PIVTONS: Pose Invariant Virtual Try-on Shoe with Conditional Image Completion

Chao-Te Chou, Cheng-Han Lee, Kaipeng Zhang,

Hu-Cheng Lee, and Winston H. Hsu

National Taiwan University, Taipei, Taiwan

Microsoft Research NVIDIA

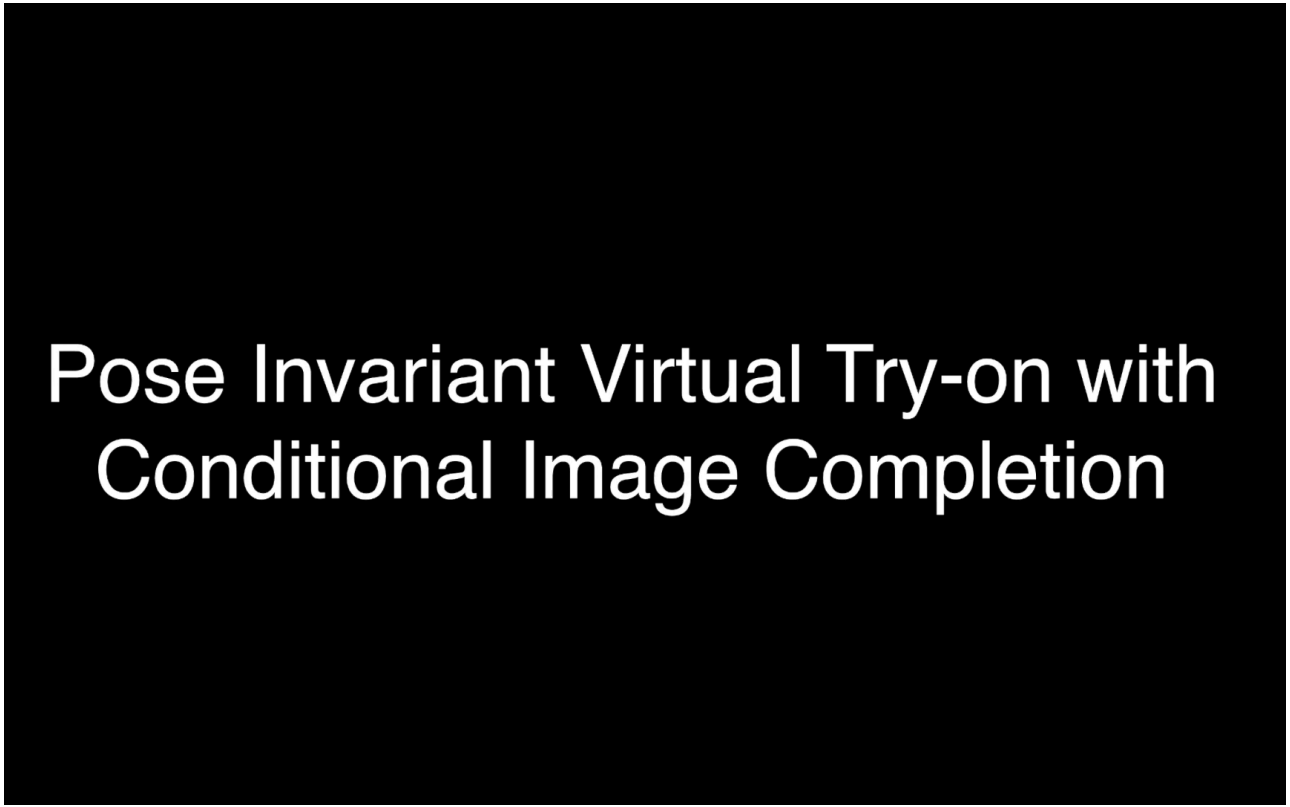
ACCV 2018

2 – 6 December 2018 Perth Western Australia

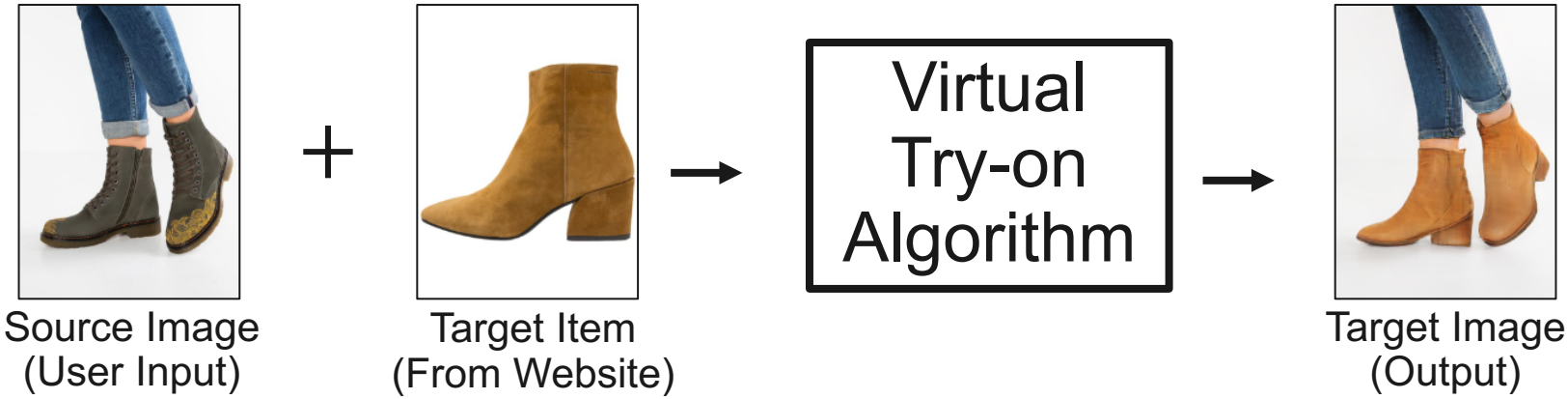
December 6, 2018

What is Virtual Try-on Shoe?

