INSTAGAN: INSTANCE-AWARE IMAGE-TO-IMAGE TRANSLATION

Sangwoo Mo*, **Minsu Cho**†, **Jinwoo Shin***,‡
*Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
†Pohang University of Science and Technology (POSTECH), Pohang, Korea
‡AItrics, Seoul, Korea
*{swmo, jinwoos}@kaist.ac.kr, †mscho@postech.ac.kr

ICLR, 2019

Presented by : Kangyeol Kim
DAVIAN Lab, Korea University

Abstract

- Unsupervised image-to-image translation achieved outstanding improvement via generative adversarial networks (GANs)
- However, the task which demands dramatic changes in shape or multiple target instance, existing methods often fail.
- To tackle this problem, InstaGAN uses instance information(e.g. object segmentation masks) for overcoming aforementioned limitations
- Also, new techniques used to improve performance:
  - Context preserving loss
  - Sequential mini-batch inference/training
- Funny dataset and experiment results
  - At first glance, we can guess that the authors are korean :)

Key contributions

- An instance-augmented neural architecture is developed that translates an image and a corresponding set of attributes (segmentation masks).

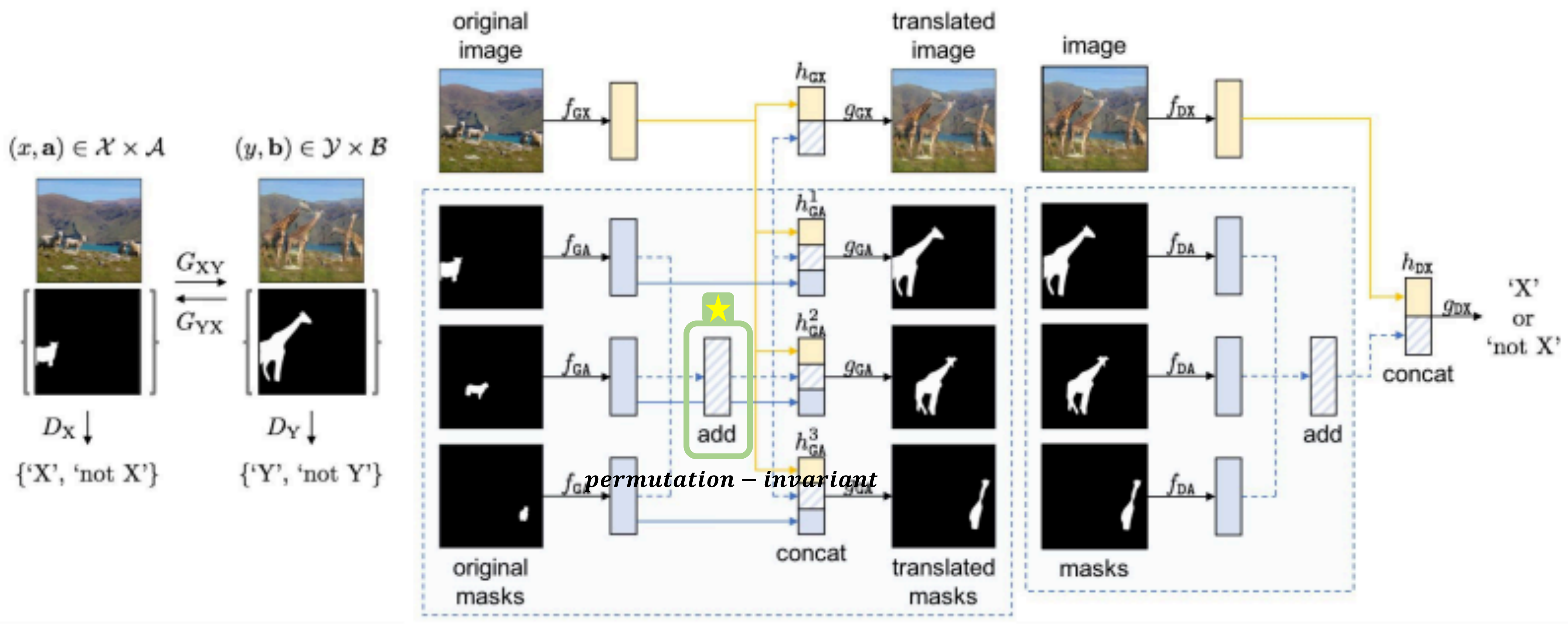
- A context preserving loss that preserves the background while transforming the target instances

