Computer Vision and Deep Learning

Lecture 13

Final exam

CVDL (coggle.it)

EEML summer school

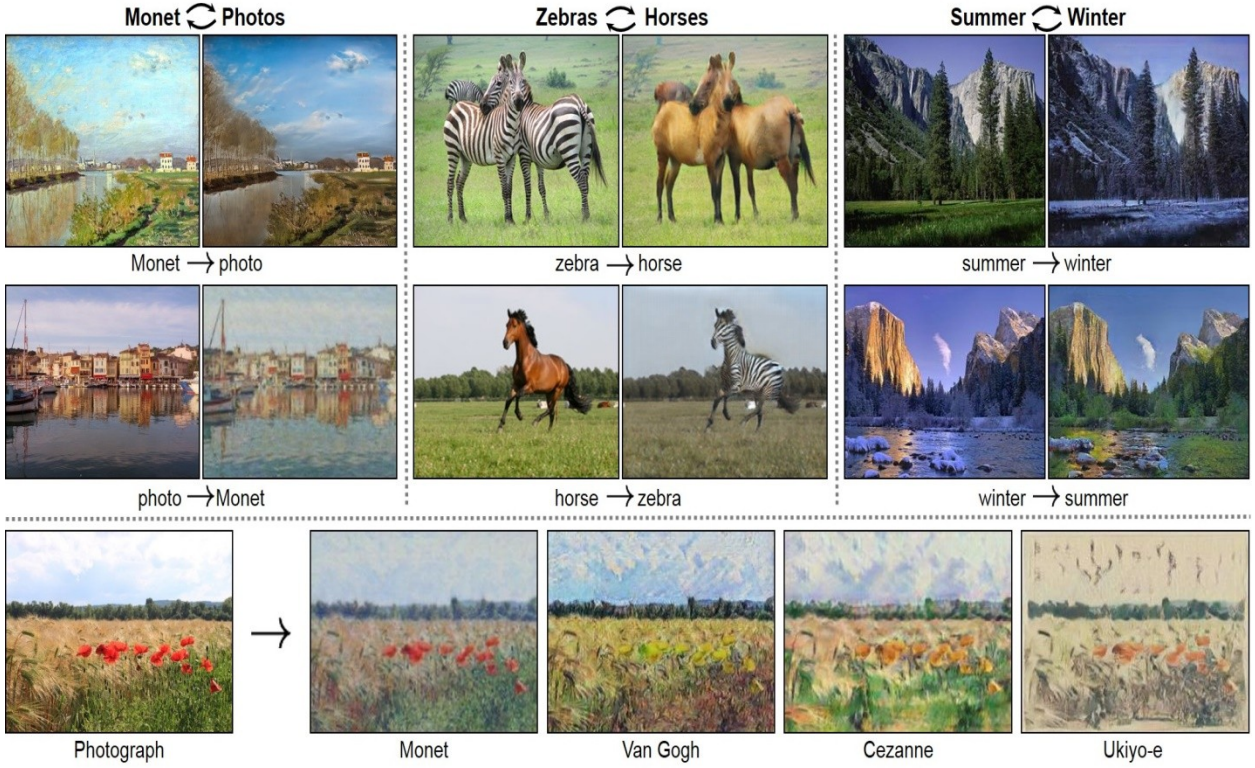
https://www.eeml.eu/home

Next time: virtual conference

Don’t forget to finish your teaser videos by 10th of Jan here

Cycle GAN

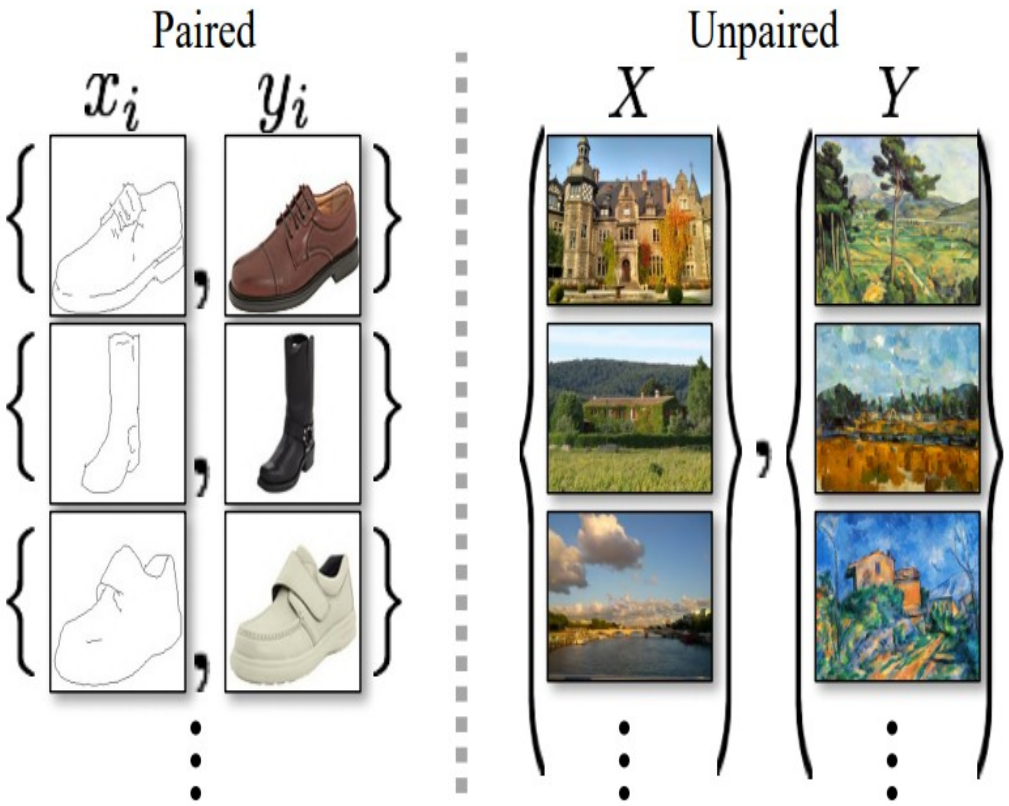
https://www.youtube.com/watch?v=D4C1dB9UheQ&ab_channel=TwoMinutePapers



Cycle GAN

Goal: capturing special characteristics of one image collection and figuring out how these characteristics could be translated into the other image collection, all in the *absence of any paired training example*

supervision is exploited at the level of sets: we are given one set of images in domain X and a different set in domain Y



Original text:

There was a feller here once by the name of Jim Smiley, in the winter of '49 or may be it was the spring of '50 I don't recollect exactly, somehow, though what makes me think it was one or the other is because I remember the big flume wasn't finished when he first came to the camp; but any way, he was the curiosest man about always betting on any