Learning Co-segmentation by Segment Swapping for Retrieval and Discovery

Xi Shen[1] Alexei A. Efros[2] Armand Joulin[3] Mathieu Aubry[1]




[1] LIGM (UMR 8049) - Ecole des Ponts, UPE

[2] UC Berkeley

[3] Meta AI

Task: co-segmentation in a pair of images

Input

predicted masks

Problem: no training data available

Roadmap

- 1. Co-segmentation by Segment Swapping
- 2. Application to Image Retrieval
- 3. Application to Discovery on a Database
- 4. Summary

Roadmap

- 1. Co-segmentation by Segment Swapping
- 2. Application to Image Retrieval
- 3. Application to Discovery on a Database
- 4. Summary

Key idea: synthetic pairs with duplicated patterns

Source and selected segment

Background

Key idea: direct copy-paste

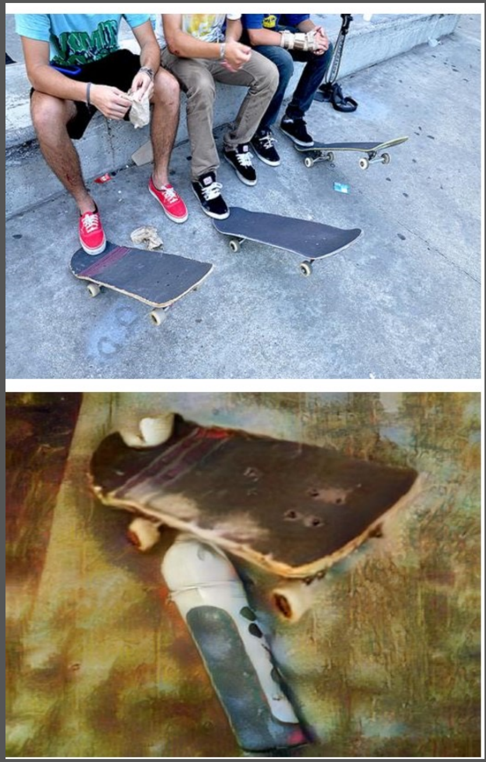
Direct
Copy-paste

Key idea: our blending

Our blending

Poisson blending
[Pérez et al. 2003]
+
Style transfer
[Huang and Belongie 2017]

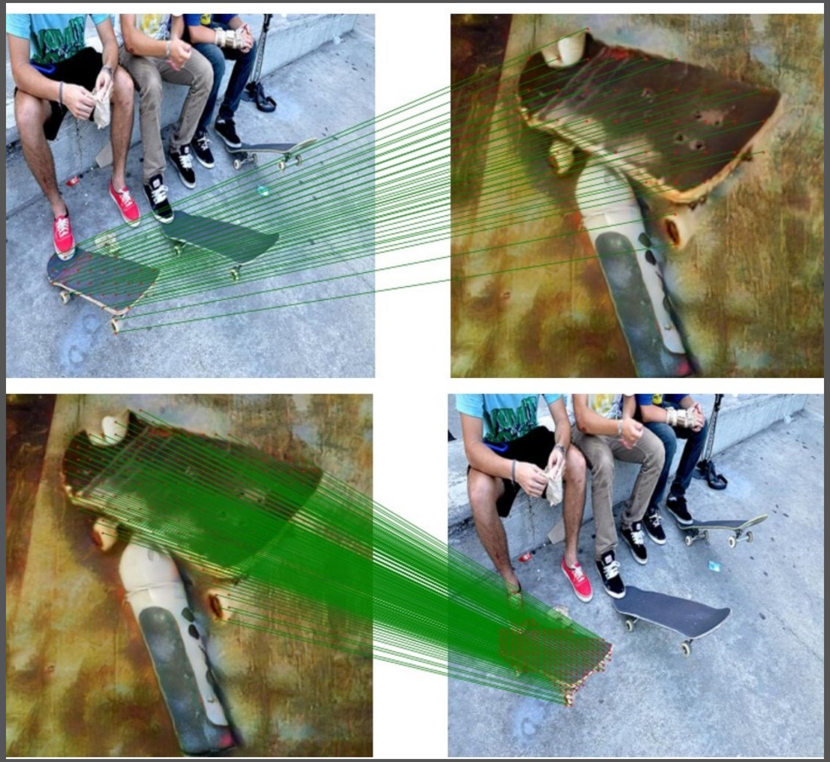
Annotations



Generated images

Annotations

Masks



Correspondences

Training data

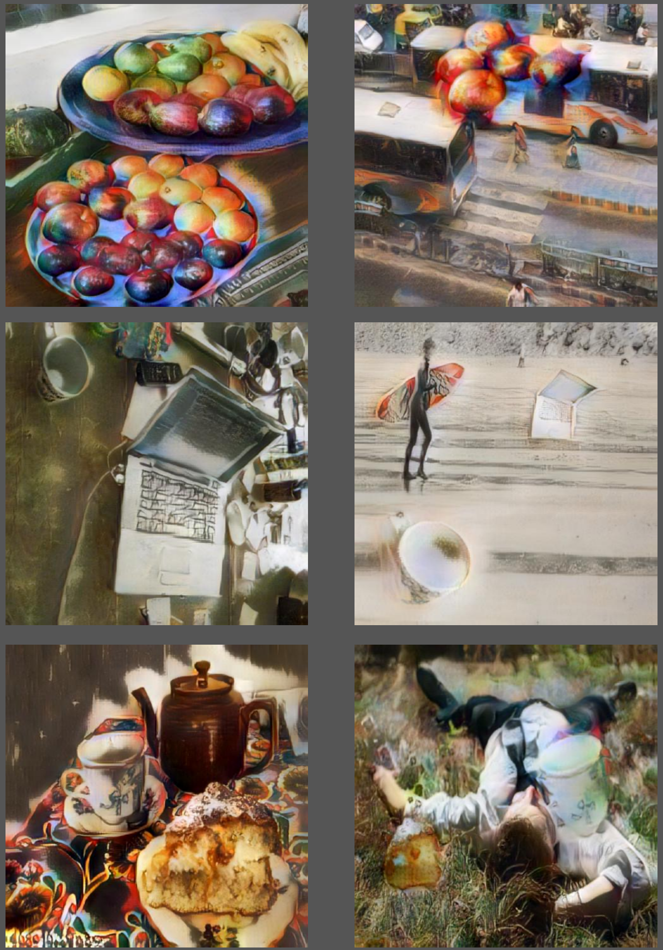
One object
COCO Seg.

Two objects
COCO Seg.

Two objects
Unsup. Seg.

