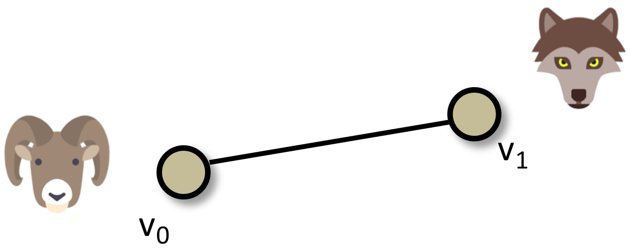


$$y = f(x,\theta) = \phi\left(W_y \sum_{ij} \phi\left(W_v \begin{bmatrix} \phi(W_u x)_i \\ \phi(W_u x)_j \end{bmatrix} \right) \right)$$

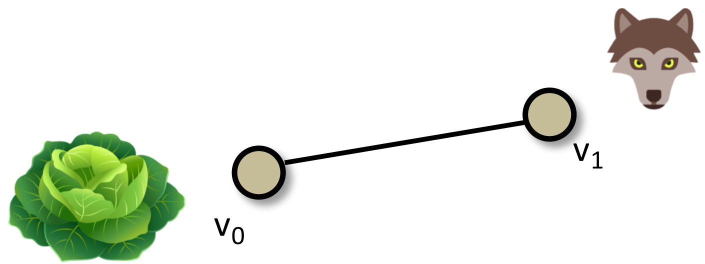
$$\theta = \{W_y, W_v, W_u\}$$

$$\begin{aligned} \hat{\theta} &= \min_\theta \quad \mathcal{L}(y^*, y) \\ &= \min_\theta \quad \mathcal{L}(y^*, f(x,\theta)) \end{aligned}$$

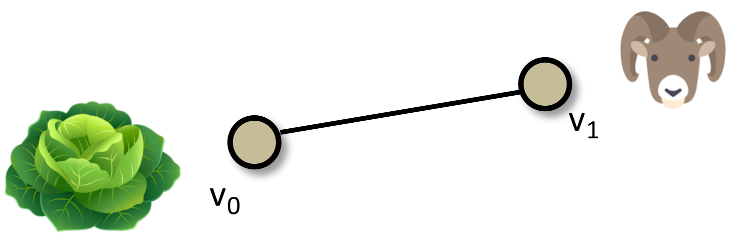
A toy problem: will somebody get eaten?



