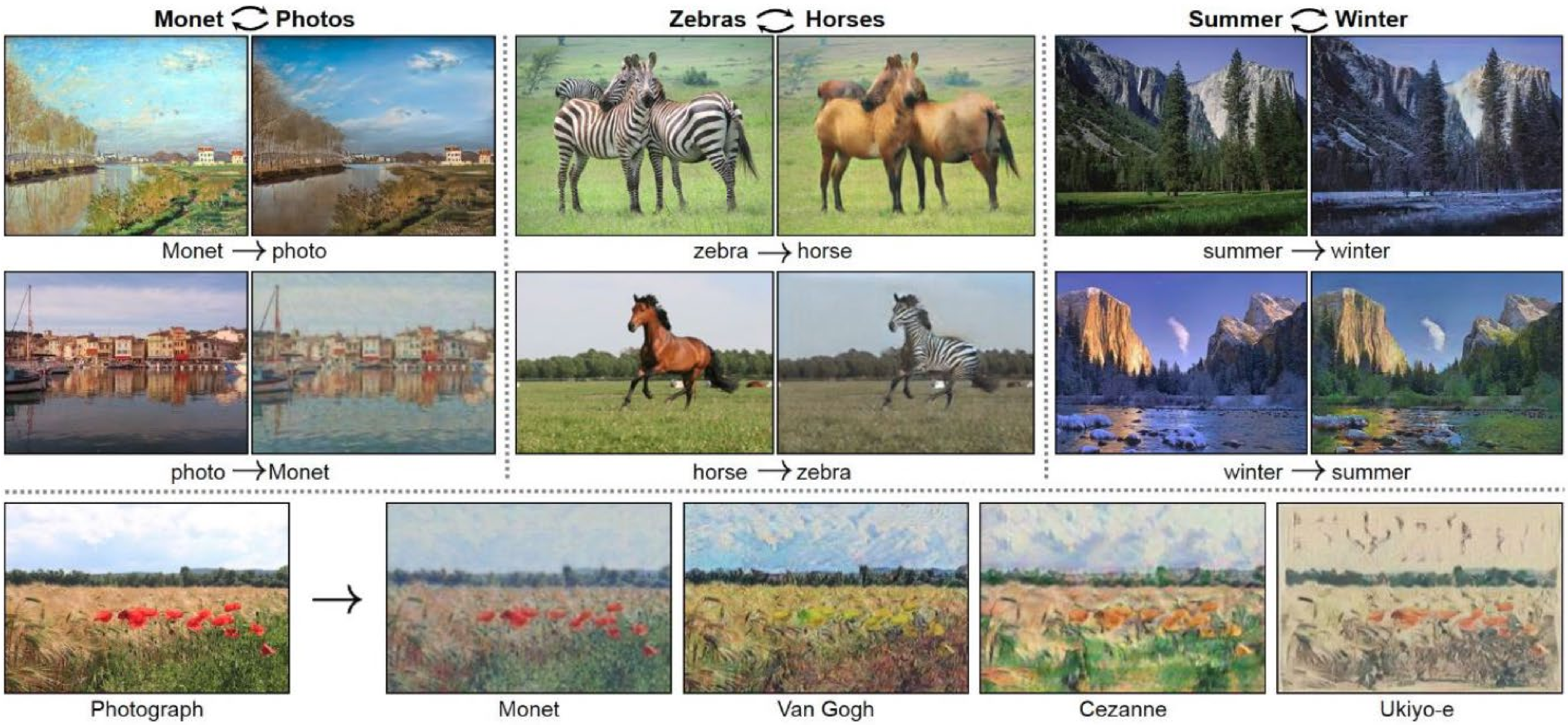
Domain Intersection and Domain Difference

Sagie Benaim[1], Michael Khaitov[1], Tomer Galanti[1], and Lior Wolf[1,2]