- A sequential mini-batch inference/training technique that allows the system to work o subsets of data rather than requiring the full set of data

Proposed methods (1/7)

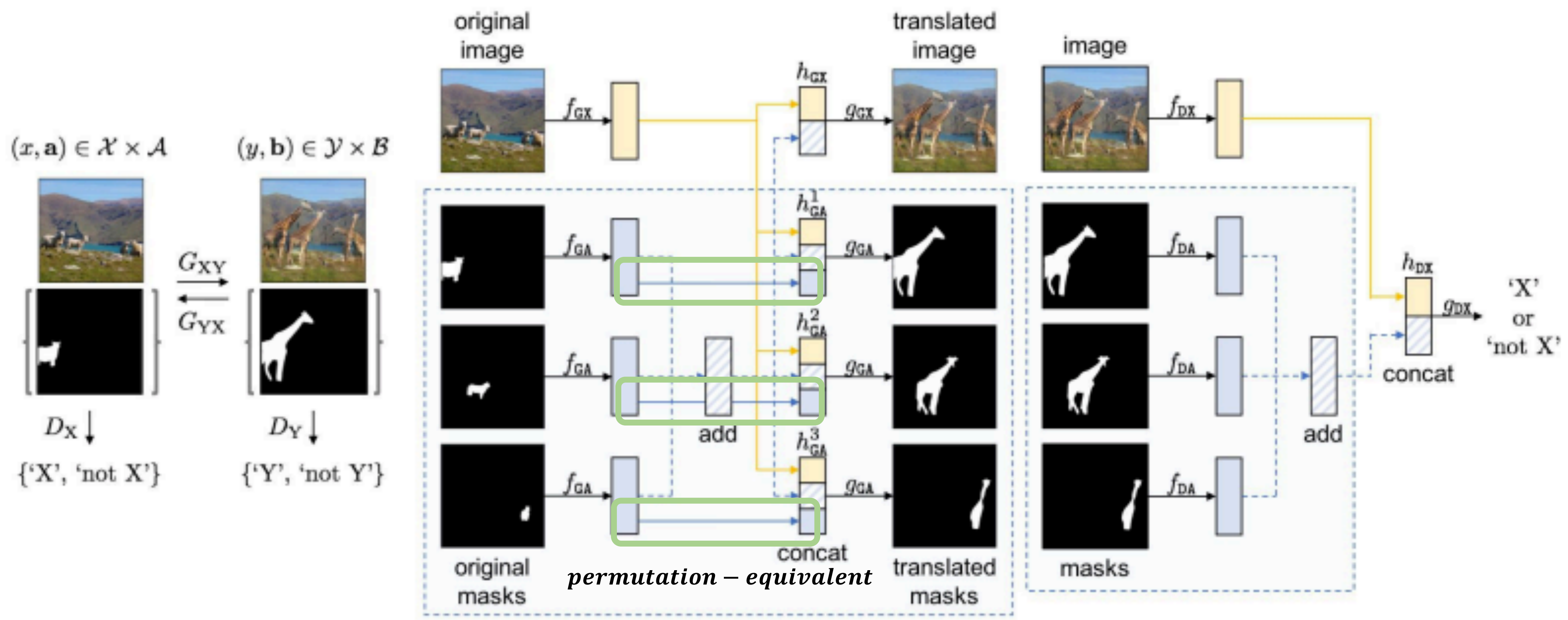
- Original setting (image to image)
  - $G_{XY}: X \rightarrow Y$ & $G_{YX}: Y \rightarrow X$
- These mappings can be reformulated as finding conditional distributions $p(y|x), p(x|y)$ when we have marginal distributions $p(x), p(y)$
- The authors argue that aforementioned information is insufficient to approximate conditional distributions when sampling one is too complex.
- Author's setting (image x attribute to image x attribute)
  - $G_{XY}: X \times A \rightarrow Y \times B$ & $G_{YX}: Y \times B \rightarrow X \times A$

Proposed methods (2/7)