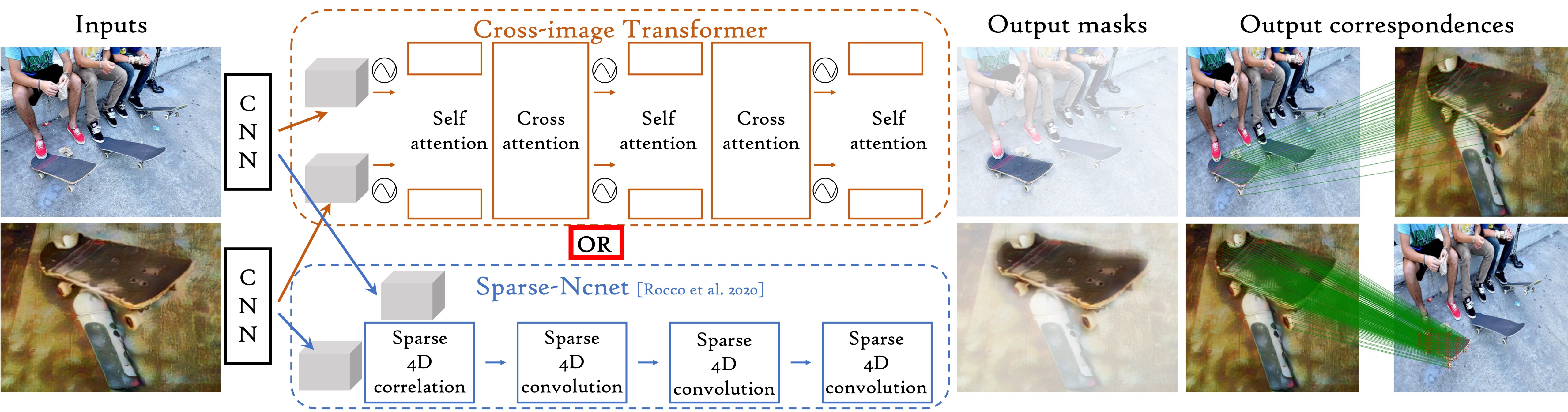
Source Blended

Style transfer

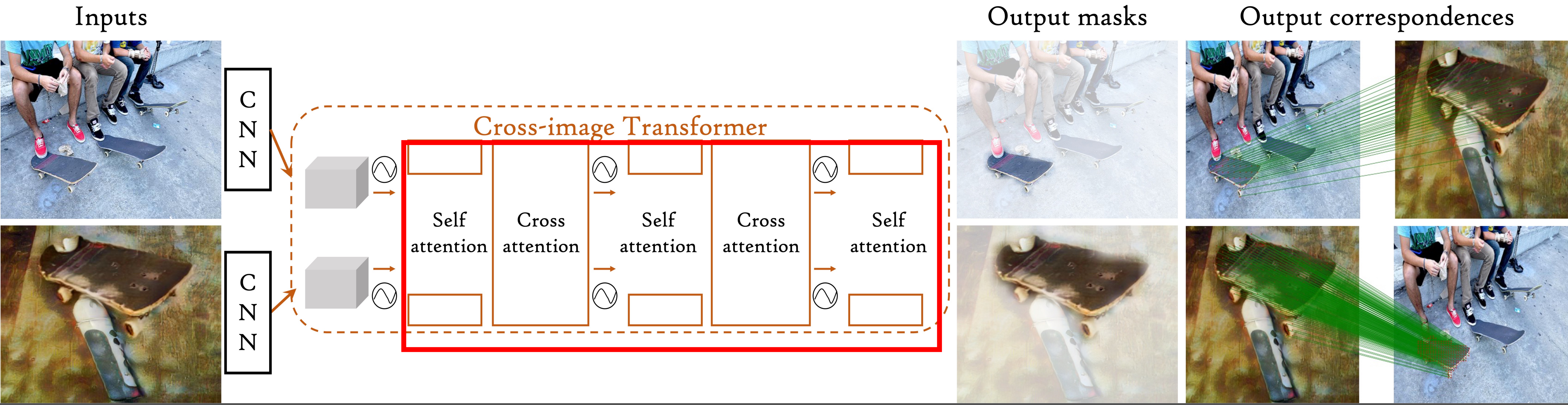


Source Blended

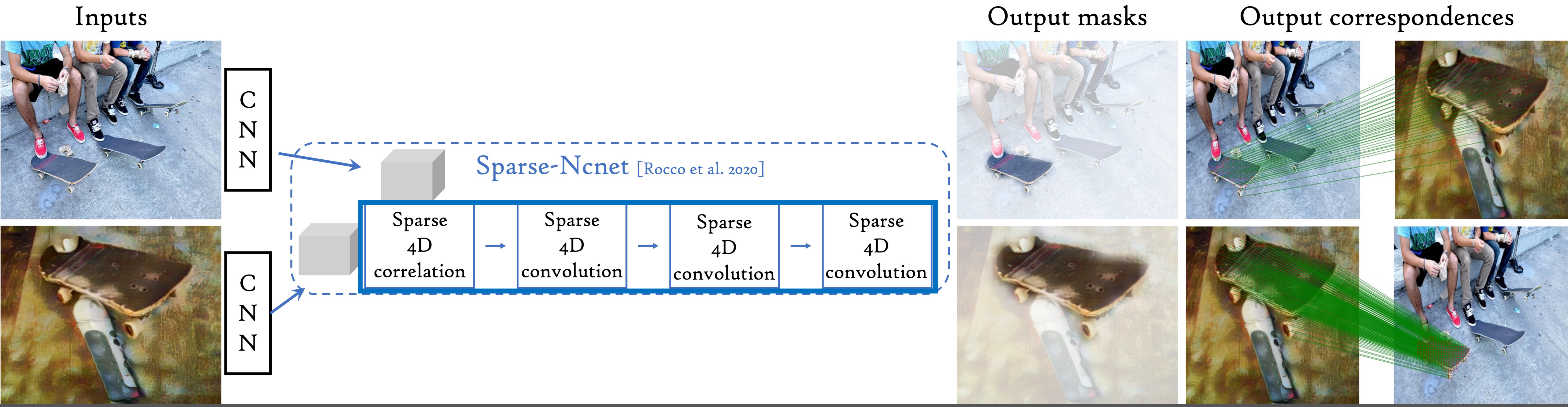
Learning co-segmentation



Learning co-segmentation



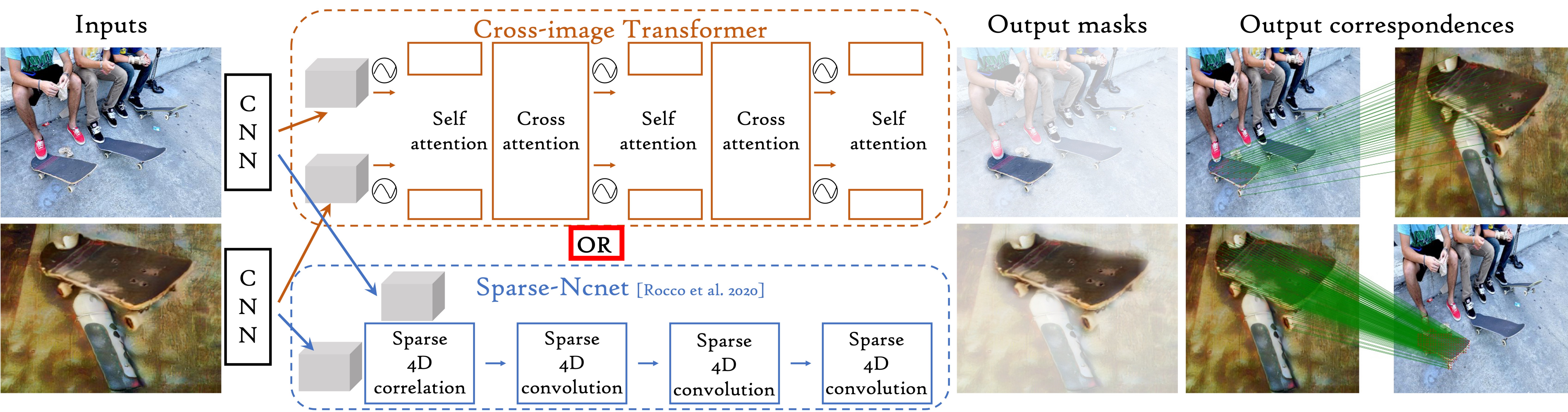
Learning co-segmentation



Objective function:

$$\mathcal{L}^s_{sup} = \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^s)}_{\mathcal{L}_{mask}} + \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^t(\mathbf{C}^{s\to t}))}_{\mathcal{L}_{tmask}} + \eta \underbrace{\frac{1}{\sum_{i,j}\mathbf{M}^s_{gt}(i,j)} \sum_{i,j} \mathbf{M}^s_{gt}(i,j)\|\mathbf{C}^{s\to t}(i,j) - \mathbf{C}^{s\to t}_{gt}(i,j)\|}_{\mathcal{L}_{corr}}$$