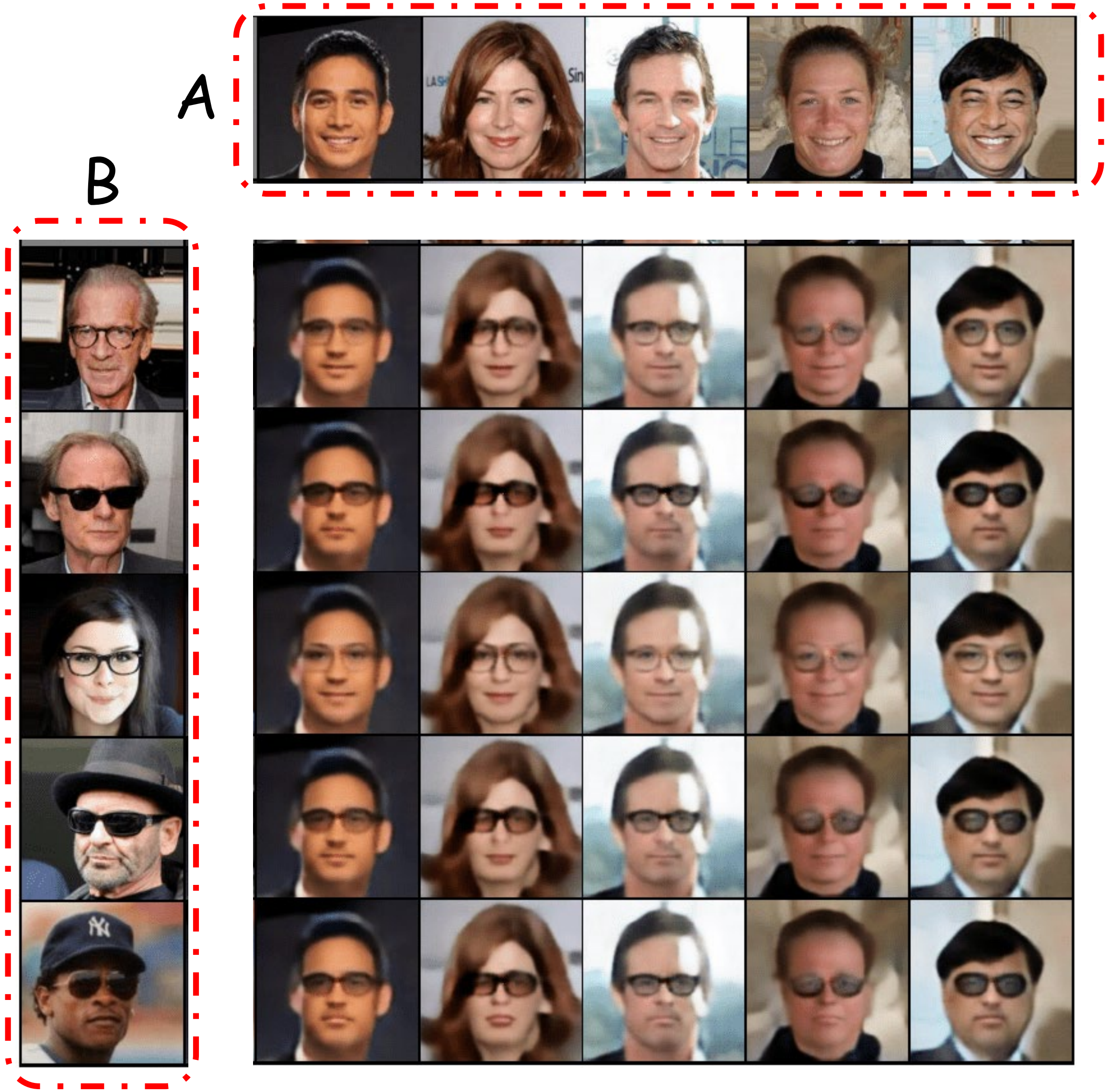
[1]School of Computer Science, Tel Aviv University
[2]Facebook AI Research

