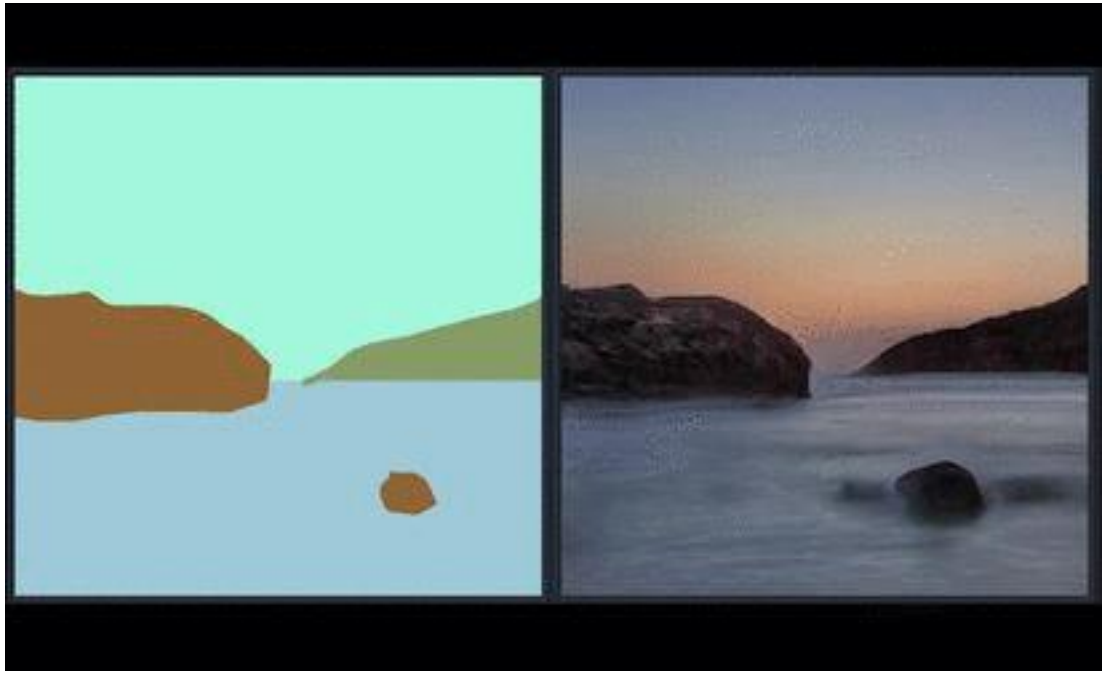
# CoCosNet v2: Full-Resolution Correspondence Learning for Image Translation

**Xingran Zhou[1], Bo Zhang[2], Ting Zhang[2], Pan Zhang[4], Jianmin Bao[2], Dong Chen[2], Zhongfei Zhang[3], Fang Wen[2]**

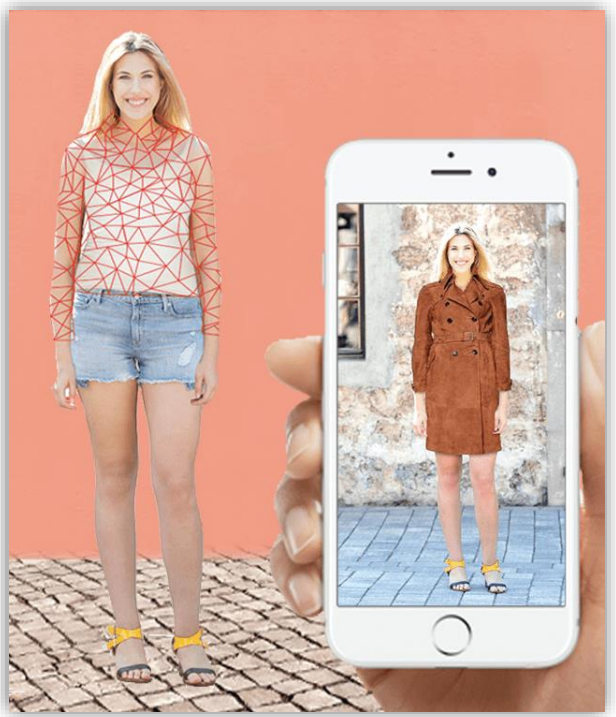
1 Zhejiang University 2 Microsoft Research Asia 3 Binghamton University 4 USTC

# *Image-to-image translation*

old photos restoration



semantic editing

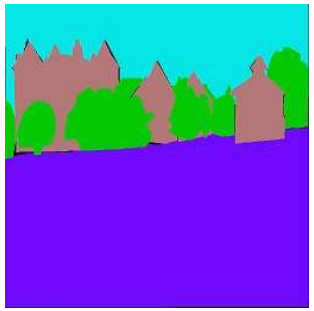


virtual try-on

Semantic Image Synthesis with Spatially-Adaptive Normalization. Taesung Park, Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu. CVPR 2019.
Bringing Old Photos Back to Life. Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, Fang Wen. CVPR 2020.

# *Exemplar-based translation*

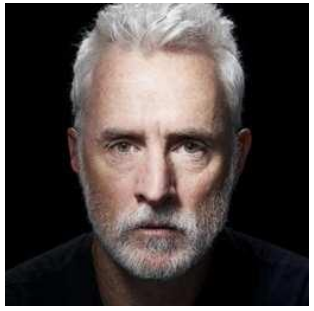
exemplar



segmentation

synthesis

- **Pros**
  - **Flexible user control**
  - **Improved generation quality**



exemplar

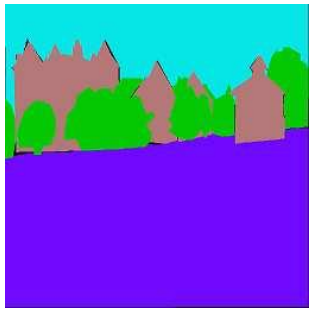
edge

synthesis

Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

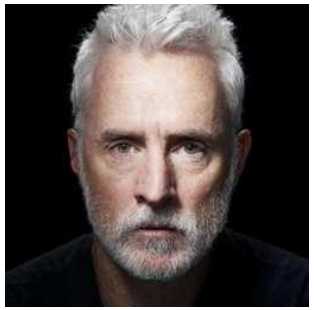
# *Exemplar-based translation*

exemplar



segmentation

synthesis



exemplar

edge

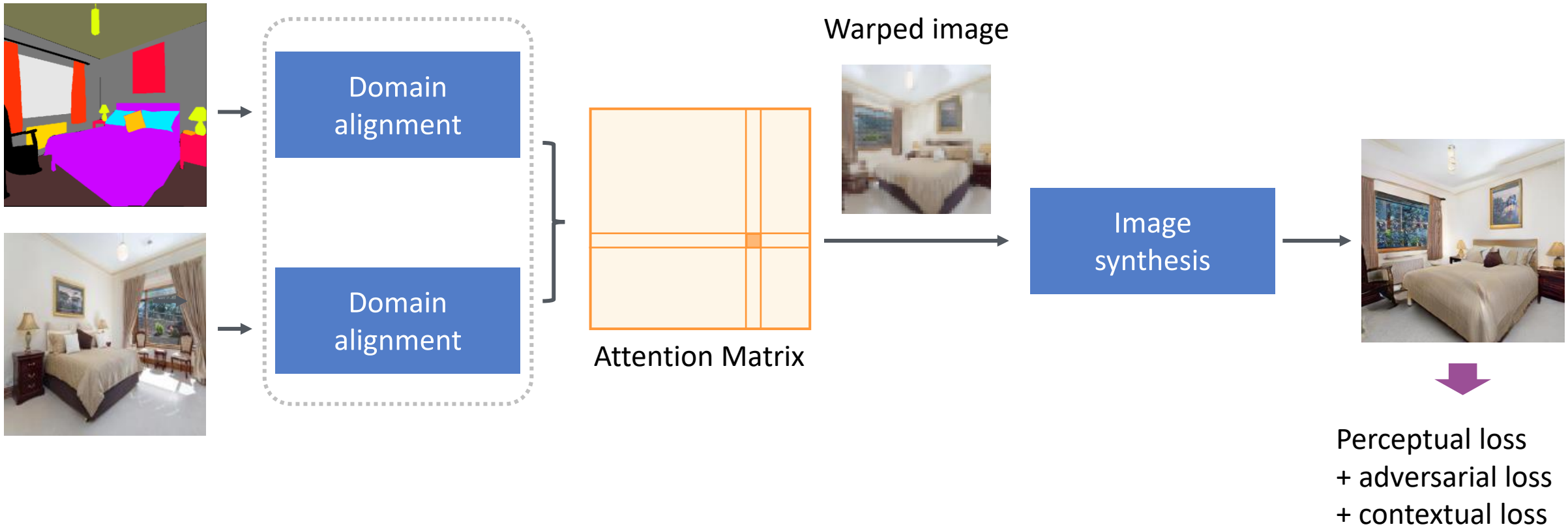
synthesis

- **Pros**
  - Flexible user control
  - Improved generation quality
- **Cons**
  - Significant artifacts for complex scenes
  - Lack of fine-grained style controllability
  - Lack of using fine textures from exemplar

Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

# *CoCosNet*

## Cross-domain correspondence learning with image synthesis

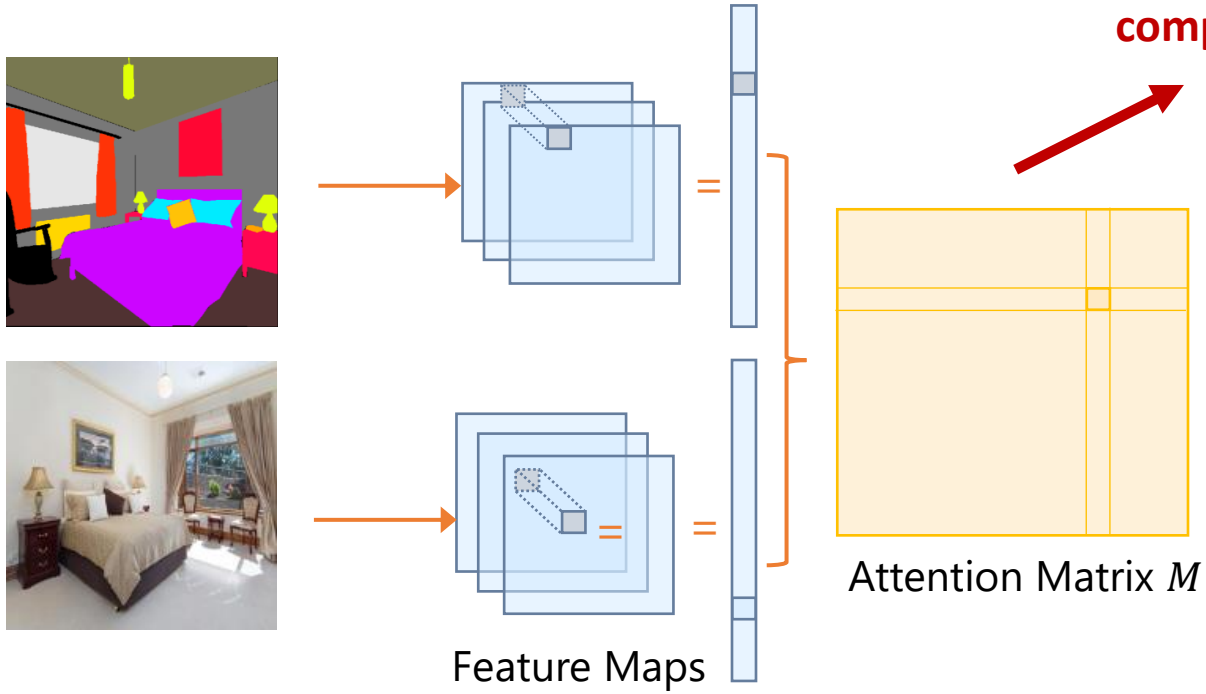



Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

# *CoCosNet*

## Full resolution attention is computationally inhibitive

**For feature contains $h \times w$ pixels,**
**computation complexity is $\mathcal{O}((h \times w)^2)$.**

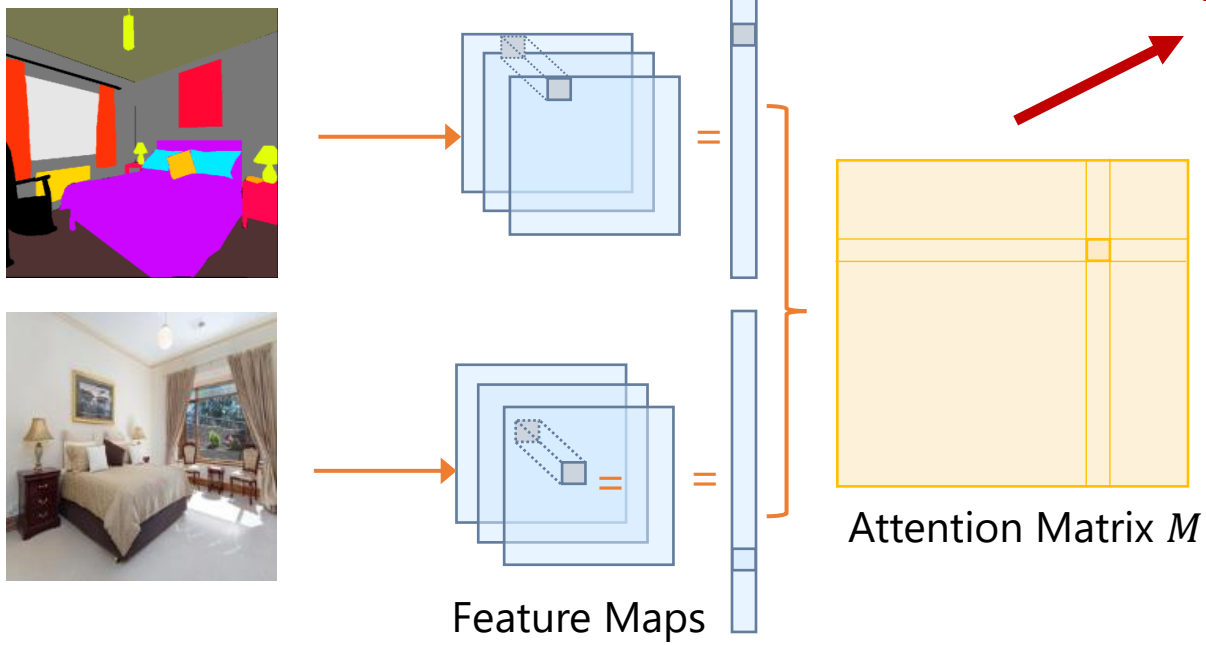





Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

# CoCosNet

## Full resolution attention is computationally inhibitive




**For feature contains $h \times w$ pixels,**
**computation complexity is $\mathcal{O}((h \times w)^2)$.**

- **Quadratic memory footprint**
  - **256x256 resolution, $\mathcal{O}(M) = 4\mathrm{G}$**
  - **512x512 resolution, $\mathcal{O}(M) = 68\mathrm{G}$**
  - **1024x1024 resolution, $\mathcal{O}(M) = 1\mathrm{T}$**

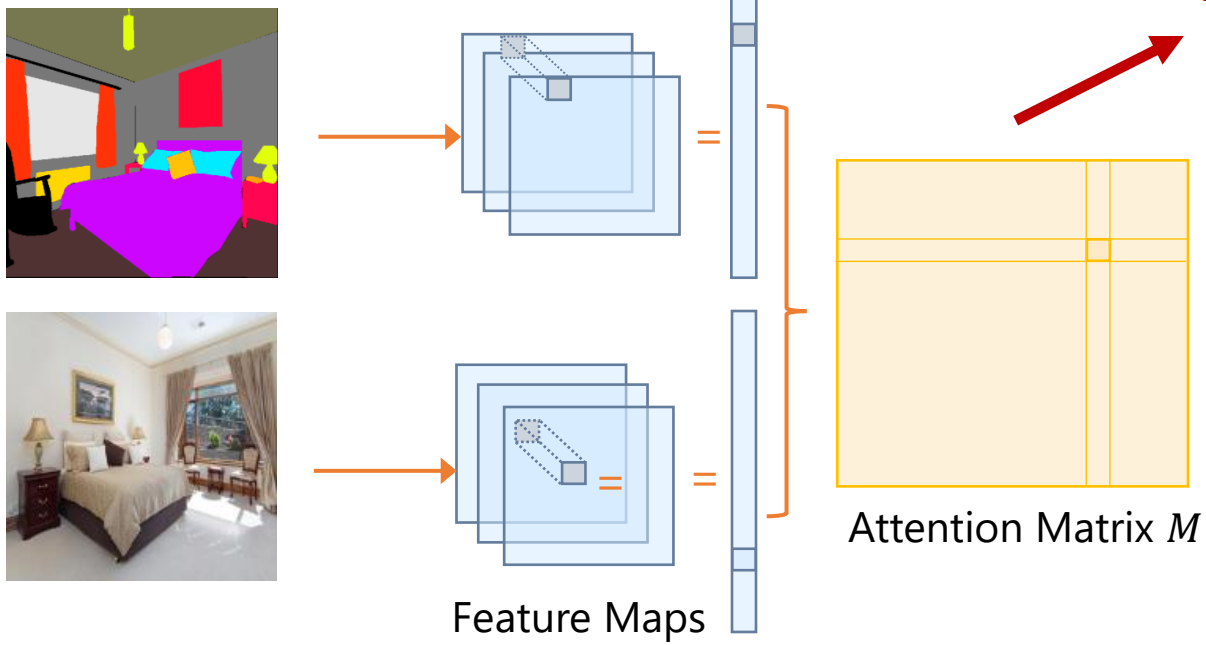
- **CoCosNet only establishes coarse correspondence (64x64)**
  - **fail to transfer detailed textures from the exemplar**



Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

# CoCosNet

## Full resolution attention is computationally inhibitive




**For feature contains $h \times w$ pixels,**
**computation complexity is $\mathcal{O}((h \times w)^2)$.**

- **Quadratic memory footprint**
  - **256x256 resolution, $\mathcal{O}(M) = 4\mathrm{G}$**
  - **512x512 resolution, $\mathcal{O}(M) = 68\mathrm{G}$**
  - **1024x1024 resolution, $\mathcal{O}(M) = 1\mathrm{T}$**
- **CoCosNet only establishes coarse correspondence (64x64)**
  - **fail to transfer detailed textures from the exemplar**

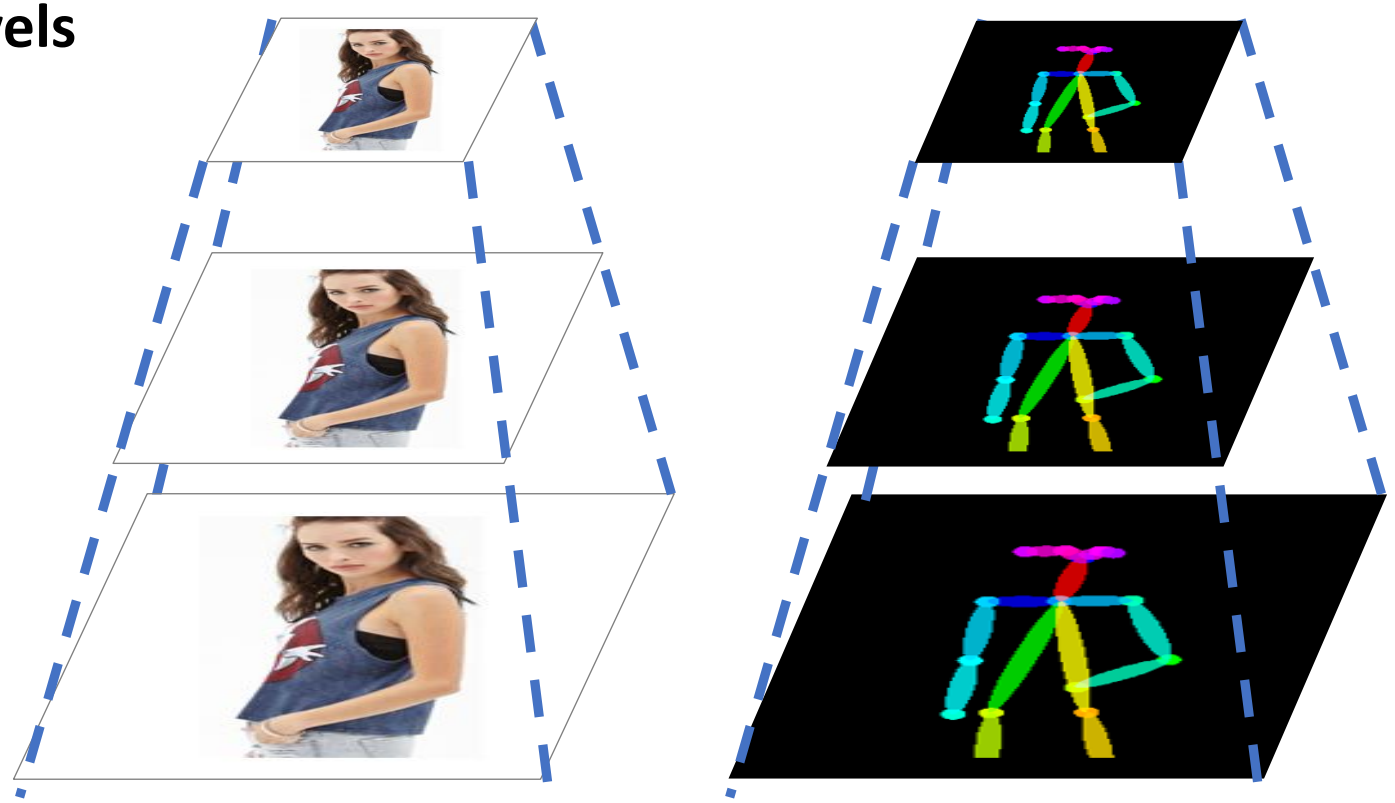
**How can we compute the correspondence on high resolution?**



Cross-domain Correspondence Learning for Exemplar-based Image Translation. Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, Fang Wen. CVPR 2020.