Example


Pose Invariant Virtual Try-on with
Conditional Image Completion

Problem Definition



Intuitive Method with Deep Learning

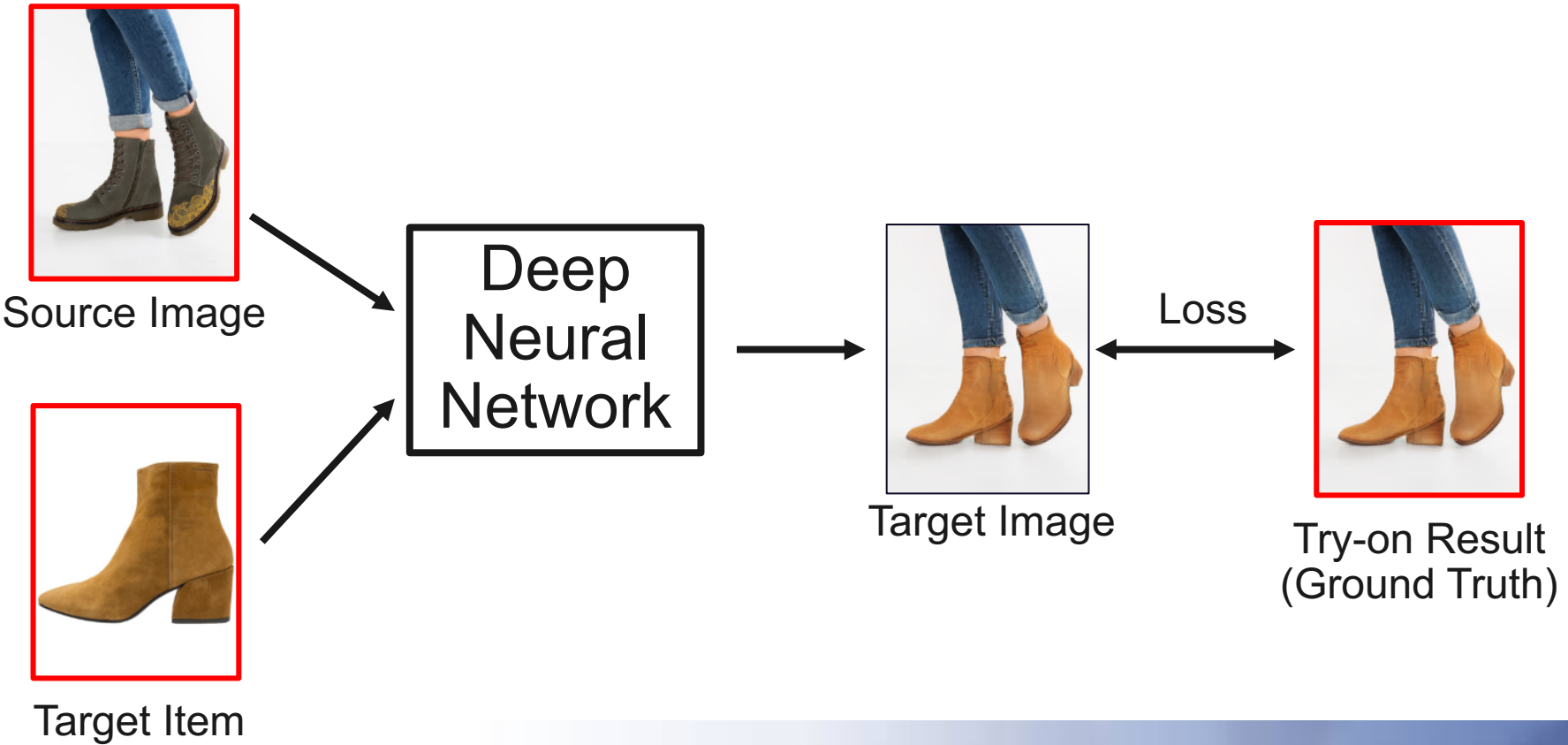
Intuitive Method

Source Image

Target Item

Try-on Result

Intuitive Method



Challenge

Source Image

Target Item

Deep Neural Network

Target Image

Loss

Try-on Result (Ground Truth)

Data collection will be expensive.

Data – In Reality

Fashion Item

Try-on Example

Image from Zalando (http://www.zalando.com/).

Intuitive method

Source Image

Target Item

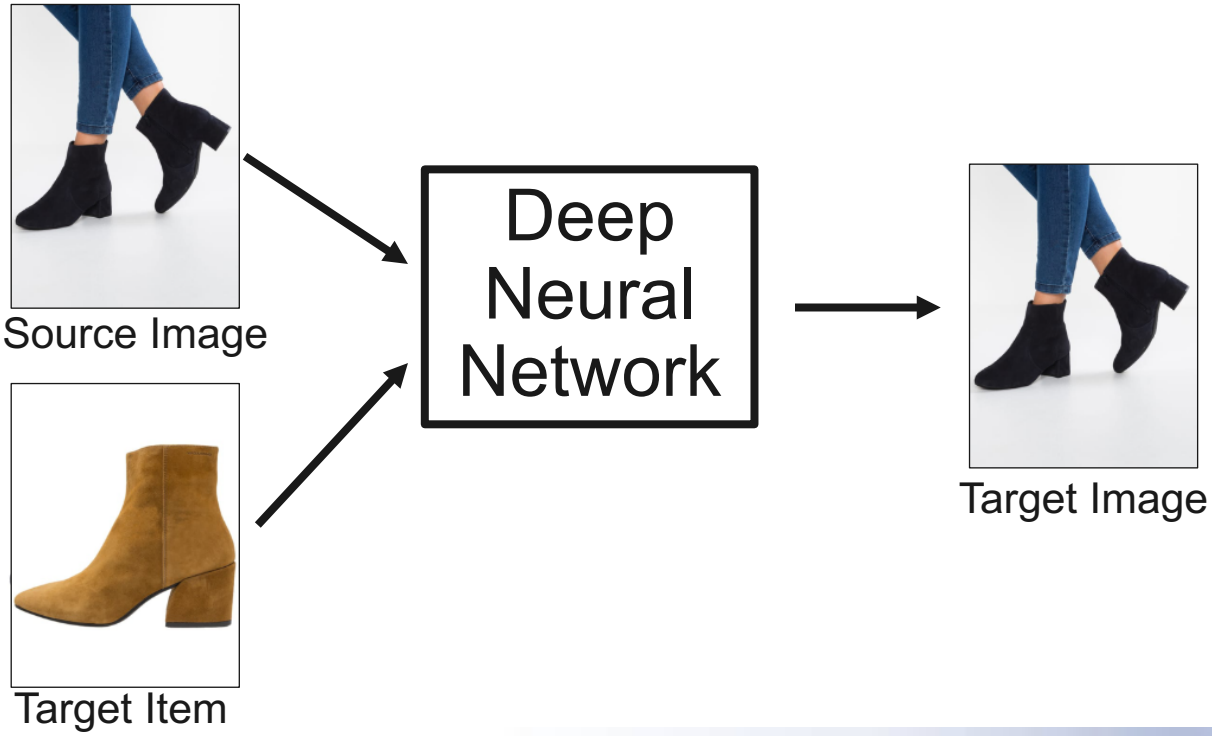
Deep Neural Network

Target Image

Loss

Source Image (Ground Truth)

Problem



Formulate as Conditional Image Completion

Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item.
2. The region outside the fashion item in the target image should be the same as that of the source image.
3. The target image should be realistic and consistent.