ICCV 2019

Image to Image Translation



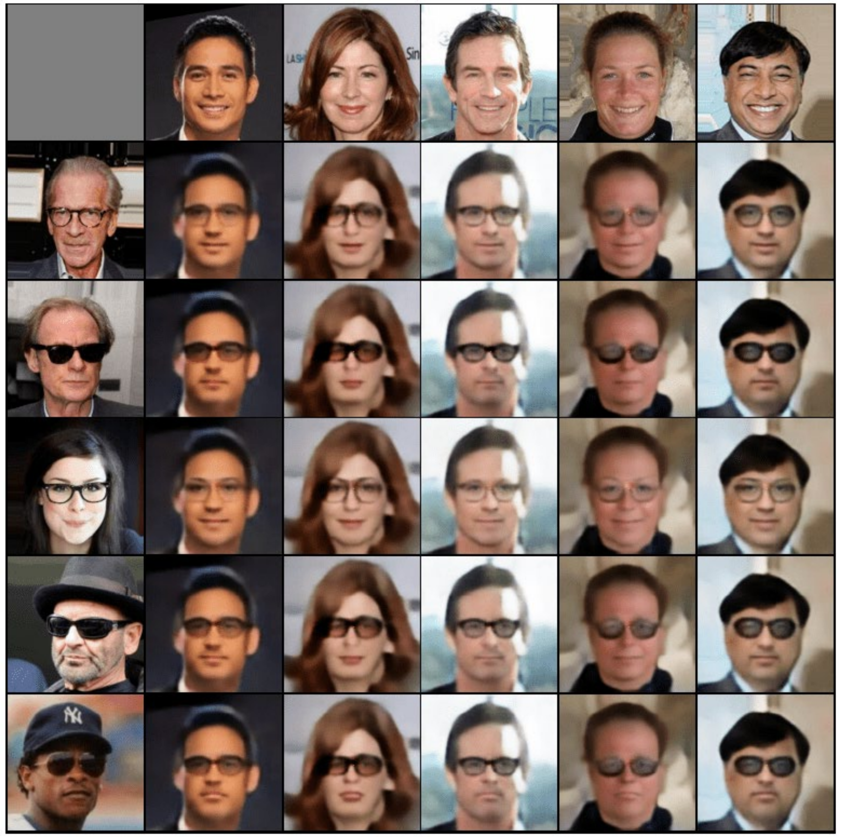
A Running Example



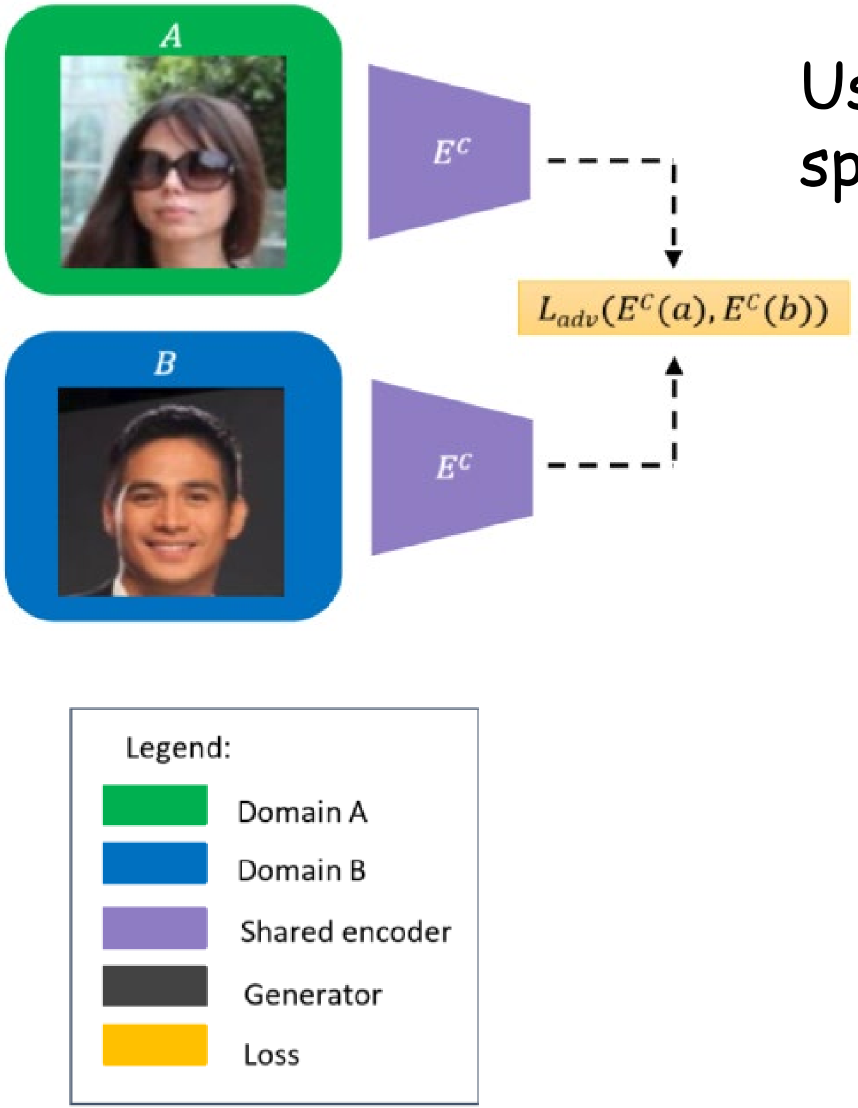
Domain Intersection and Domain Difference