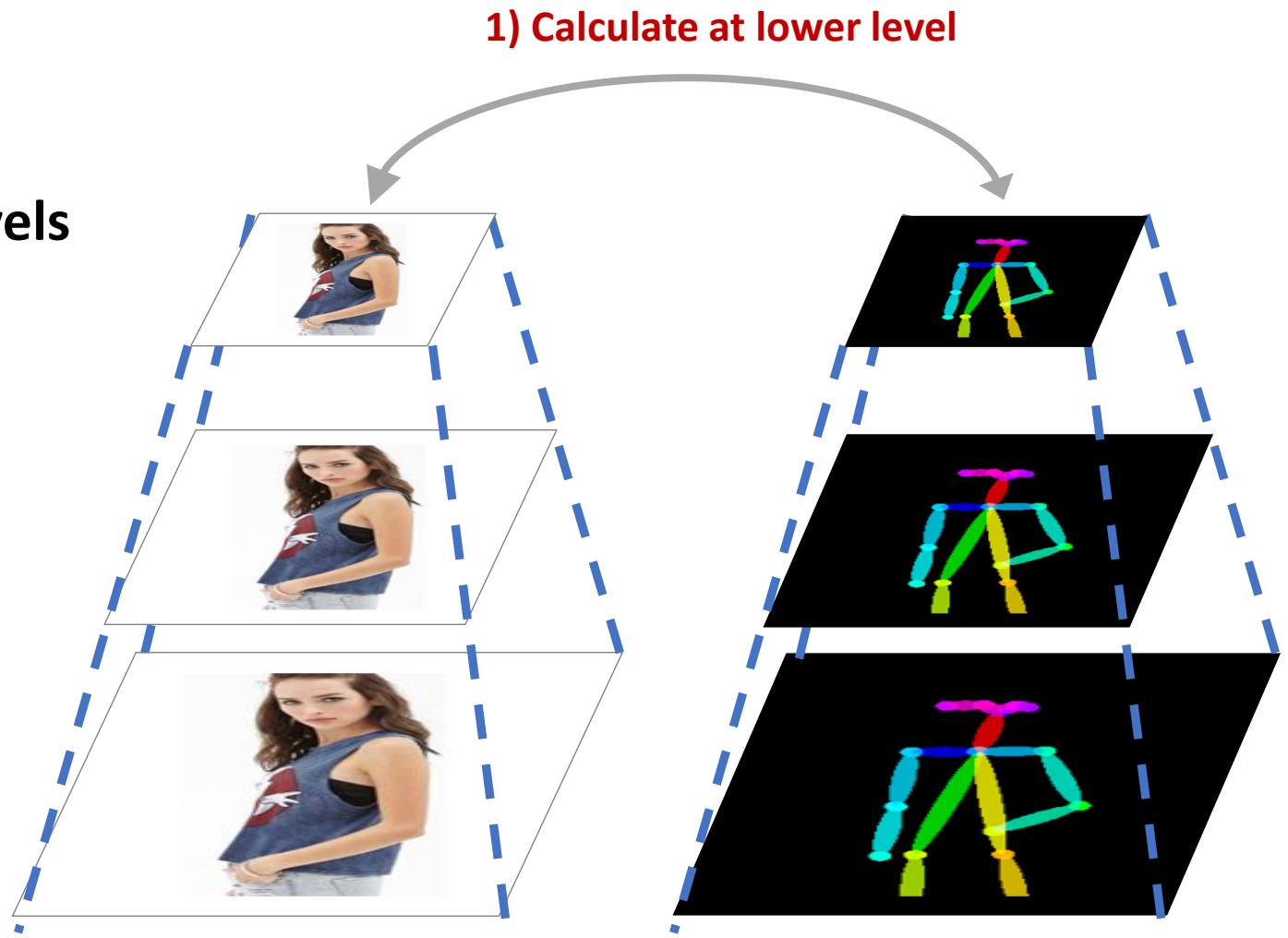
# *CoCosNet v2*

- **Coarse-to-fine strategy**
  - **Coarse level guides the finer levels**



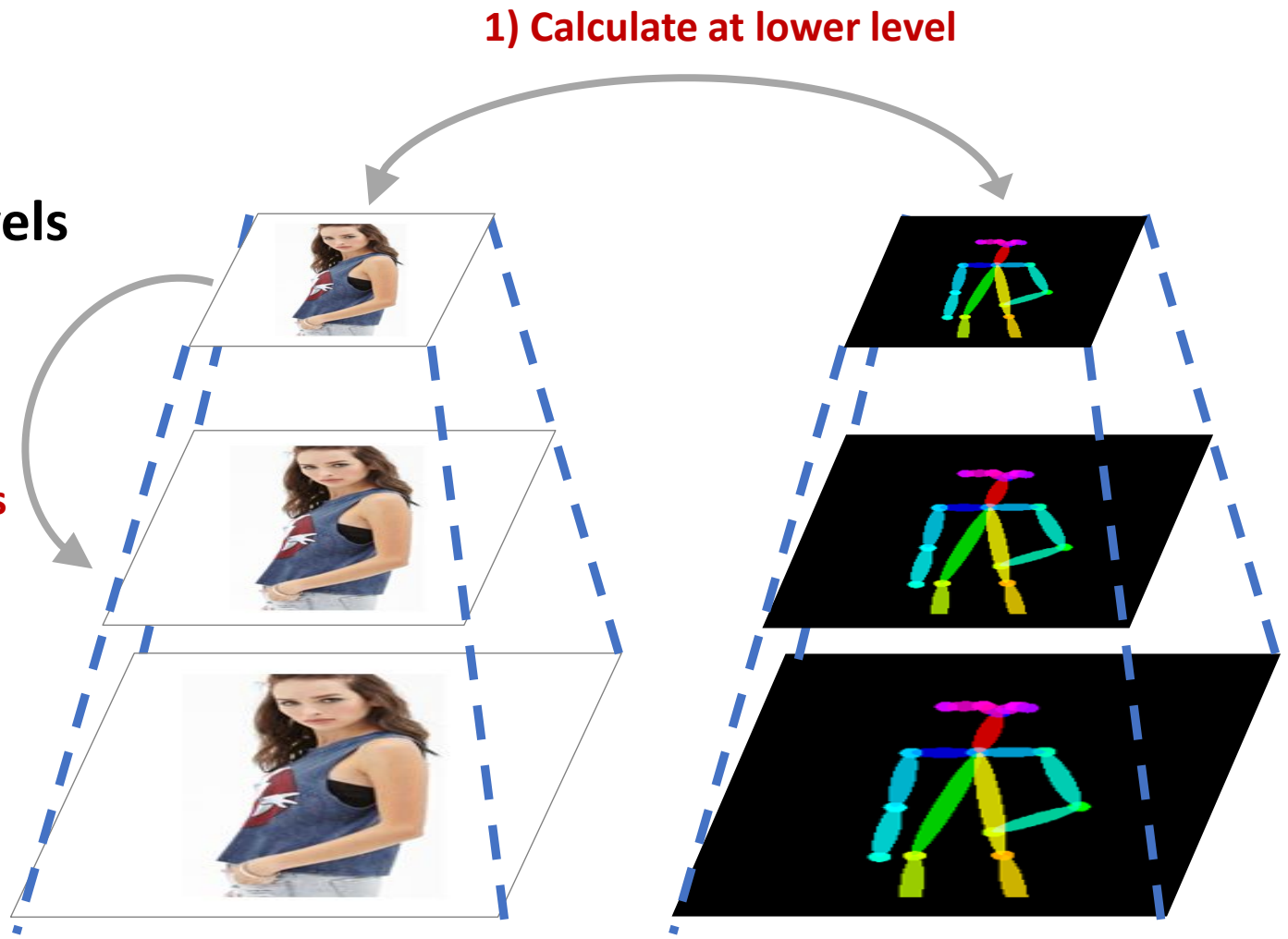
# *CoCosNet v2*

- **Coarse-to-fine strategy**
  - **Coarse level guides the finer levels**



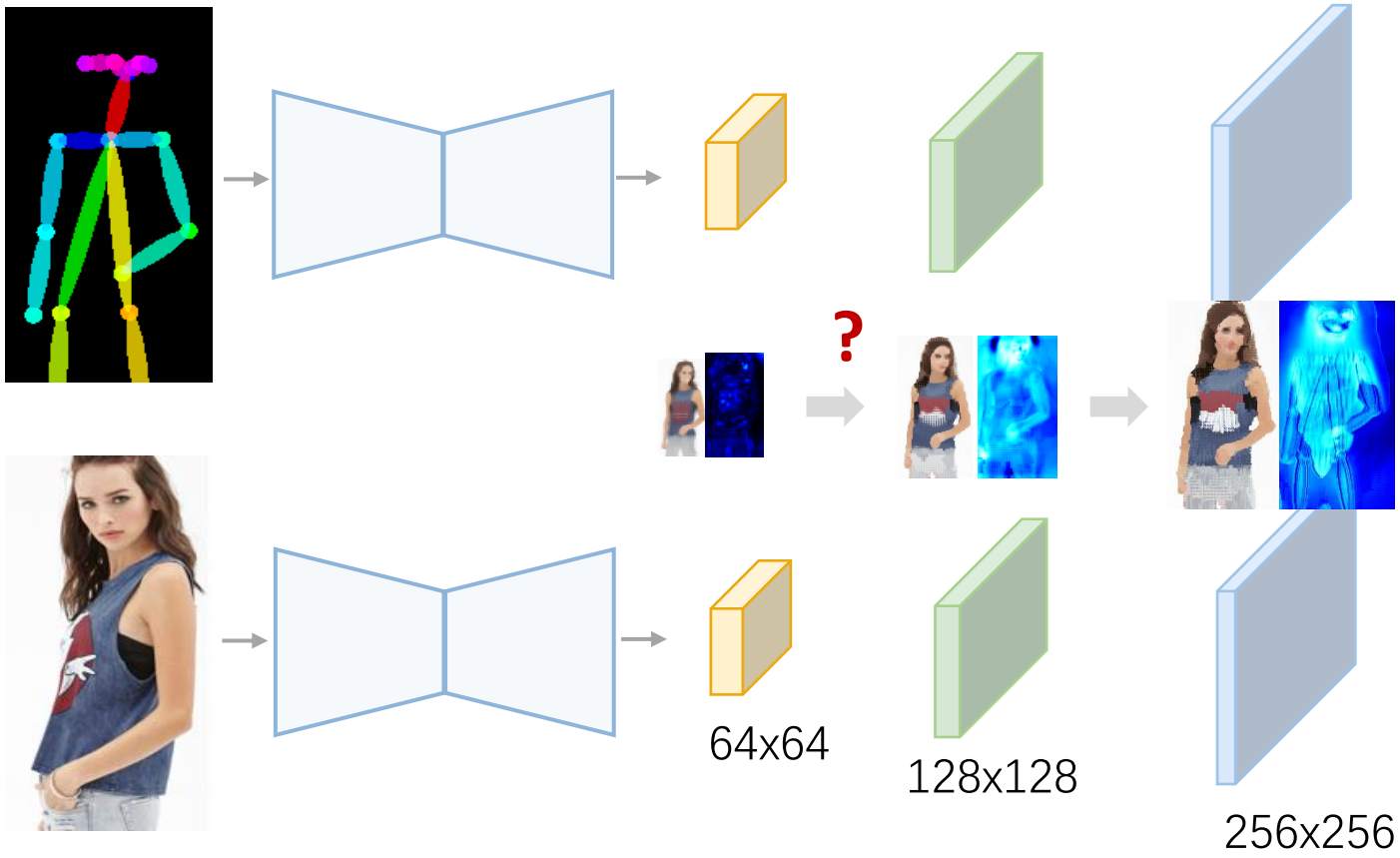
# *CoCosNet v2*

- **Coarse-to-fine strategy**
  - **Coarse level guides the finer levels**



# *CoCosNet v2*

- **Coarse-to-fine strategy**
  - **Coarse level guides the finer levels**

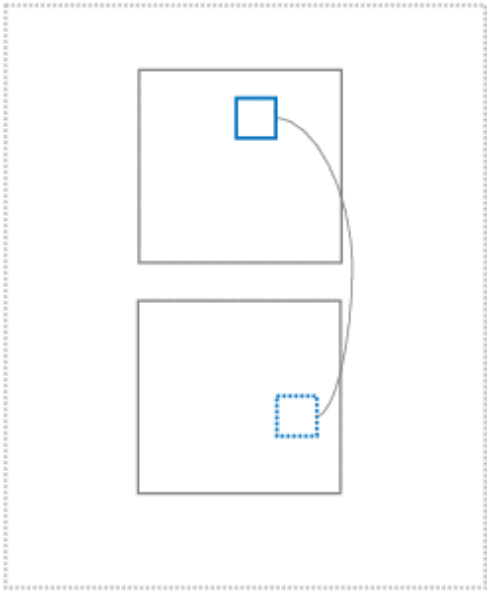


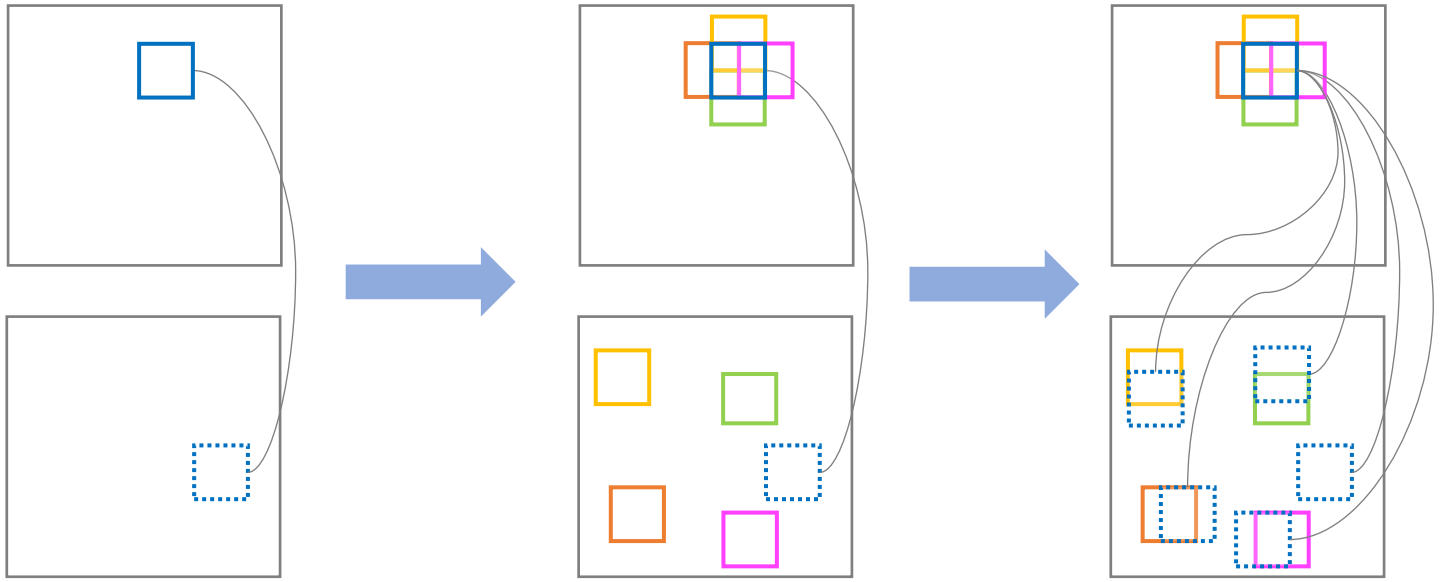

**How can we make use of the initialization from the lower level?**

# *PatchMatch*

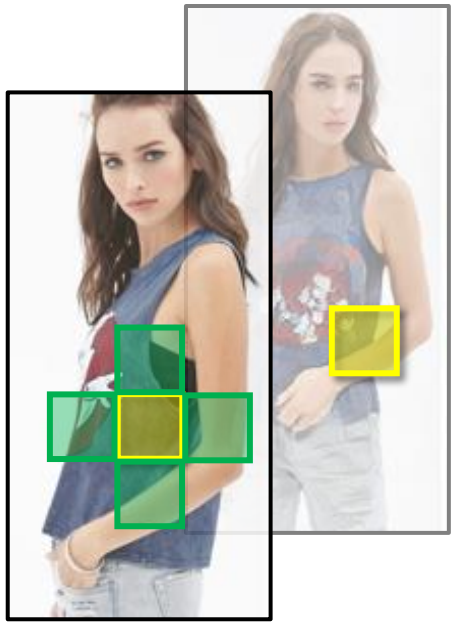
**Global ➡ Local : PatchMatch**

PatchMatch searches from the **neighborhood** rather than searching **globally**.





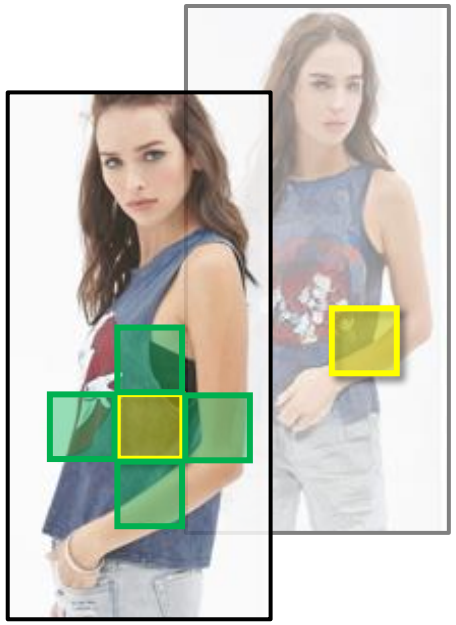
# *Differentiable PatchMatch*



- Our supervision comes from the image warping, $i.e.$,