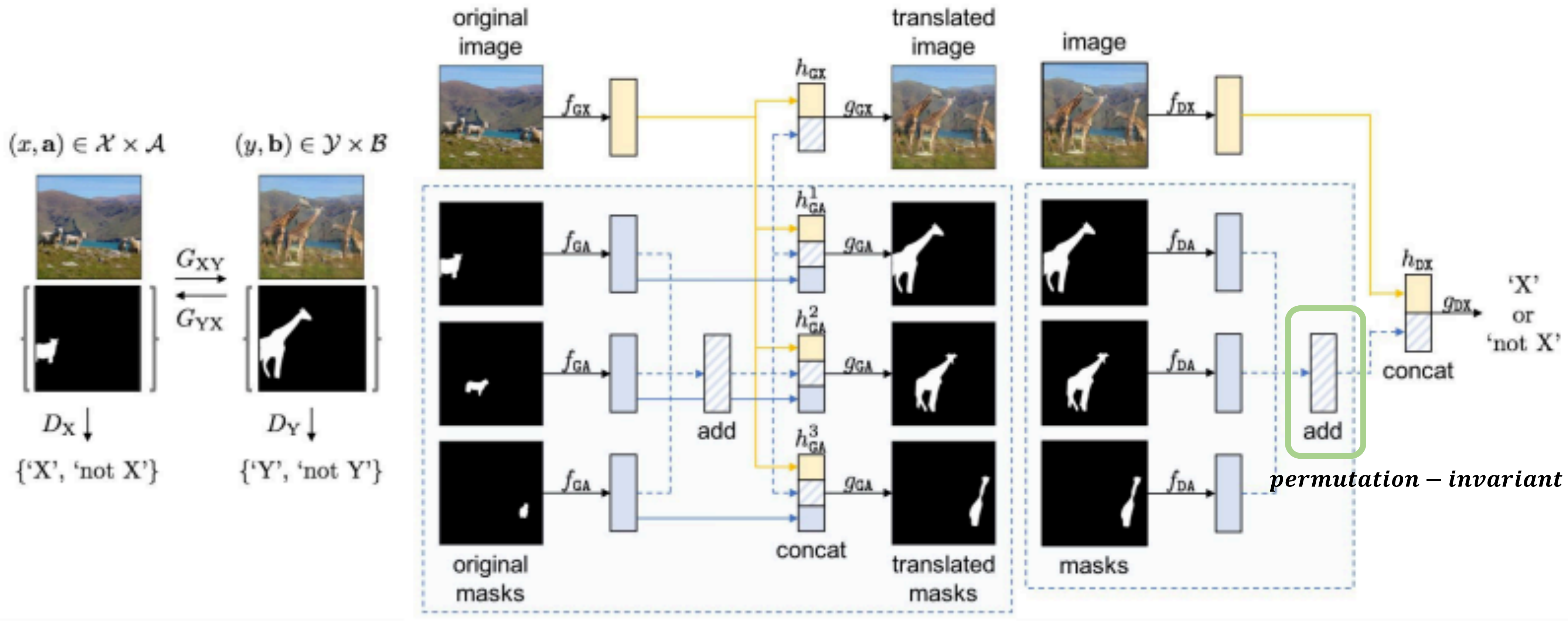
(a) Overview (b) Generator G (c) Discriminator D

Proposed methods (3/7)



(a) Overview (b) Generator G (c) Discriminator D

Proposed methods (4/7)



