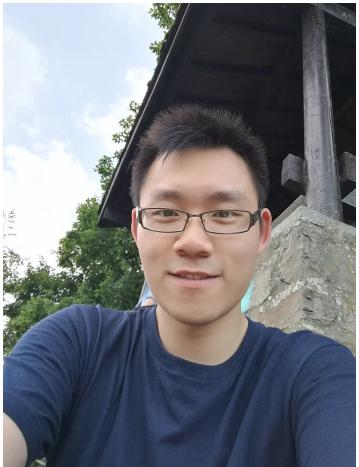


Unselfie: Translating Selfies to Neutral-pose Portraits in the Wild

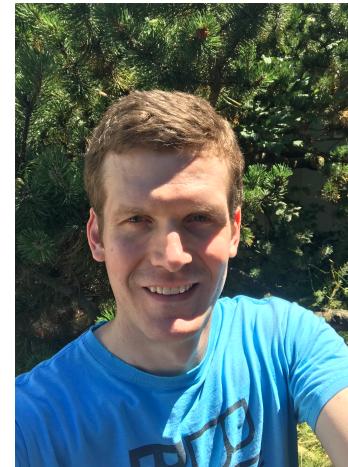
Code: <https://github.com/charliememory/Unselfie>



Liqian Ma¹



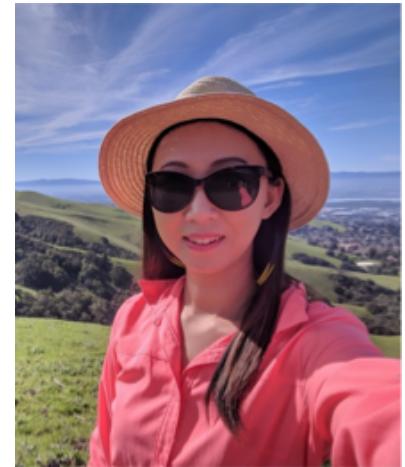
Zhe Lin²



Connelly Barnes²



Alyosha Efros^{2,3}



Jingwan Lu²

¹KU Leuven

²Adobe Research

³UC Berkeley

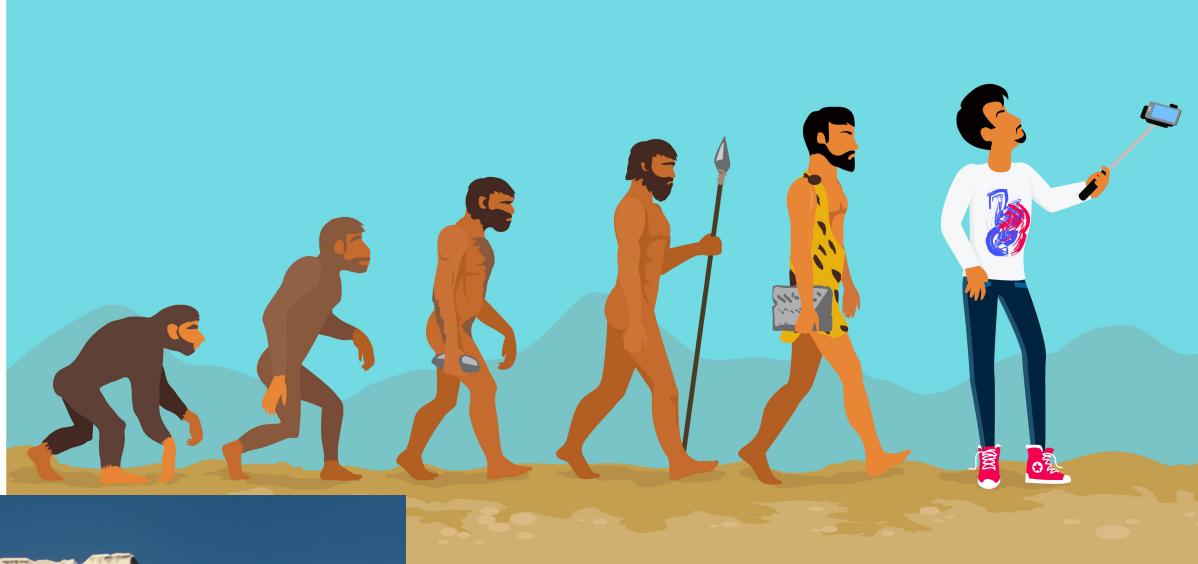


Taking selfies is convenient.



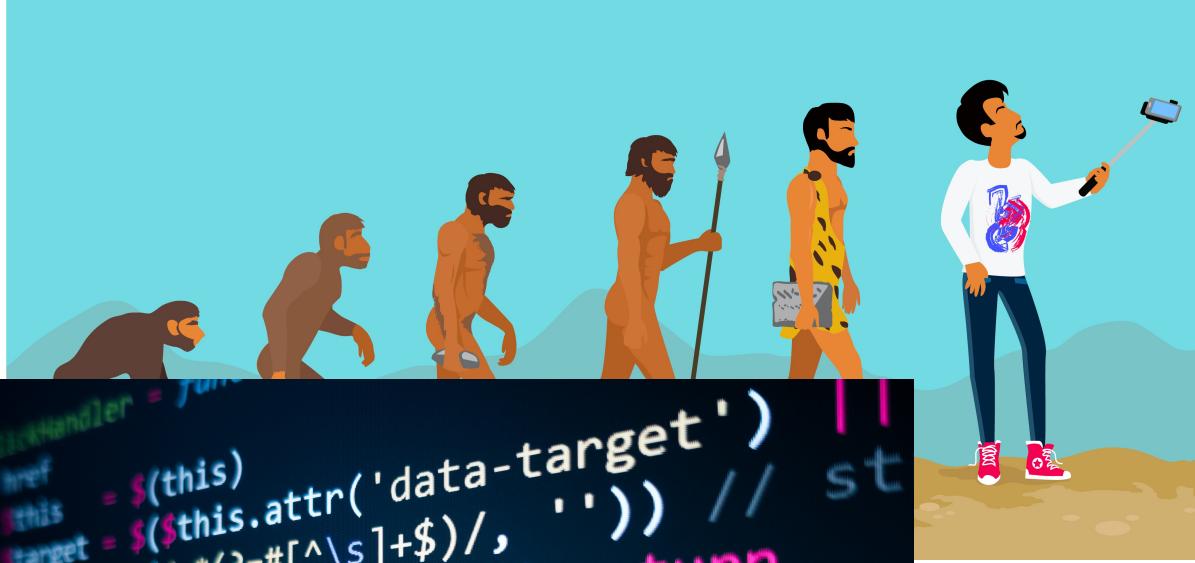
However, the body poses in selfies are usually unnatural.







```
    clickHandler = function() {
        var href = $(this).attr('href');
        var target = $($this.attr('data-target'));
        href.replace(/.*(?=#[^\s]+$/)/, '');
        if (!target.hasClass('carousel')) return;
        var options = $.extend({}, $target.data(), $);
        var slideIndex = $this.attr('data-slide-to');
        if (slideIndex) options.interval = false;
        Plugin.call($target, options);
        if (slideIndex) {
            $target.data('bs.carousel').to(slideIndex);
        }
    };
}(jQuery);
```