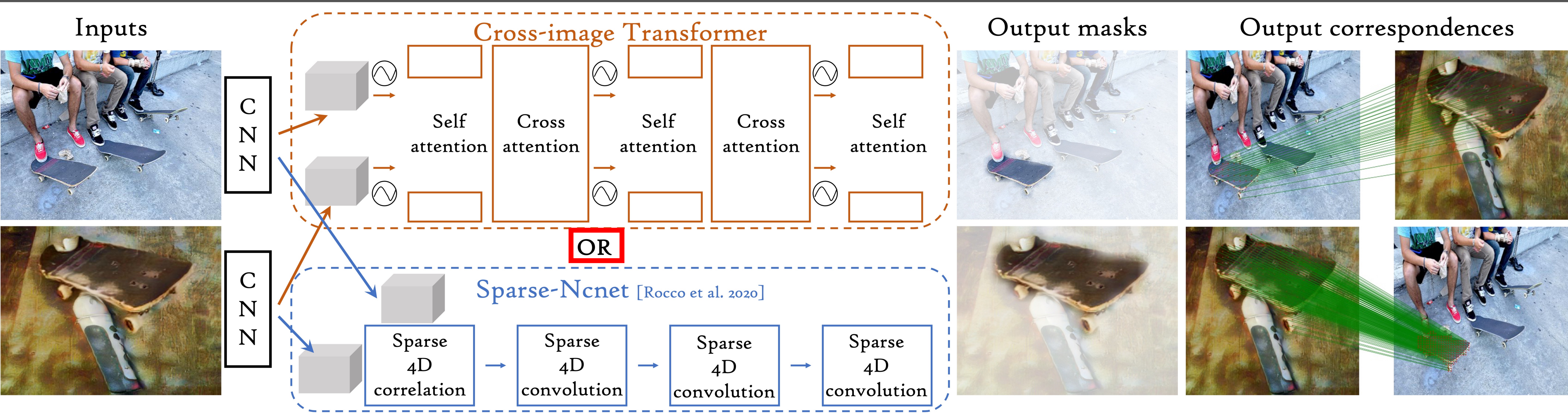
Learning co-segmentation



Objective function:

$$\mathcal{L}^s_{sup} = \boxed{\underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^s)}_{\mathcal{L}_{mask}}} + \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^t(\mathbf{C}^{s\to t}))}_{\mathcal{L}_{tmask}} + \underbrace{\eta \frac{1}{\sum_{i,j} \mathbf{M}^s_{gt}(i,j)} \sum_{i,j} \mathbf{M}^s_{gt}(i,j) \| \mathbf{C}^{s\to t}(i,j) - \mathbf{C}^{s\to t}_{gt}(i,j) \|}_{\mathcal{L}_{corr}}$$

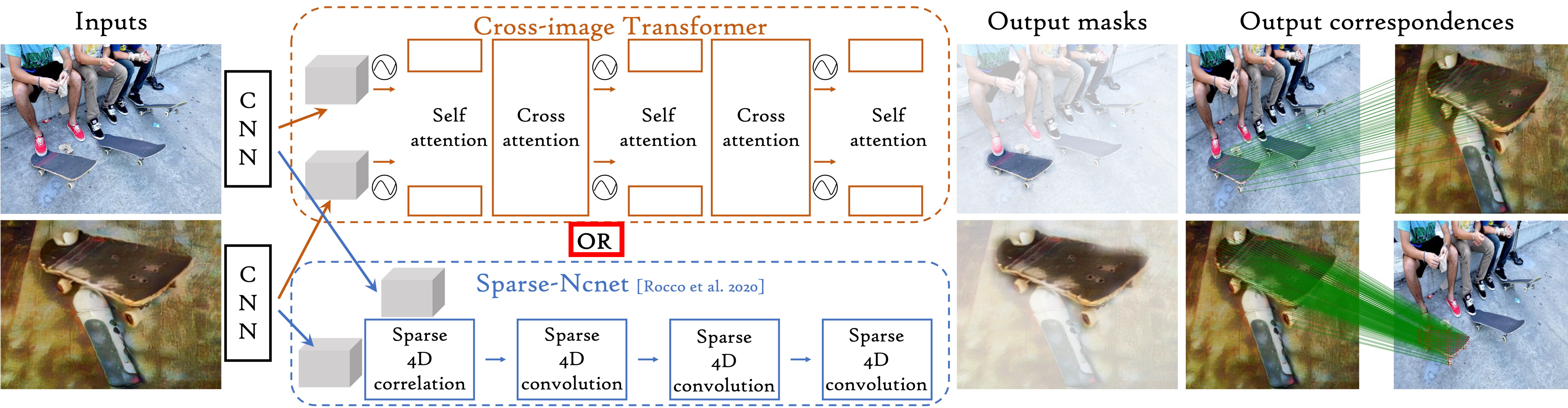
Learning co-segmentation



Objective function:

$$\mathcal{L}^s_{sup} = \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^s)}_{\mathcal{L}_{mask}} + \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^t(\mathbf{C}^{s\to t}))}_{\mathcal{L}_{tmask}} + \boxed{\underbrace{\eta \frac{1}{\sum_{i,j} \mathbf{M}^s_{gt}(i,j)} \sum_{i,j} \mathbf{M}^s_{gt}(i,j) \|\mathbf{C}^{s\to t}(i,j) - \mathbf{C}^{s\to t}_{gt}(i,j)\|}_{\mathcal{L}_{corr}}}$$

Learning co-segmentation



Objective function:

$$\mathcal{L}^s_{sup} = \underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^s)}_{\mathcal{L}_{mask}} + \boxed{\underbrace{CE(\mathbf{M}^s_{gt}, \mathbf{M}^t(\mathbf{C}^{s\to t}))}_{\mathcal{L}_{tmask}}} + \underbrace{\eta \frac{1}{\sum_{i,j} \mathbf{M}^s_{gt}(i,j)} \sum_{i,j} \mathbf{M}^s_{gt}(i,j) \| \mathbf{C}^{s\to t}(i,j) - \mathbf{C}^{s\to t}_{gt}(i,j) \|}_{\mathcal{L}_{corr}}$$

Roadmap

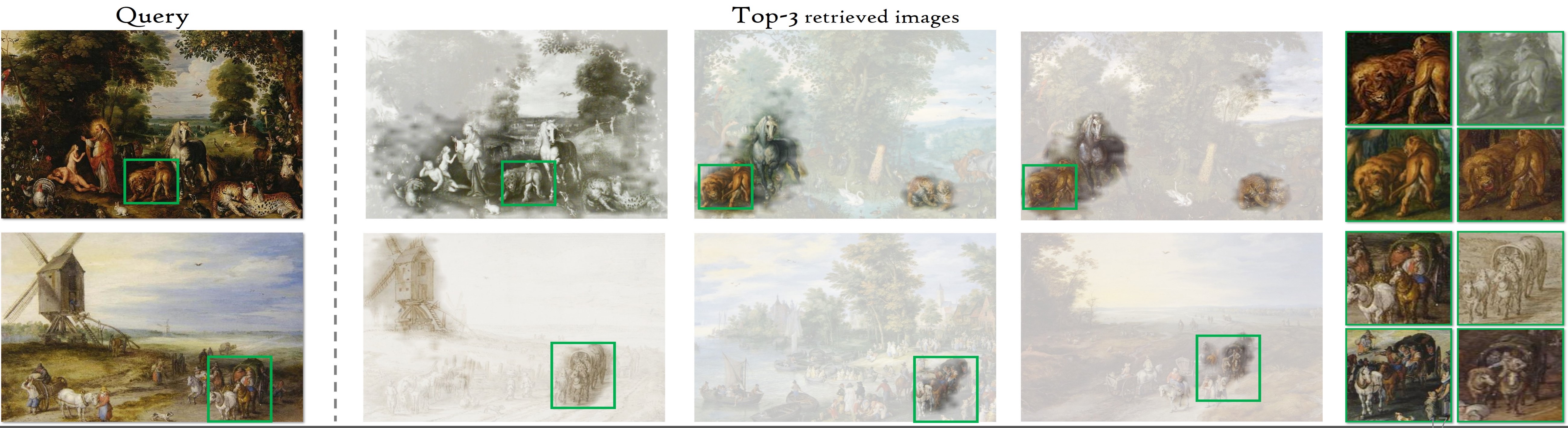
- 1. Co-segmentation by Segment Swapping
- 2. Application to Image Retrieval
- 3. Application to Discovery on a Database
- 4. Summary

One-shot detection on Brueghel

Score between a pair of images

$$\mathcal{S}(\mathbf{I}^s, \mathbf{I}^t) = \sum_{i,j} \underbrace{\mathbf{M}^s_{joint}(i,j)}_{\text{Mask}} \boxed{\underbrace{cos(\mathbf{F}^s(i,j), \mathbf{F}^t(\mathbf{C}^{s\to t}(i,j)))}_{\text{Feat. similarity}}}$$

$$\mathbf{M}^s_{joint}(i,j) = \mathbf{M}^t(\mathbf{C}^{s\to t}(i,j))\mathbf{M}^s(i,j)$$



One-shot detection on Brueghel

Feat. + Methods	mAP Retrieval	mAP Det.(IoU > 0.3)
Shen et al. [53] + cos [53]	75.5	75.3
Shen et al. [53] + discovery [53]	76.6	76.4
MocoV2 [8] + cos [53]	79.0	78.7
MocoV2 [8] + discovery [53]	80.8	79.6
Ours + Unsupervised segments		
Transformer	81.8	79.4
Sparse-Ncnet	82.8	73.4
Ours + COCO segments [35]		
Transformer	**84.4**	**81.8**
Sparse-Ncnet	83.3	73.7

Table 1. Art detail retrieval and detection on Brueghel [53]. For detection, we employ ArtMiner (Brueghel [53] + cos [53]) as a post-processing and reports results with IoU > 0.3 [53]