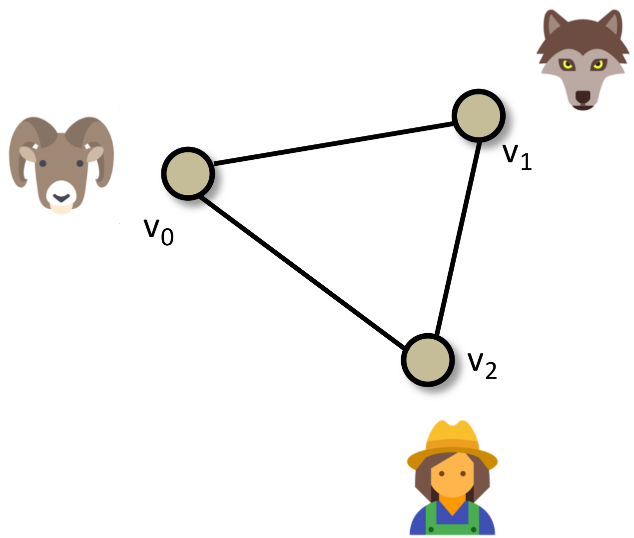
Yes



No

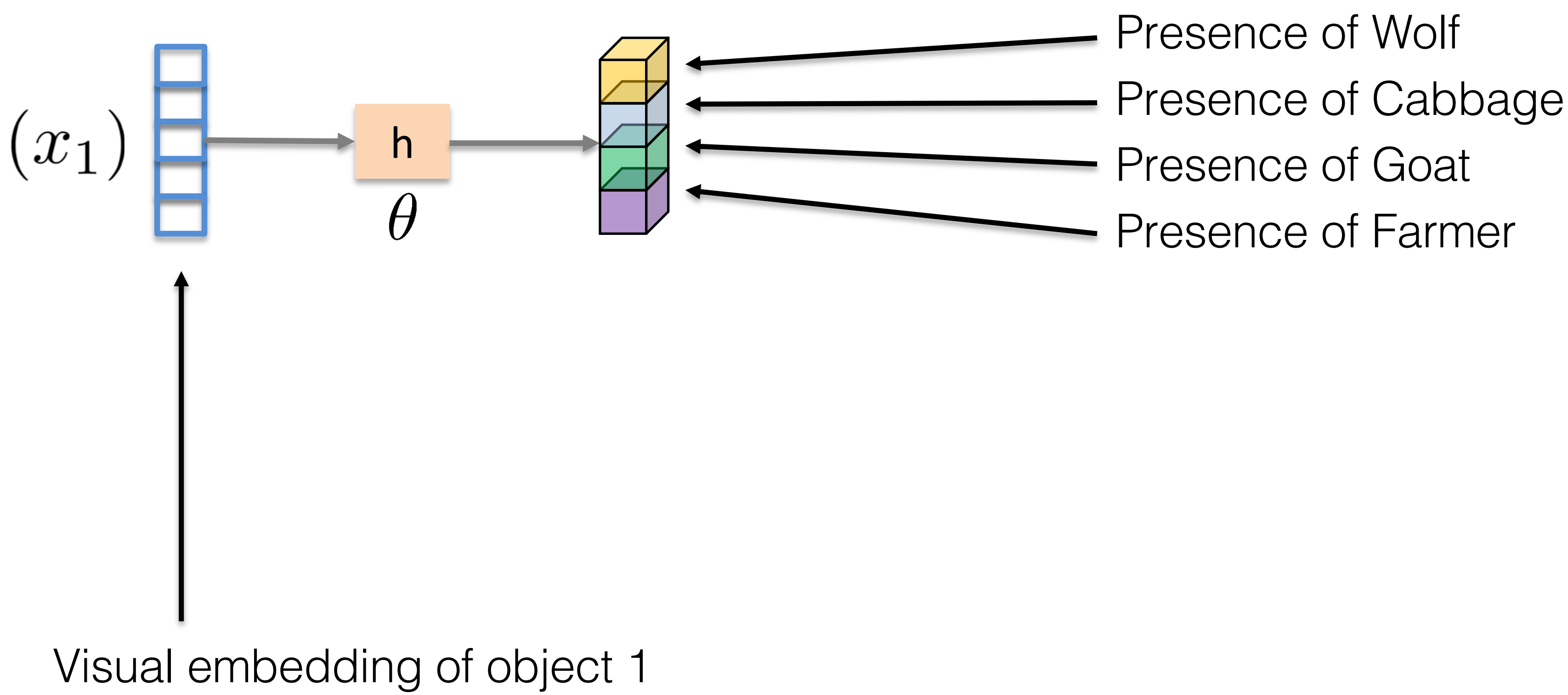


Yes

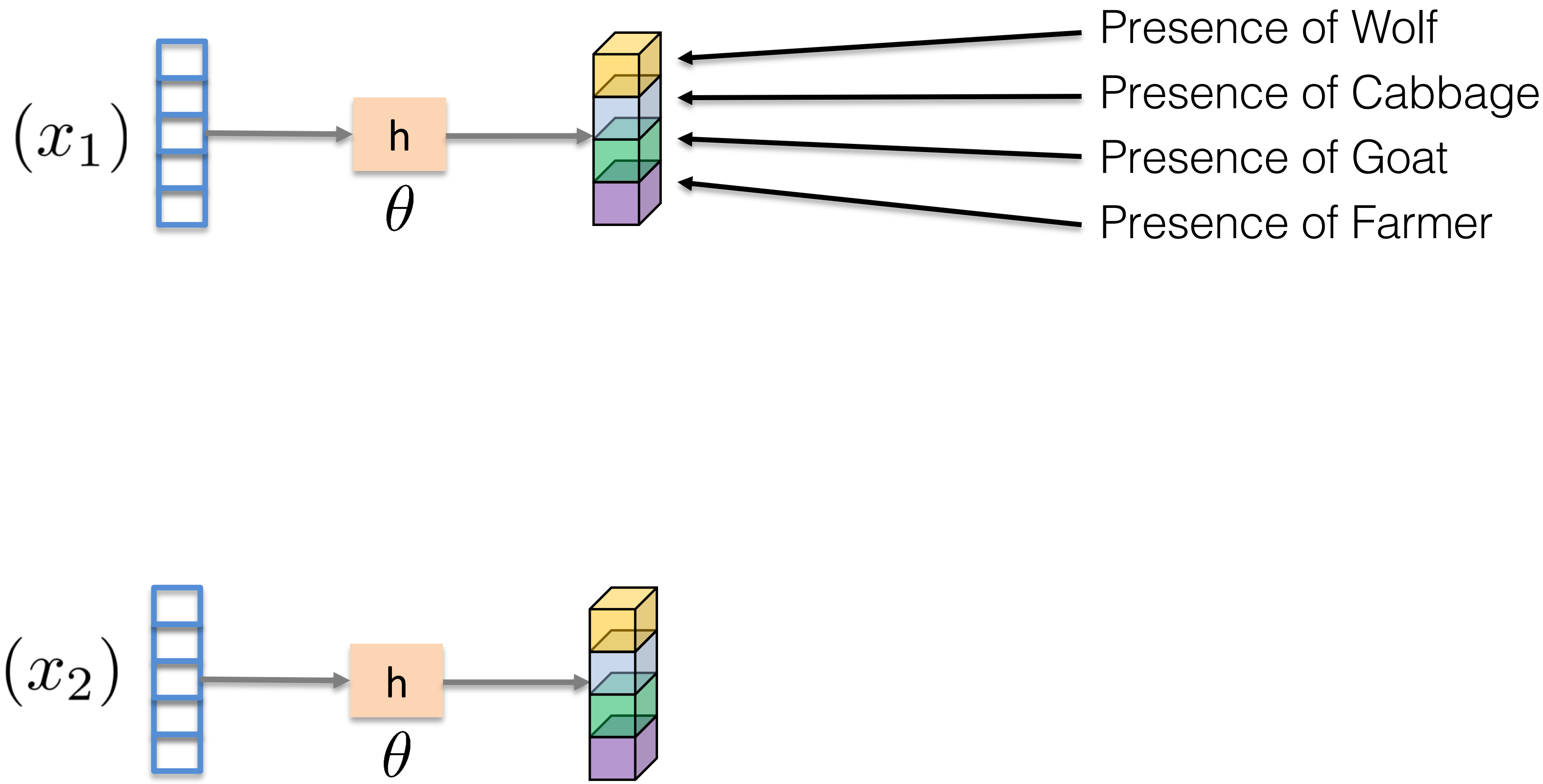


No

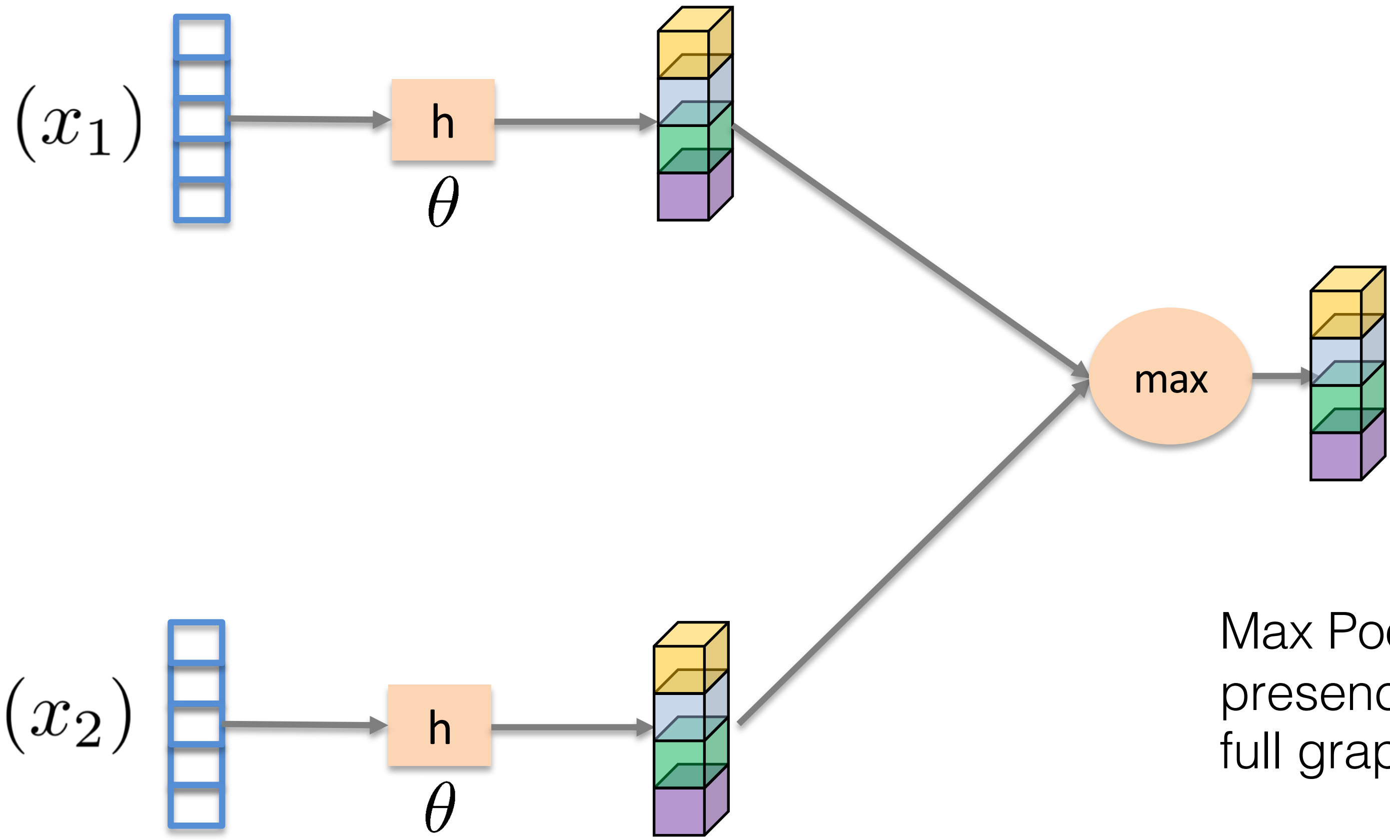
Representation by PointNet



Representation by PointNet

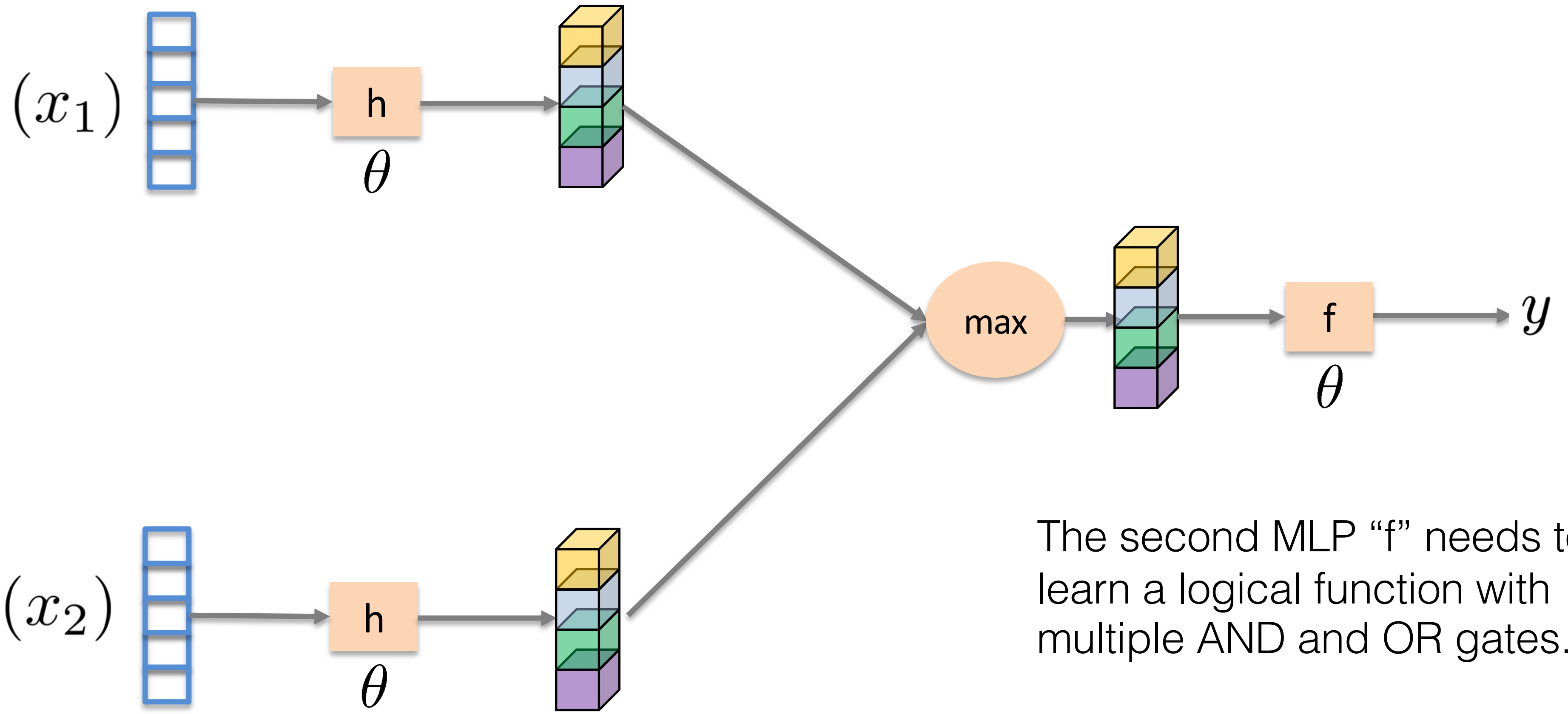


Representation by PointNet



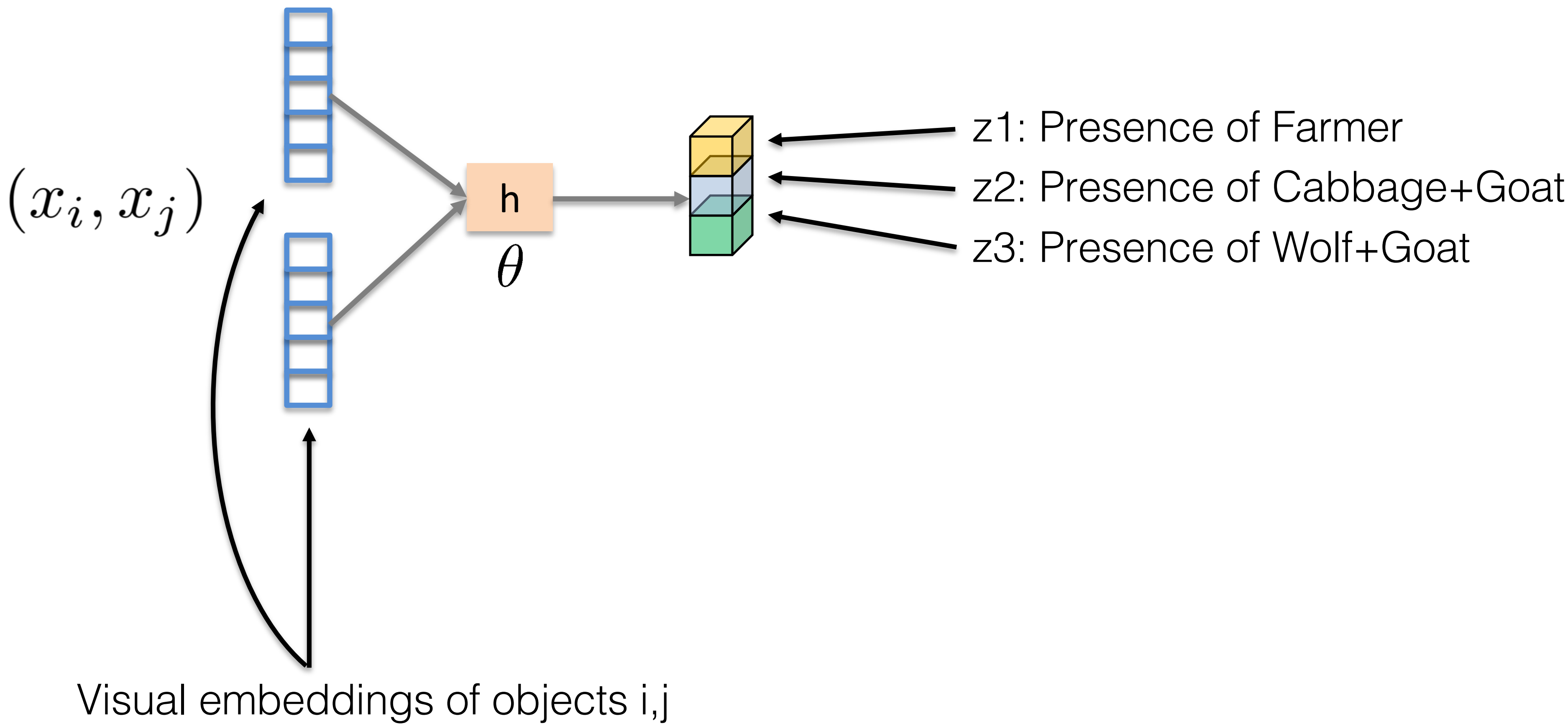
Max Pooling will cumulate the presence information from the full graph into a single vector