thing that turned up you ever see, if he could get any body to bet on the other side; and if he couldn't, he'd change sides. Any way that suited the other man would suit him any way just so's he got a bet, he was satisfied. But still he was lucky, uncommon lucky; he most always come out winner.

Text translated into French:

Il y avait une fois ici un individu connu sous le nom de Jim Smiley: c'était dans l'hiver de 49, peut-être bien au printemps de 50, je ne me rappelle pas exactement. Ce qui me fait croire que c'était l'un ou l'autre, c'est que je me souviens que le grand bief n'était pas achevé lorsqu'il arriva au camp pour la première fois, mais de toutes façons il était l'homme le plus friand de paris qui se put voir, pariant sur tout ce qui se présentait, quand il pouvait trouver un adversaire, et, quand il n'en trouvait pas, il passait du côté opposé. Tout ce qui convenait à l'autre lui convenait; [illegible]fait. Et il avait une chance! une chance inouïe: presque toujours il gagnait.

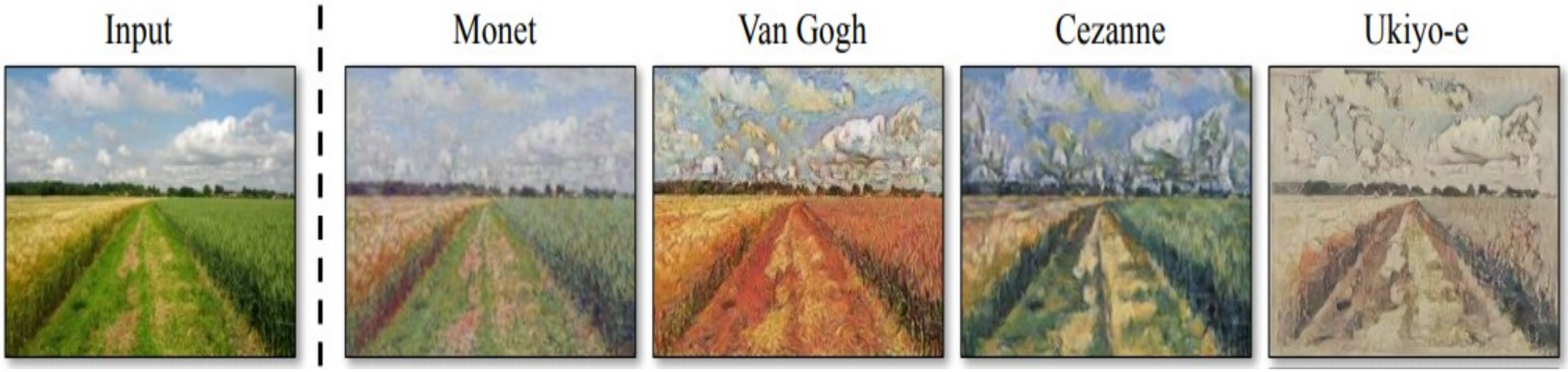
Text re-translated into English by Mark Twain:

It there was one time here an individual known under the name of Jim Smiley; it was in the winter of '49, possibly well at the spring of '50, I no me recollect not exactly. This which me makes to believe that it was the one or the other, it is that I shall remember that the grand flume is not achieved when he arrives at the camp for the first time, but of all sides he was the man the most fond of to bet which one have seen, betting upon all that which is presented, when he could find an adversary; and when he not of it could not, he passed to the side opposed. All that which convenienced to the other to him convenienced also; seeing that he had a bet Smiley was satisfied. And he had a chance! a chance even worthless; nearly always he gained.

The Jumping Frog: in English, then in French, and then Clawed Back into a Civilized Language Once More by Patient, Unremunerated Toil,
Mark Twain

Cycle GAN

Unpaired image to image translation

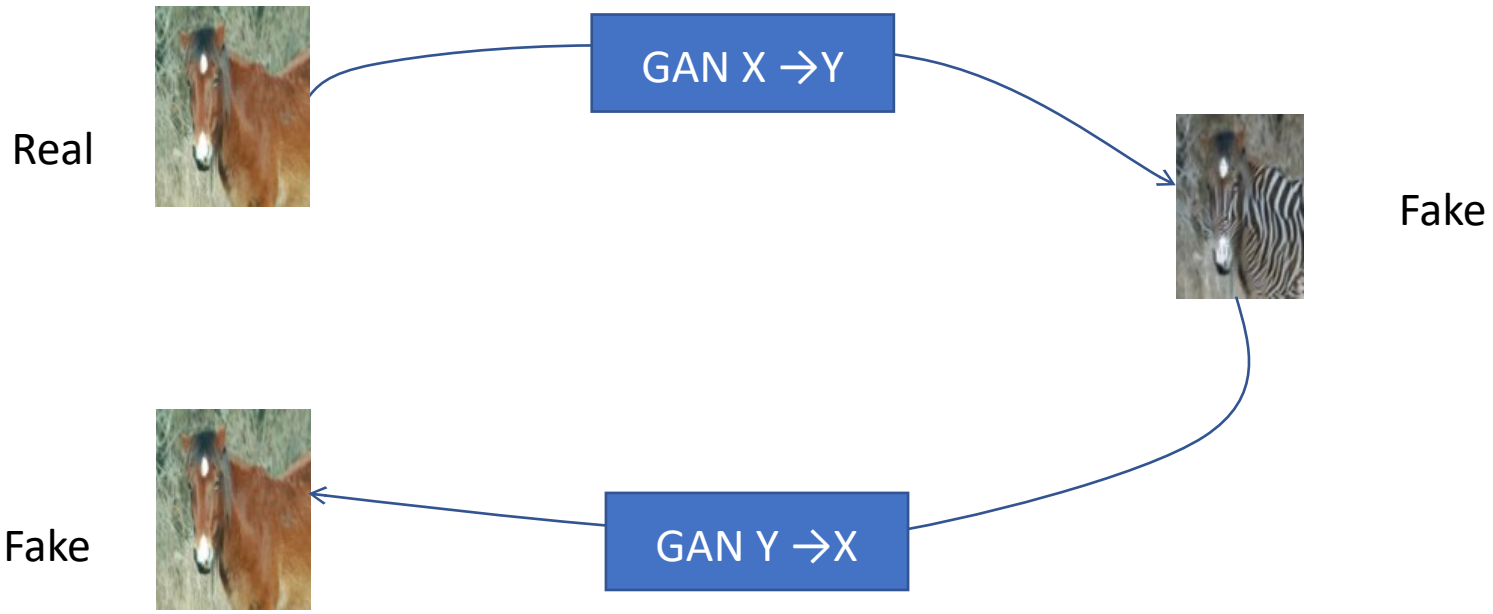


- Content - common elements, must be kept in the generated image
- Style – unique elements, must be transferred to the generated image

Cycle GAN

Unpaired image to image translation

- Use two GANs
- Cycle consistency: getting the content to be preserved while only changing the styles