Unselfie: translates a selfie into a neutral-pose portrait.

Input



Baikal360 - stock.adobe.com



Result



MaximBeykov - stock.adobe.com



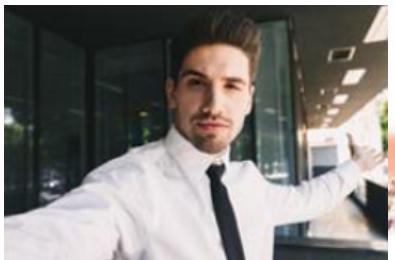
Unselfie Challenges

- Lack of paired selfie-portrait data.
- Multi-modal results with different target neutral poses.
- Dis-occlusion and seamless composition.

Unselfie Challenges

- Lack of paired selfie-portrait data.
- Multi-modal results with different target neutral poses.
- Dis-occlusion and seamless composition.

Selfie



Drobot Dean - stock.adobe.com



rh2010 - stock.adobe.com



Oleg Shelomentsev - stock.adobe.com



travnikovstudio - stock.adobe.com

Portrait

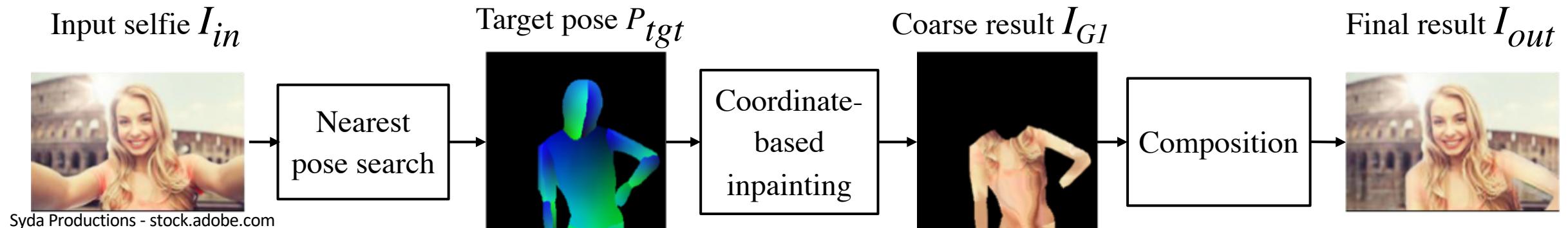


- from ATR, DeepFashion, DeepFashion2 datasets

An unpaired dataset with a self-supervised learning strategy

Unselfie Challenges

- Lack of paired selfie-portrait data.
- Multi-modal results with different target neutral poses.
- Dis-occlusion and seamless composition.



Our three-stage pipeline

Nearest
pose search

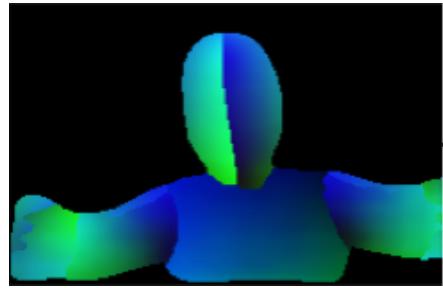
>>

Coordinate-based
inpainting

>>

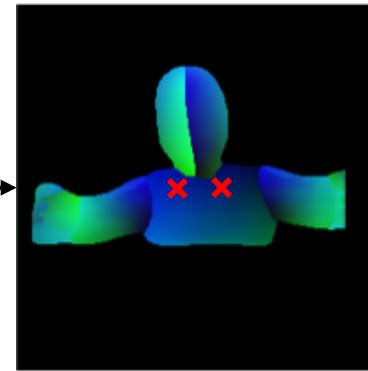
Composition

Input selfie pose P_{in}



Densepose
alignment

Source pose P_{src}

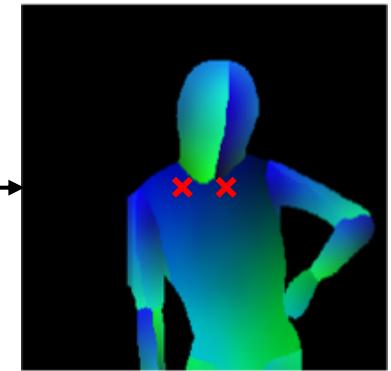


Pose database

$\{P^i\}_N$

Nearest
Search Eq. (1)(2)