Representation by PointNet

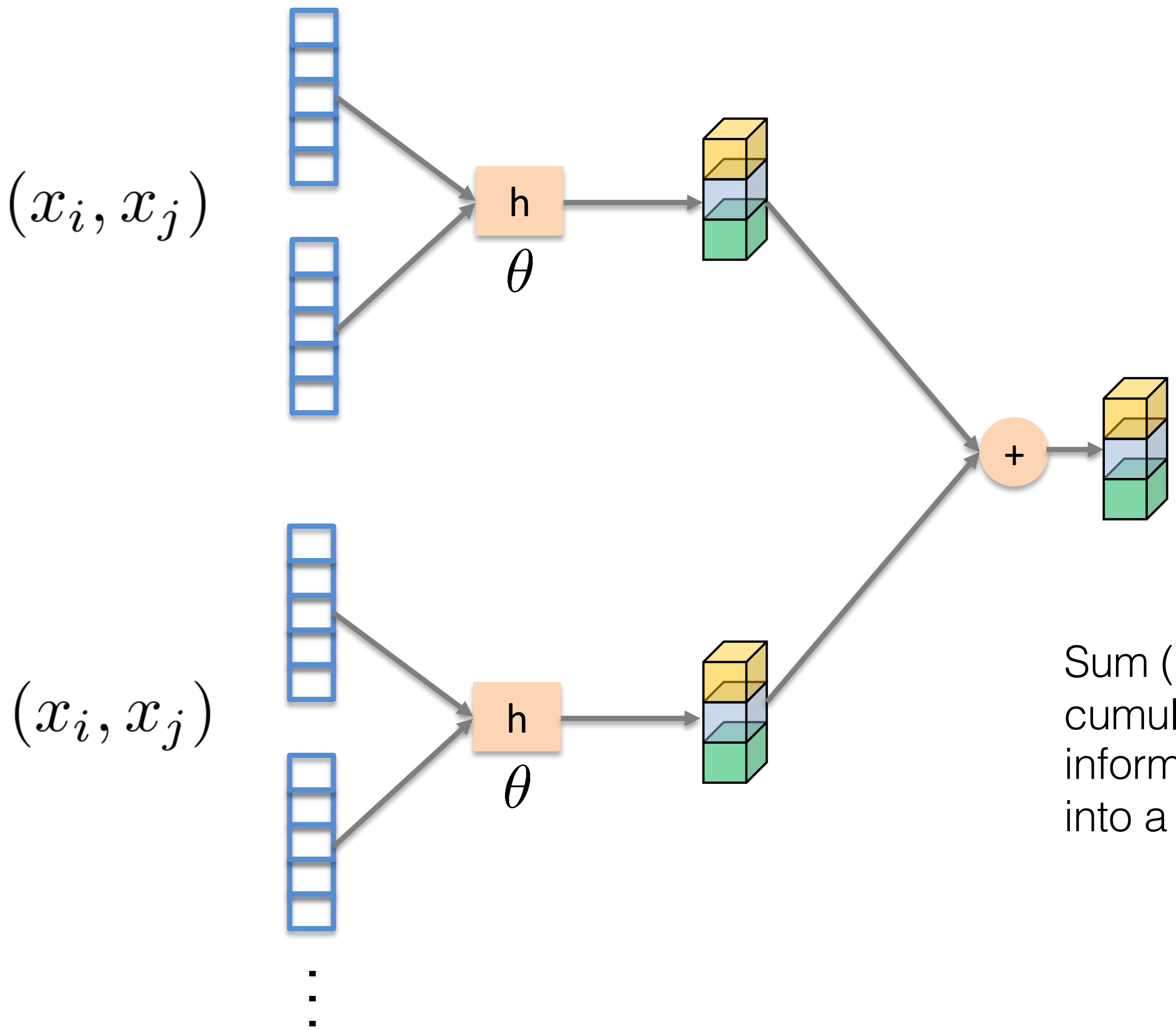


The second MLP "f" needs to learn a logical function with multiple AND and OR gates.

Representation with pairwise terms

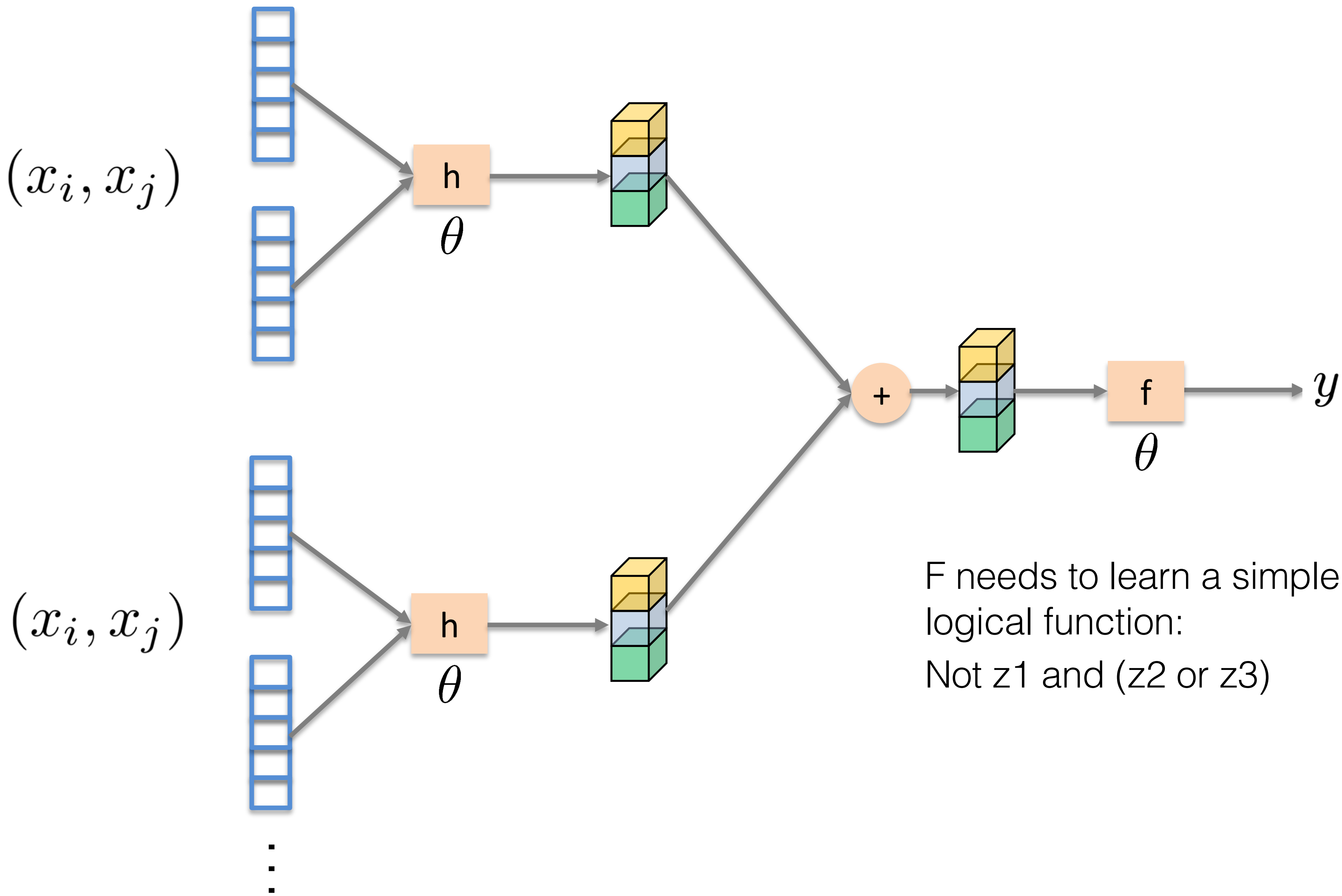


Representation with pairwise terms



Sum (or Max Pooling) will cumulate the logical information from the full graph into a single vector