(a) Overview (b) Generator G (c) Discriminator D

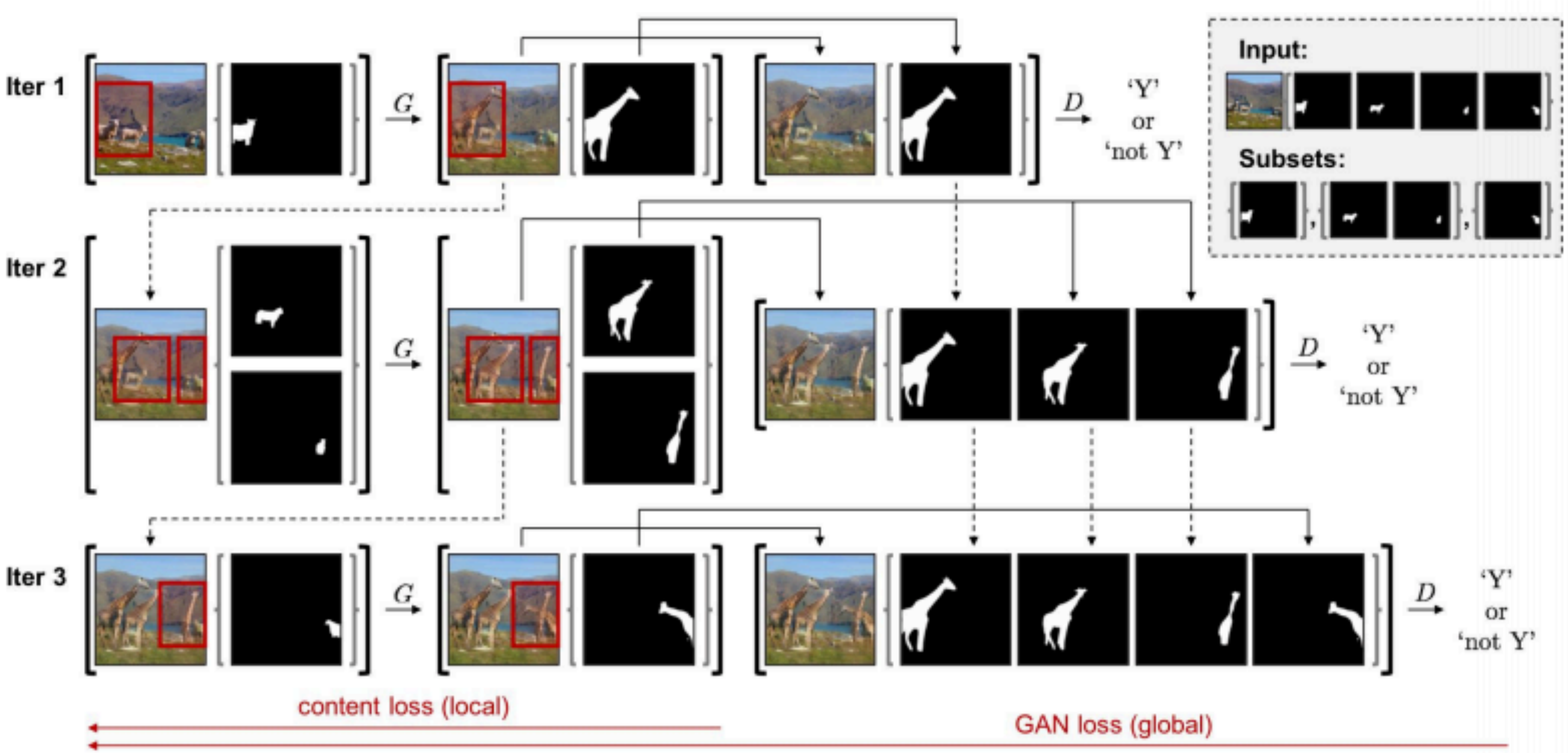
Proposed methods (5/7)

- **Domain loss which makes the generated outputs to follow the style of a target domain**
  - LSGAN loss

- **Content loss which makes the outputs to keep the original contents**
  - Cycle-consistency loss ($i.e. G_{YX}(G_{XY}(x,a))$ – (x,a))
  - Identity mapping loss ($i.e. G_{XY}(y,b)$ – (y,b))
  - Context preserving loss ($i.e. w(a,b') \circ (x-y') \Leftrightarrow (1-max(a,b') \circ (x-y')$)