Target pose P_{tgt}



Nearest
pose search

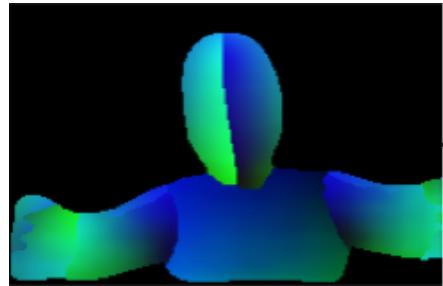
>>

Coordinate-based
inpainting

>>

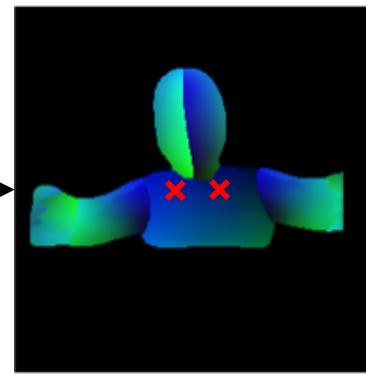
Composition

Input selfie pose P_{in}



Densepose
alignment

Source pose P_{src}

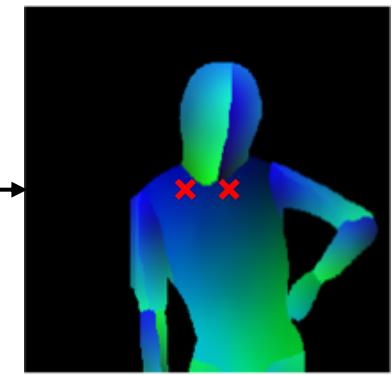


Pose database

$\{P^i\}_N$

Nearest
Search Eq. (1)(2)

Target pose P_{tgt}



Step 1: Global shape

$$d^I(P_1, P_2) = \sum_{x \in R_1 \cup R_2} \mathbb{1}(P_1^I(x) \neq P_2^I(x)) \quad (1)$$

Step 2: Local coordinate

$$d^{UV}(P_1, P_2) = \sum_{x \in R_1 \cap R_2} \|P_1^{UV}(x) - P_2^{UV}(x)\|_2 \quad (2)$$

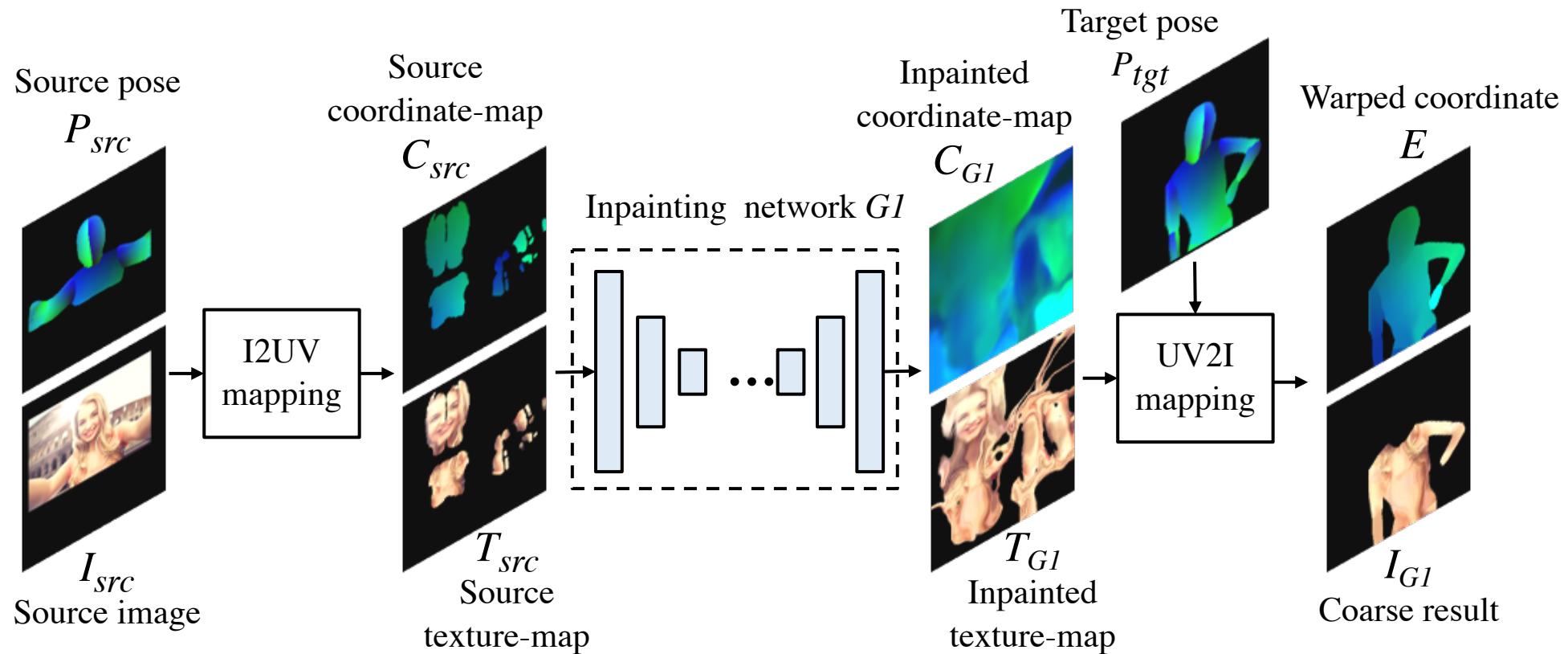
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