Source Image
(User Input)

+

Target Item
(From Website)

→

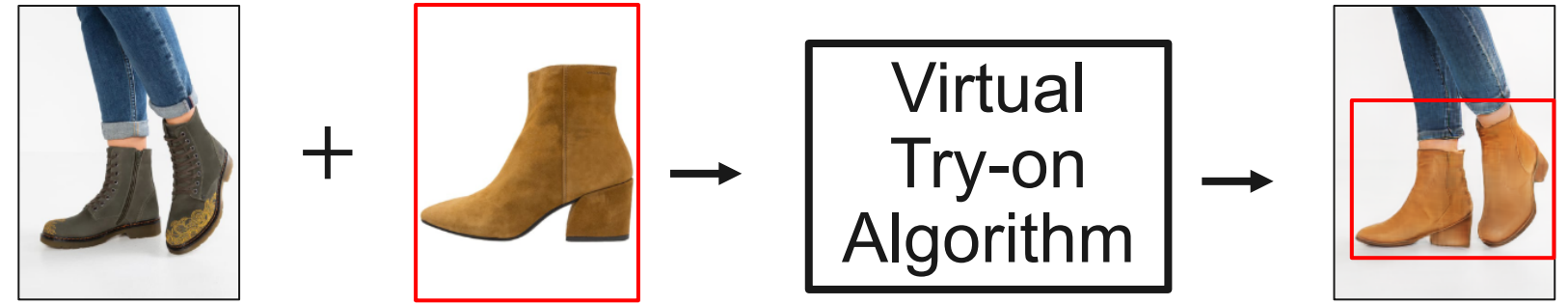
Virtual
Try-on
Algorithm

→

Target Image
(Output)

Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item.
2. The region outside the fashion item in the target image should be the same as that of the source image.
3. The target image should be realistic and consistent.



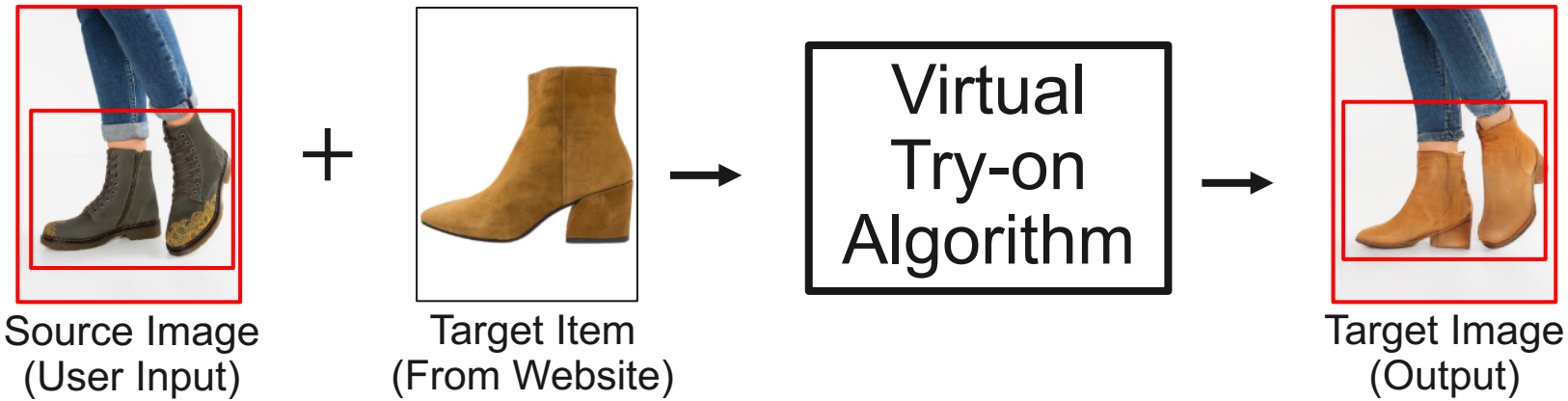
Source Image (User Input)

Target Item (From Website)

Target Image (Output)

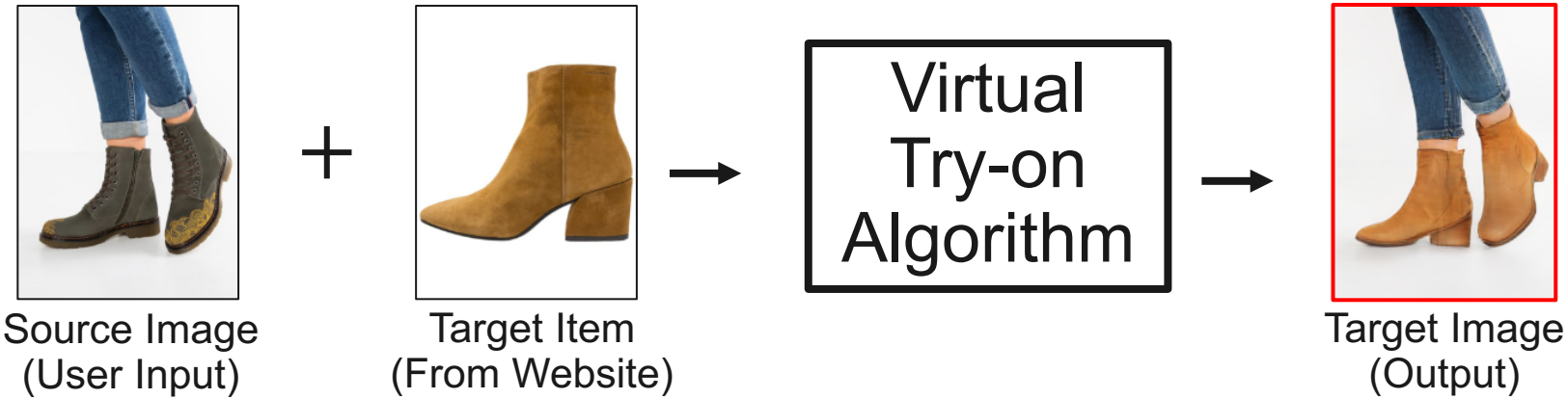
Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item.
2. The region outside the fashion item in the target image should be the same as that of the source image.
3. The target image should be realistic and consistent.