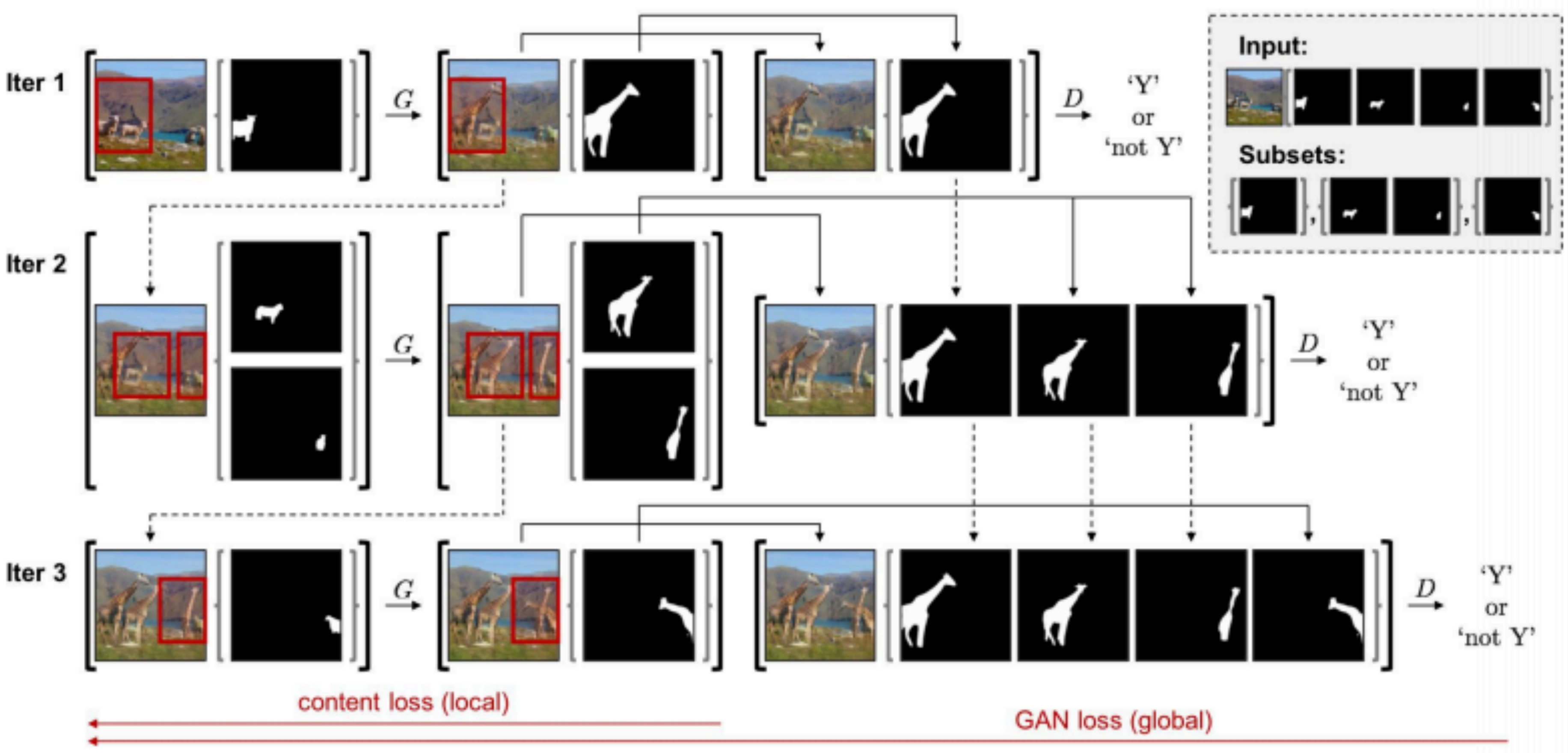
$$\mathcal{L}_{\texttt{InstaGAN}} = \underbrace{\mathcal{L}_{\texttt{LSGAN}}}_{\texttt{GAN (domain) loss}} + \underbrace{\lambda_{\texttt{cyc}}\mathcal{L}_{\texttt{cyc}} + \lambda_{\texttt{idt}}\mathcal{L}_{\texttt{idt}} + \lambda_{\texttt{ctx}}\mathcal{L}_{\texttt{ctx}}}_{\texttt{content loss}},$$

Proposed methods (6/7)



$$\mathcal{L}_{\texttt{InstaGAN-SM}} = \sum_{m=1}^{M} \mathcal{L}_{\texttt{LSGAN}}((x, \boldsymbol{a}), (y'_m, \boldsymbol{b}'_{1:m})) + \mathcal{L}_{\texttt{content}}((x_m, \boldsymbol{a}_m), (y'_m, \boldsymbol{b}'_m))$$

Proposed methods (7/7)