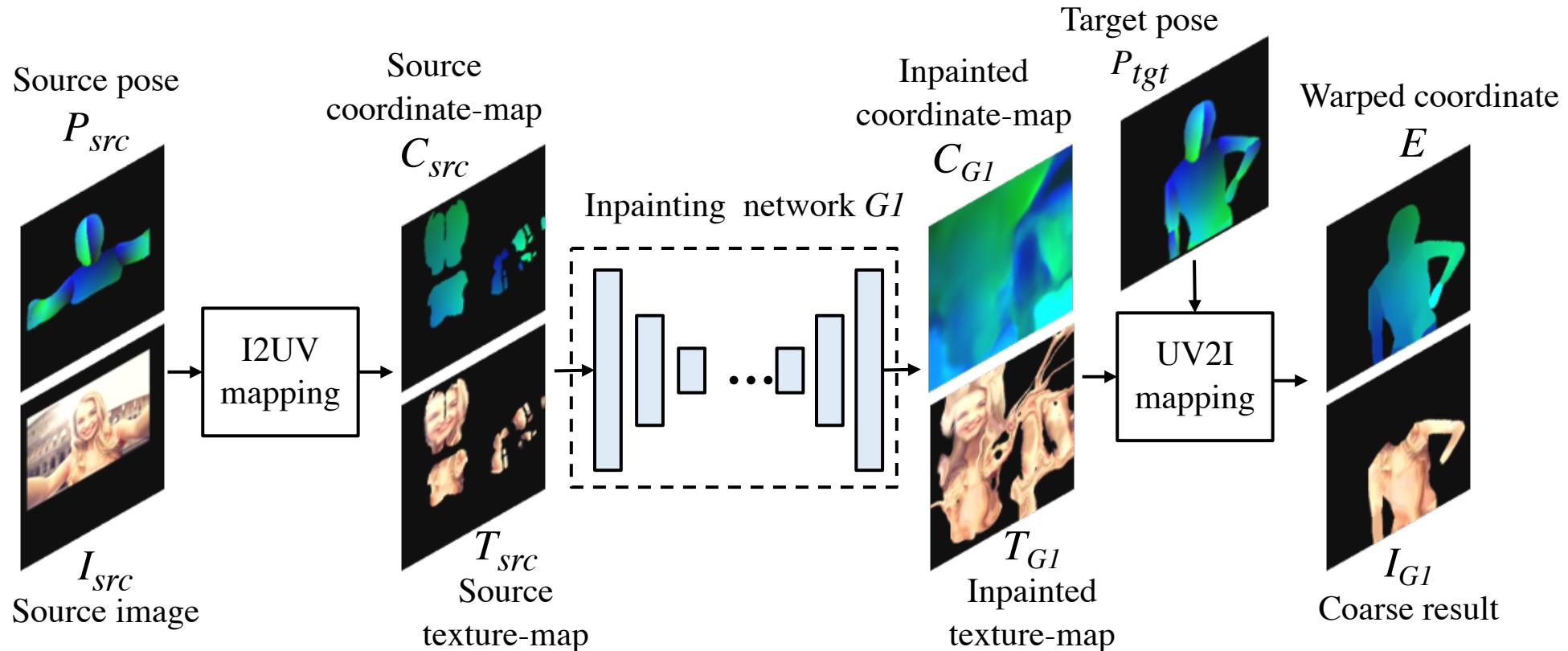
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



- Preserve source coordinate

$$L_{idt}^{G_1} = \mathbb{E}[\|C_{G_1} - C_{src}\|_2^2 V_{src}] \quad (3)$$

- Target texture supervision

$$L_1^{G_1} = \mathbb{E}[\|T_{G_1} - T_{tgt}\|_1 V_{tgt}], \quad (4)$$

$$L_P^{G_1} = \mathbb{E}[\|\phi(T_{G_1}) - \phi(T_{tgt})\|_2^2 V_{tgt}] \quad (5)$$

Nearest
pose search

>>

Coordinate-based
inpainting

>>

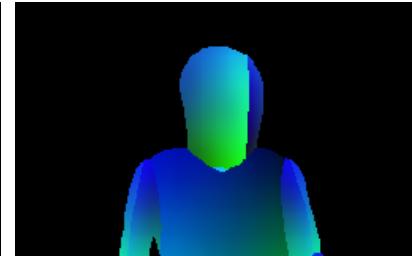
Composition

- Synthesized (portrait, selfie) pair in image space

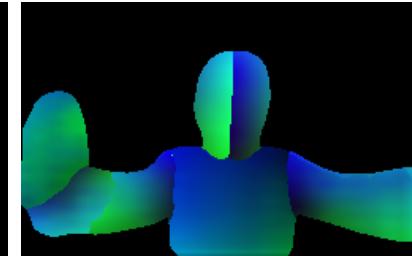
Ground truth
portrait image I_{tgt}



Portrait pose
 P_{tgt}



Nearest selfie pose P_{src}
from $\{P_{selfie}^i\}$



Synthesized selfie
image I_{src}

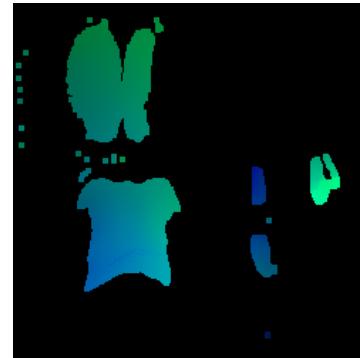


- Synthesized (portrait, selfie) pair in UV space (Our choice)