The paper:

- Disentangles the Common (Intersection) and Specific (Difference) parts of each domain.
- Utilize a Encoder-Decoder architecture:
  - Three encoders E: produces three latent spaces:
    - $\mathrm{E}^c(\mathrm{A}) = \mathrm{E}^c(\mathrm{B})$: Common to A and B
    - $\mathrm{E}^s_{\mathrm{A}}(\mathrm{A})$ (or $\mathrm{E}^s_{\mathrm{B}}(\mathrm{B})$): Specific to A (respectively to B)
  - Decoder G: generate images according codes.



The "Common" (or shared) Loss



Use an adversarial loss to do not encode the specific part of each of two domains

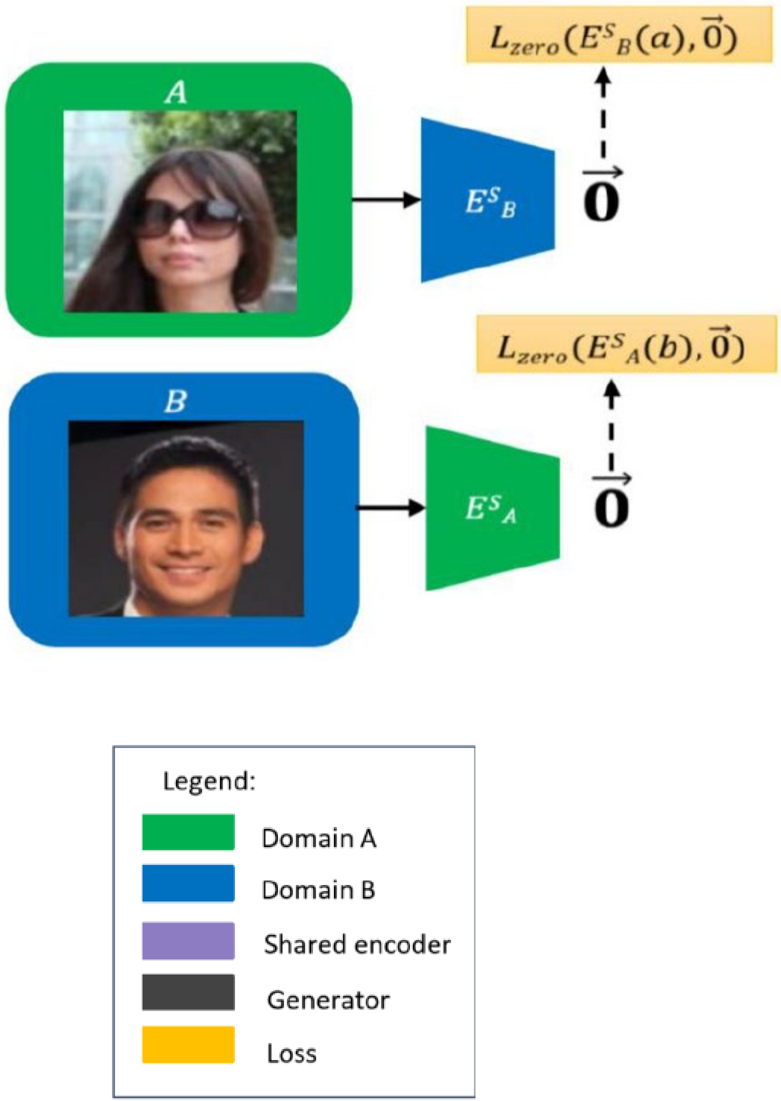
For discriminator d:

$$\mathcal{L}_d := \frac{1}{m_1}\sum_{i=1}^{m_1} l(d(E^c(a_i)), 0) + \frac{1}{m_2}\sum_{j=1}^{m_2} l(d(E^c(b_j)), 1)$$