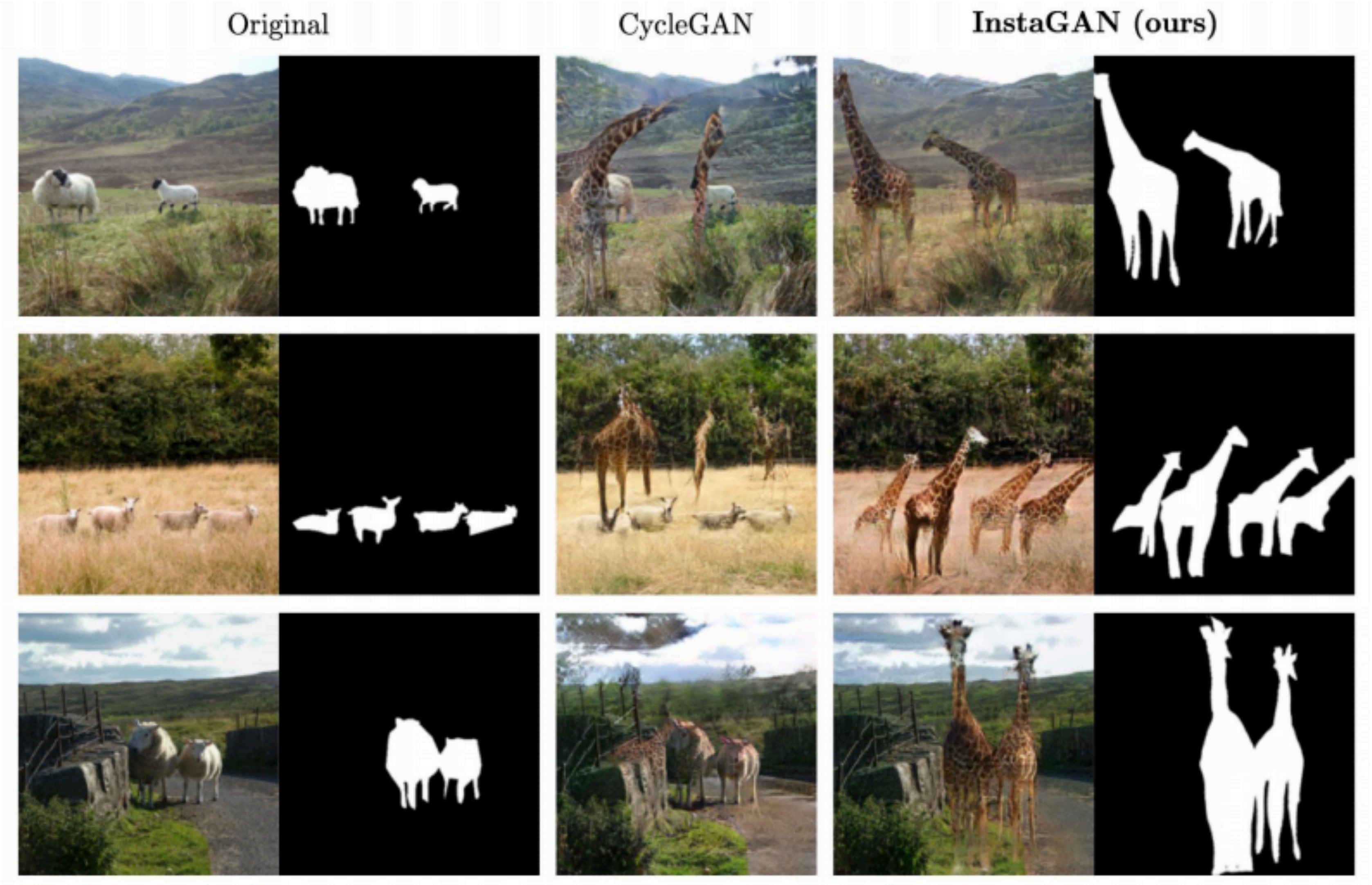


- Divided the instances into mini-batches a_1 … a_M according to the decreasing order of the spatial sizes of intances -> Better performance than random order

- Small instances tend to be occluded by other instances in image s, thus often losing their intrinsic shape information

Experiments (1/5)

COCO (sheep -> giraffe)