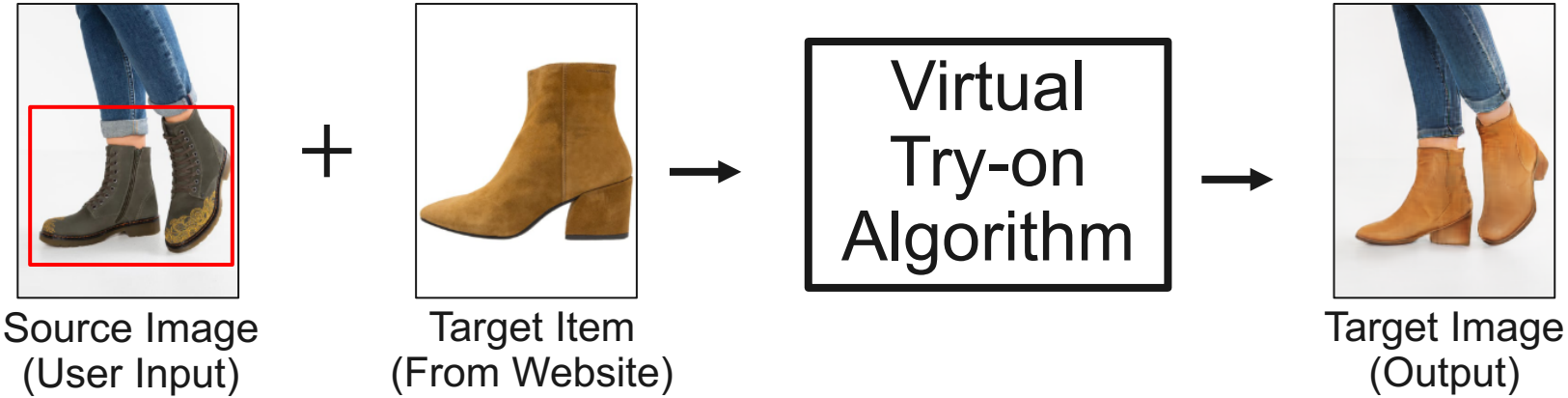
Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item.
2. The region outside the fashion item in the target image should be the same as that of the source image.
3. The target image should be realistic and consistent.



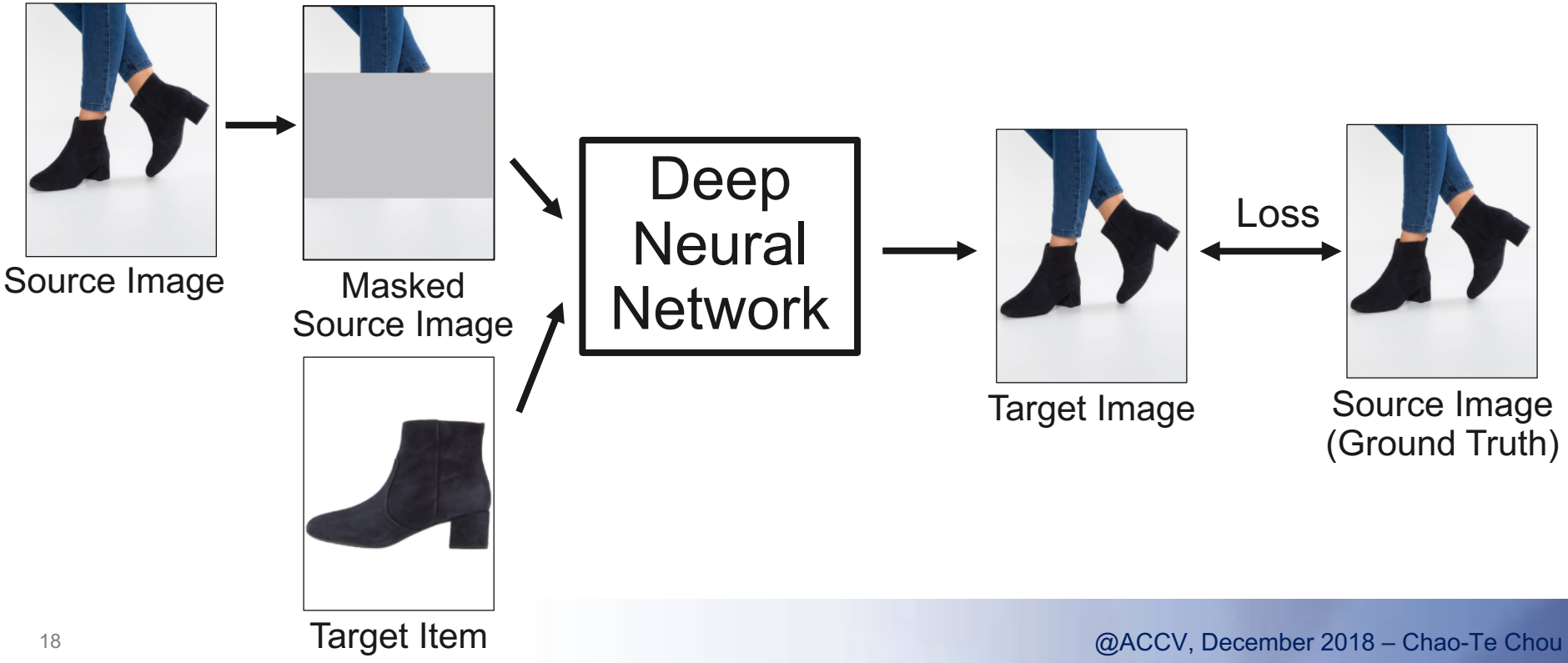
Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item.
2. The region outside the fashion item in the target image should be the same as that of the source image.
3. The target image should be realistic and consistent.