$$y_B(H(p)) \quad \text{where,} \quad H(p) = \arg\min_q ||f_x(p) - f_y(q)||$$

# *Differentiable PatchMatch*



- Our supervision comes from the image warping, $i.e.$,

$$y_B(H(p)) \quad \text{where,} \quad H(p) = \arg\min_q ||f_x(p) - f_y(q)||$$

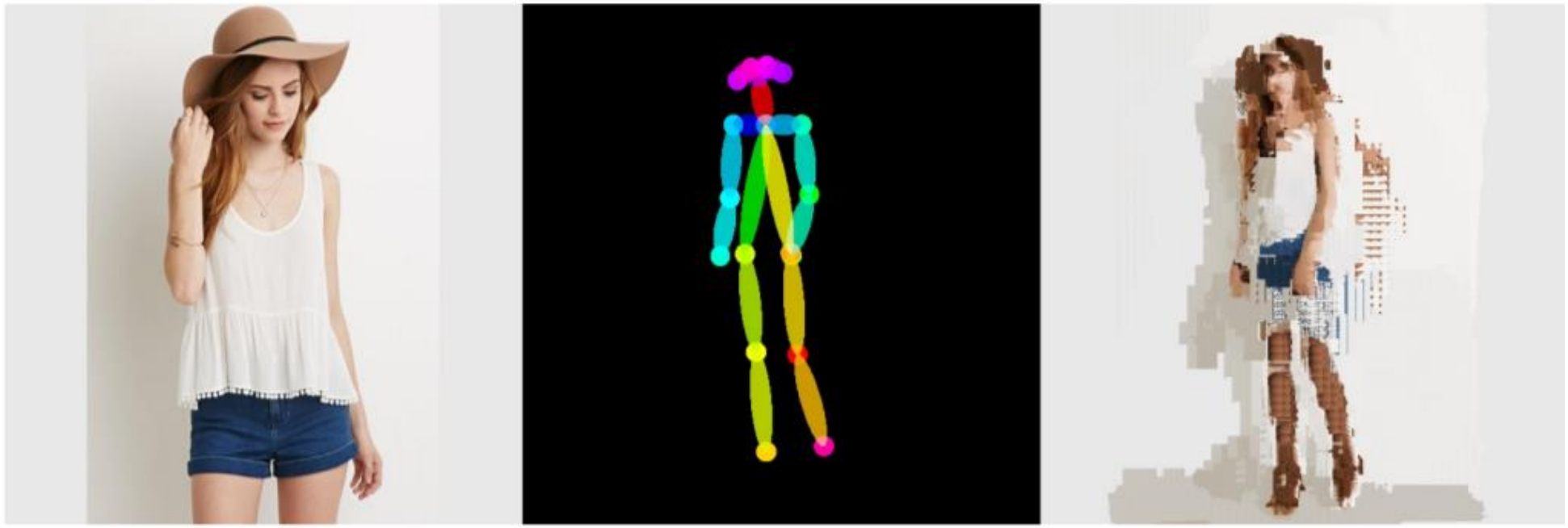
- To make it differentiable, we consider the K possible matchings. Now the warping becomes:

$$w^{y \to x}(p) = \sum_{k=1}^{K} softmax\{S(p;k)\ y_B(H(p;k)\}$$

**confidence** — **$\mathrm{k}^{th}$ matching**

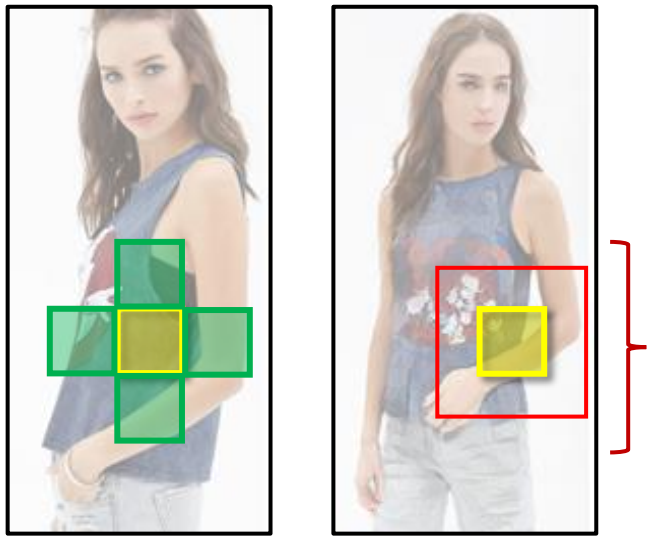
# *Differentiable PatchMatch*

- PatchMatch implicitly assume local smoothness, which is true to natural images
- However, this is violated because deep features are not well-trained at the beginning



warped image
(only PatchMatch propagation)

# *ConvGRU-assisted PatchMatch*



- PatchMatch only considers the adjacent patches
- Conv makes the propagation consider the distant patches
- The gradient can be propagated to more locations

