Not just Compete, but Collaborate:
Local Image-to-Image Translation via Cooperative Mask Prediction

Daejin Kim[1]   Mohammad Azam Khan[2]   Jaegul Choo[1]

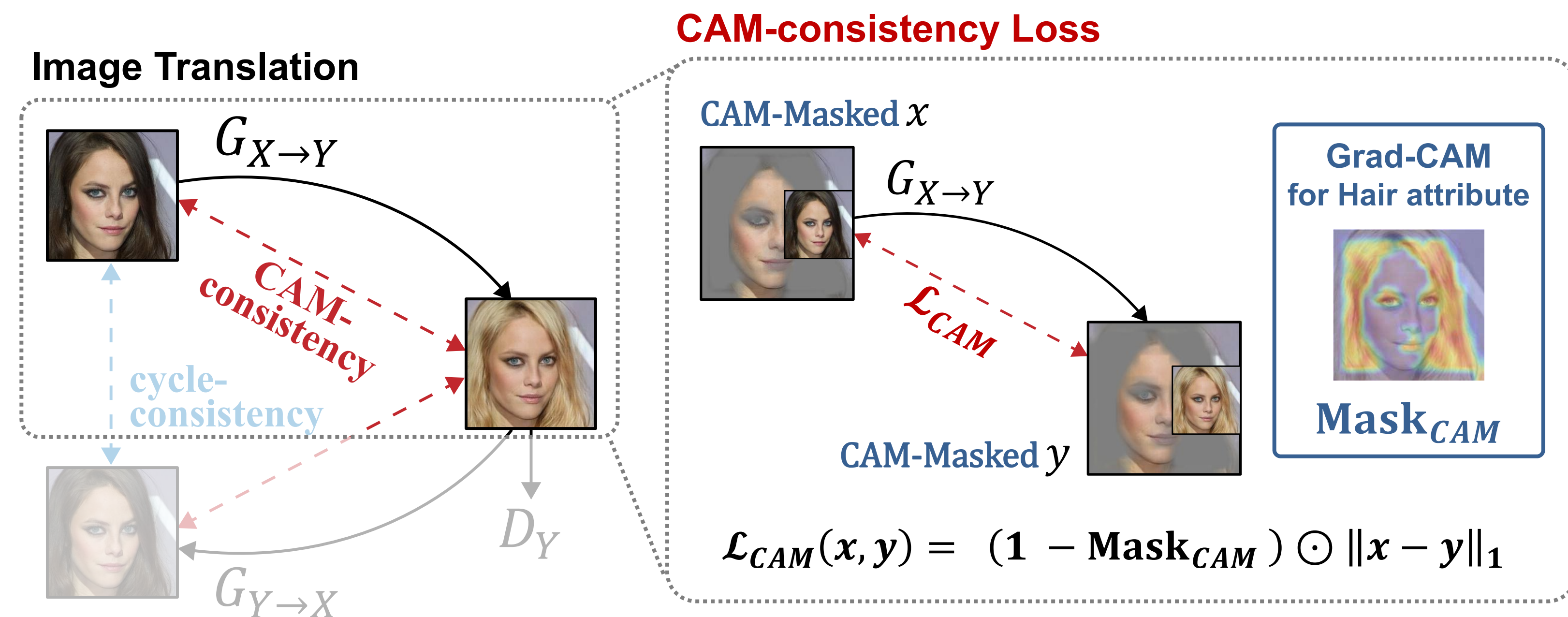[1] KAIST   [2] Dhaka Power Distribution Company Ltd.

CVPR VIRTUAL JUNE 19-25

Problems

Existing facial editing methods cannot well preserve the attribute-irrelevant regions due to the absence of GT of translated images.