For encoder E^c:

$$\frac{1}{m_1}\sum_{i=1}^{m_1} l(d(E^c(a_i)), 1) + \frac{1}{m_2}\sum_{j=1}^{m_2} l(d(E^c(b_j)), 1)$$

'Zero' Loss



Encourage $E_A^S(E_B^S)$ to do not encode the specific part of domain B(A) :

For A domain:

$$\mathcal{L}_{zero}^{A} := \frac{1}{m_2}\sum_{j=1}^{m_2} \|E_A^s(b_j)\|_1$$

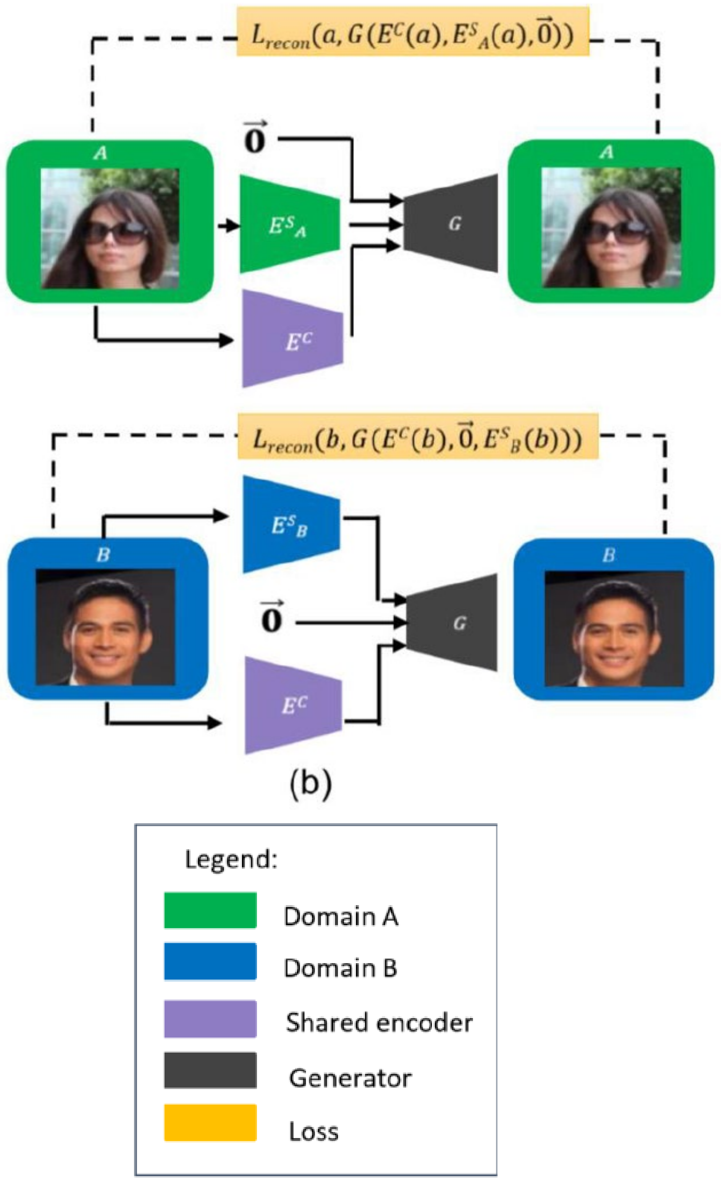
For B domain:

$$\mathcal{L}_{zero}^{B} := \frac{1}{m_1}\sum_{i=1}^{m_1} \|E_B^s(a_i)\|_1$$

Final loss:

$$\mathcal{L}_{zero} := \mathcal{L}_{zero}^{A} + \mathcal{L}_{zero}^{B}$$

Reconstruction Loss



(b)

The loss mentioned above ensure that no encoder encode more information .