**Receptive field of ConvGRU-assist propagation**

# *CoCosNet v2*

**Warped images via different variants of our method.**

***CoCosNet v2* produces the most faithful warped image.**

**only PatchMatch propagation**

**only ConvGRU**

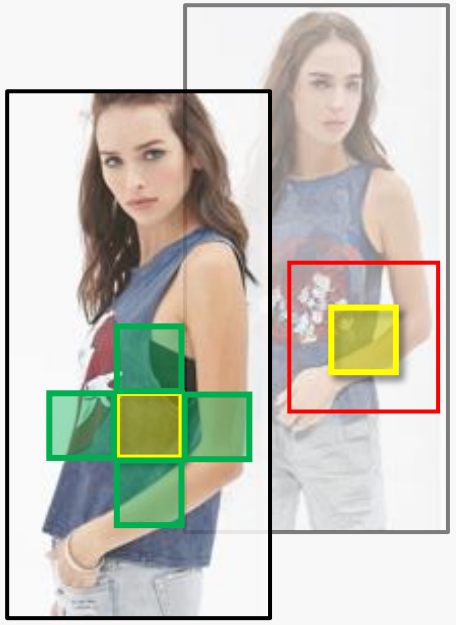
**PatchMatch propagation with Conv**

**CoCosNet v2**

# *CoCosNet v2*

## ConvGRU-assisted Patch Match

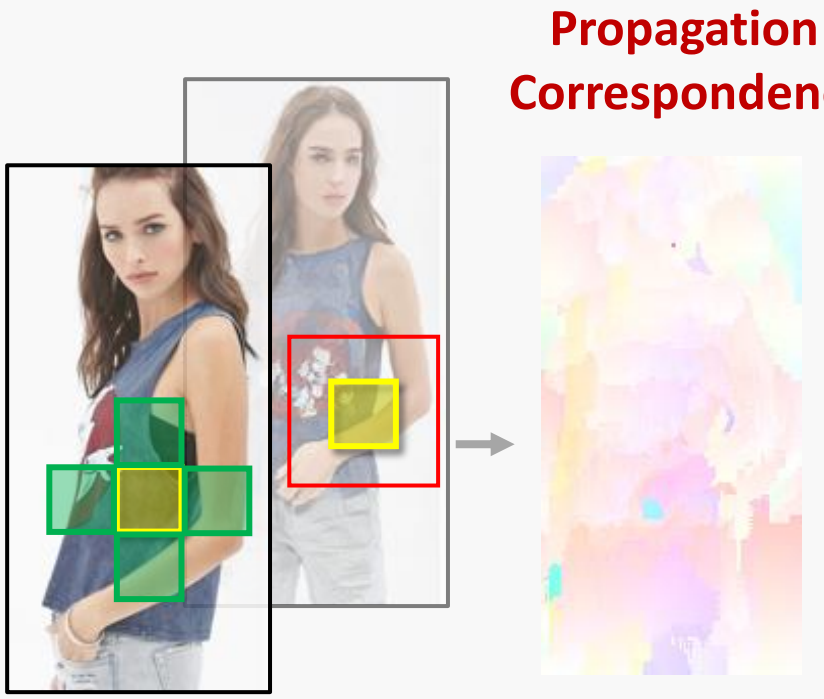
**Propagation from neighborhood**



ConvGRU-assisted PatchMatch interation at level $L$

# *CoCosNet v2*

## ConvGRU-assisted Patch Match



Propagation

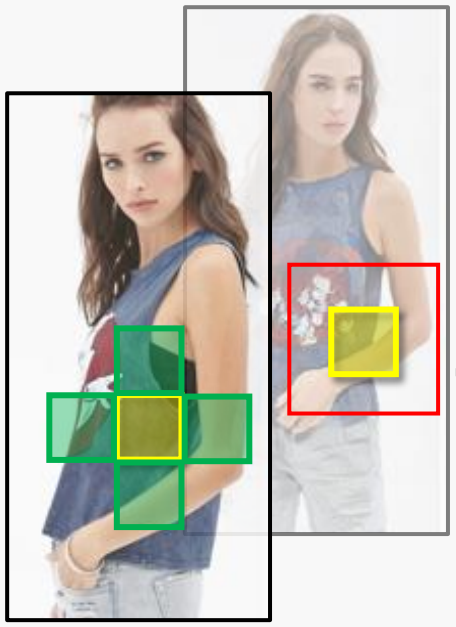
**Propagation**
**Correspondence**



ConvGRU-assisted PatchMatch interation at level $L$