Ground truth portrait
texture-map T_{tgt}



Synthesized selfie
coordinate-map C_{src}



Synthesized selfie
texture-map T_{src}



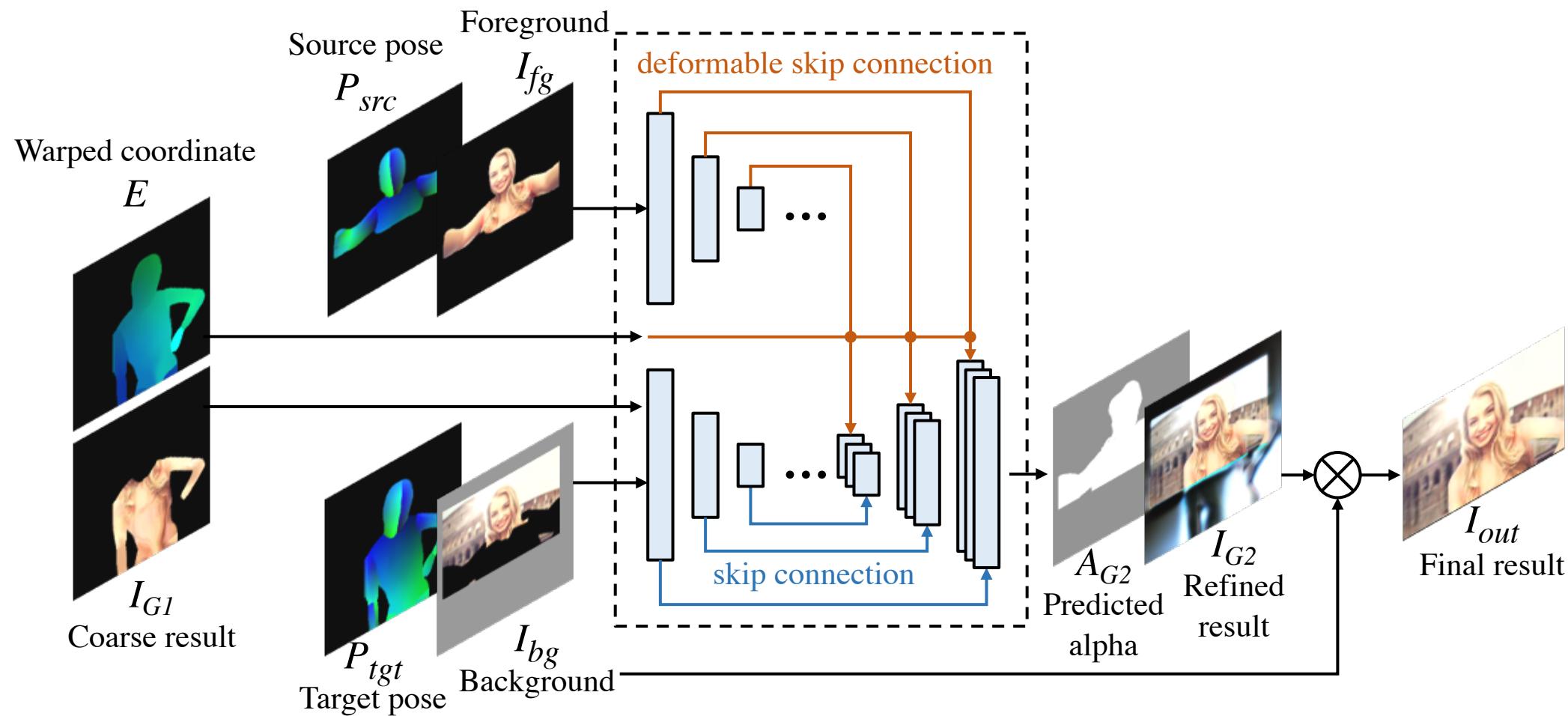
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



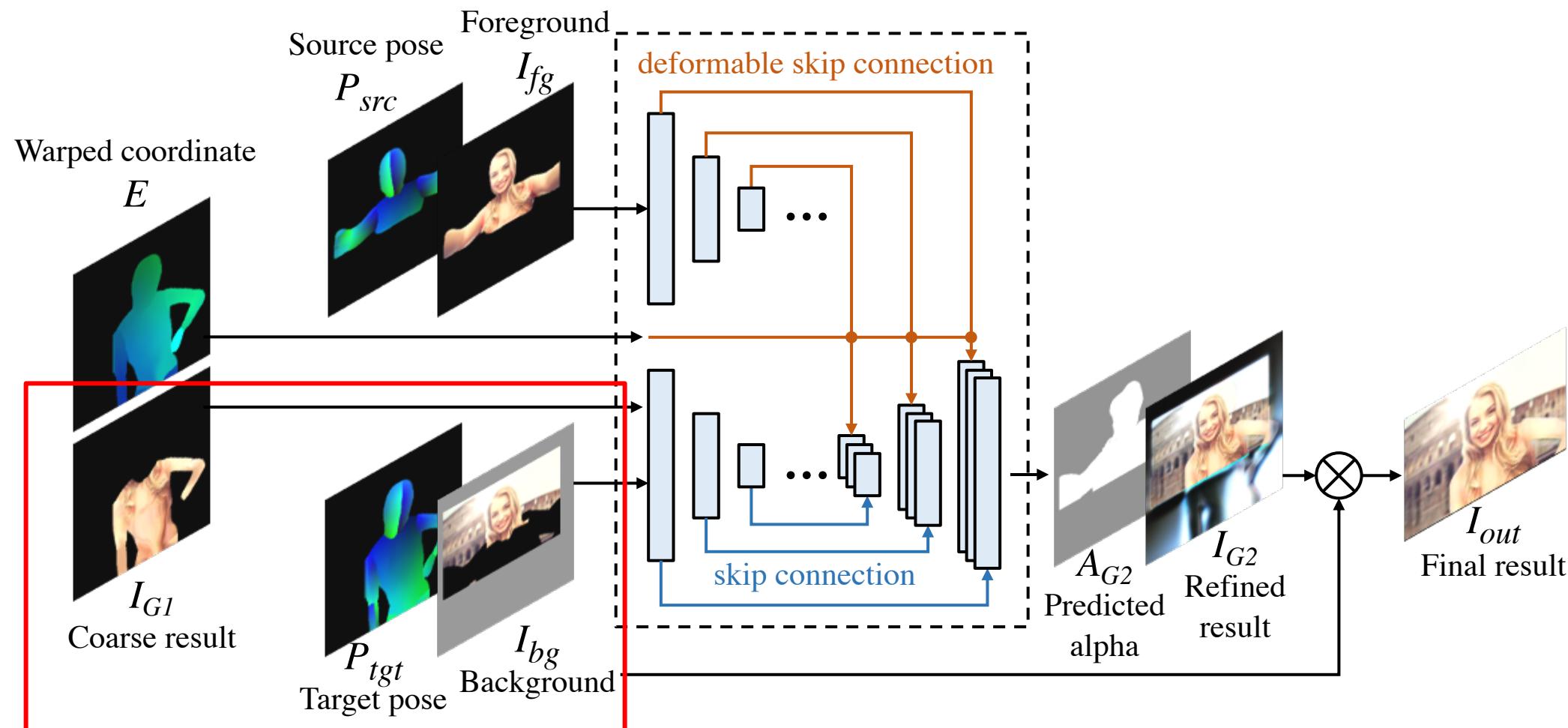
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