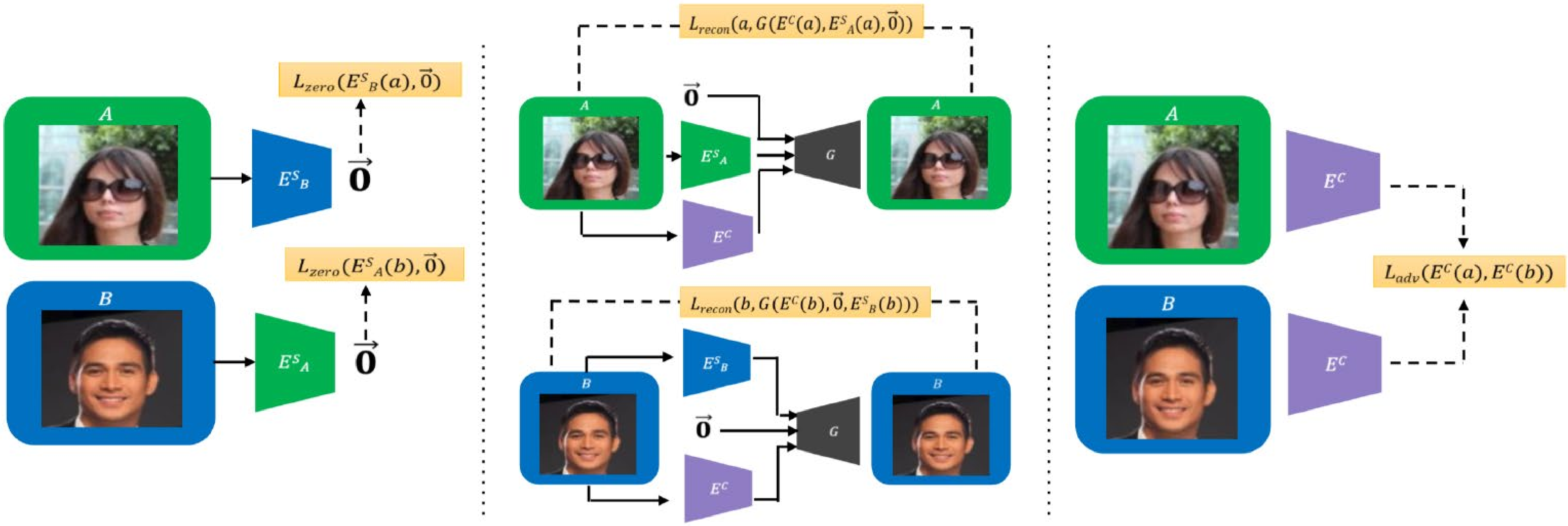
We need to also ensure that all the needed information is encoded.

$$\mathcal{L}^A_{recon} := \frac{1}{m_1}\sum_{i=1}^{m_1} \|G(E^c(a_i), E^s_A(a_i), 0) - a_i\|_1$$

$$\mathcal{L}^B_{recon} := \frac{1}{m_2}\sum_{j=1}^{m_2} \|G(E^c(b_i), 0, E^s_B(b_j)) - b_j\|_1$$