Cycle GAN

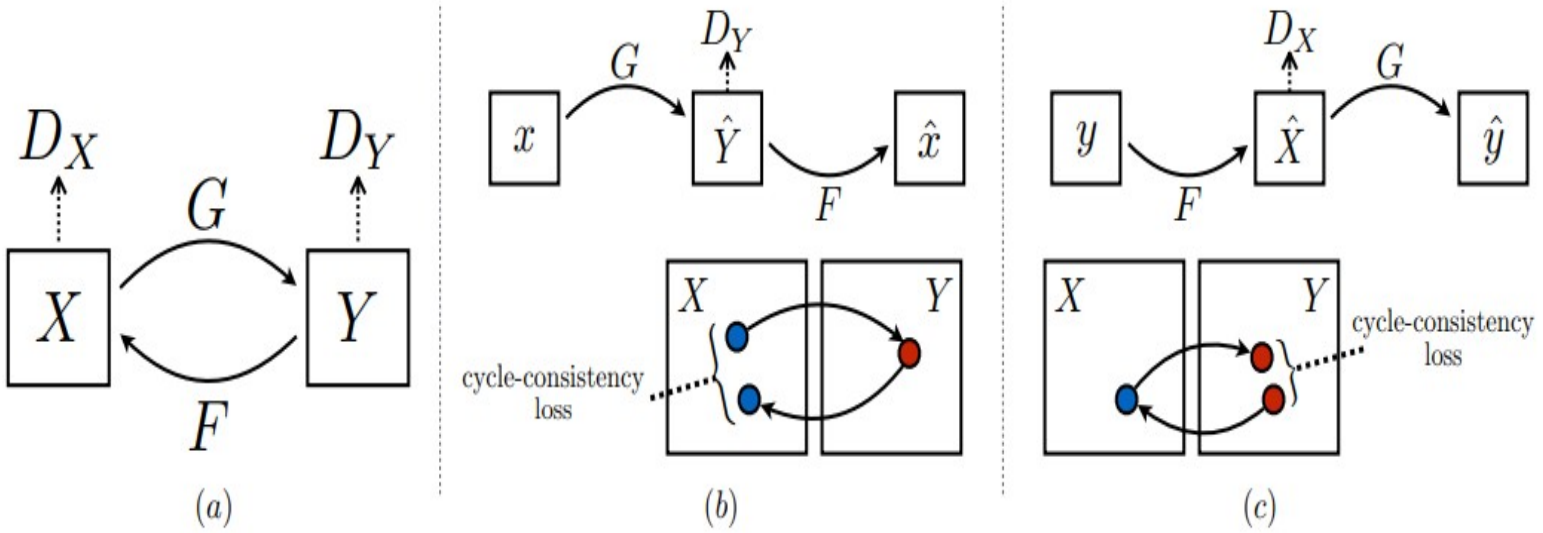
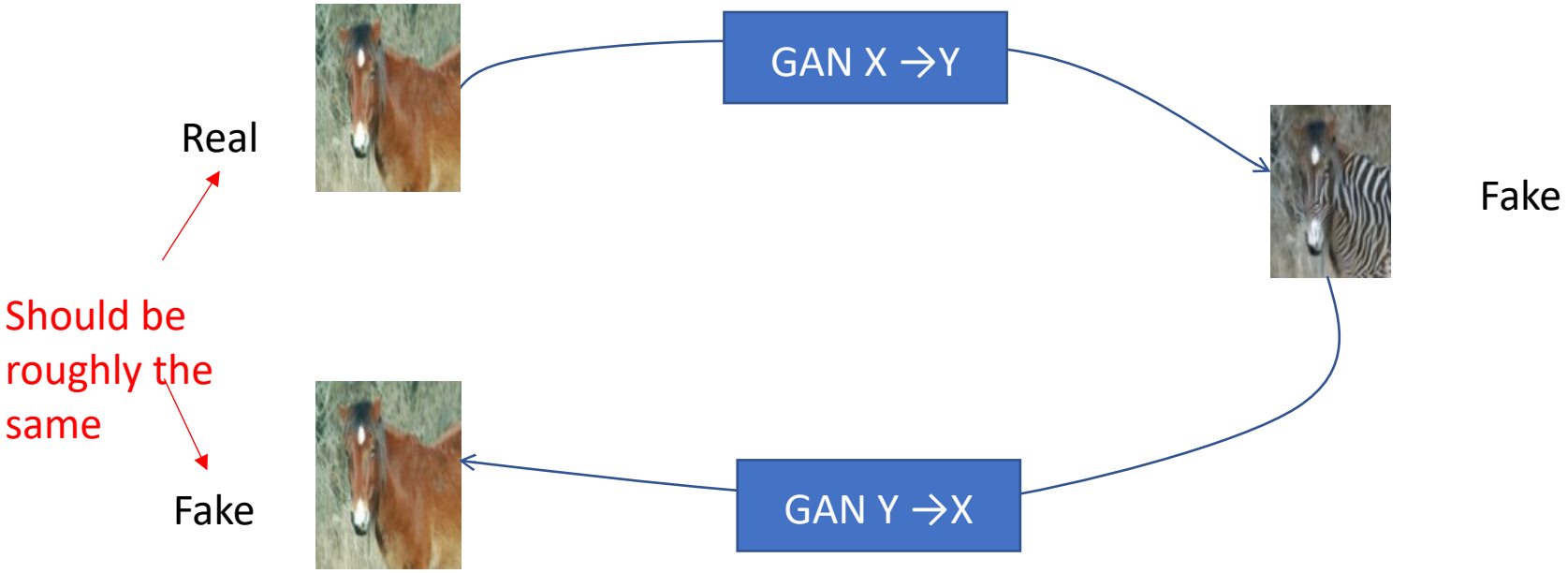