# *CoCosNet v2*

## ConvGRU-assisted Patch Match



Propagation

Large kernel Convs

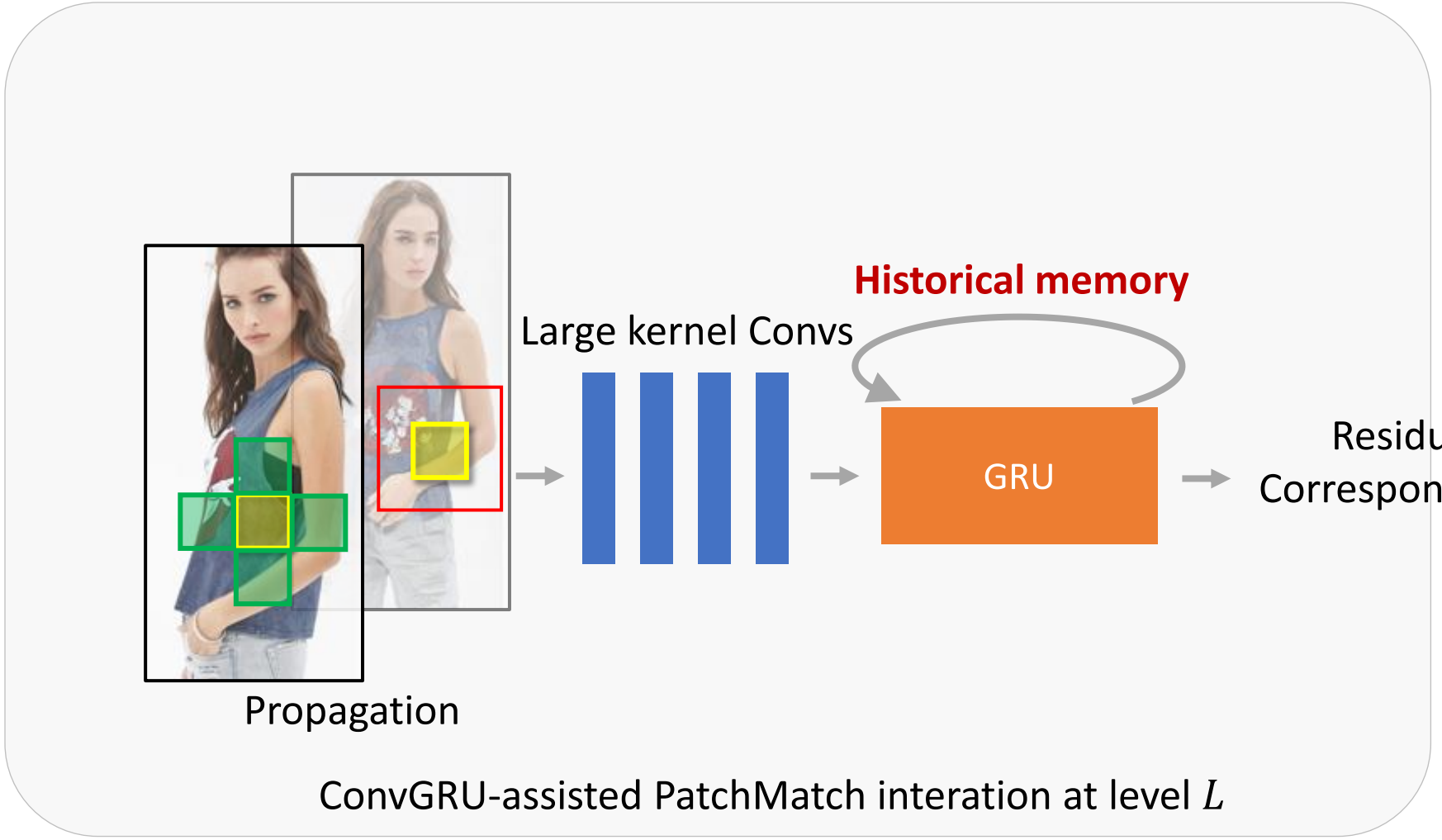
GRU

**Residual Correspondence**

ConvGRU-assisted PatchMatch interation at level $L$

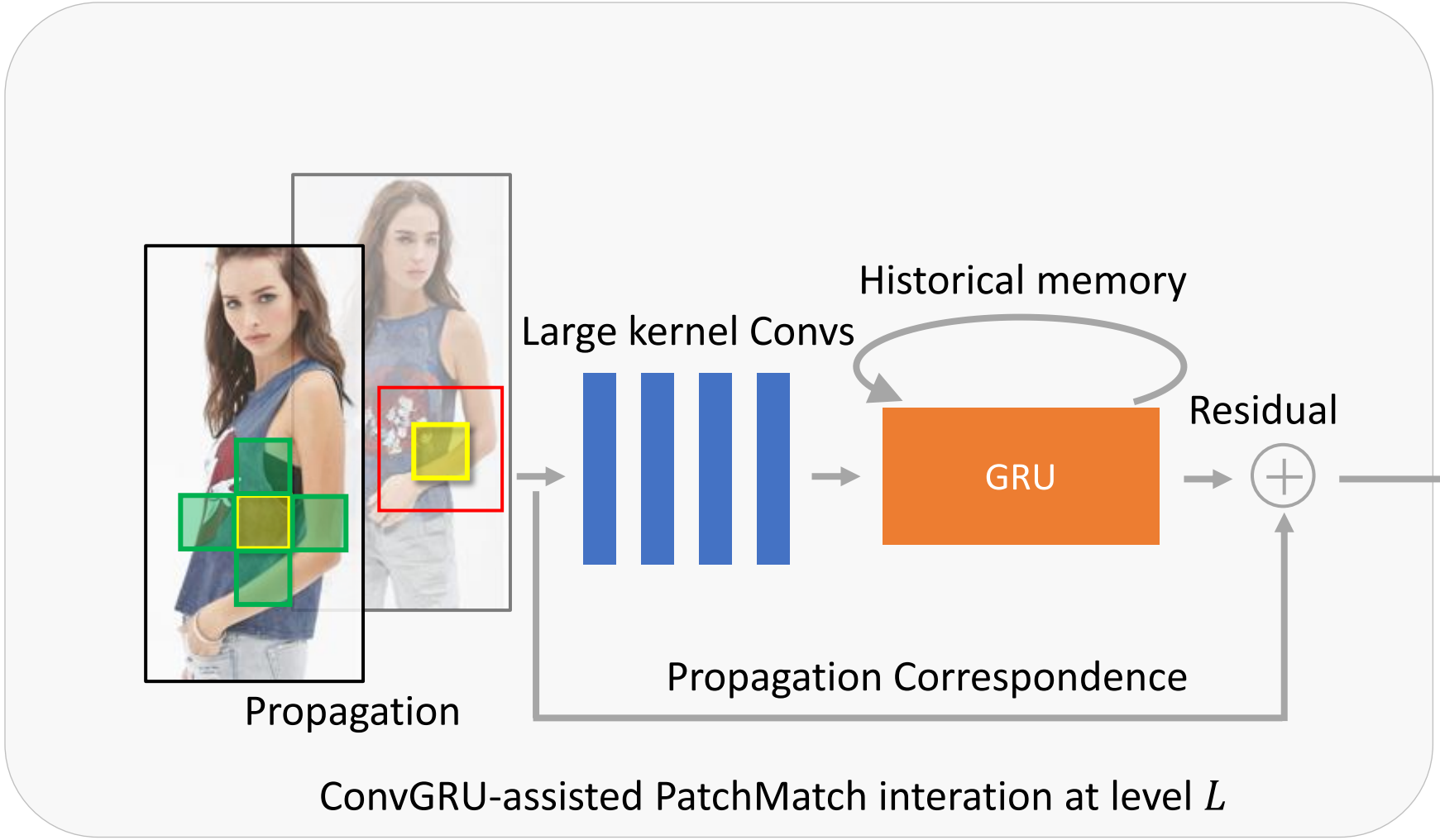
# *CoCosNet v2*

## ConvGRU-assisted Patch Match



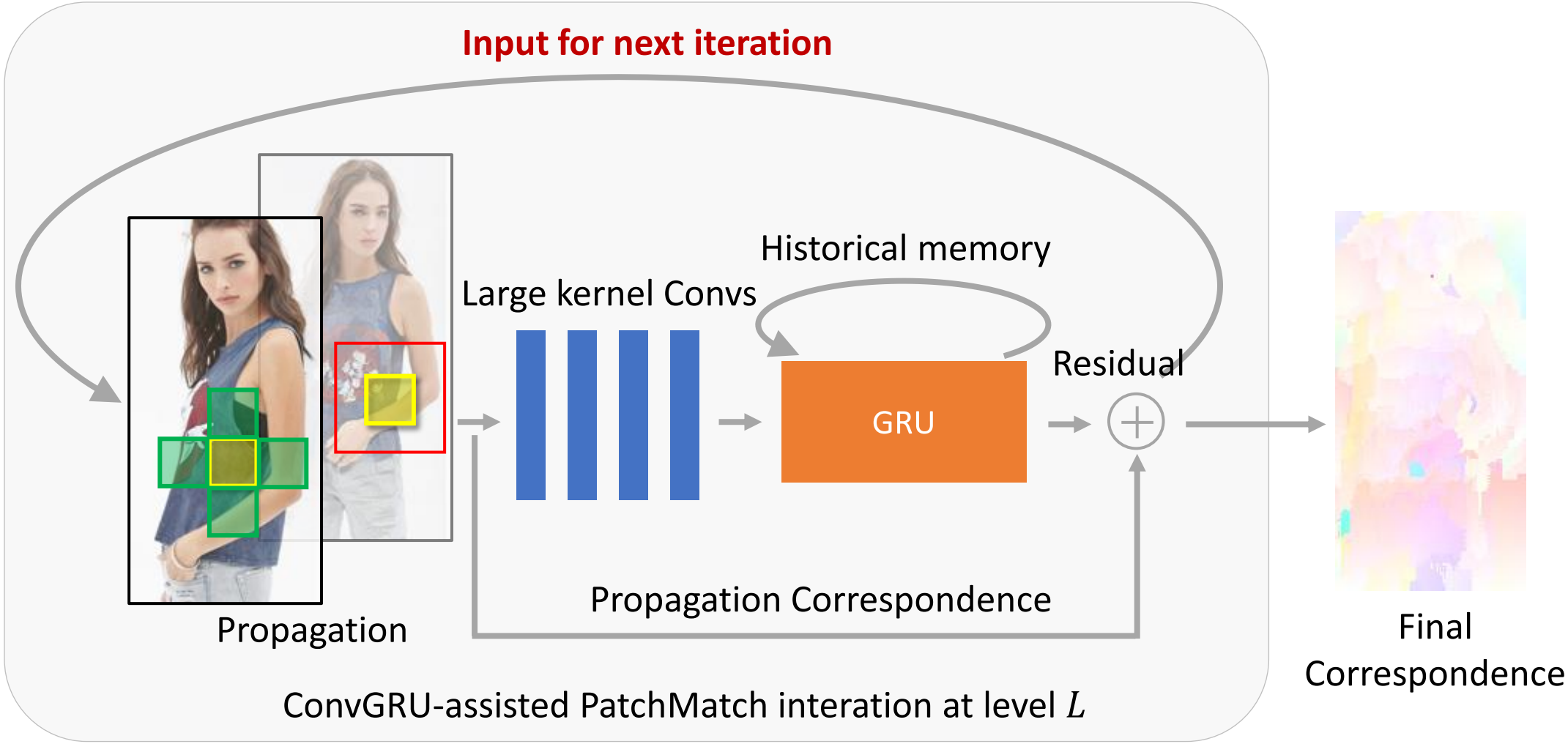
# *CoCosNet v2*

## ConvGRU-assisted Patch Match



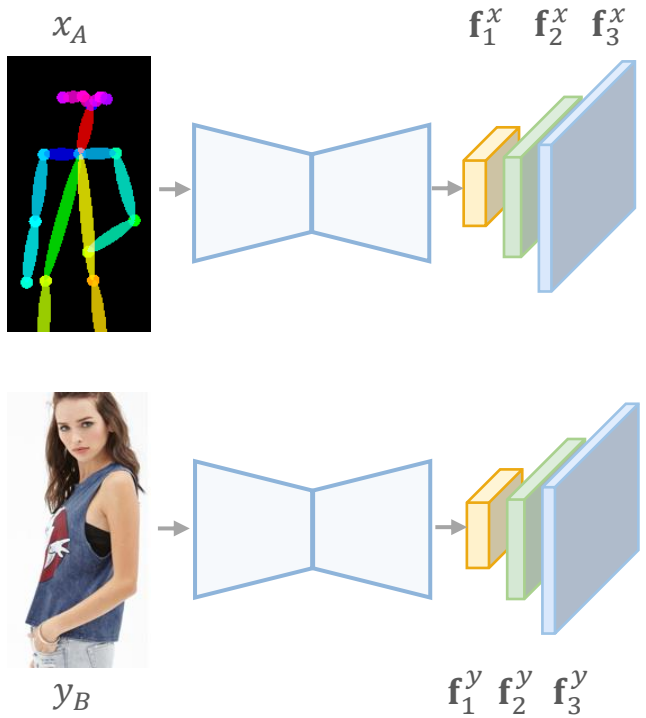
# *CoCosNet v2*

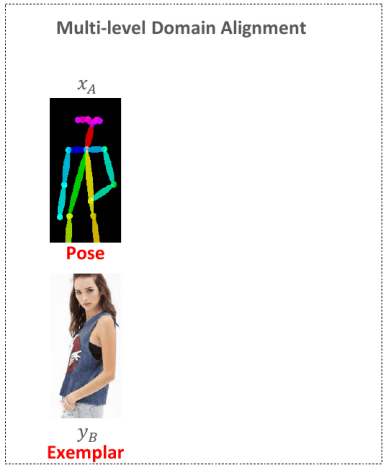
## ConvGRU-assisted Patch Match



# CoCosNet v2

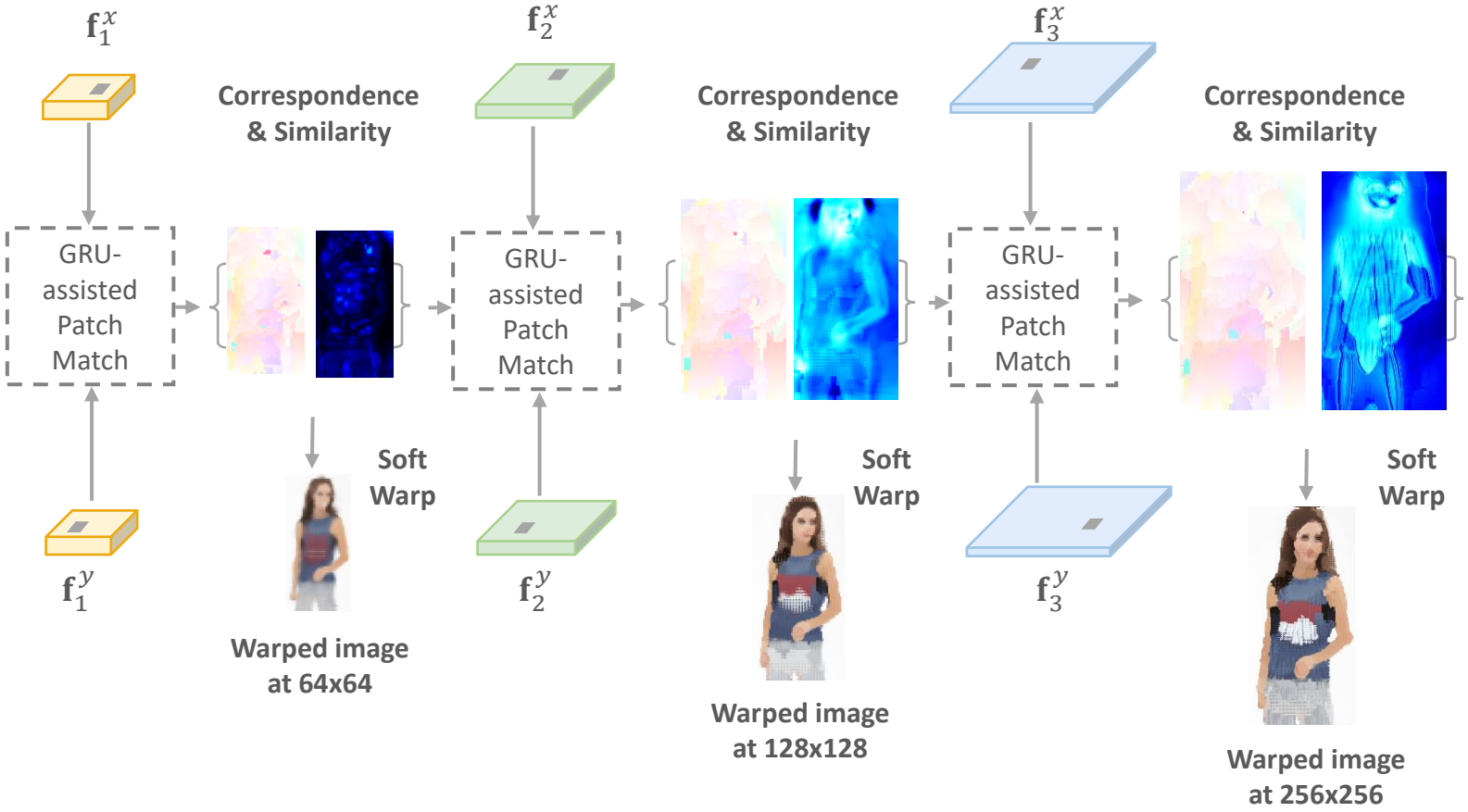
## Multi-level Domain Alignment
## Feature Extraction