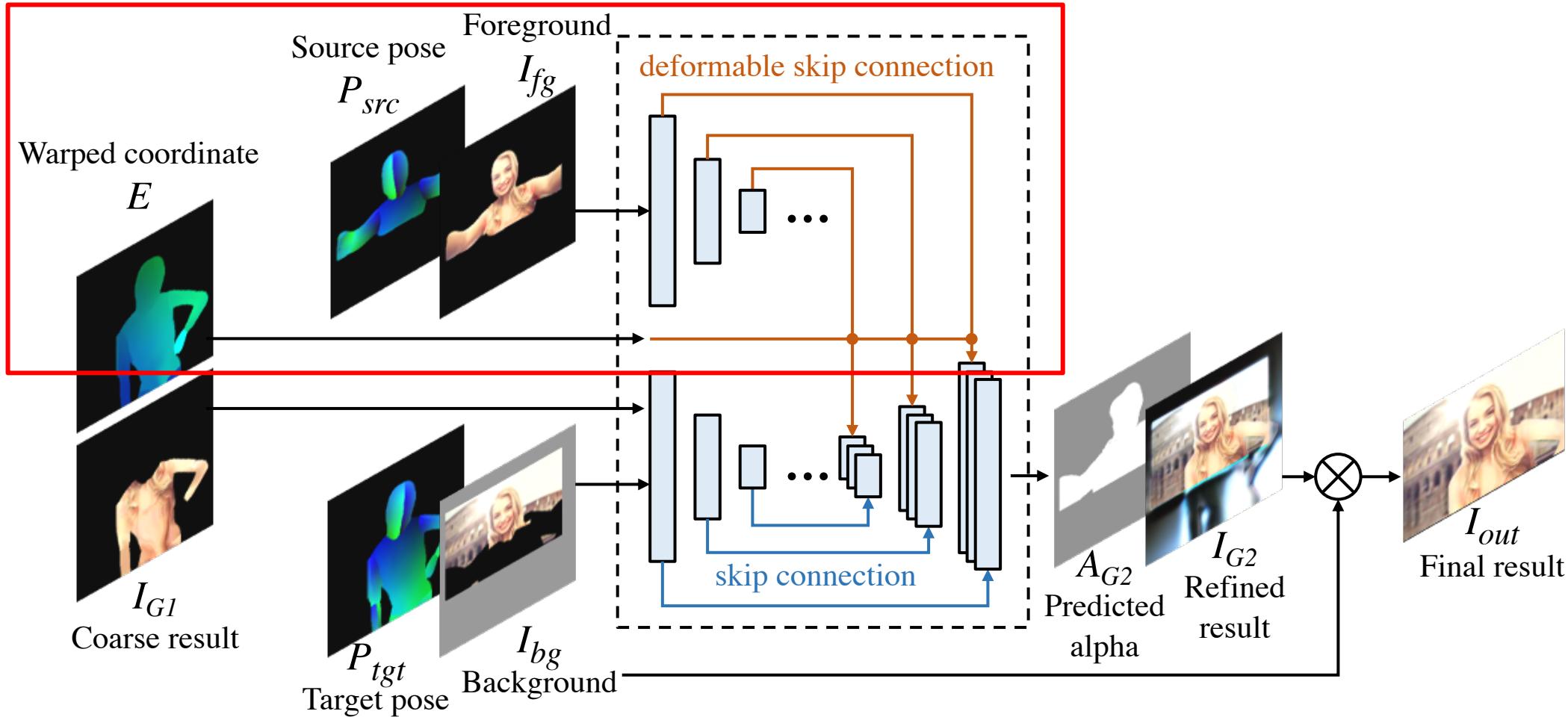
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