Representation with pairwise terms

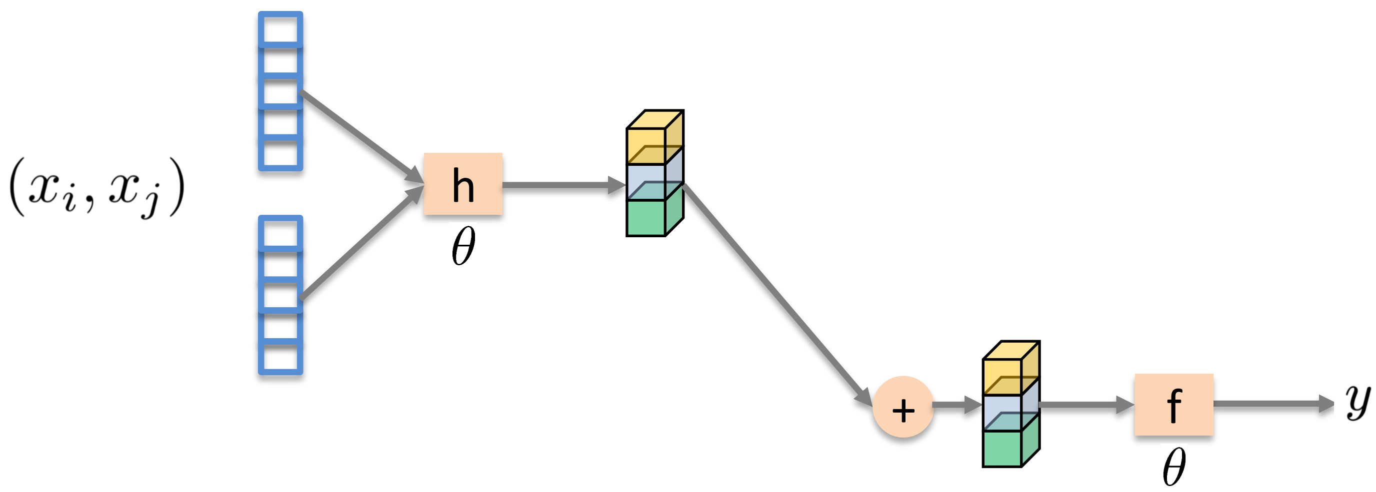


F needs to learn a simpler logical function:

Not z1 and (z2 or z3)

Comparison

- Increasing the complexity of h may allow to decrease the complexity of f.
- There is no known rule which determines the best trad-off between h and f for a given problem.
- Example: there are problems dominated by pairwies relationships in the data where models without pairwise terms work better.



Graph networks

Graph networks describe graphs with sets of embeddings:

- $\{x_0, \ldots, x_{|\mathcal{V}|}\}$ are node embeddings.
- $\{e_0, \ldots, e_{|\mathcal{E}|}\}$ are edge embeddings.
- $\boldsymbol{u}$ is an embedding of global graph information.

Graph networks update these embeddings by iteratively passing messages:

$$(x, e, \boldsymbol{u})' \leftarrow \phi(x, e, \boldsymbol{u})$$

Graph networks

A large class of **graph networks** (GN) exist. In a paper by Deepmind, a general class of models has been proposed, which generalizes a majority of known models:

`Battaglia et al., ‘‘Relational inductive biases, deep learning, and graph network’’, ICLR 2019`

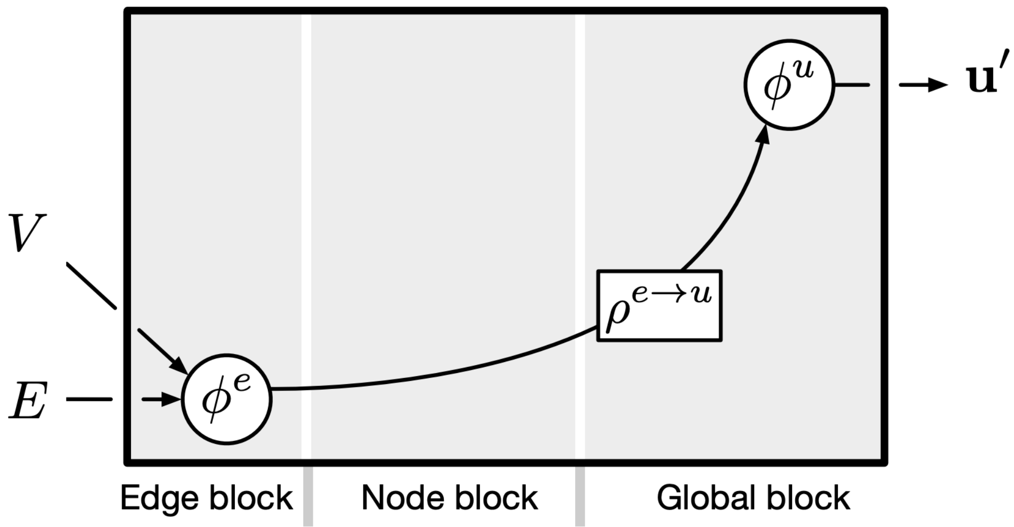
The different models differ in the way in which the function ϕ factorizes:

$$(x, e, \boldsymbol{u})' \leftarrow \phi(x, e, \boldsymbol{u})$$

Relational reasoning as GN

Relational reasoning can be expressed as graph networks:

$$df$$



Battaglia et al., ICLR 2019

GN: the general case

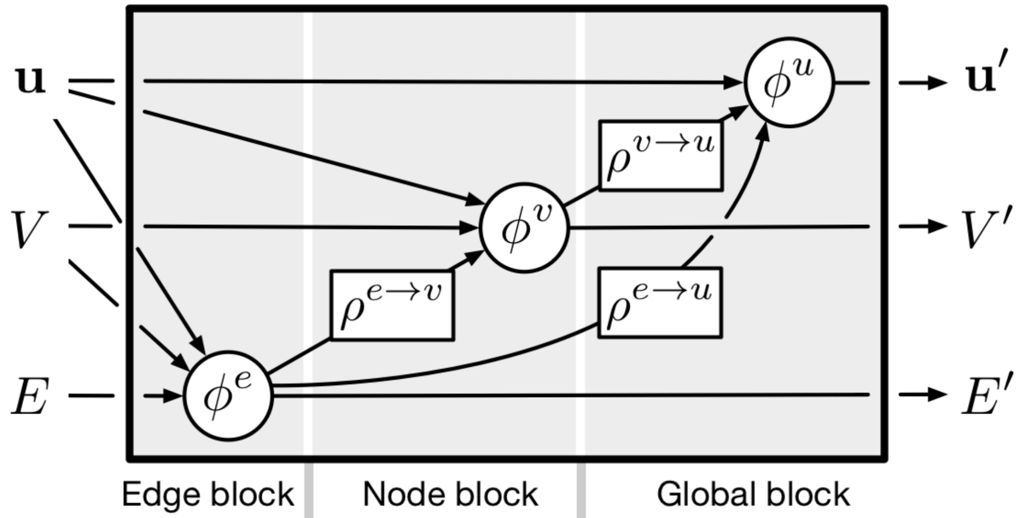
$$\mathbf{e}'_k = \phi^e\left(\mathbf{e}_k, \mathbf{x}_{r_k}, \mathbf{x}_{s_k}, \mathbf{u}\right) \qquad \bar{\mathbf{e}}'_i = \rho^{e\to x}\left(E'_i\right)$$

$$\mathbf{x}'_i = \phi^x\left(\bar{\mathbf{e}}'_i, \mathbf{x}_i, \mathbf{u}\right) \qquad \bar{\mathbf{e}}' = \rho^{e\to u}\left(E'\right)$$

$$\mathbf{u}' = \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{x}}', \mathbf{u}\right) \qquad \bar{\mathbf{x}}' = \rho^{x\to u}\left(X'\right)$$

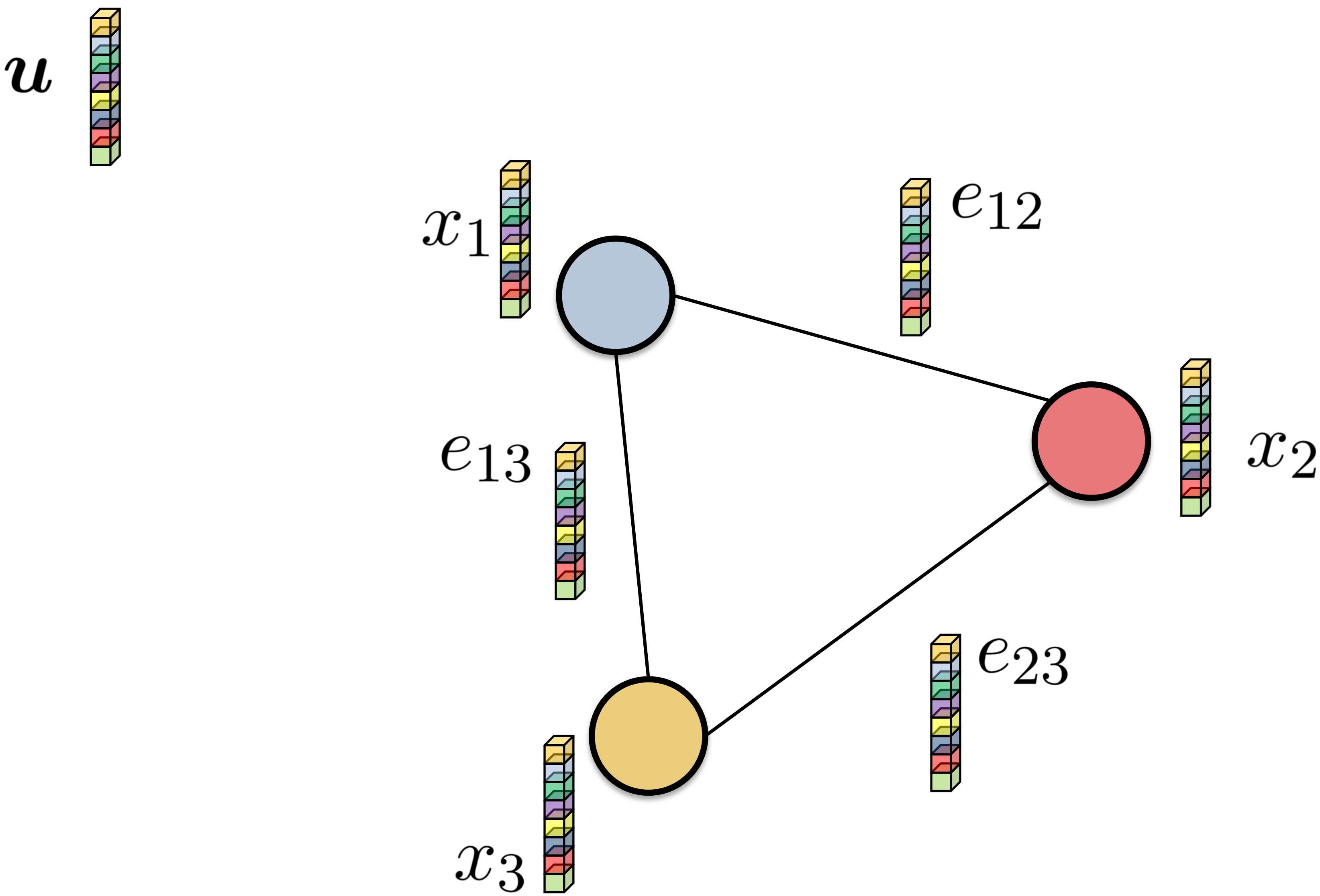
$$E'_i = \left\{\left(\mathbf{e}'_k, r_k, s_k\right)\right\}_{r_k=i,k=1:N^e}, \qquad X' = \left\{\mathbf{x}'_i\right\}_{i=1:N^x}$$

$$E' = \bigcup_i E'_i = \left\{\left(\mathbf{e}'_k, r_k, s_k\right)\right\}_{k=1:N^e}$$