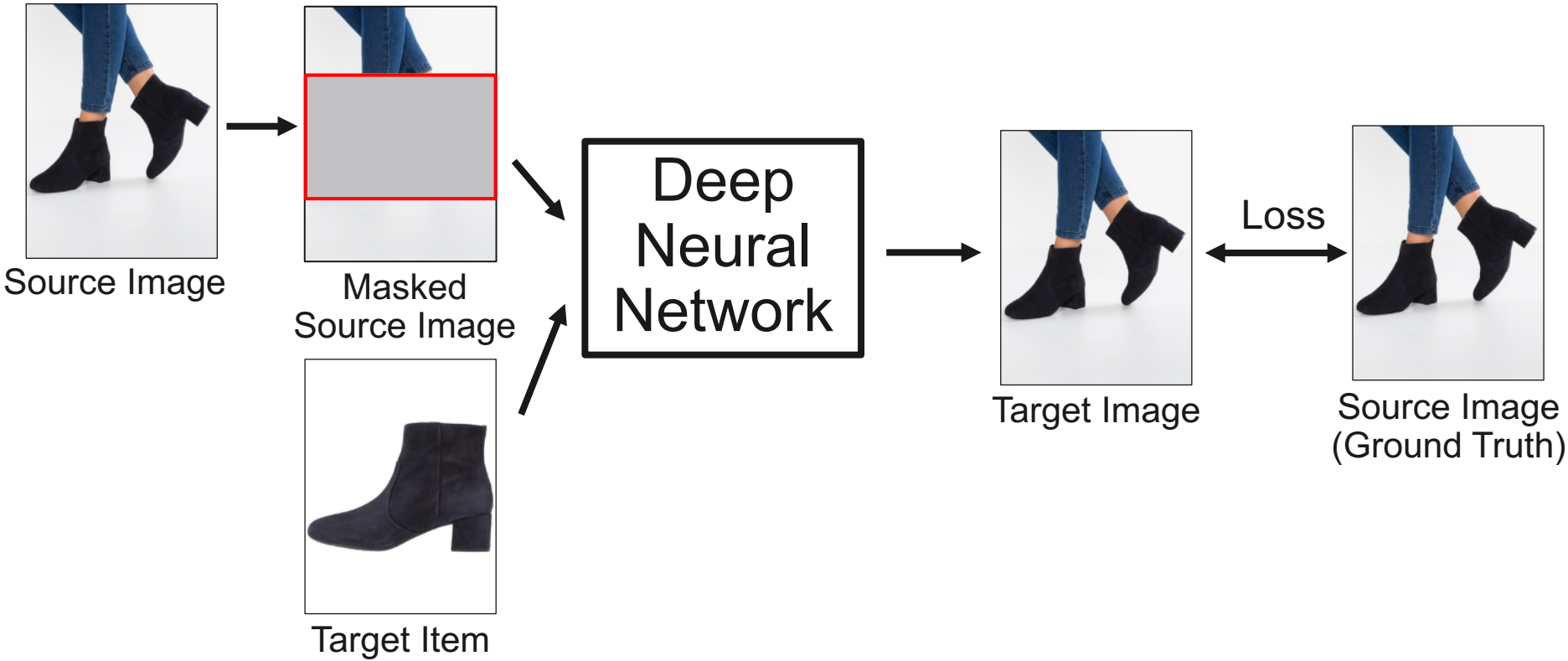


Source Image (User Input) + Target Item (From Website) → Virtual Try-on Algorithm → Target Image (Output)