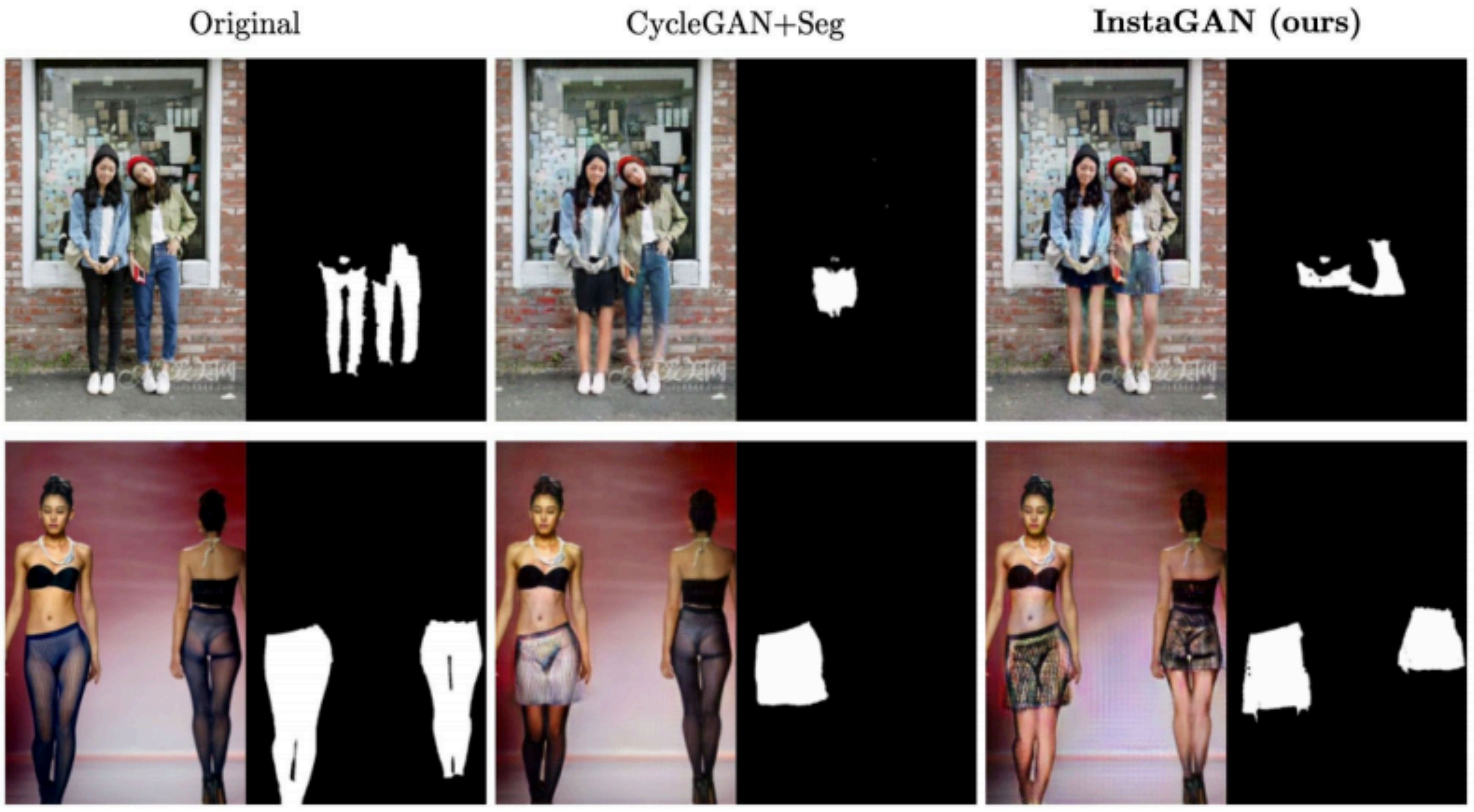


Experiments (2/5)

MHP (skirt ->pants)

Experiments (3/5)

MHP (pants ->skirt)



Experiments (4/5)