Battaglia et al., ICLR 2019

Visualization

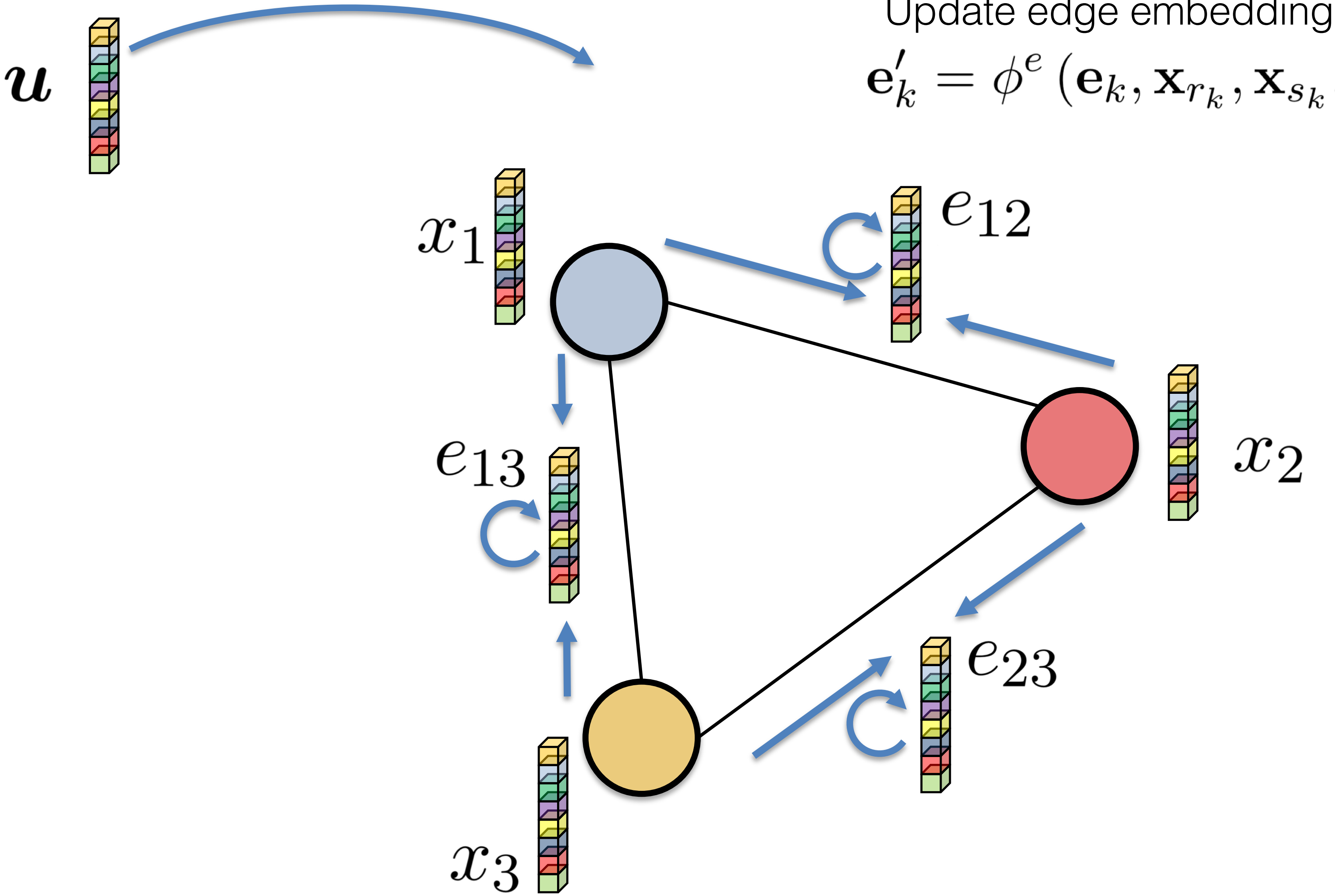


Visualization

All messages are influenced by the global (context) embedding

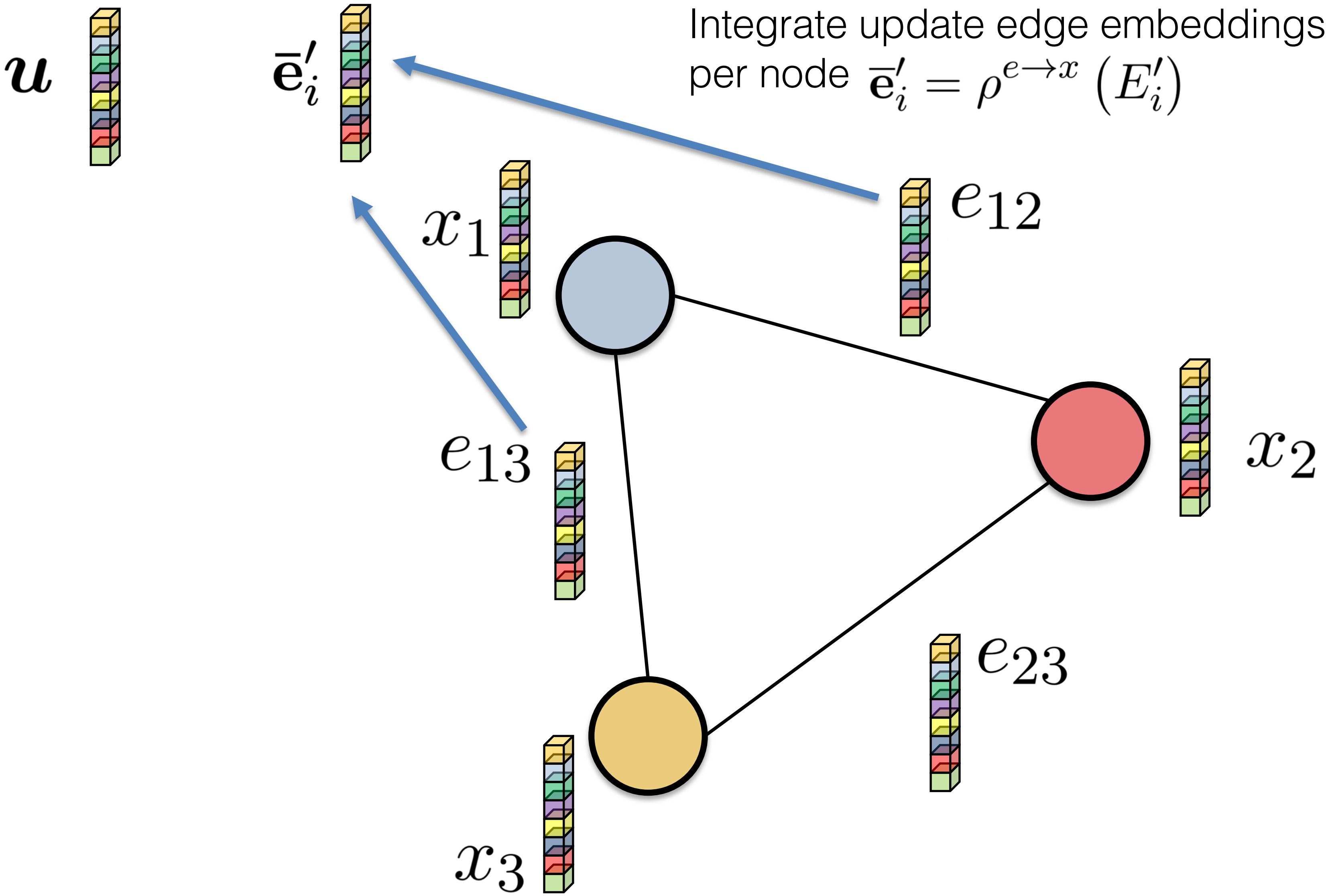
Update edge embeddings

$$\mathbf{e}'_k = \phi^e\left(\mathbf{e}_k, \mathbf{x}_{r_k}, \mathbf{x}_{s_k}, \mathbf{u}\right)$$