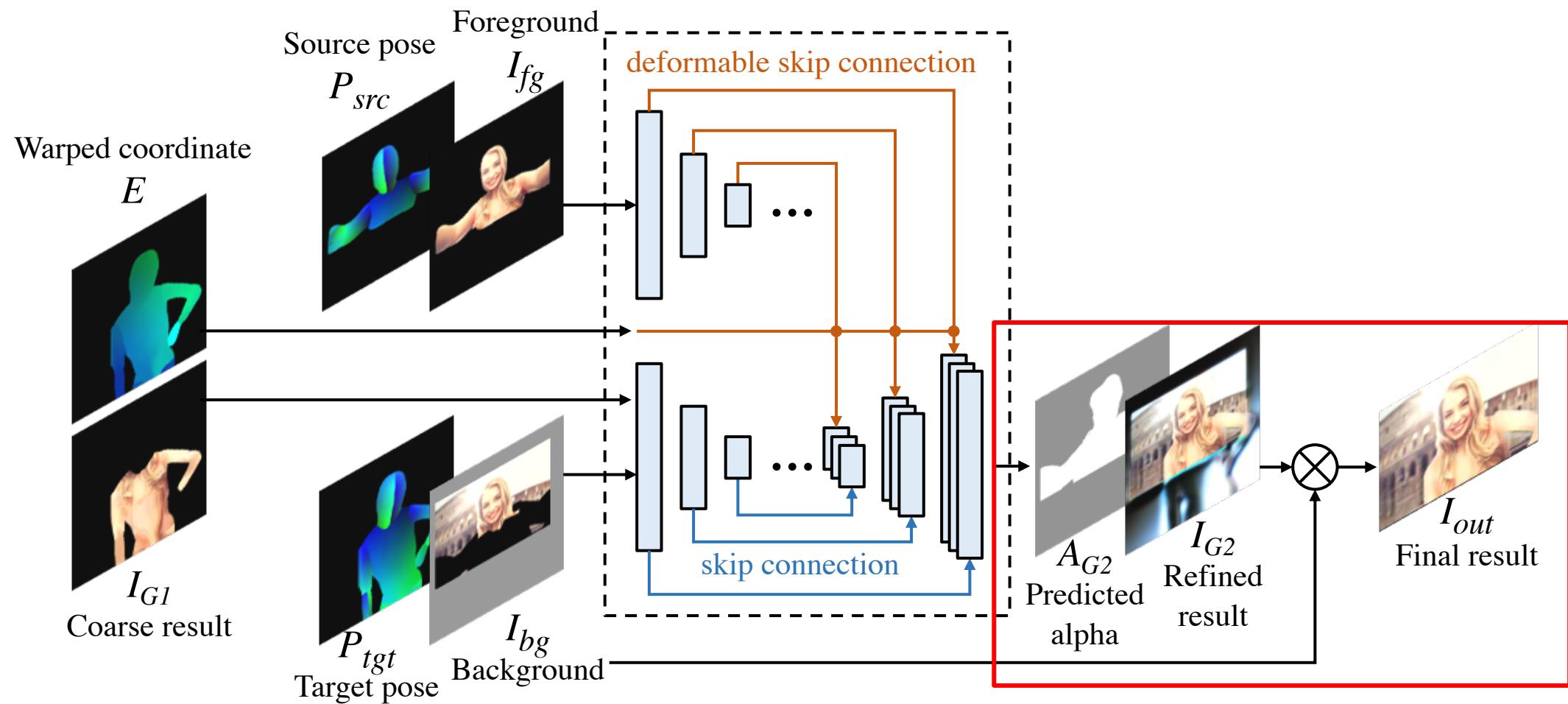
Nearest
pose search

>>

Coordinate-based
inpainting

>>

Composition



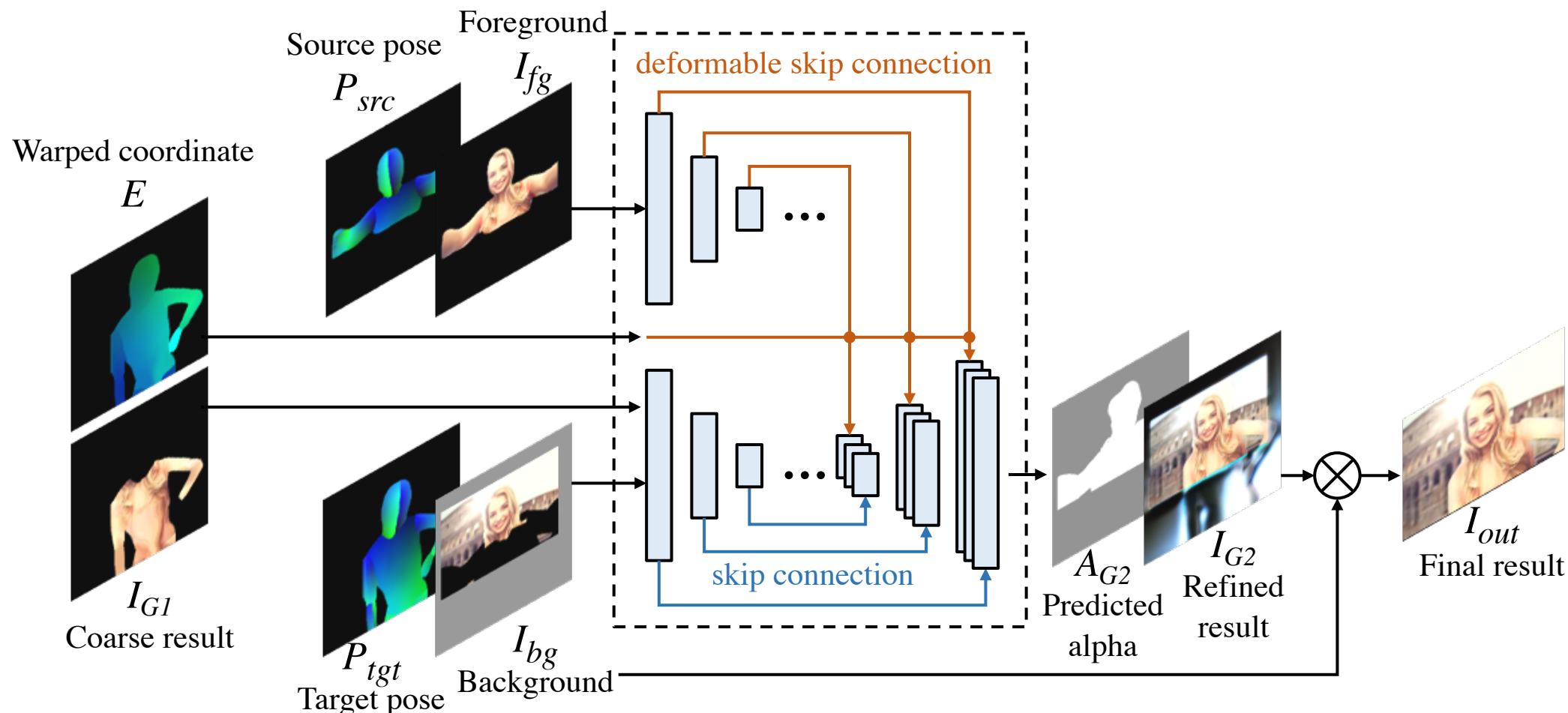
Nearest
pose search

>>

Coordinate-based
inpainting

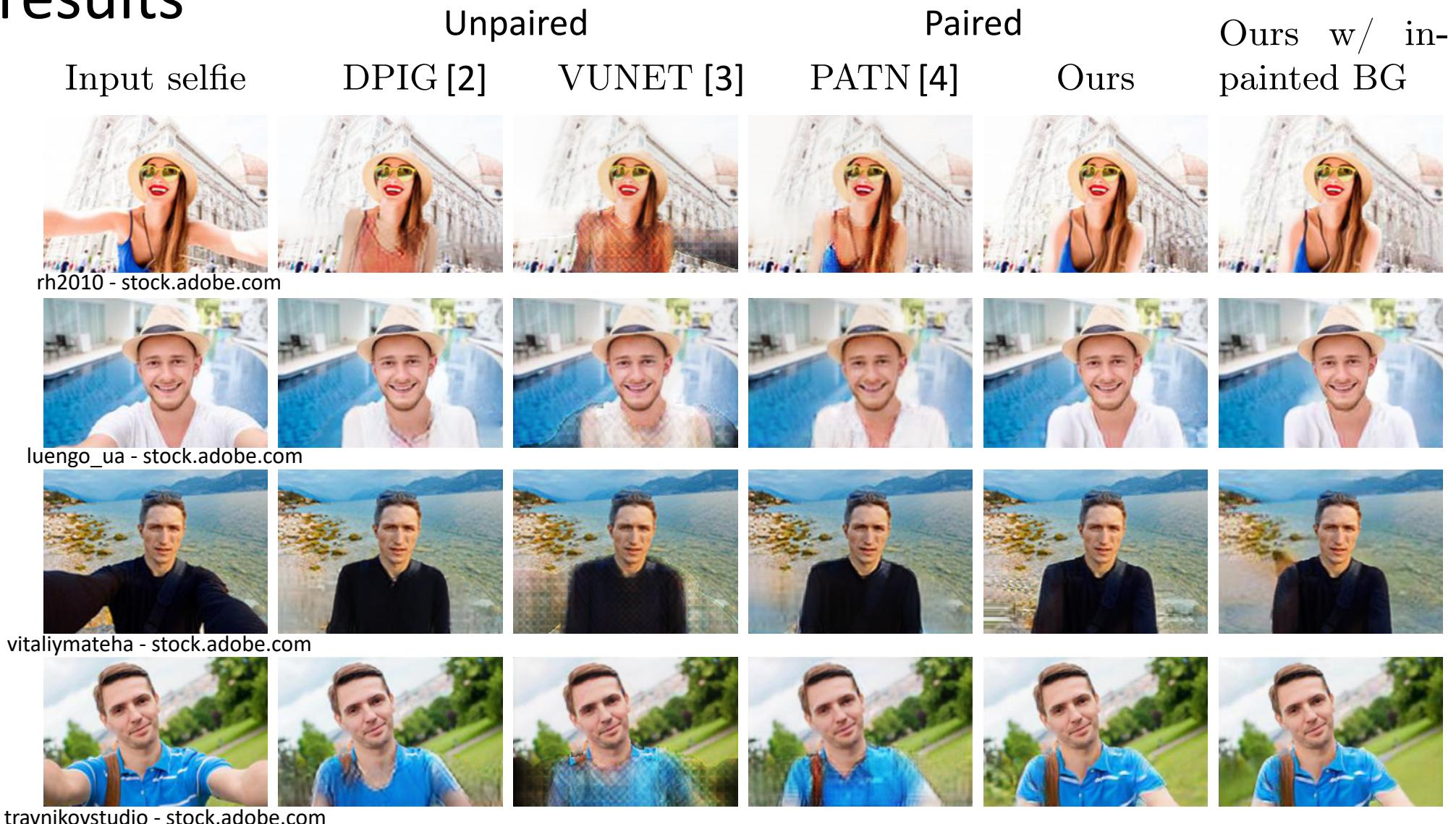
>>

Composition



$$\min_{G_2} \max_D L^{G_2} = \underbrace{\lambda_3 L_1^{G_2}}_{\text{Reconstruction}} + \underbrace{\lambda_4 L_P^{G_2}}_{\text{Adversarial}} + \underbrace{\lambda_5 L_{adv}^{G_2, D}}_{\text{Alpha}} + L_A^{G_2}$$

Our results



[2] Ma, L., et al. Disentangled person image generation. CVPR'18.

[3] Esser, P., et al. A variational u-net for conditional appearance and shape generation. CVPR'18.

[4] Zhu, Z., et al. Progressive pose attention transfer for person image generation. CVPR'19.

Our results

Model	Human Prefers Ours	FID↓	KID↓
DPIG [2]	0.798	88.27	0.026
VUNET [3]	0.851	135.90	0.077
PATN [4]	0.822	104.74	0.041
Ours	N/A	71.93	0.014

FID/KID measures how realistic the generated image looks like.
Real domain: real selfies and neutral-pose portraits.
Fake domain: generated results.