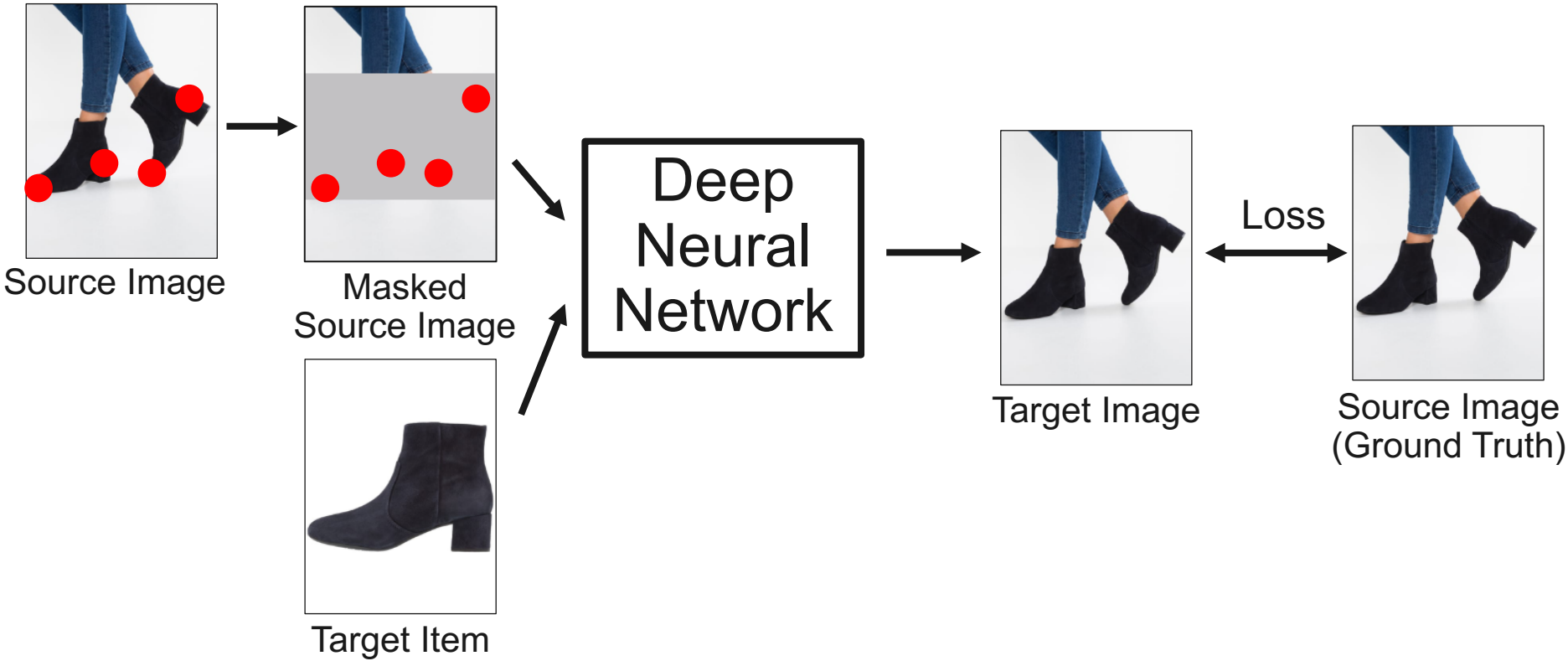
Conditional Image Completion



Pose Information missed

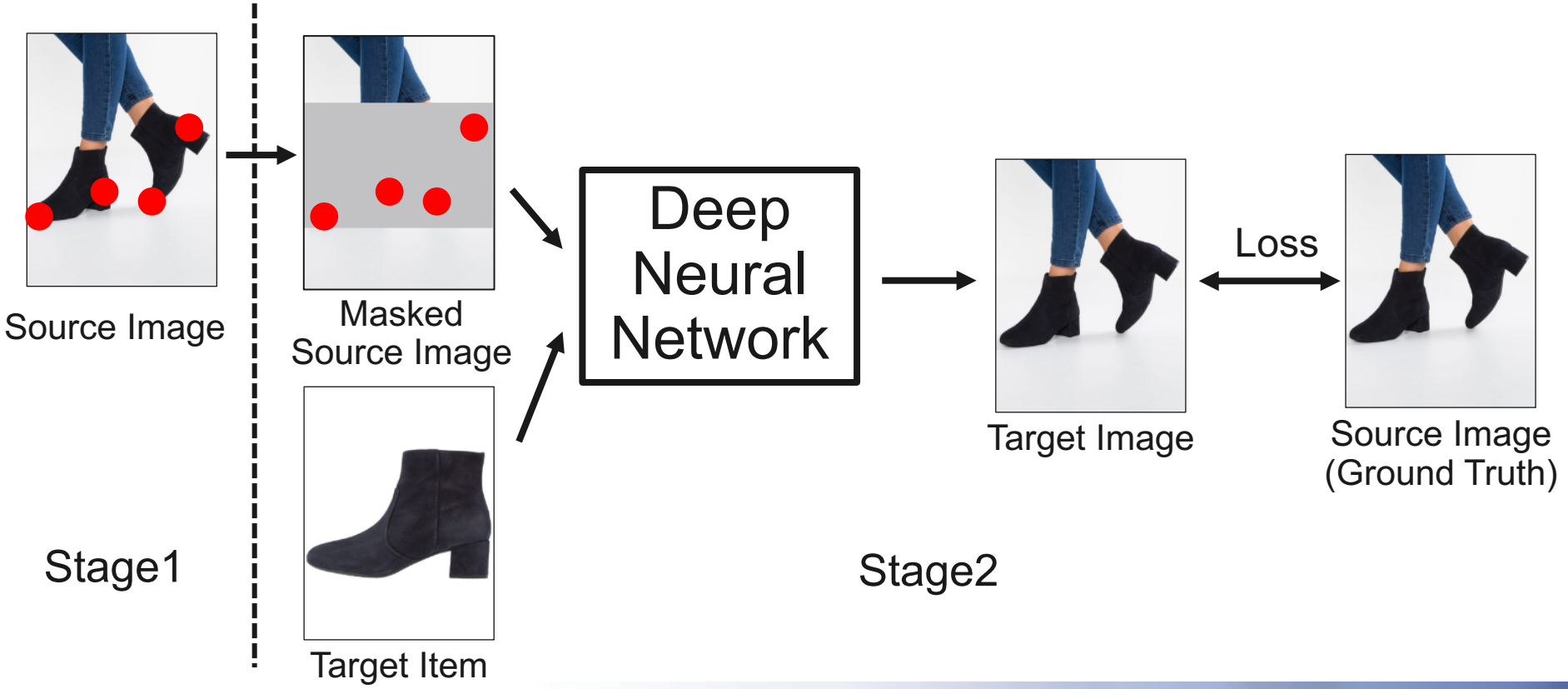


Key-points

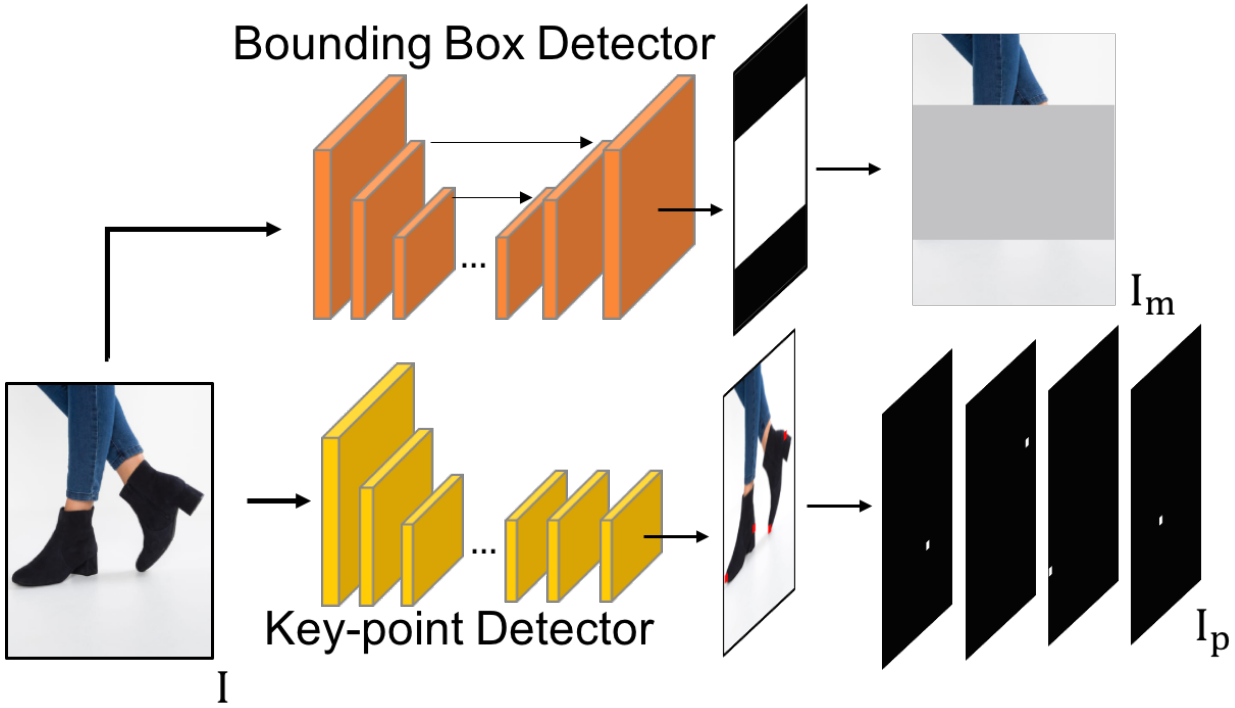

Source Image
Masked Source Image
Target Item
Deep Neural Network
Target Image
Loss
Source Image (Ground Truth)