One-shot detection on Brueghel

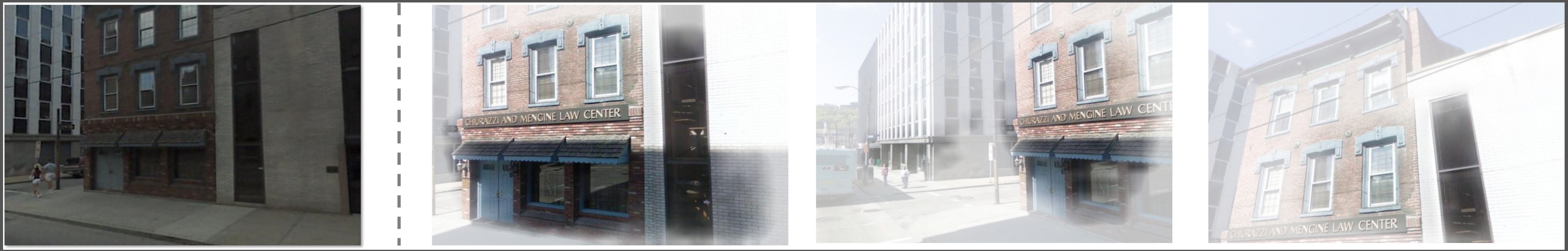
Feat. + Methods	mAP Retrieval	mAP Det.(IoU > 0.3)
Shen et al. [53] + cos [53]	75.5	75.3
Shen et al. [53] + discovery [53]	76.6	76.4
MocoV2 [8] + cos [53]	79.0	78.7
MocoV2 [8] + discovery [53]	80.8	79.6
Ours + Unsupervised segments		
Transformer	81.8	79.4
Sparse-Ncnet	82.8	73.4
Ours + COCO segments [35]		
Transformer	**84.4**	**81.8**
Sparse-Ncnet	83.3	73.7

Table 1. Art detail retrieval and detection on Brueghel [53]. For detection, we employ ArtMiner (Brueghel [53] + cos [53]) as a post-processing and reports results with IoU > 0.3 [53]

Place recognition on Pitts30K and Tokyo 24/7

Place recognition Tokyo24/7 [Torii et al. 2015]

Place recognition Pitts30K [Torri et al. 2013]



Ablation on the architectures and losses: http://imagine.enpc.fr/~shenx/SegSwap/

Place recognition on Pitts30K and Tokyo 24/7

Method	Supervision	Tokyo 24/7 [61]			Pitts30k-test [62]		
		R@1	R@5	R@10	R@1	R@5	R@10
AP-GEM [20,44]	Image location	40.3	55.6	65.4	75.3	89.3	92.5
DenseVLAD [20,61]	Image location	59.4	67.3	72.1	77.7	88.3	91.6
NetVLAD [3,20]	Image location	73.3	82.9	86.0	86.0	93.2	95.1
CRN [16,29]	Image location	75.2	83.8	87.3	-	-	-
SARE [16,38]	Image location	79.7	86.7	90.5	-	-	-
IBL [16]	Image location	**85.4**	**91.1**	**93.3**	-	-	-
Re-ranking Top-100 from NetVLAD [3,20]							
Patch-NetVLAD [20]	Image location	81.9	85.7	87.9	88.6	94.5	95.8
Patch-NetVLAD [20] + RANSAC	Image location	86.0	88.6	**90.5**	**88.7**	94.5	95.9
SuperGlue [20,50]*	Pose+Depth	**88.2**	**90.2**	90.2	**88.7**	**95.1**	**96.4**
Ours + Unsupervised segments							
Transformer	Segment swapping	74.0	82.9	86.0	85.2	93.5	95.4
Nc-Net	Segment swapping	84.1	87.0	88.9	86.4	94.3	95.6
Ours + COCO segments [36]							
Transformer	Segment swapping	80.0	86.0	87.9	84.7	93.5	95.6
Nc-Net	Segment swapping	85.4	88.3	89.2	86.8	94.4	95.8

* uses learnt keypoint detector Superpoint [14]

Table 2. Image-based localization on Tokyo 24/7 [61] and Pitts30k [62]. We follow Patch-NetVLAD [20] and re-rank the top-100 images ranked by NetVLAD [3] features.

Roadmap

- 1. Co-segmentation by Segment Swapping
- 2. Application to Image Retrieval
- 3. Application to Discovery on a Database
- 4. Summary

Discovery on Brueghel: Correspondences graph

Correspondences $\mathcal{V} = \{v_1, v_2, ..., v_i, ..., v_j, ...\}$

Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Mask Pred.

Corr. Pred.