Ablation study

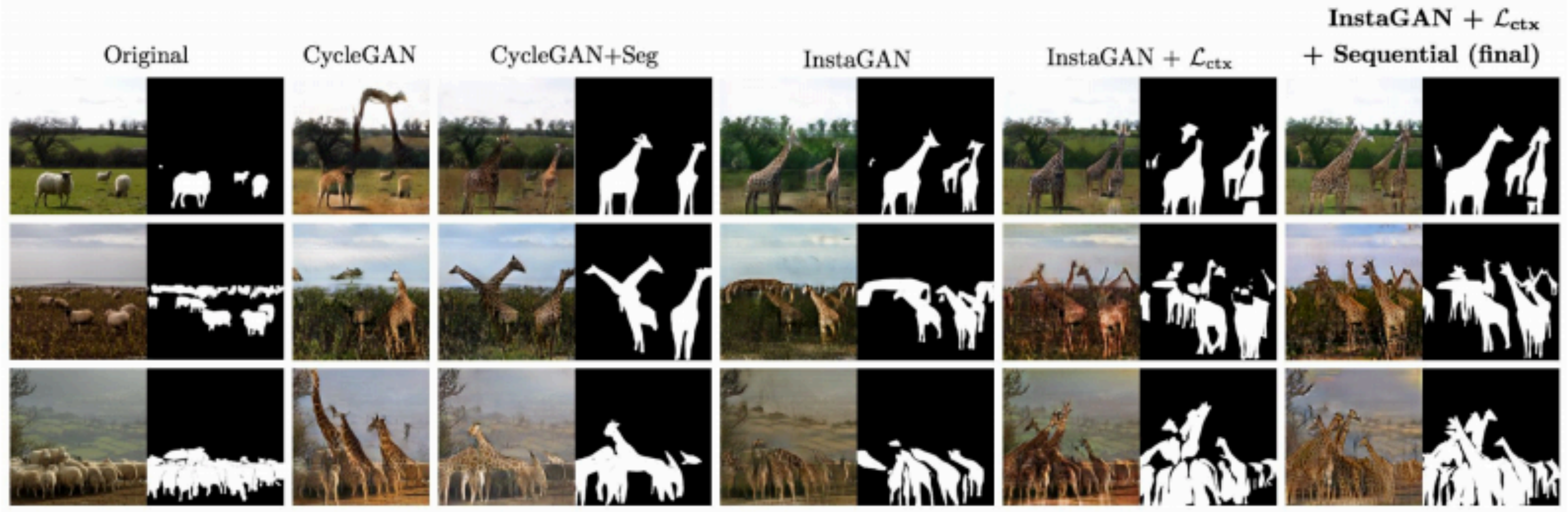


Figure 9: Ablation study on the effect of each component of our method: the InstaGAN architecture, the context preserving loss, and the sequential mini-batch inference/training algorithm, which are denoted as InstaGAN, $\mathcal{L}_{ctx}$, and Sequential, respectively.

Experiments (5/5)

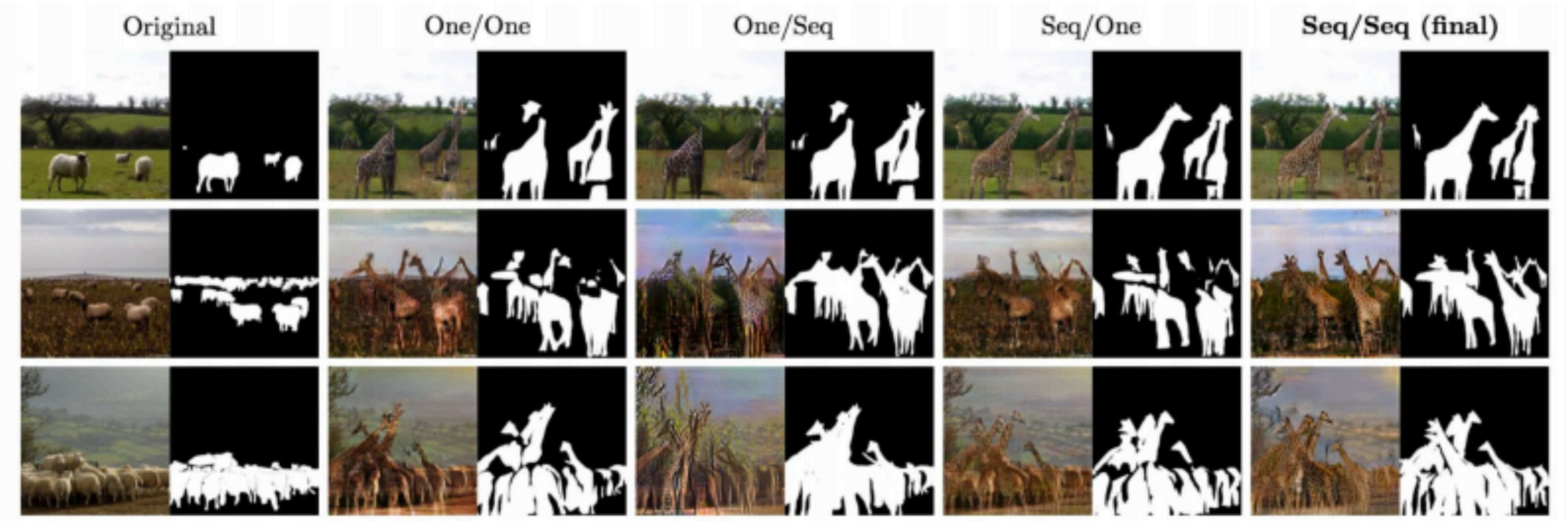


Figure 10: Ablation study on the effects of the sequential mini-batch inference/training technique. The left and right side of title indicates which method used for training and inference, respectively, where “One” and “Seq” indicate the one-step and sequential schemes, respectively.