Proposed Method

Image Translation

CAM-consistency Loss

$$\mathcal{L}_{CAM}(x, y) = (1 - \text{Mask}_{CAM}) \odot \|x - y\|_1$$

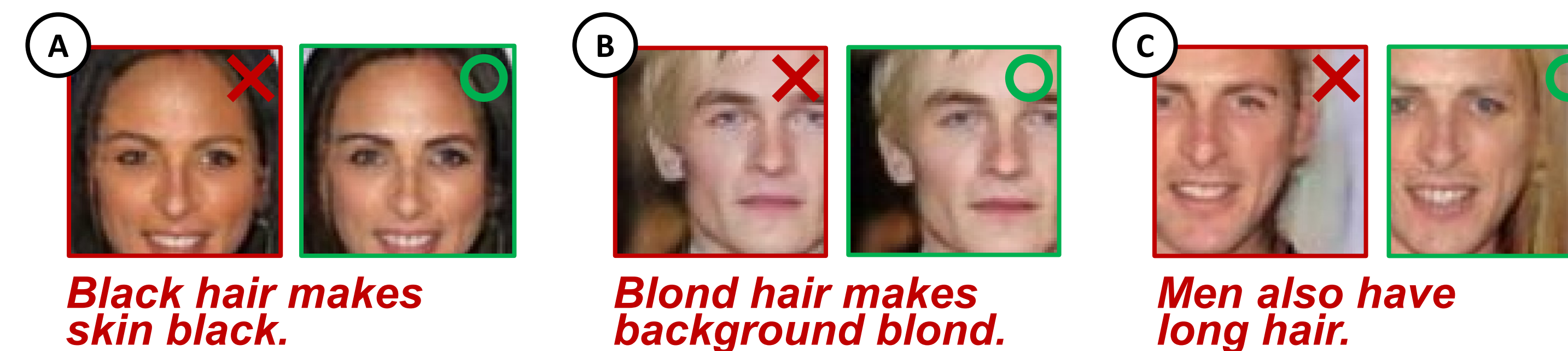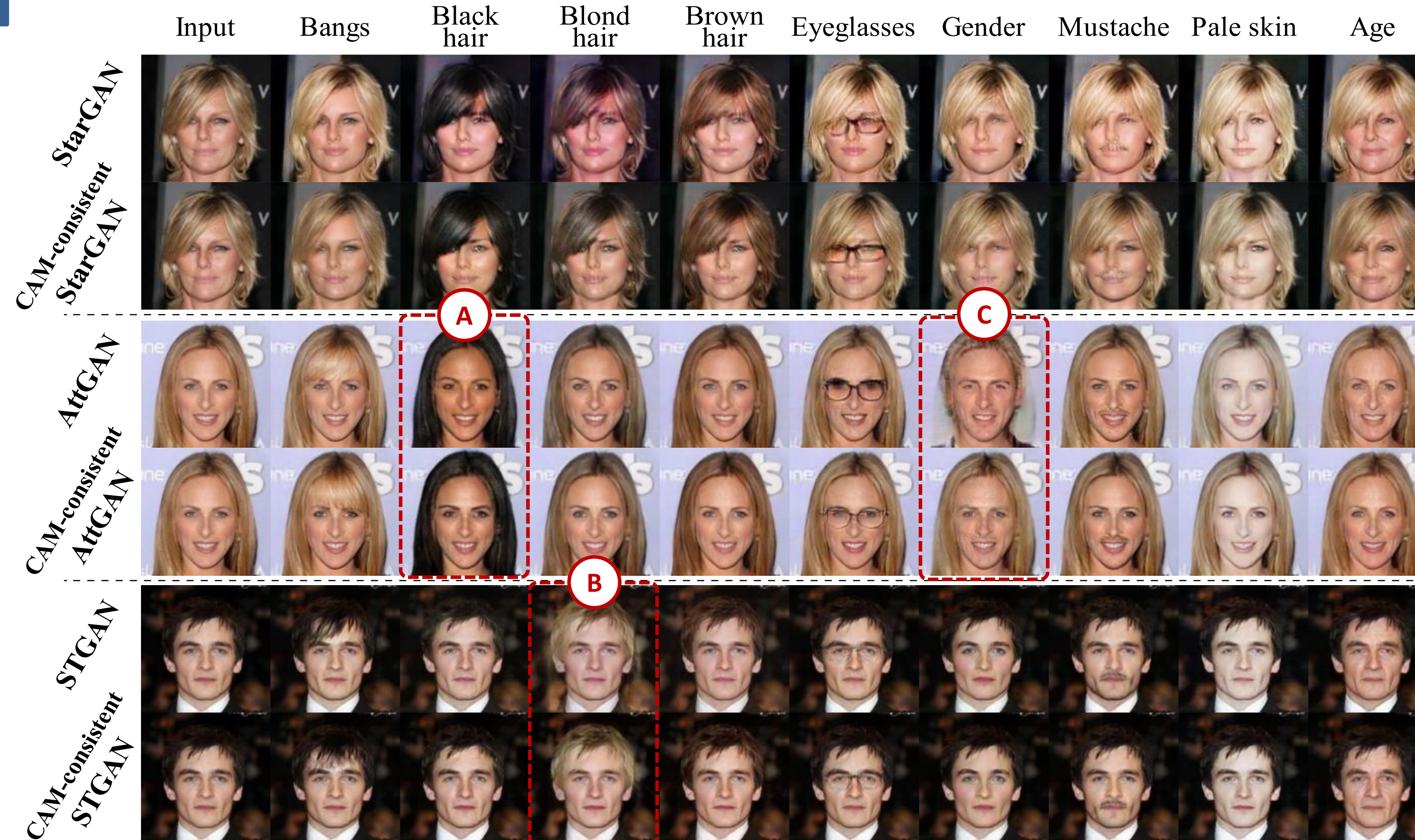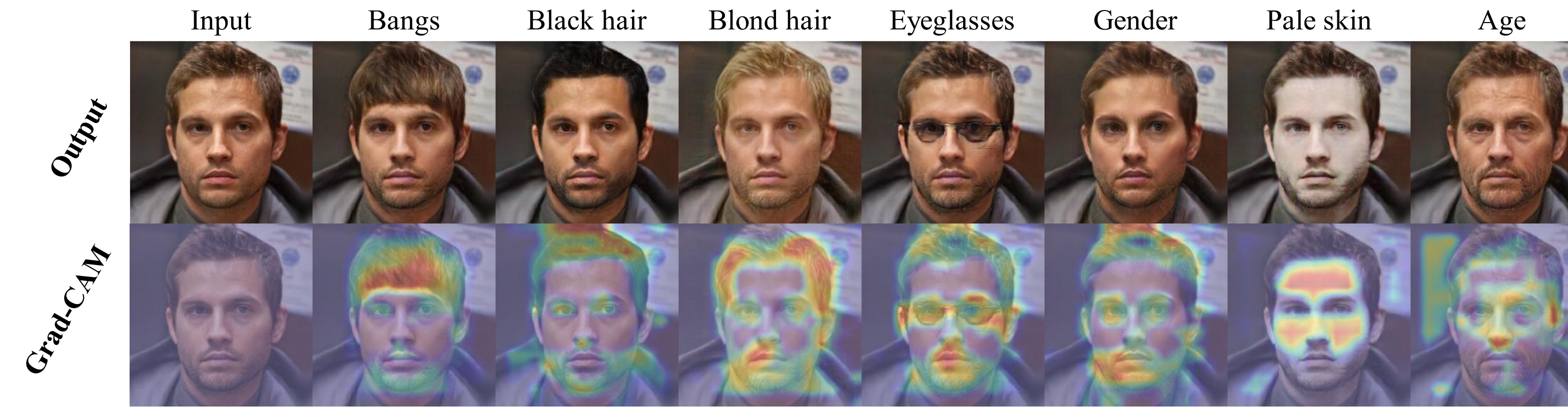Grad-CAM for Hair attribute — Mask_{CAM}