[2] Ma, L., et al. Disentangled person image generation. CVPR'18.

[3] Esser, P., et al. A variational u-net for conditional appearance and shape generation. CVPR'18.

[4] Zhu, Z., et al. Progressive pose attention transfer for person image generation. CVPR'19.

Our multi-modal results

Input selfie



ilovemayorova - stock.adobe.com

Top-1



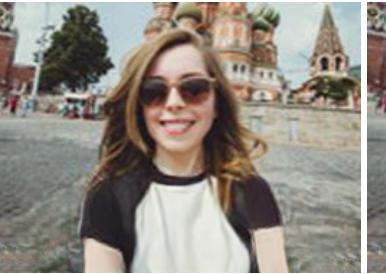
Top-2



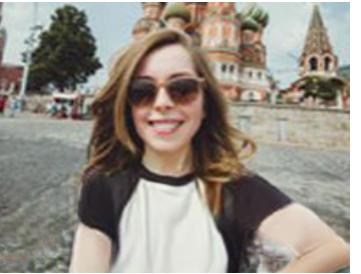
Top-3



Top-4



Top-5



Rido - stock.adobe.com



deagrezz - stock.adobe.com



luengo_ua - stock.adobe.com



Limitation and future work

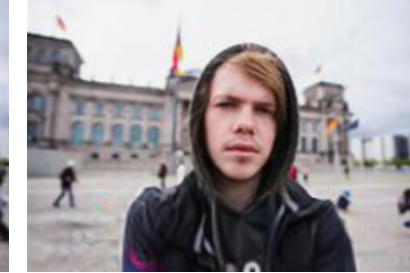
- Pose search failure

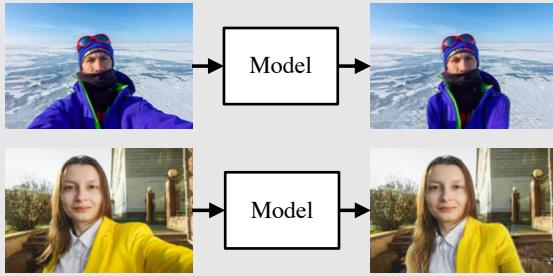


- DensePose/Mask detection failure



Pros and cons of BG inpainting

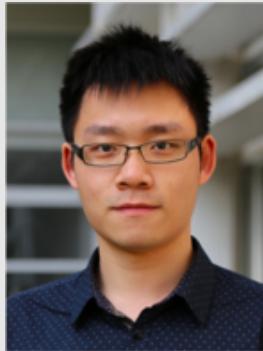
	Input selfie	Input BG I_{bg}	Input inpainted BG I_{bg}	Result (inpainted BG)
Pros				
Pros				
Cons				



Unselfie: Translating Selfies to Neutral-pose Portraits in the Wild

Code: <https://github.com/charliememory/Unselfie>

Thank you!



Homepage: <https://homes.esat.kuleuven.be/~liqianma/>

Email: liqian.ma@esat.kuleuven.be