Method

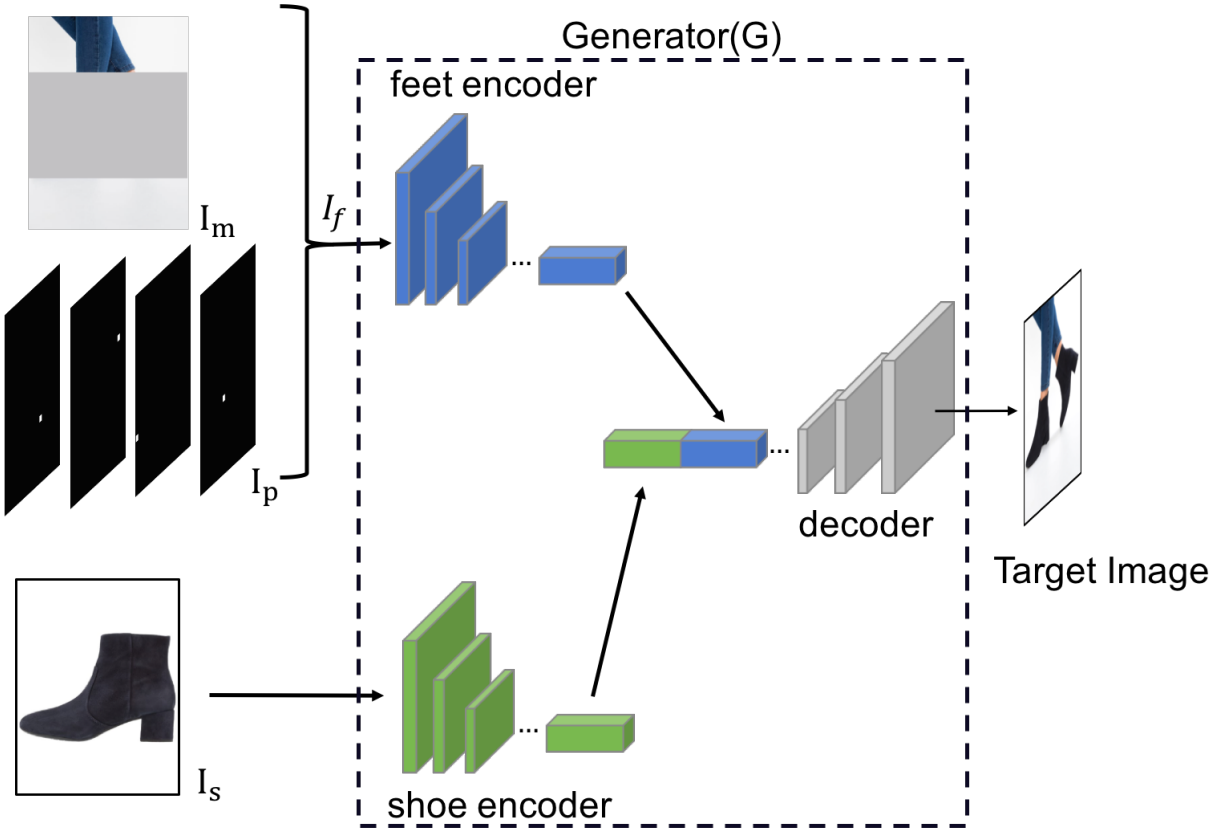
Method


Source Image
Masked Source Image
Target Item
Deep Neural Network
Target Image
Loss
Source Image (Ground Truth)
Stage1
Stage2

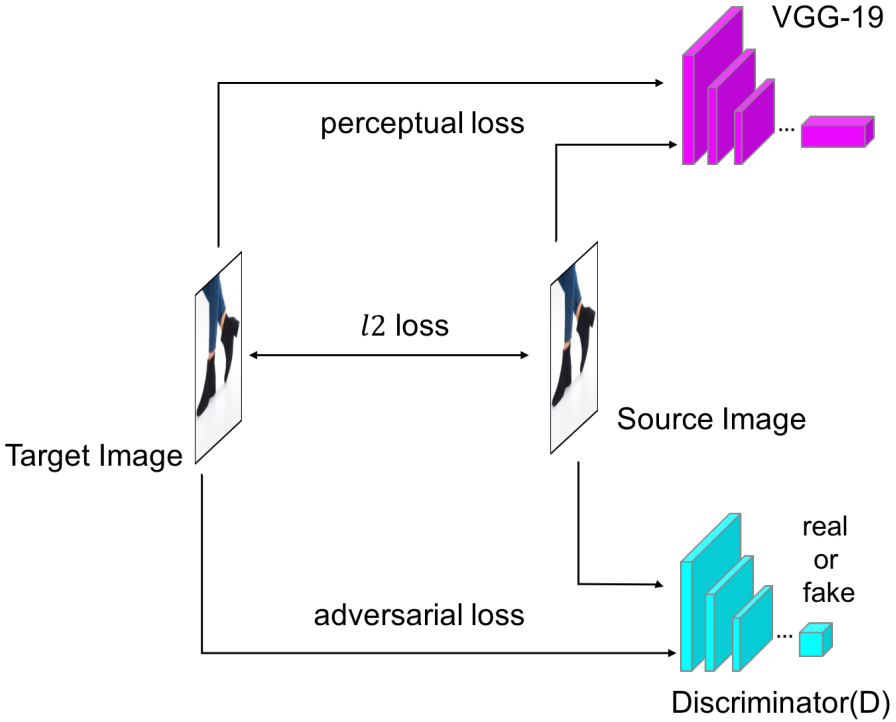
Stage1 – Detectors



Stage2 – Conditional Image Completion



Loss Function



$$L = L_{l2} + \alpha \cdot L_{perc} + \beta \cdot L_{adv}$$

We choose $\alpha = 1$ and β = 1.

Goodfellow et al., Generative Adversarial Nets., NIPS14

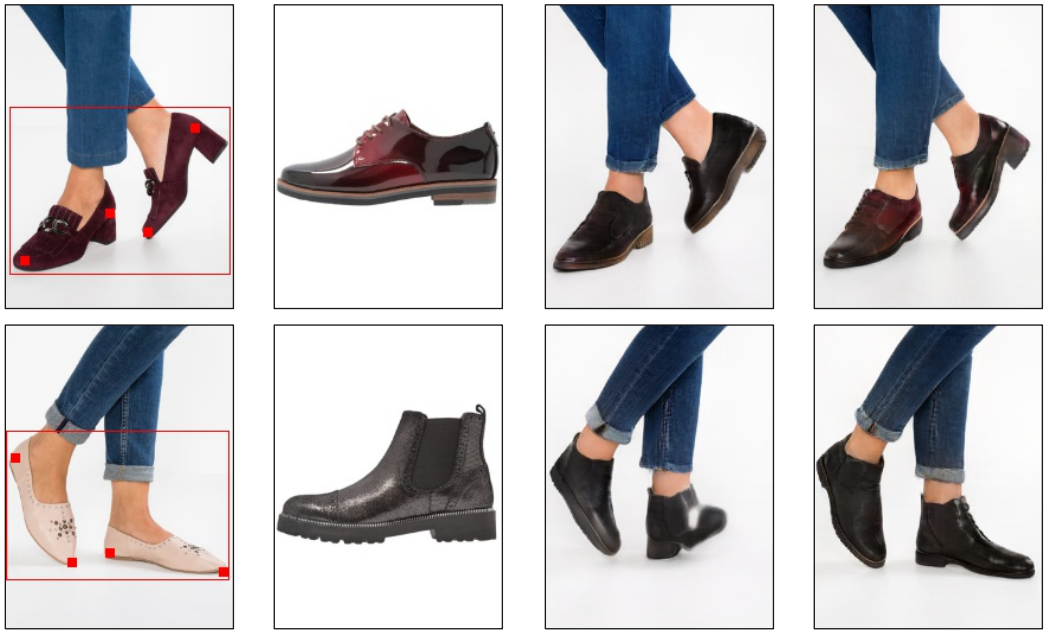
Johnson et al., Perceptual Losses for Real-Tim Style Transfer and Super-Resolution., ECCV16

Experimental Results

The effectiveness of the masked source image $\mathrm{I_m}$

Source Image | Target Item | Input I | Input $\mathrm{I_m}$

The importance of key-points



Source Image | Target Item | w/o key-points | w/ key-points

The combination of losses

Source Image | Target Item | L_{l2} | $L_{l2} + L_{perc}$ | $L_{l2} + L_{perc} + L_{adv}$

Comparison with pix2pix-m

Isola et al., Image-to-Image Translation with Conditional Adversarial Networks., CVPR17

User Study – Pairwise Comparison

	pix2pix −m	PIVTONS w/o key − points	PIVTONS w/o $L_{perc} + L_{adv}$	PIVTONS w/o L_{adv}
PIVTONS	0.9675	0.9775	0.9775	0.52

The ratio that PIVTONS wins.