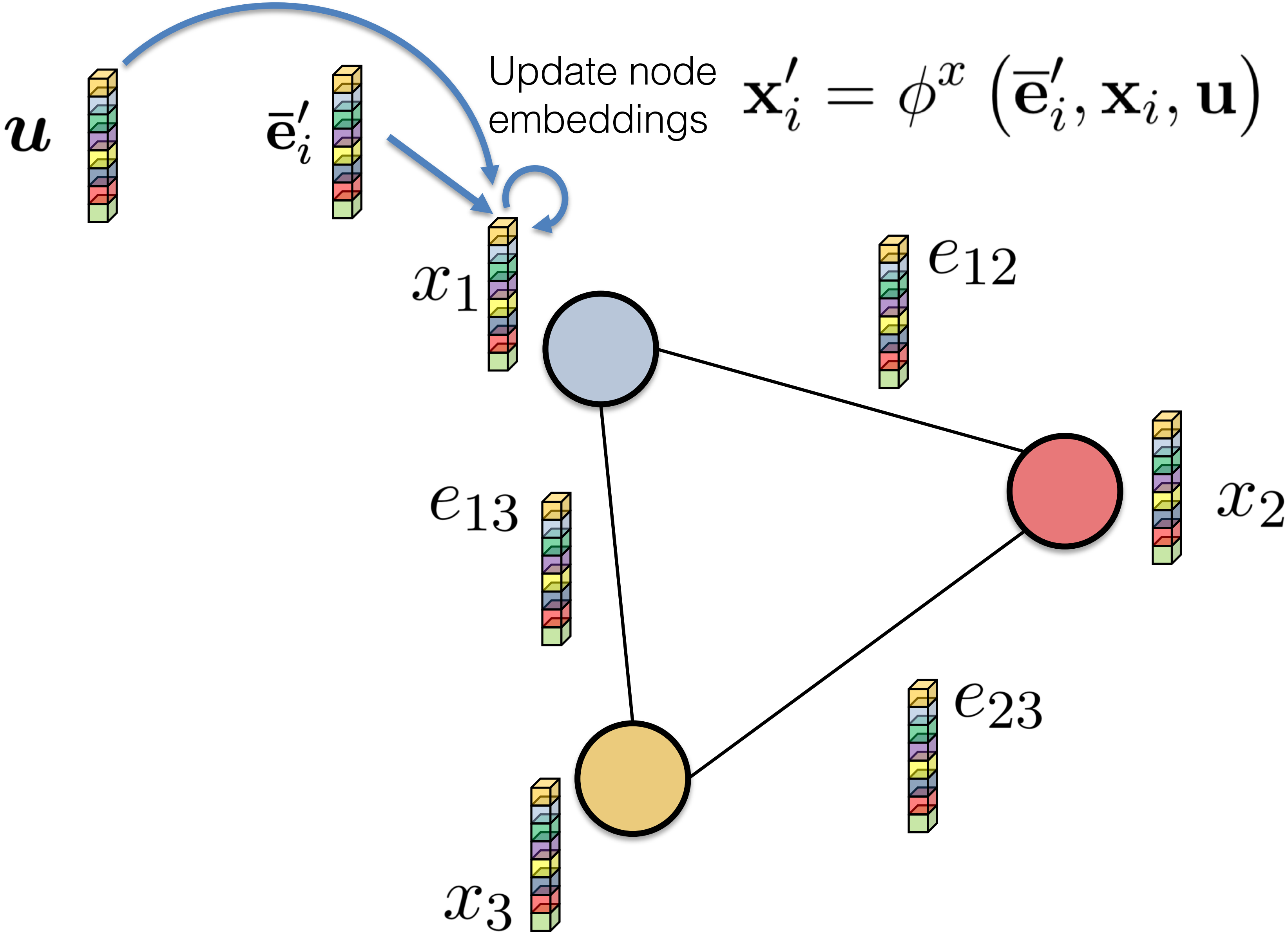


Visualization

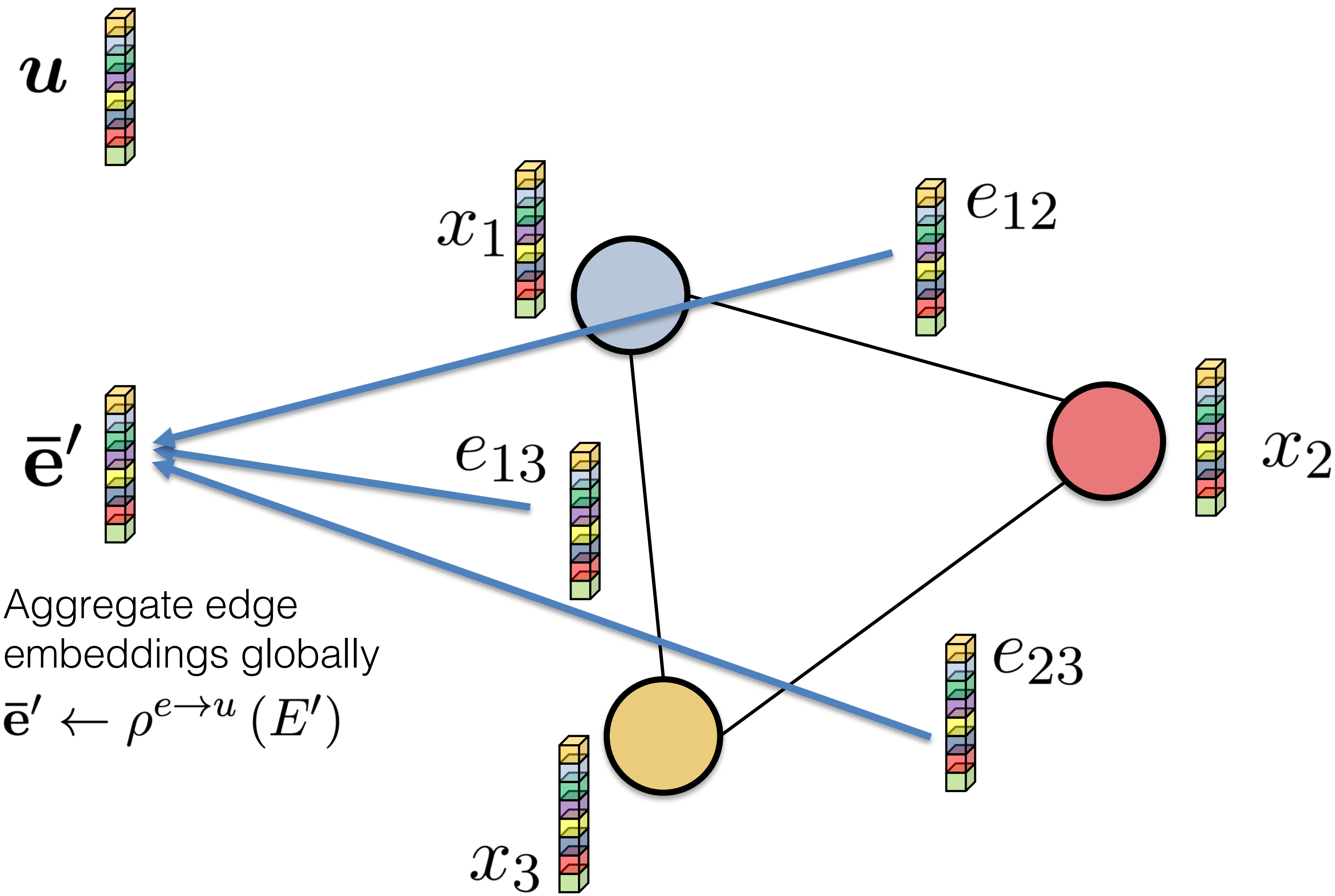
Integrate update edge embeddings per node $\bar{\mathbf{e}}'_i = \rho^{e \to x}\left(E'_i\right)$