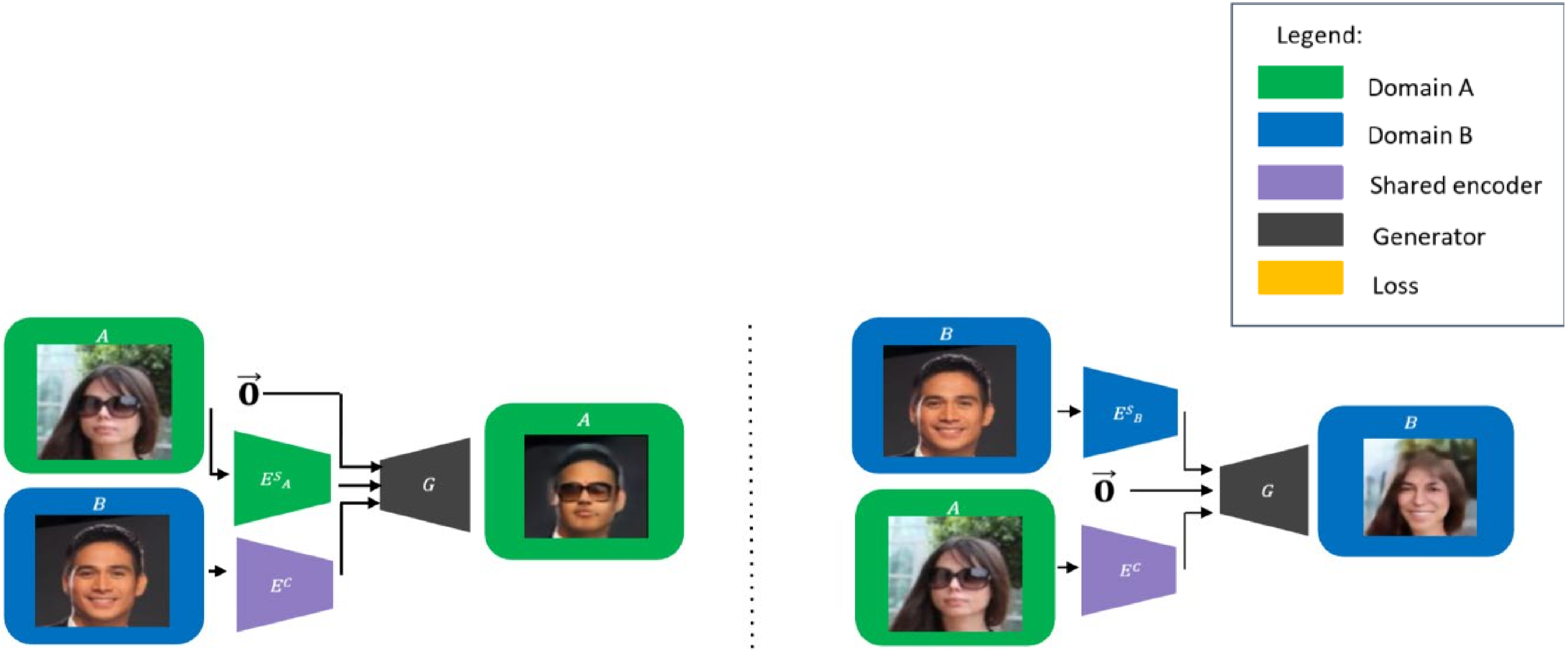
$$\mathcal{L}_{recon} := \mathcal{L}^A_{recon} + \mathcal{L}^B_{recon}$$

Training



A
B
$E^S{}_B$
$E^S{}_A$
$\vec{0}$
$L_{zero}(E^S{}_B(a), \vec{0})$
$L_{zero}(E^S{}_A(b), \vec{0})$
$L_{recon}(a, G(E^C(a), E^S{}_A(a), \vec{0}))$
$L_{recon}(b, G(E^C(b), \vec{0}, E^S{}_B(b)))$
E^C
G
$L_{adv}(E^C(a), E^C(b))$

Inference


Legend:
Domain A
Domain B
Shared encoder
Generator
Loss
A
B
$\vec{0}$
$E^S{}_A$
E^C
G
A
B
$E^S{}_B$
$\vec{0}$
A
E^C
G
B

Experiment

	Smile To Glasses	Glasses To Smile	Facial Hair To Smile	Smile To Facial Hair	Facial Hair To Glasses	Glasses To Facial Hair
Fader networks [15]	76.8%	97.3%	95.4%	84.2%	77.8 %	85.2%
Guided content transfer [20]	45.8%	92.7%	85.6%	85.1%	38.6%	82.2%
MUNIT [12]	7.3%	9.2%	9.3%	8.4%	7.3%	8.5%
DRIT [16]	8.5%	6.3%	6.3%	10.3%	8.6%	10.1%
Ours	91.8%	99.3%	93.7%	87.1%	93.1%	97.2%

Table 1. We pretrain a classifier to distinguish between samples in A (e.g. images of persons with glasses) and samples in B (e.g. images of persons with smile). We then sample $a \in A$, $b \in B$ from the test samples and check the membership of the generated image $G(E^c(b), E^s_A(a), 0))$ in A. Similarly, in the reverse direction, we check the membership of $G(E^c(a), 0, E^s_B(b))$ in B.

	Smile To Glasses	Glasses To Smile	Facial Hair To Smile	Smile To Facial Hair	Facial Hair To Glasses	Glasses To Facial Hair
Question (1) ours	4.74 ±0.13	4.30 ±0.21	4.26 ±0.20	4.30 ±0.15	4.18 ±0.17	4.50 ±0.18
Question (2) ours	3.92 ±0.16	4.45 ±0.12	4.03 ±0.15	3.34 ±0.17	3.85 ±0.20	3.95 ±0.22
Question (3) ours	3.95 ±0.23	3.20 ±0.24	3.24 ±0.25	3.22 ±0.27	3.49 ±0.22	3.39 ±0.23
Question (1) for [20]	3.67 ±0.17	4.16 ±0.18	3.39 ±0.19	3.34 ±0.13	4.24 ±0.12	3.15 ±0.15
Question (2) for [20]	1.87 ±0.35	4.42 ±0.22	3.00 ±0.32	2.67 ±0.33	2.20 ±0.42	3.30 ±0.22
Question (3) for [20]	3.95 ±0.15	2.93 ±0.22	3.37 ±0.25	3.40 ±0.27	3.43 ±0.28	3.75 ±0.20