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, we introduce two *cycle consistency losses* that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

Cycle GAN
Cycle consistency loss



$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

Cycle GAN

- Full loss function:

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

FSGAN

- Deep learning–based approach to face swapping and reenactment in images and videos
- Subject agnostic: applied to faces of different subjects **without requiring subject specific training**

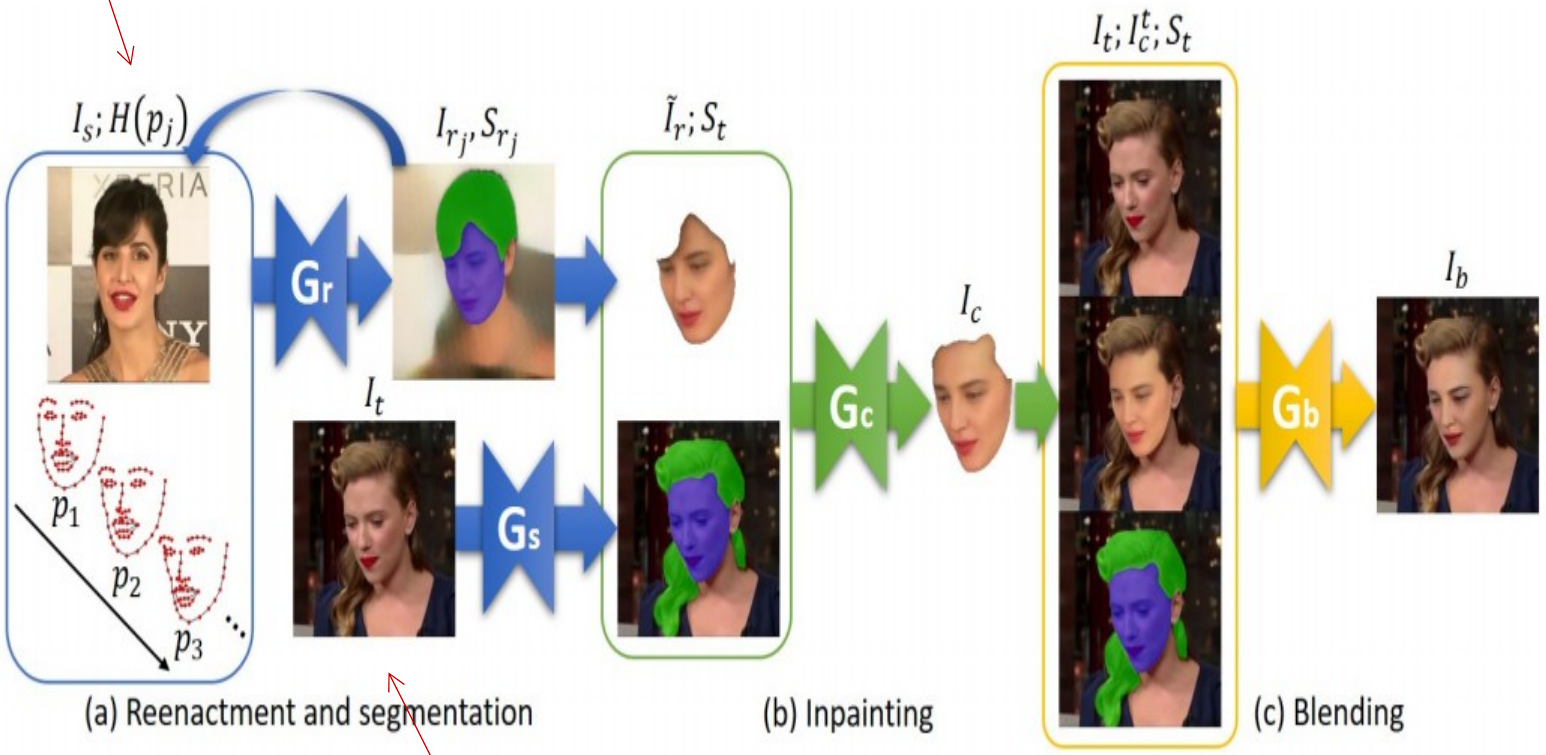
https://youtu.be/BsITEVX6hkE?t=7

https://www.youtube.com/watch?v=duo-tHbSdMk

FSGAN

Goal: create a new image based on the target image, such that the face in this image is replaced by the face in the source image, while maintaining the pose and expression

Source image
Fs – face in the source image



Target image
Ft – face in the target image

Given an image $I \in \mathbb{R}^{3\times H\times W}$ and a heatmap representation $H(p) \in \mathbb{R}^{70\times H\times W}$ of facial landmarks, $p \in \mathbb{R}^{70\times 2}$, we define the face reenactment generator, G_r, as the mapping $G_r : \{\mathbb{R}^{3\times H\times W}, \mathbb{R}^{70\times H\times W}\} \rightarrow \mathbb{R}^{3\times H\times W}$.

Let $v_s, v_t \in \mathbb{R}^{70\times 3}$ and $e_s, e_t \in \mathbb{R}^3$, be the 3D landmarks and Euler angles corresponding to F_s and F_t. We generate intermediate 2D landmark positions p_j by interpolating between e_s and e_t, and the centroids of v_s and v_t, using intermediate points for which we project v_s back to I_s. We define the reenactment output recursively for each iteration $1 \leq j \leq n$ as