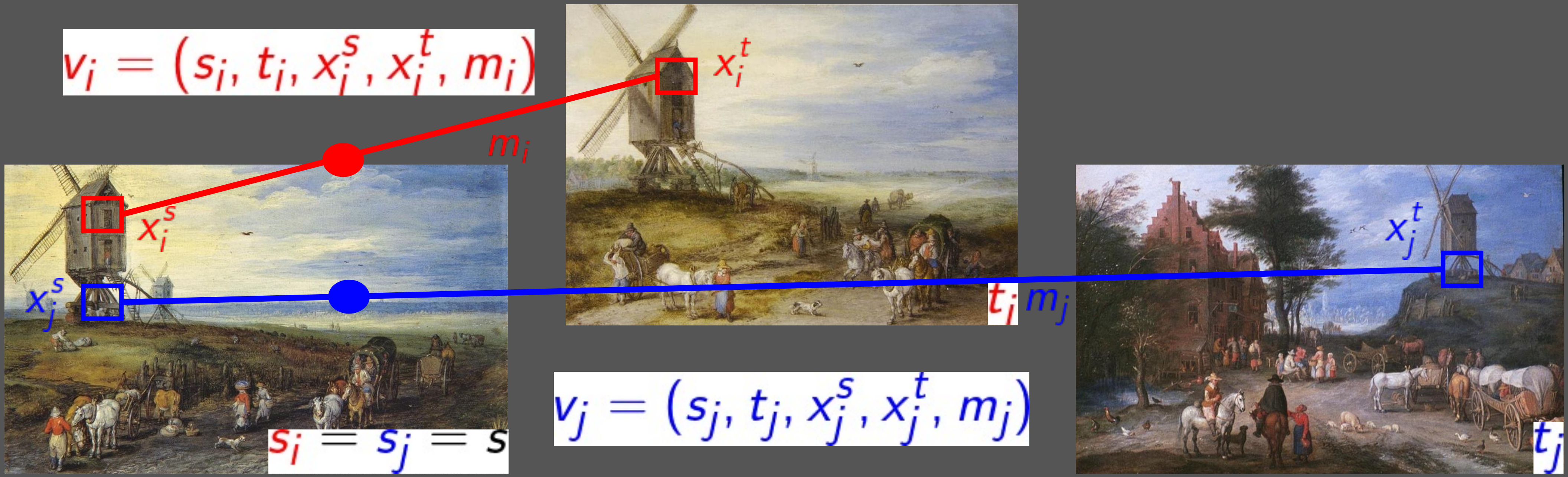
$v_i = (s_i, t_i, x_i^s, x_i^t, m_i)$

Indexes
of
source and targe images

Discovery on Brueghel: Correspondences graph

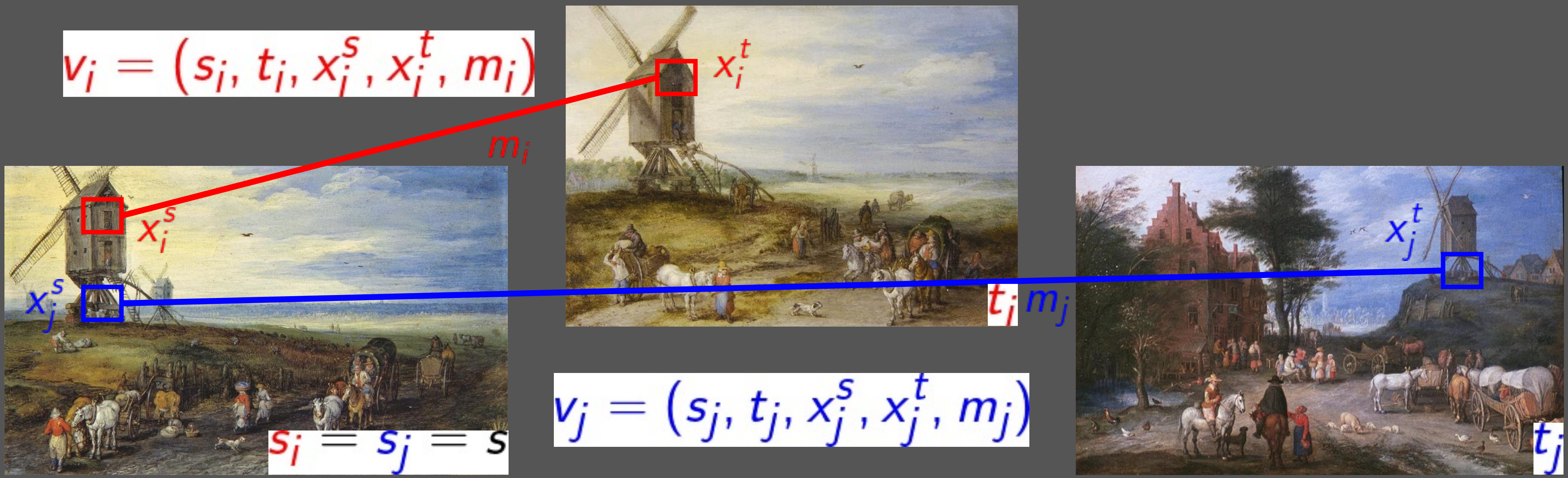
3-cycle Consistency $\mathcal{E} = \{e_{i,j}\}$

Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$



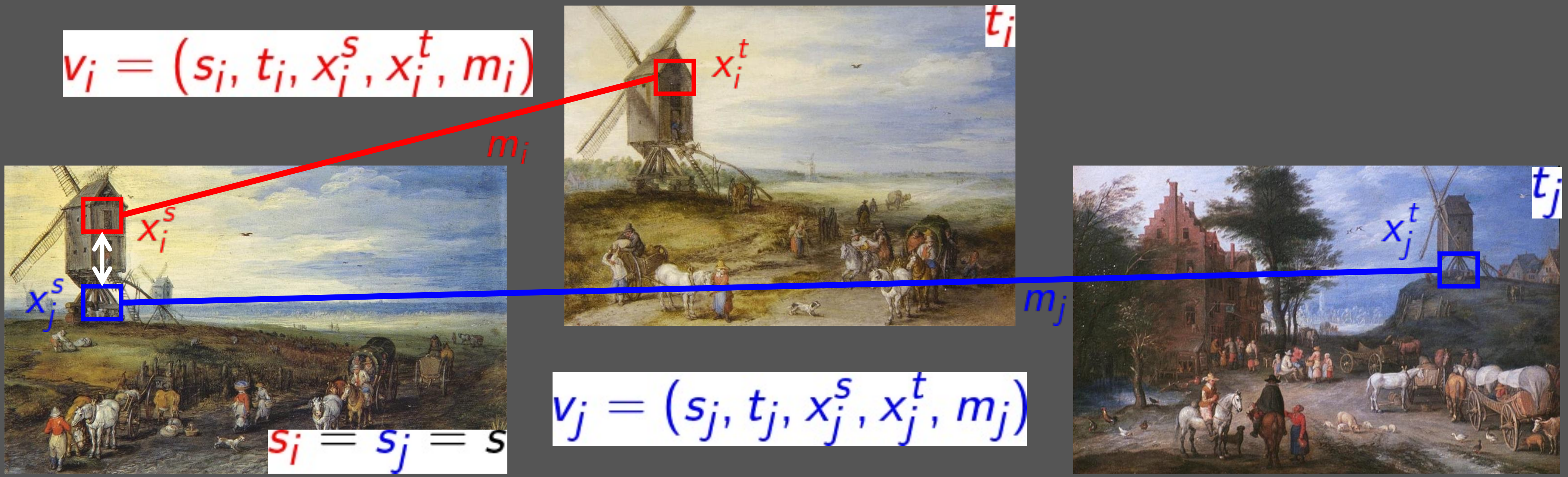
Discovery on Brueghel: Correspondences graph

$$\frac{1}{2}\boxed{m_i m_j}\exp\left(\frac{||x_i^s - x_j^s||}{\sigma}\right)\left[\exp\left(\frac{||x_i^t - C^{t_j \rightarrow t_i}(x_j^t)||}{\sigma}\right) + \exp\left(\frac{||x_j^t - C^{t_i \rightarrow t_j}(x_i^t)||}{\sigma}\right)\right]$$