Visualization



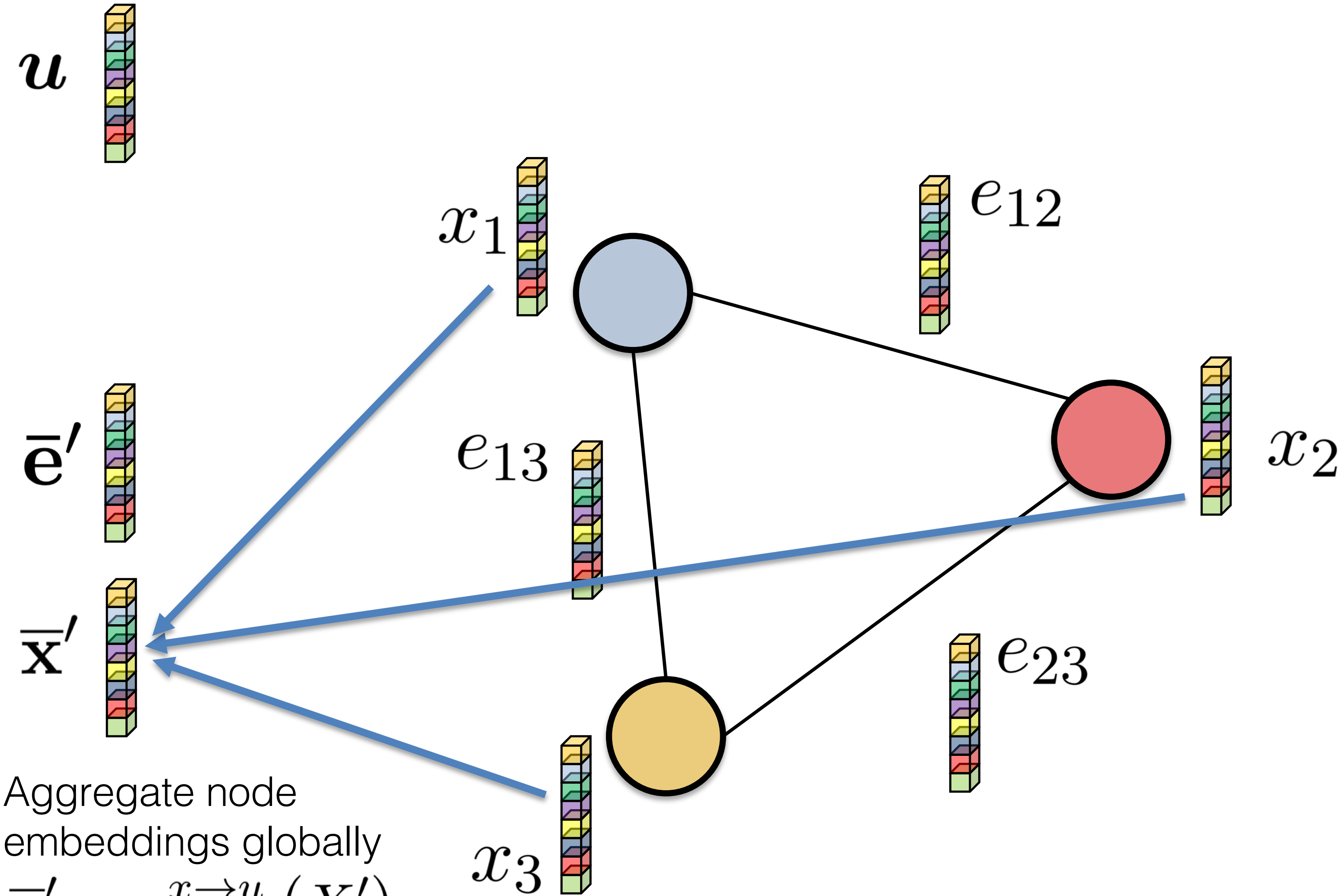
Visualization



Aggregate edge embeddings globally

$\bar{\mathbf{e}}' \leftarrow \rho^{e \rightarrow u}(E')$

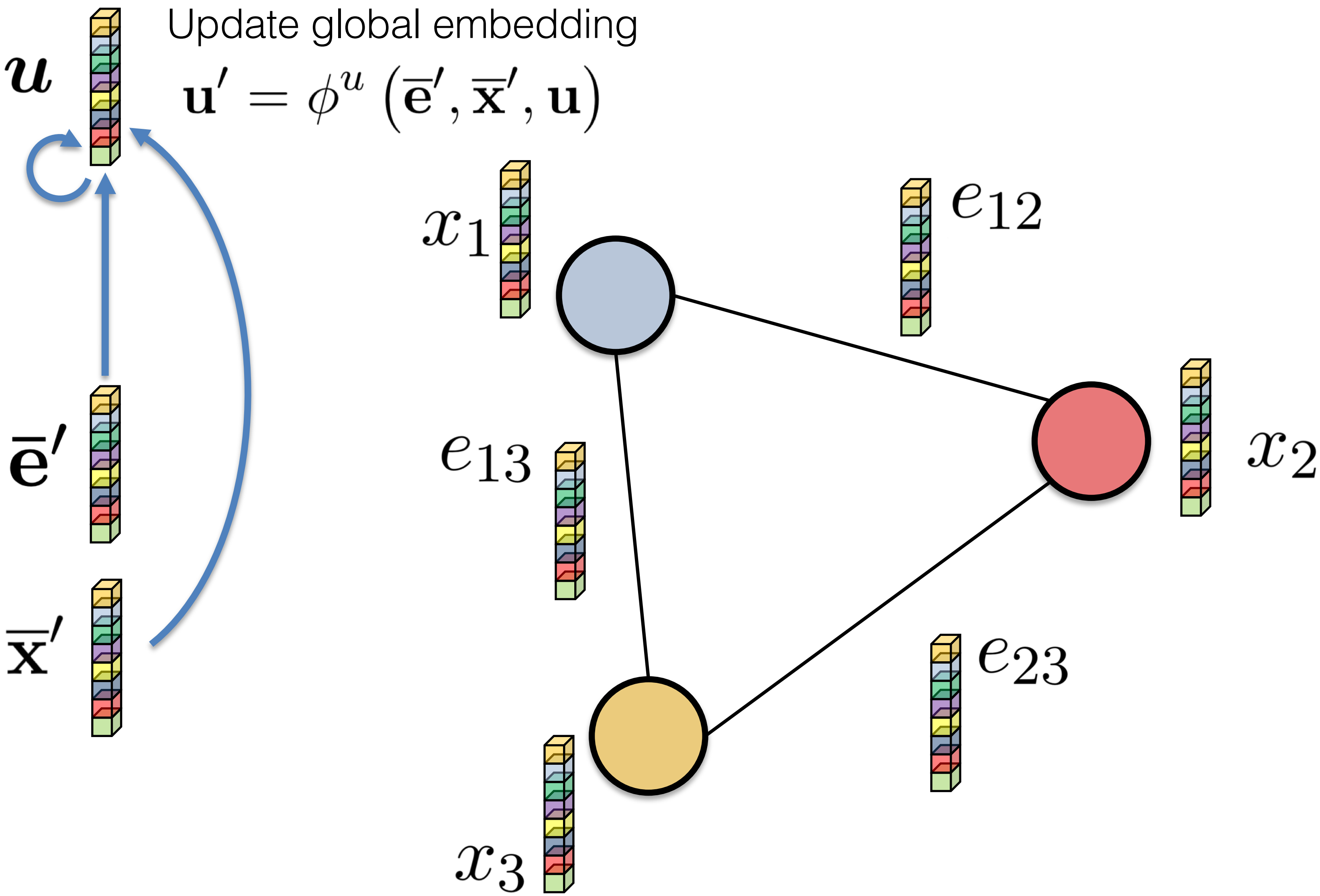
Visualization



Aggregate node embeddings globally

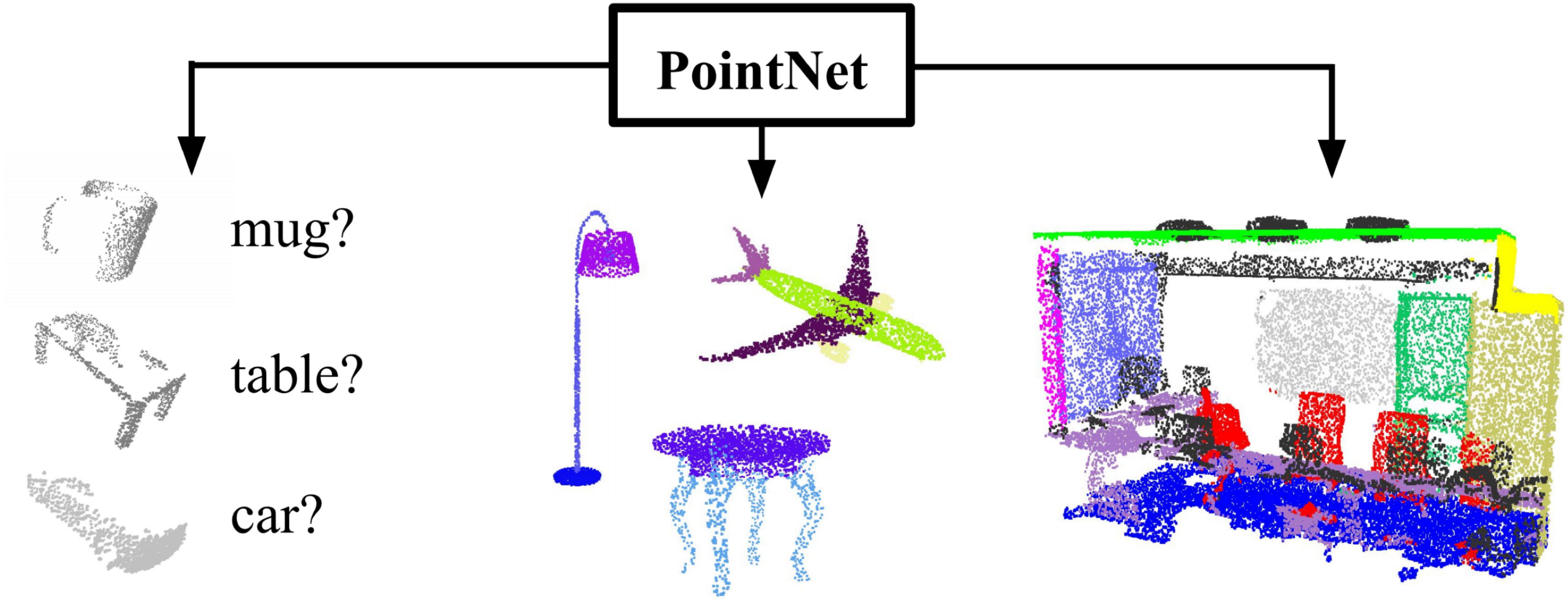
$$\overline{\mathbf{x}}' = \rho^{x \to u}\left(X'\right)$$

Visualization



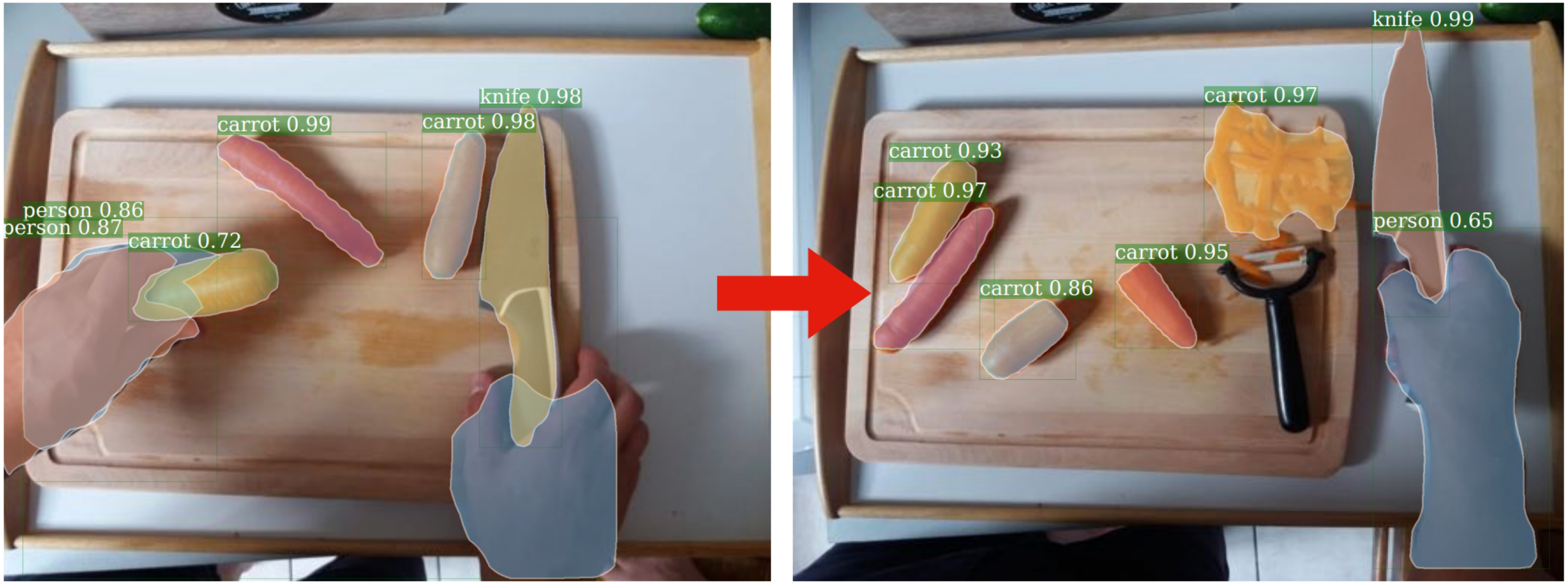
Example applications

PointNet



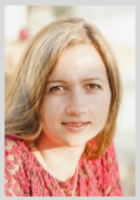
[Qi, Su, Mo, Guibas, CVPR 2017]