$$I_{r_j}, S_{r_j} = G_r(I_{r_{j-1}}; H(p_j)), \quad (6)$$

$$I_{r_0} = I_s.$$

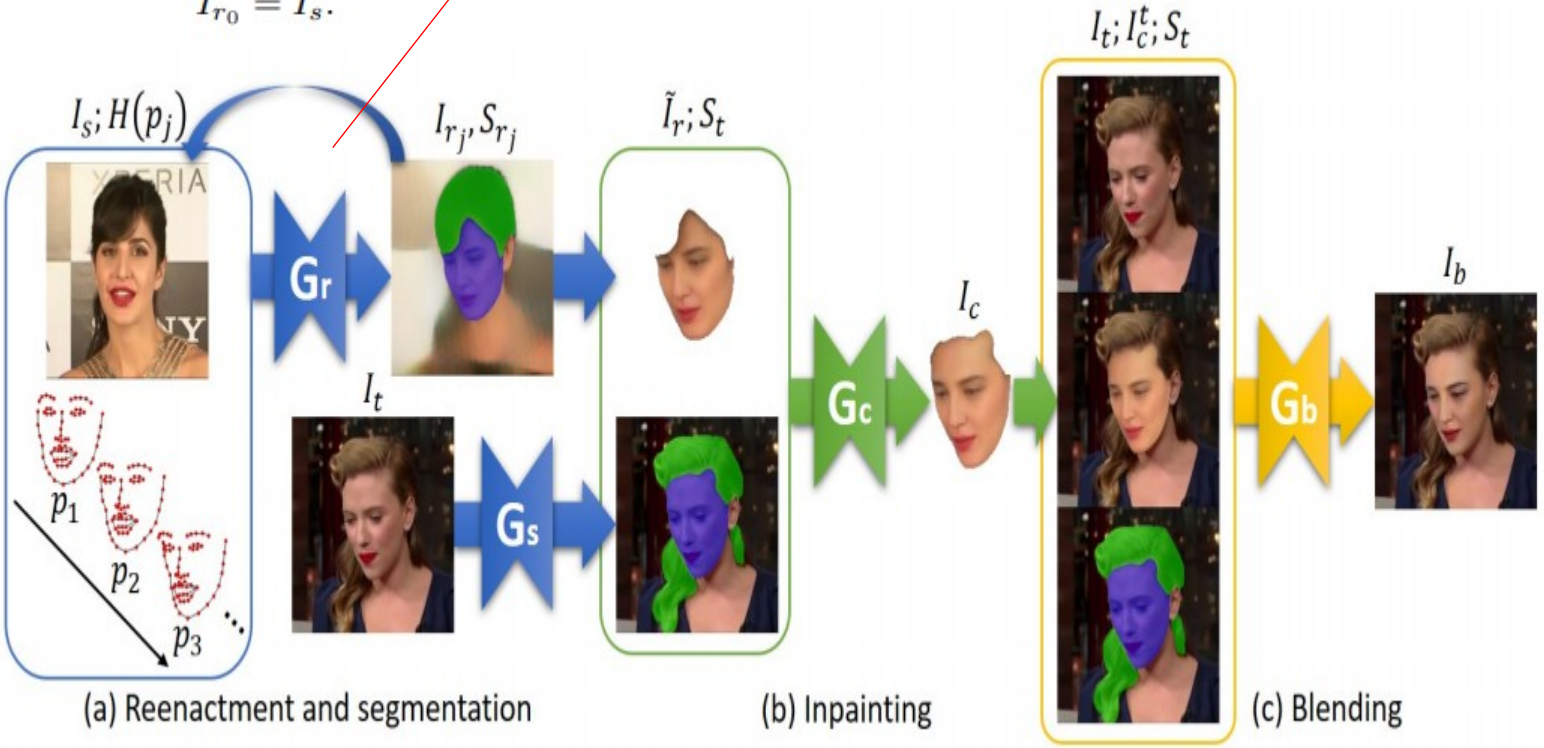
Reenactment generator