Table 2. Given 20 randomly selected images $a \in A$ and $b \in B$, we consider the generated image $G(E^c(a), 0, E^s_B(b)))$ and ask if (1) a's separate part is removed (2) b's separate part is added (3) a's common part is preserved (similarly in the reverse direction). Mean opinion scores in the range of 1 to 5 are reported, where higher is better.

Experiment

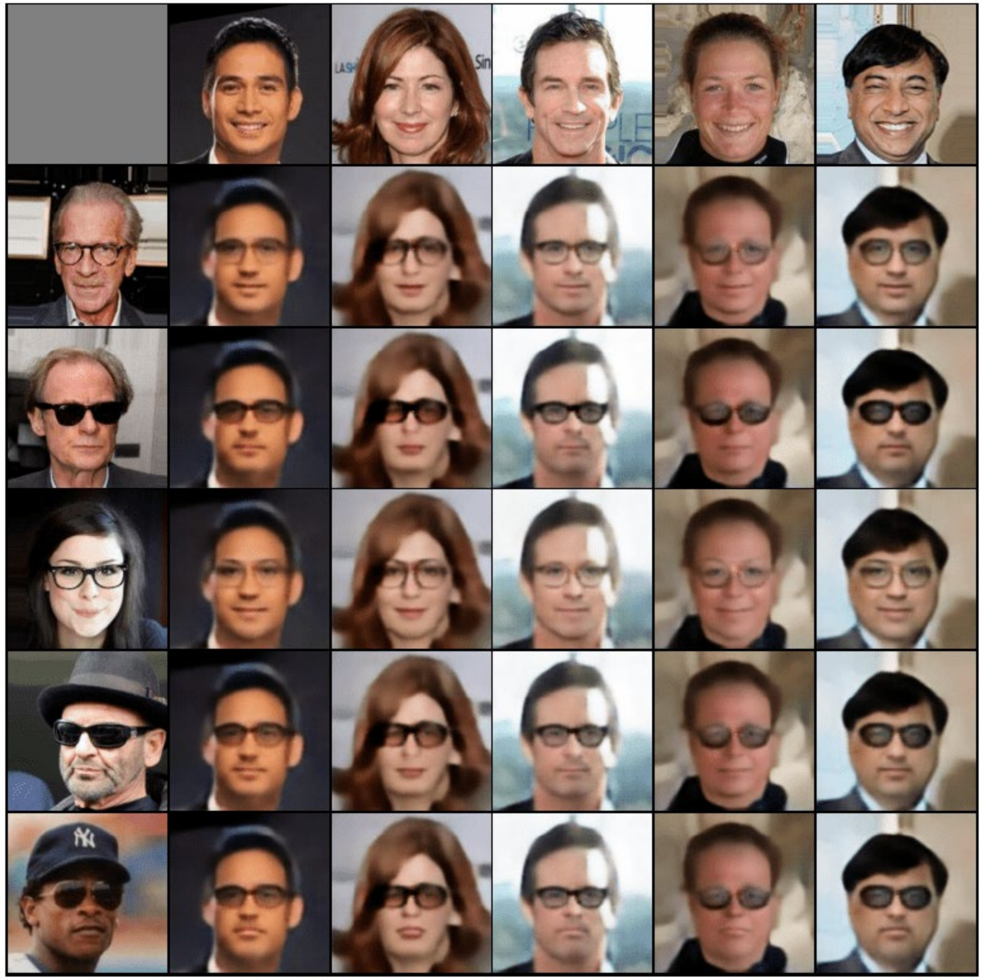


Figure 2. Images $a \in A$ are in the top row and $b \in B$ in the left column. The images constructed are $G(E^c(a), 0, E^s_B(b)))$, consisting of the common parts of a and separate part of b (smile is removed and glasses added).

Thanks