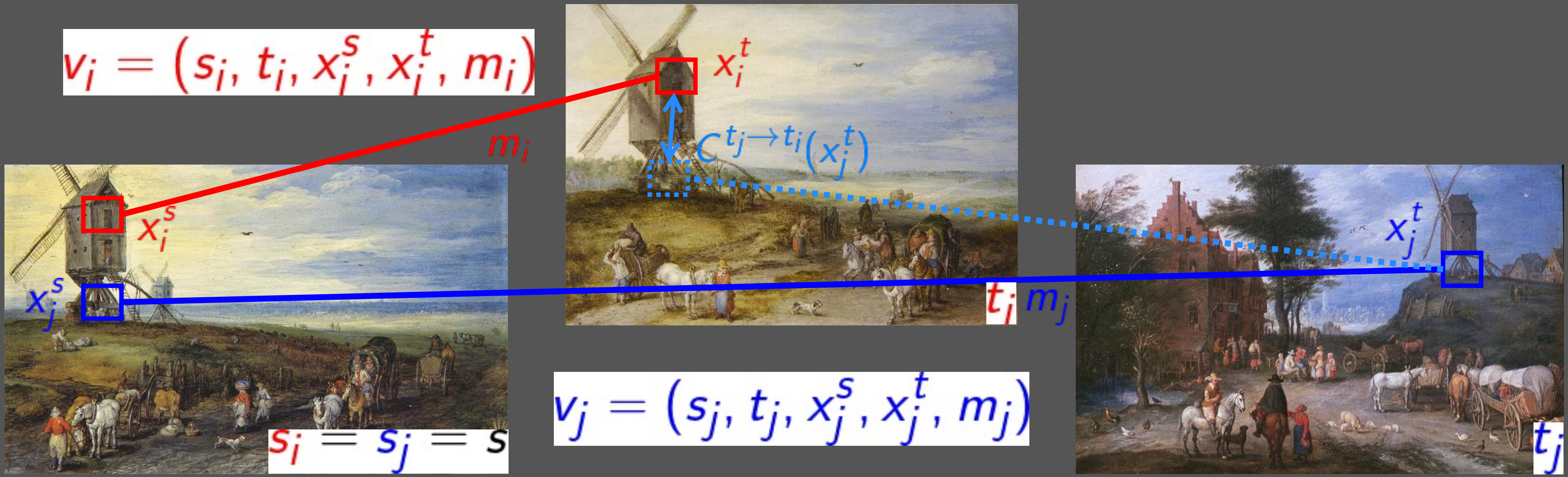
Discovery on Brueghel: Correspondences graph

$$\frac{1}{2} m_i m_j \boxed{\exp\left(\frac{||x_i^s - x_j^s||}{\sigma}\right)} \left[\exp\left(\frac{||x_i^t - C^{t_j \to t_i}(x_j^t)||}{\sigma}\right) + \exp\left(\frac{||x_j^t - C^{t_i \to t_j}(x_i^t)||}{\sigma}\right)\right]$$



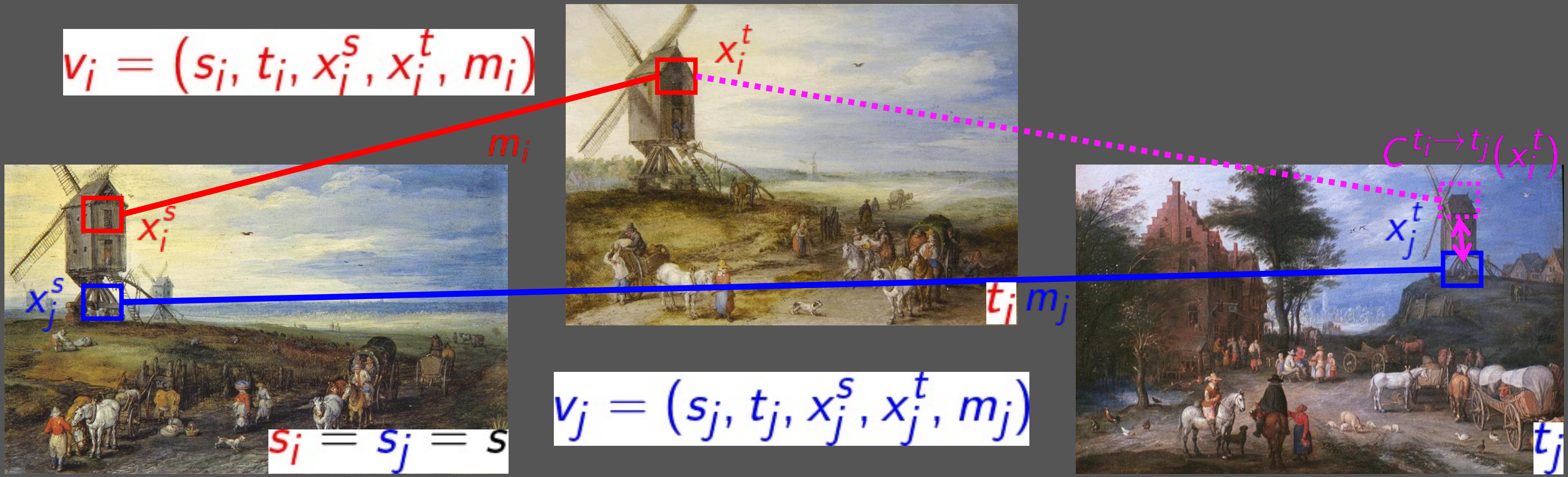
Discovery on Brueghel: Correspondences graph

$$\frac{1}{2} m_i m_j \exp\left(\frac{\|x_i^s - x_j^s\|}{\sigma}\right)\left[\boxed{\exp\left(\frac{\|x_i^t - C^{t_j \to t_i}(x_j^t)\|}{\sigma}\right)} + \exp\left(\frac{\|x_j^t - C^{t_i \to t_j}(x_i^t)\|}{\sigma}\right)\right]$$