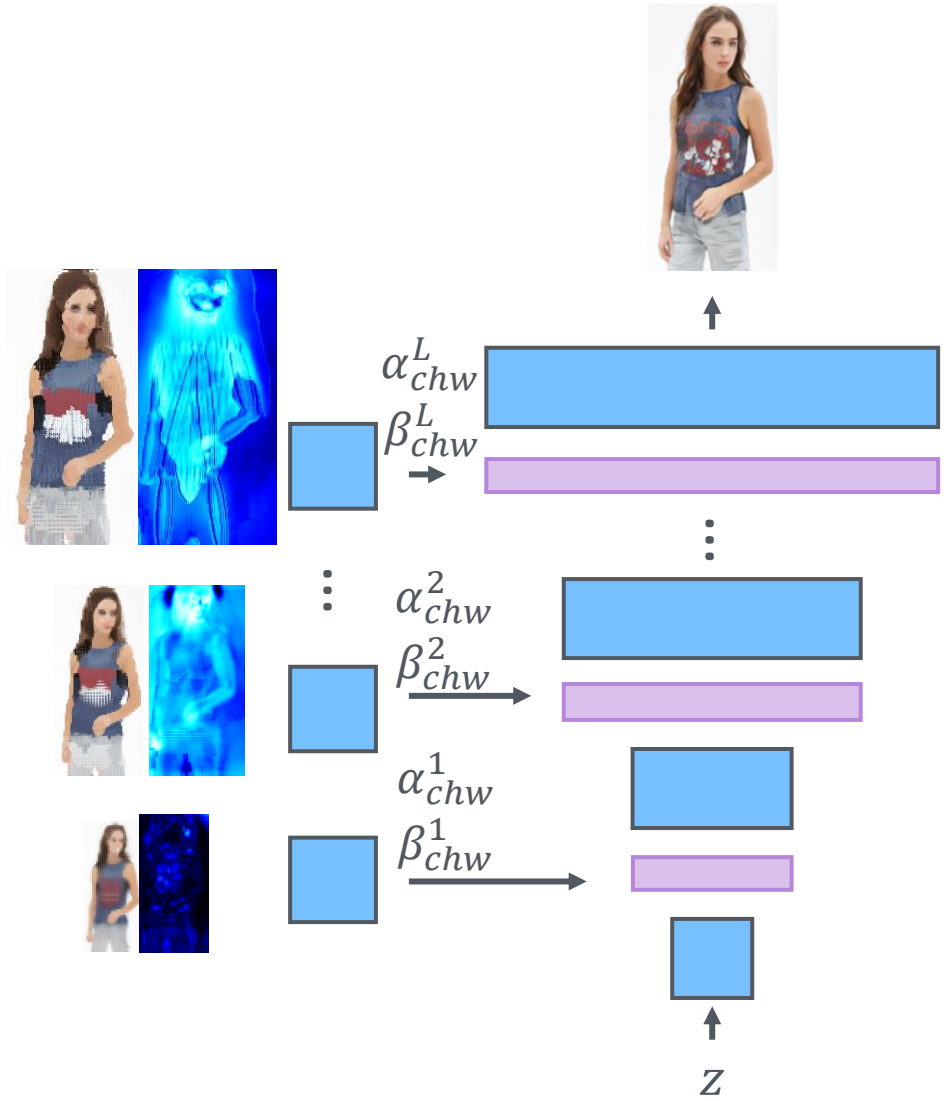
# CoCosNet v2

## Hierarchical ConvGRU-assisted Patch Match

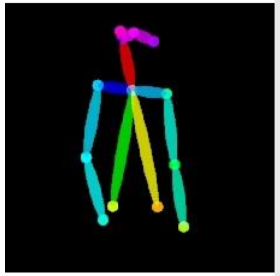



# CoCosNet v2

## Translation Network



# *Pose-to-body*



**Pose**

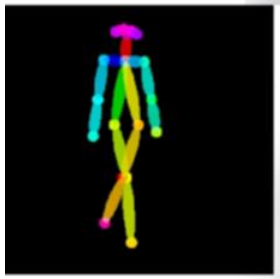
**Exemplar**

**Synthesis**

**Deepfashion dataset (512x512 resolution)**

# *Pose-to-body*

**Exemplar**

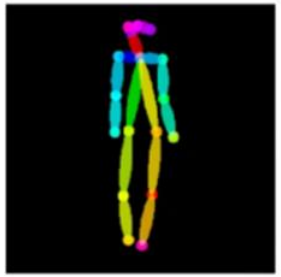


**Pose**

**Synthesis**

**Deepfashion dataset (512x512 resolution)**

# *Pose-to-body*



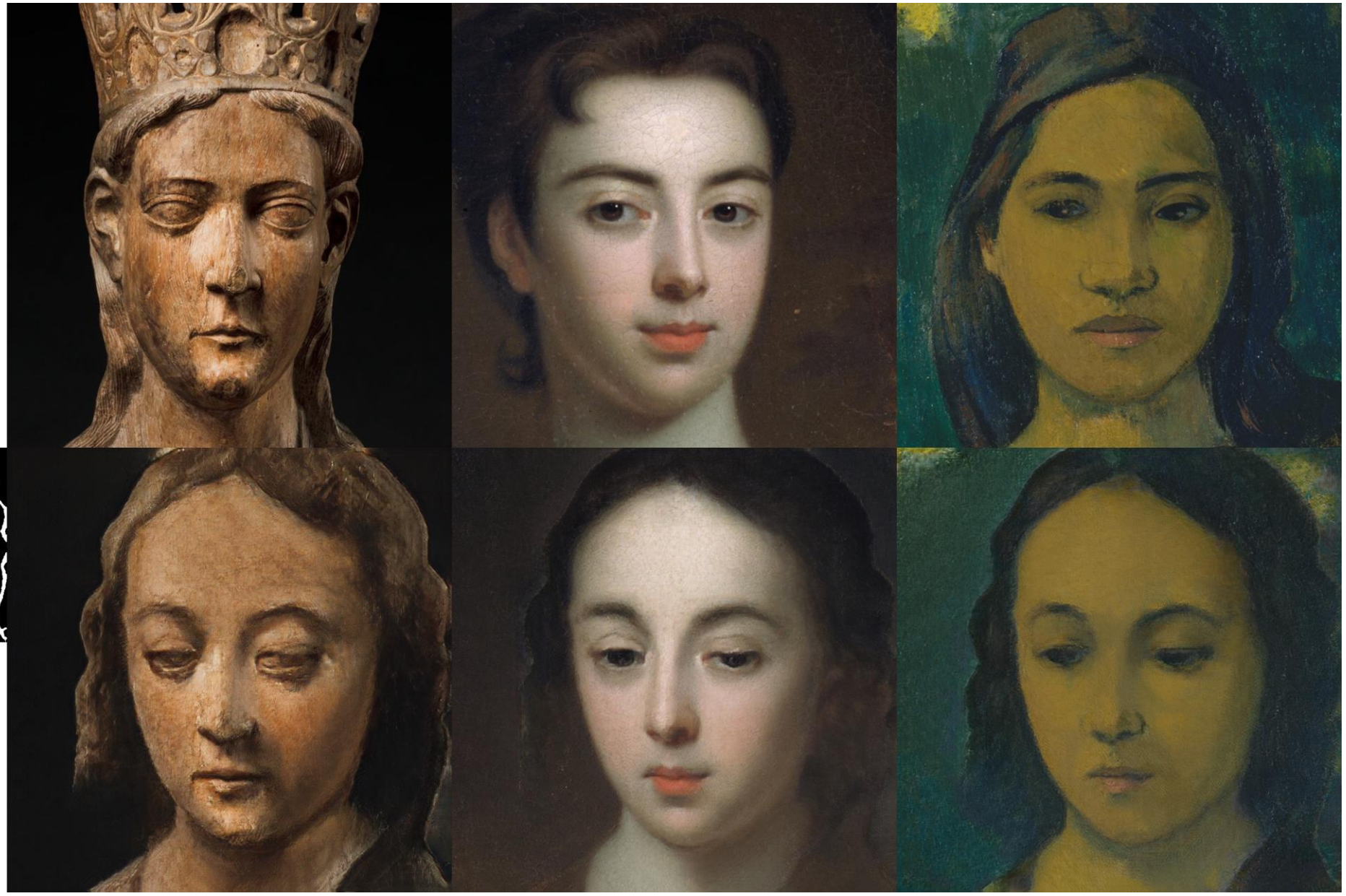
**Pose**

**Exemplar**

**Synthesis**

**Deepfashion dataset (512x512 resolution)**

# *Edge-to-face*



Exemplar

Edge

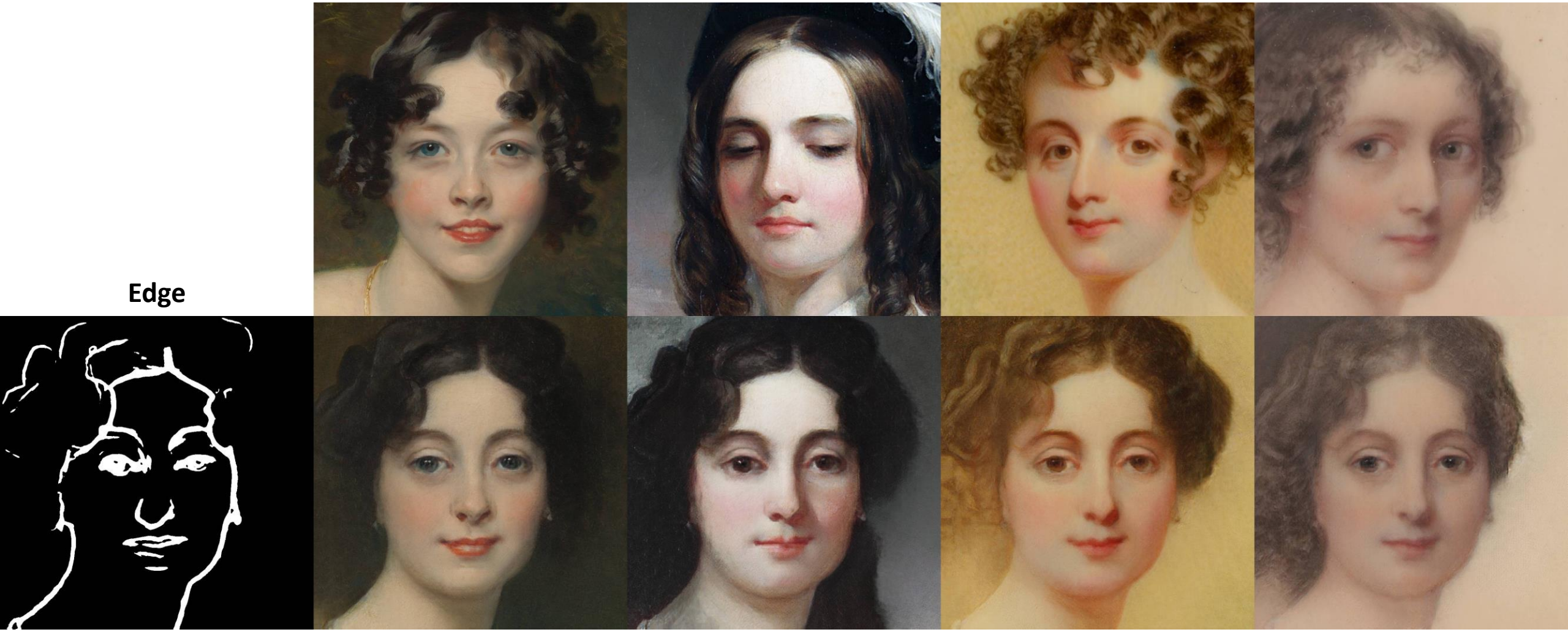