Example: Object level Visual Reasoning



[Baradel, Neverova, Wolf, Mille, Mori, ECCV 2018]

Fabien Baradel
Phd @ LIRIS,
INSA-Lyon



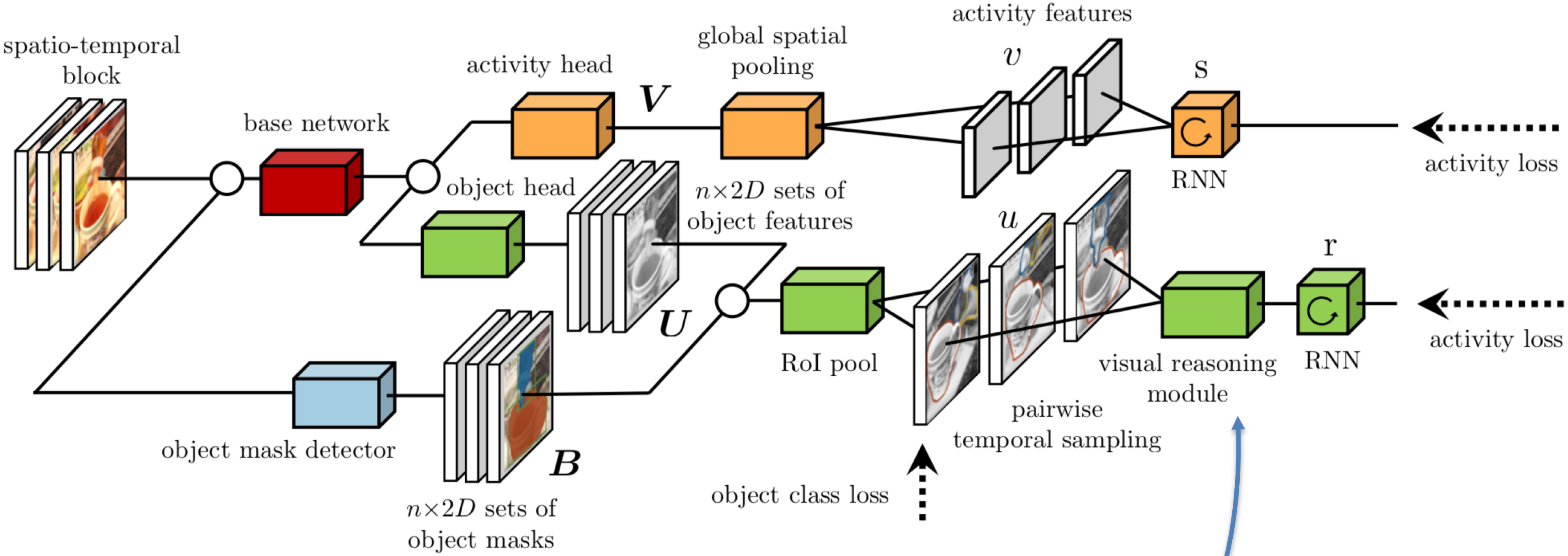
Natalia Neverova
Facebook AI
Research, Paris

Christian Wolf
Insa-Lyon,
LIRIS
INRIA Chroma

Julien Mille
LIFAT,
INSA VdL

Greg Mori
Simon
Fraser
University,
Canada

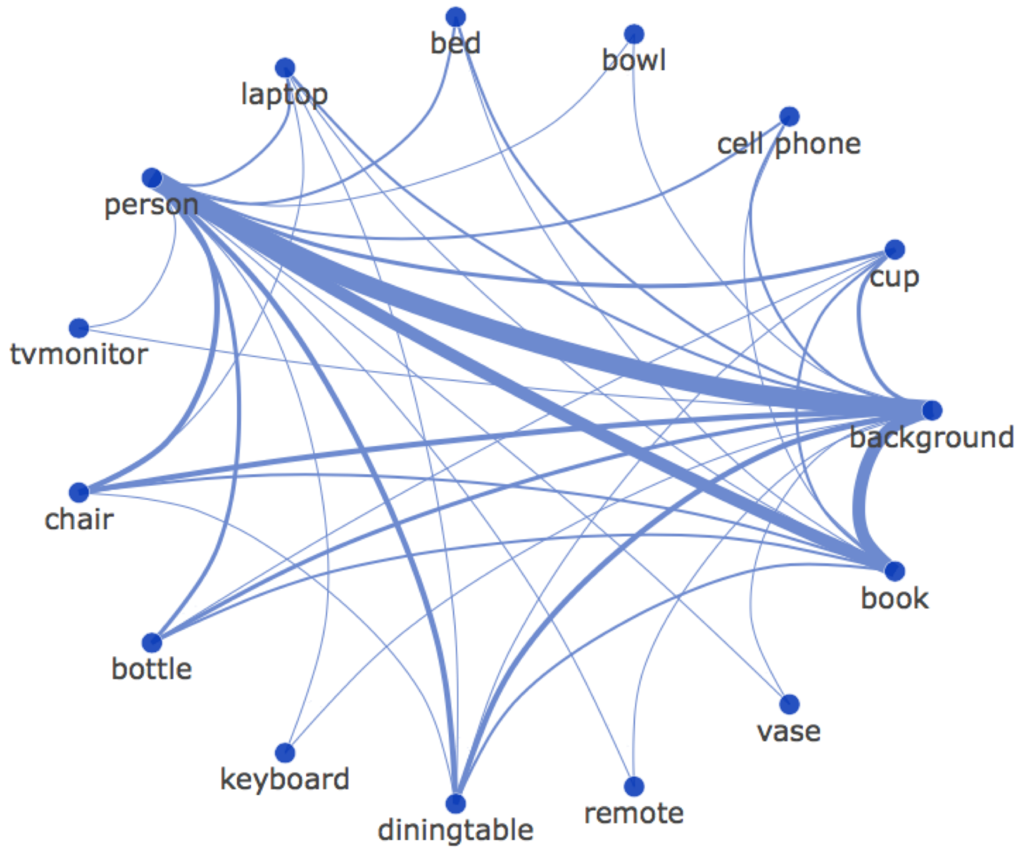
Object level Visual Reasoning

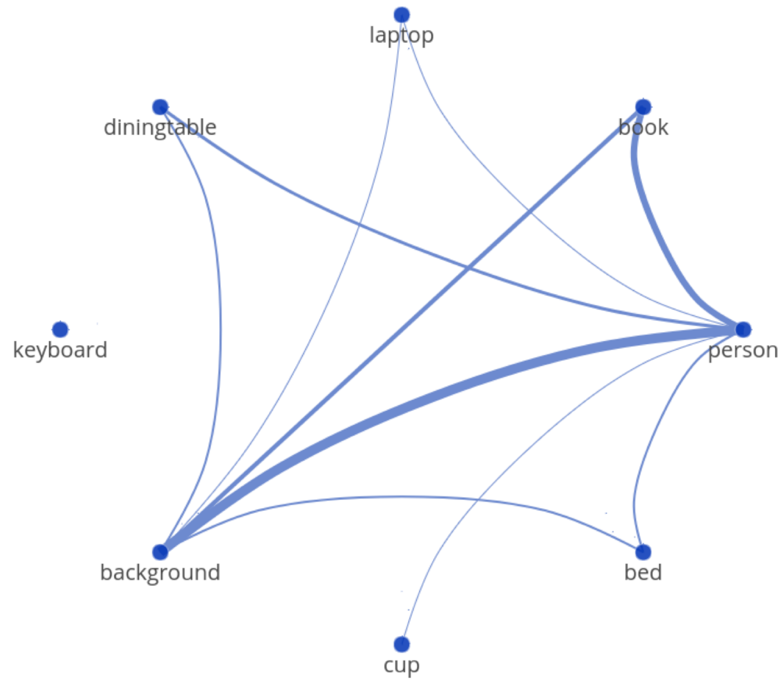


$$\mathbf{g}_t = \sum_{j,k} h_\theta(\mathbf{o}_{t'}^j, \mathbf{o}_t^k)$$

[Baradel, Neverova, Wolf, Mille, Mori, ECCV 2018]

Learned interactions





Class: person-book interaction

[Baradel, Neverova, Wolf, Mille, Mori, ECCV 2018]

Results

Methods	Top1
C3D + Avg [5]	21.50
I3D [5]	27.63
MultiScale TRN [39]	33.60
Ours	**34.32**

Something-something dataset

R50 [45]	40.5
I3D [3]	39.7
Ours	**41.7**

VLOG dataset

Methods	Top1
R18 [44]*	32.05
I3D-18 [3]*	34.20
Ours	**40.89**

EPIC Kitchen dataset

	Nb. head		Object type		f_ϕ		Pairwise relations	Results	
	1	2	Pixel	COCO	RNN	MLP		VLOG	Something
Baseline	-	-	-	-	-	-	-	29.92	33.43
Variant 1	✓	-	-	✓	✓	-	✓	32.01	35.09
Variant 2	-	✓	✓	-	✓	-	✓	31.36	35.15
Variant 3	-	✓	-	✓	-	✓	✓	32.38	34.15
Variant 4	-	✓	-	✓	✓	-	-	31.82	34.65
Ours	-	✓	-	✓	✓	-	✓	**33.75**	**36.12**

[Baradel, Neverova, Wolf, Mille, Mori, ECCV 2018]