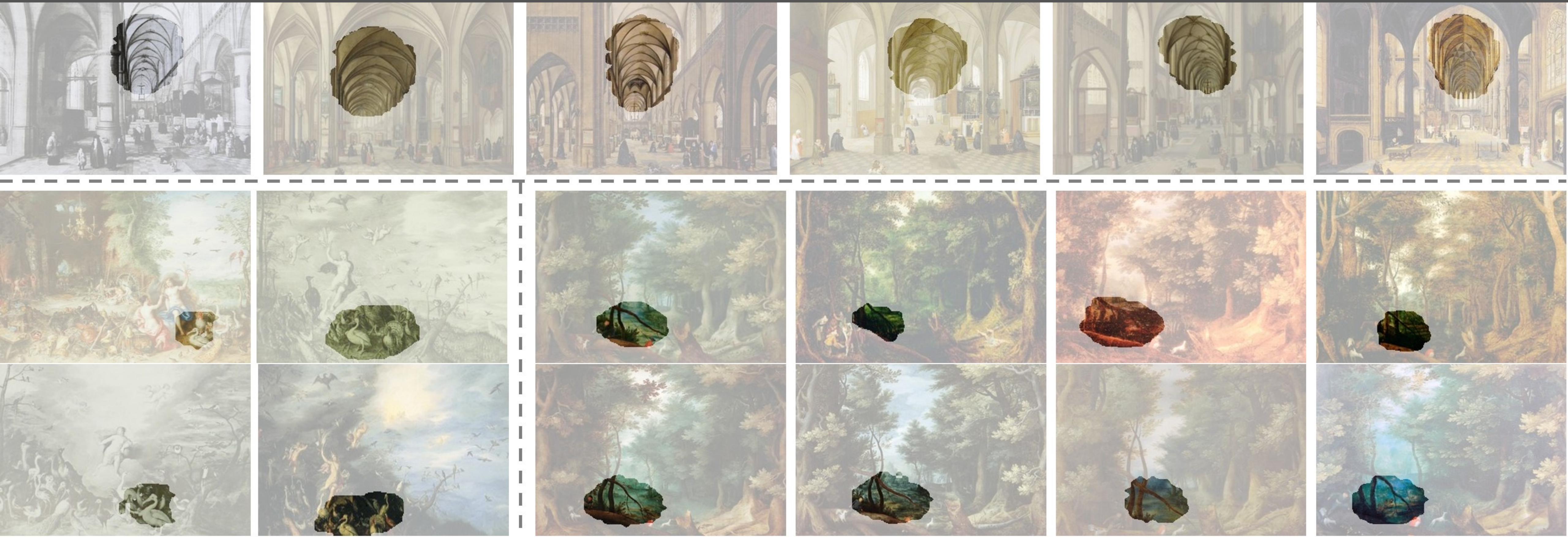


Discovery on Brueghel: Correspondences graph

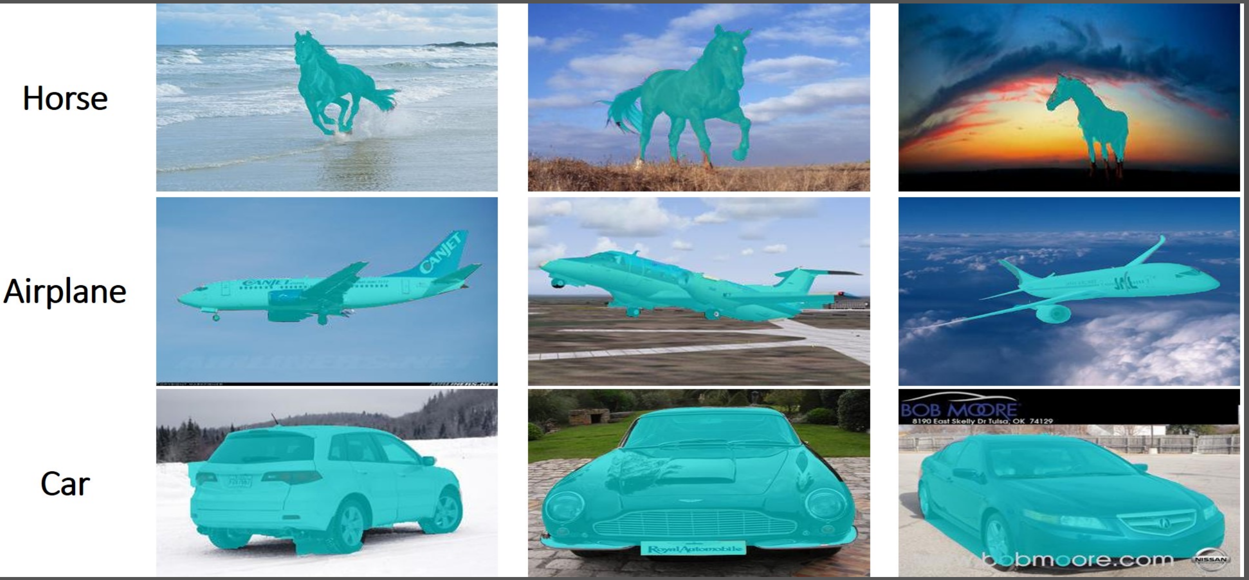
$$\frac{1}{2} m_i m_j \exp\left(\frac{\|x_i^s - x_j^s\|}{\sigma}\right)\left[\exp\left(\frac{\|x_i^t - C^{t_j \to t_i}(x_j^t)\|}{\sigma}\right) + \boxed{\exp\left(\frac{\|x_j^t - C^{t_i \to t_j}(x_i^t)\|}{\sigma}\right)}\right]$$



Experiments: discovery on Brueghel [Shen et al. 2019]



Object discovery on the dataset of [Rubinstein et al. 2013]



Method	Airplane $\mathcal{P}$	Airplane $\mathcal{J}$	Car $\mathcal{P}$	Car $\mathcal{J}$	Horse $\mathcal{P}$	Horse $\mathcal{J}$	Avg $\mathcal{P}$	Avg $\mathcal{J}$
DOCS [35]*	0.946	0.64	0.940	0.83	0.914	0.65	0.933	0.70
Sun et al. [57]	0.886	0.36	0.870	0.73	0.876	0.55	0.877	0.55
Rubinstein et al. [49]	0.880	0.56	0.854	0.64	0.828	0.52	0.827	0.43
Chen et al. [10]	0.902	0.40	0.876	0.65	0.893	0.58	0.890	0.54
Quan et al. [43]	0.910	0.56	0.885	0.67	0.893	0.58	0.896	0.60
Chang et al. [8]	0.726	0.27	0.759	0.36	0.797	0.36	0.761	0.33
Lee et al. [31]	0.528	0.36	0.647	0.42	0.701	0.39	0.625	0.39
Jerripothula et al. [25]	0.905	0.61	0.880	0.71	0.883	0.61	0.889	0.64
Hsu et al. [22]	0.936	0.66	0.914	0.79	0.876	0.59	0.909	0.68
Chen et al. [11]	**0.941**	0.65	**0.940**	**0.82**	**0.922**	**0.63**	**0.935**	**0.70**
Ours + Unsupervised segments								
transformer	0.925	0.65	0.914	0.79	0.909	0.60	0.916	0.68
Nc-Net	0.746	0.25	0.874	0.68	0.836	0.38	0.819	0.44
Ours + COCO segments [36]								
transformer	**0.941**	**0.67**	0.928	**0.82**	0.916	0.60	0.928	**0.70**
Nc-Net	0.655	0.23	0.857	0.61	0.873	0.43	0.795	0.42

* learned with strong supervision (i.e., manually annotated object masks)

Table 5. Co-segmentation on the dataset of [49]. We report pixel level precision $\mathcal{P}$ and Jaccard index $\mathcal{J}$

Summary

- Learning co-segmentation from synthetic pairs
- Discovering patterns using the correspondence graph

Learning Co-segmentation by Segment Swapping for Retrieval and Discovery

arXiv

Xi Shen[1]   Alexei A. Efros[2]   Armand Joulin[3]   Mathieu Aubry[1]

[1]LIGM (UMR 8049) - Ecole des Ponts, UPE   [2]UC Berkeley   [3]Facebook AI Research

Code   arXiv

Code, data and more experimental results can be found in our project page:
http://imagine.enpc.fr/~shenx/SegSwap/