Synthesis

**MetFaces dataset (1024x1024 resolution)**

# *Edge-to-face*

**Exemplar**

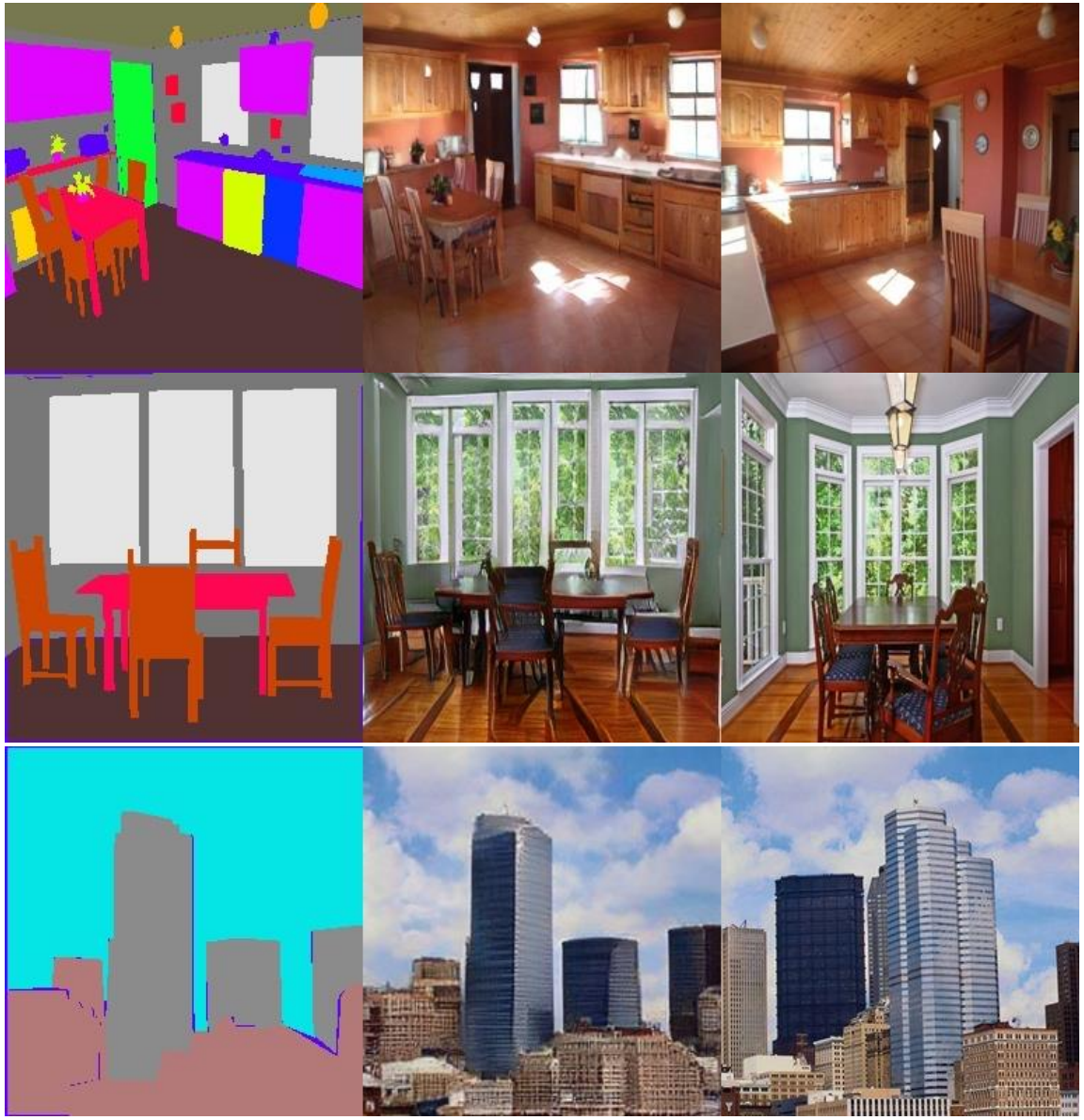
**Edge**



**Synthesis**

**MetFaces dataset (1024x1024 resolution)**

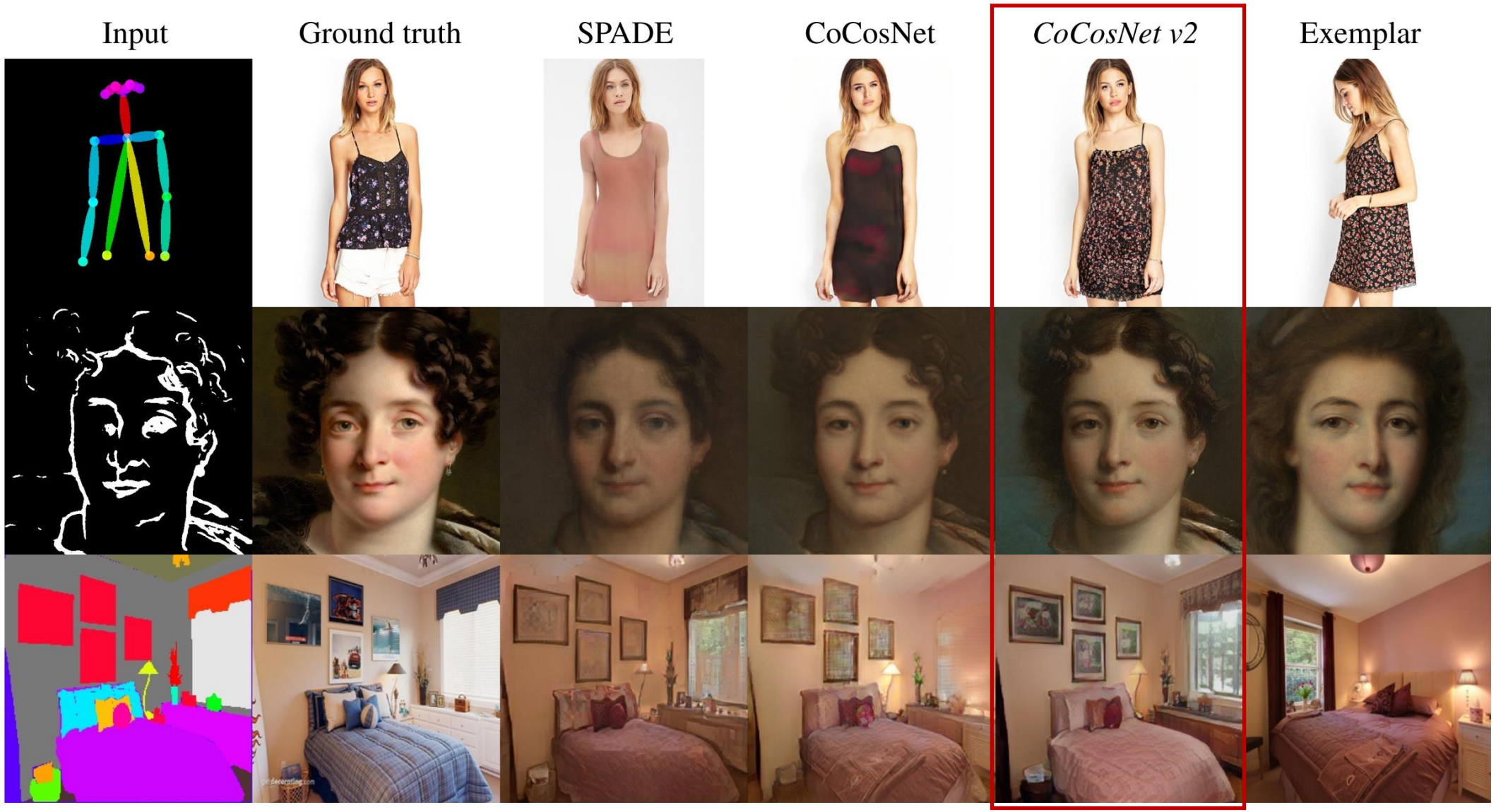
# *Mask-to-image*



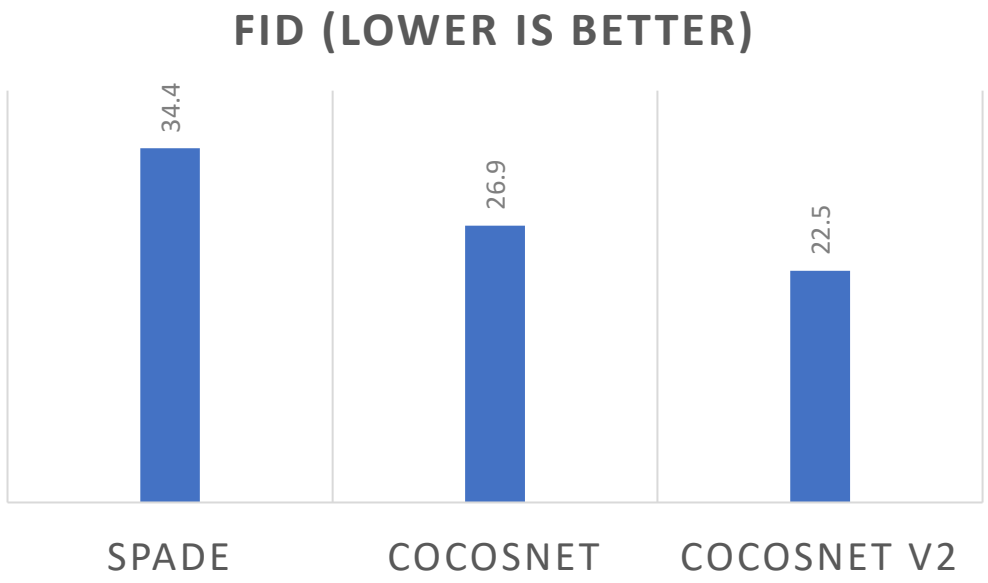
**Segmentation** **Exemplar** **Synthesis**

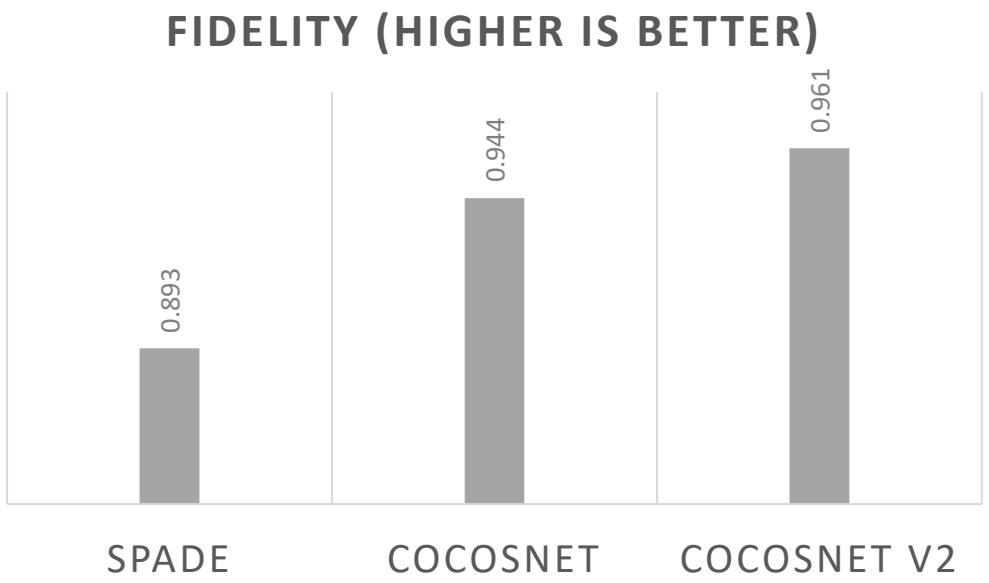
**ADE20k dataset**

# *Quantitative comparison*



# *Quantitative comparison*



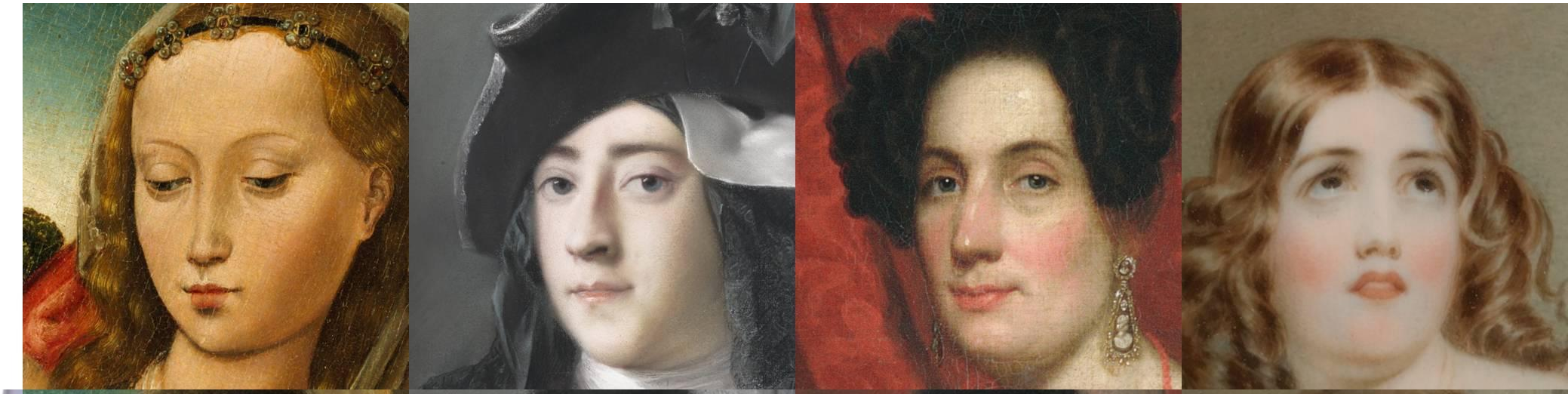



# *Application*


**Oil portrait application**

# *Application*

**Exemplar**



**Photos of real person**

**Synthesis**

**Oil portrait application**

# Thank you !