Conclusion

- To the best of our knowledge, we are the first to cope with virtual try-on shoe with deep neural network.
- We present a conditional image completion method to deal with this problem.
- Through extensive experiments conducted in the collect images, we show promising results of our method.

Thank you for your listening!

Questions?

Question Slide

Pose Invariant?

Diverse poses of feet





Diverse poses of feet





Key – points!!

Pose difference between feet and shoe



Failed Cases

Failed cases

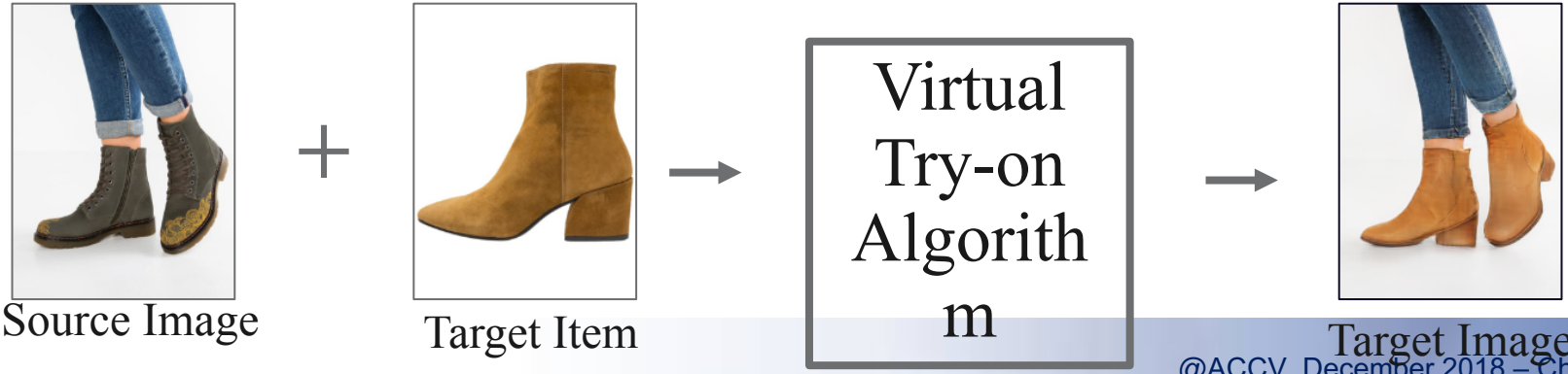
Source Image

Target Item

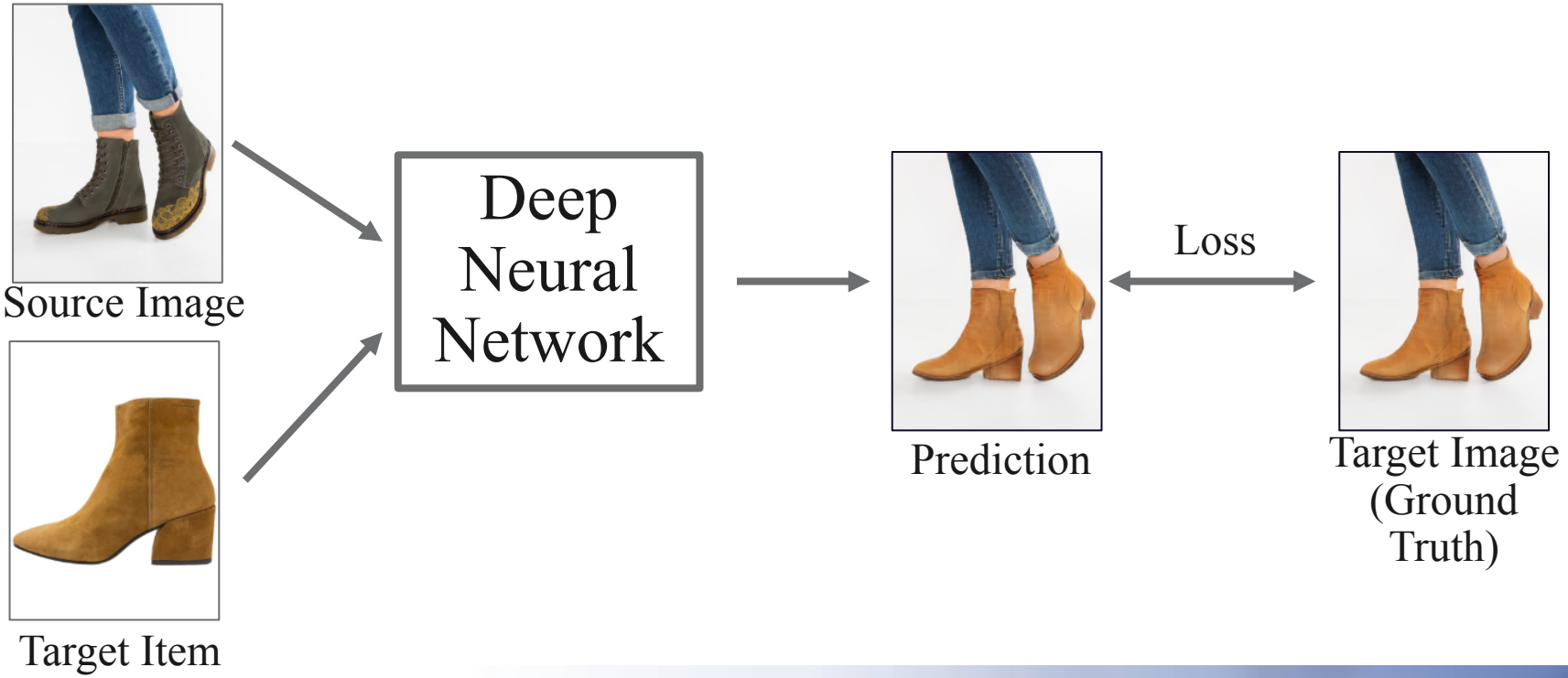
Backup

Criteria of Virtual Try-on

1. The fashion item in the target image should be the same as the target fashion item
2. The region outside the fashion item in the target image should be the same as that of the source image
3. The target image should be realistic and consistent



Method


Source Image
Target Item
Deep Neural Network
Prediction
Loss
Target Image (Ground Truth)

Thank you for your listening!