Unlike cycle-consistency loss [1], our **CAM-consistency loss** can directly preserve the attribute-irrelevant regions with the translated images via **Grad-CAM Mask.**

The generator G is trained to preserve the regions except where the discriminator D attends for the attributes. **The discriminator D** is trained to attend the regions (by *Grad-CAM* [2]) where the generator G makes the changes.
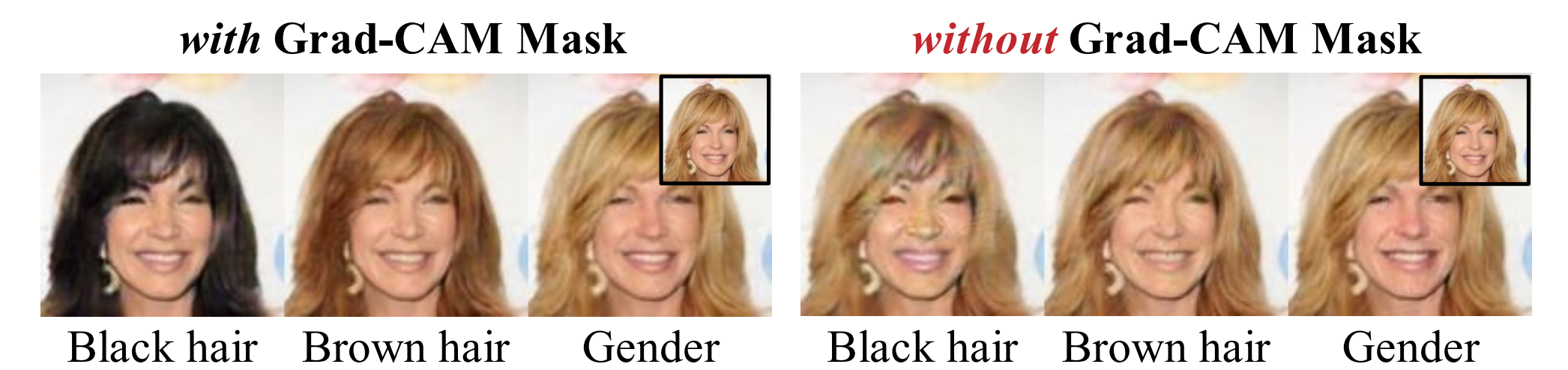
Collaborate with CAM-consistency Loss: Good G makes D to attend the correct regions while good D makes G to manipulate only the attribute-relevant regions.

[1] Zhu et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.", *ICCV*, 2017

[2] Selvaraju et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.", *ICCV*, 2017.

[3] Kim et al. "U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation.", *ICLR*, 2020.

Our Contributions

- Our **CAM-consistency loss** allows the generator and the discriminator to collaborate while improving the existing GANs to better preserve the attribute-irrelevant regions.

- We show the possibility of using the *Grad-CAM* [2] as the **trainable** objective.

Experimental Results

Black hair makes skin black. *Blond hair makes background blond.* *Men also have long hair.*

Effectiveness of Grad-CAM Mask

Grad-CAM Mask plays a critical role in manipulation.

Without Grad-CAM Mask, only *L1-loss* (indicated as *identity loss* in [3]) cannot makes manipulation correctly.

Ablation Study