Input: heatmaps encoding the face landmarks in the target image

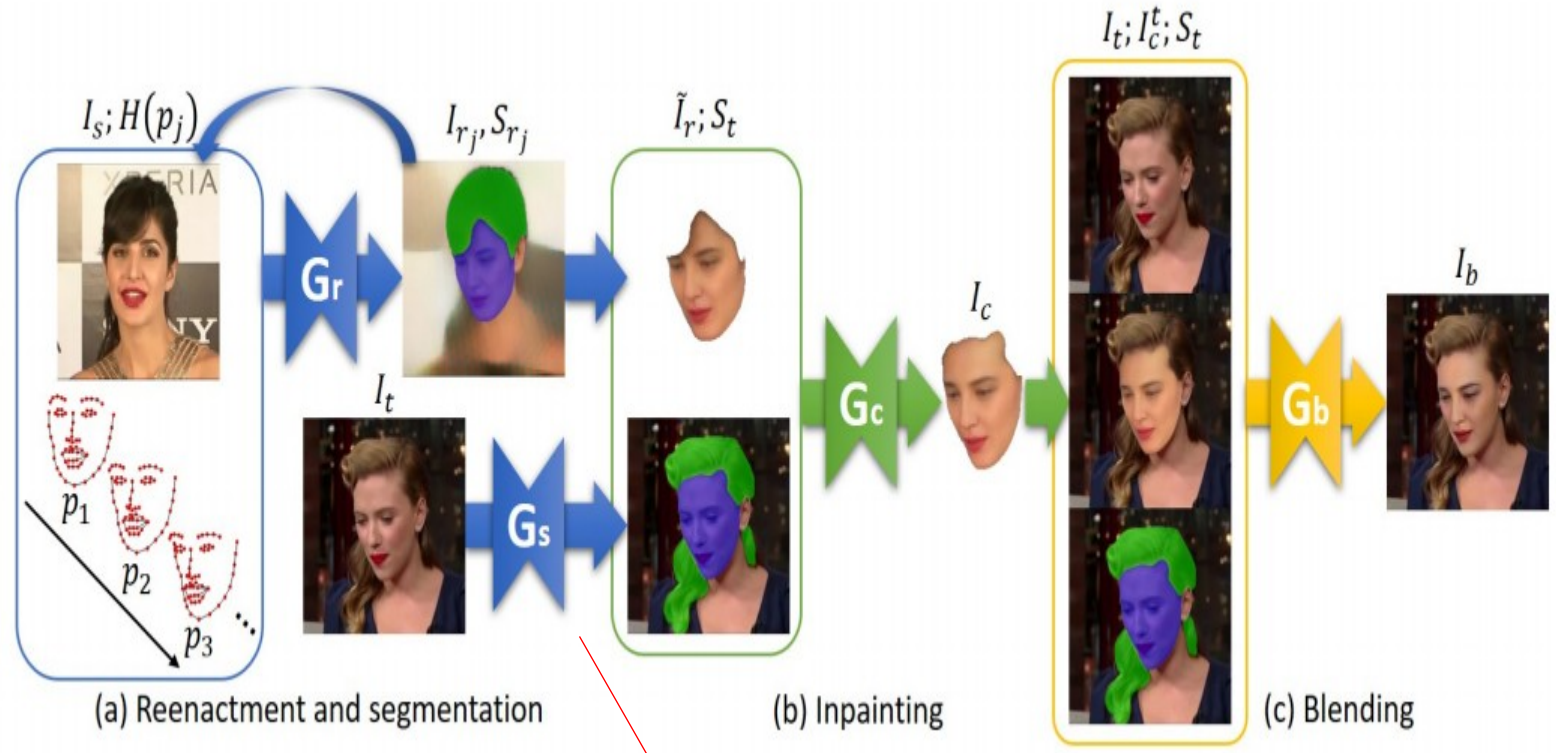
recurrent

Output:

- **Ir –** reenacted image, such that the face in this image (Fr) depicts Fs at the same pose and expression as Ft
- ***Sr –*** segmentation map of Fs – face and hair



FSGAN

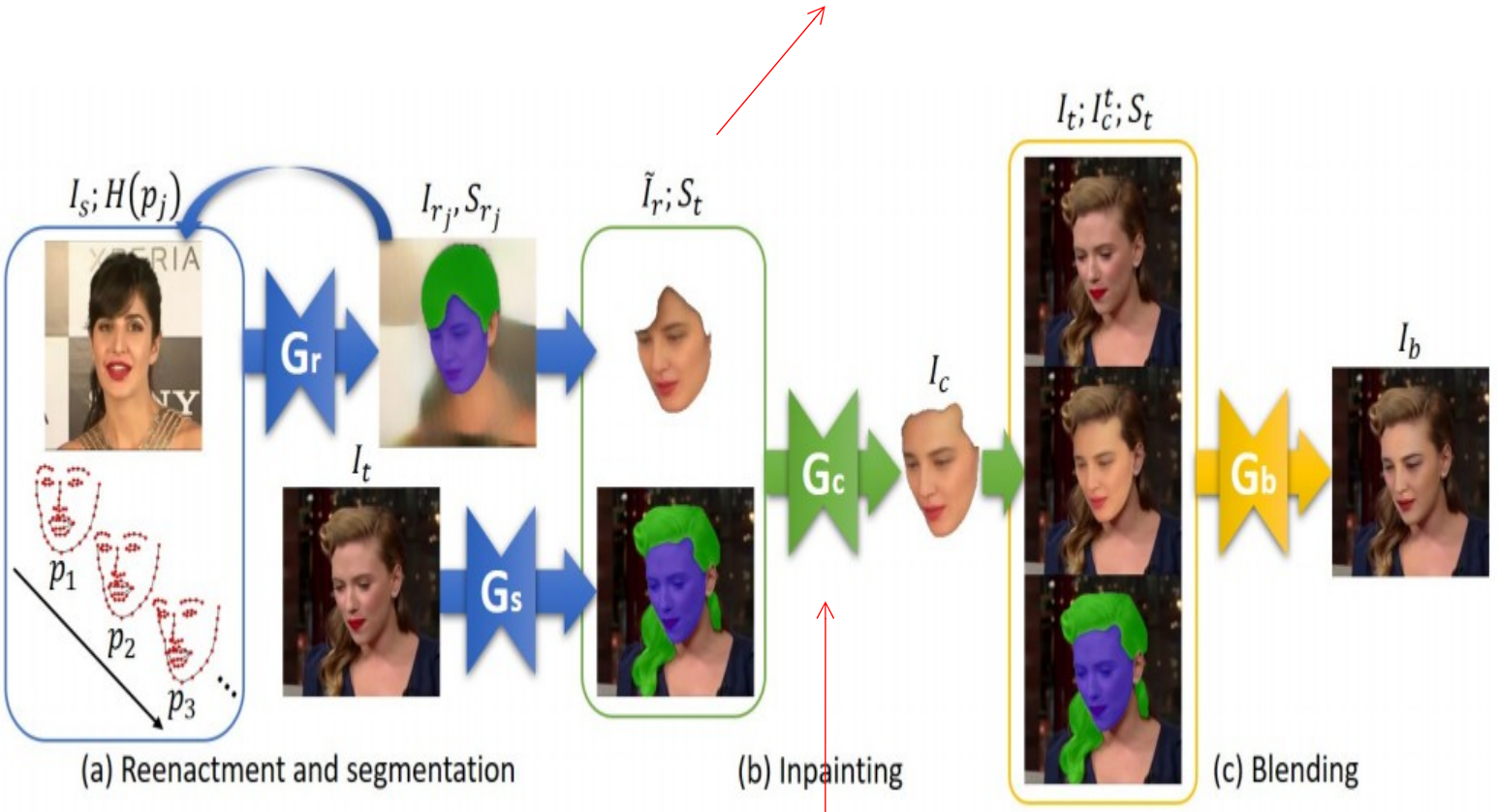


Gs: compute the source image face-hair segmentation

U-Net

FSGAN

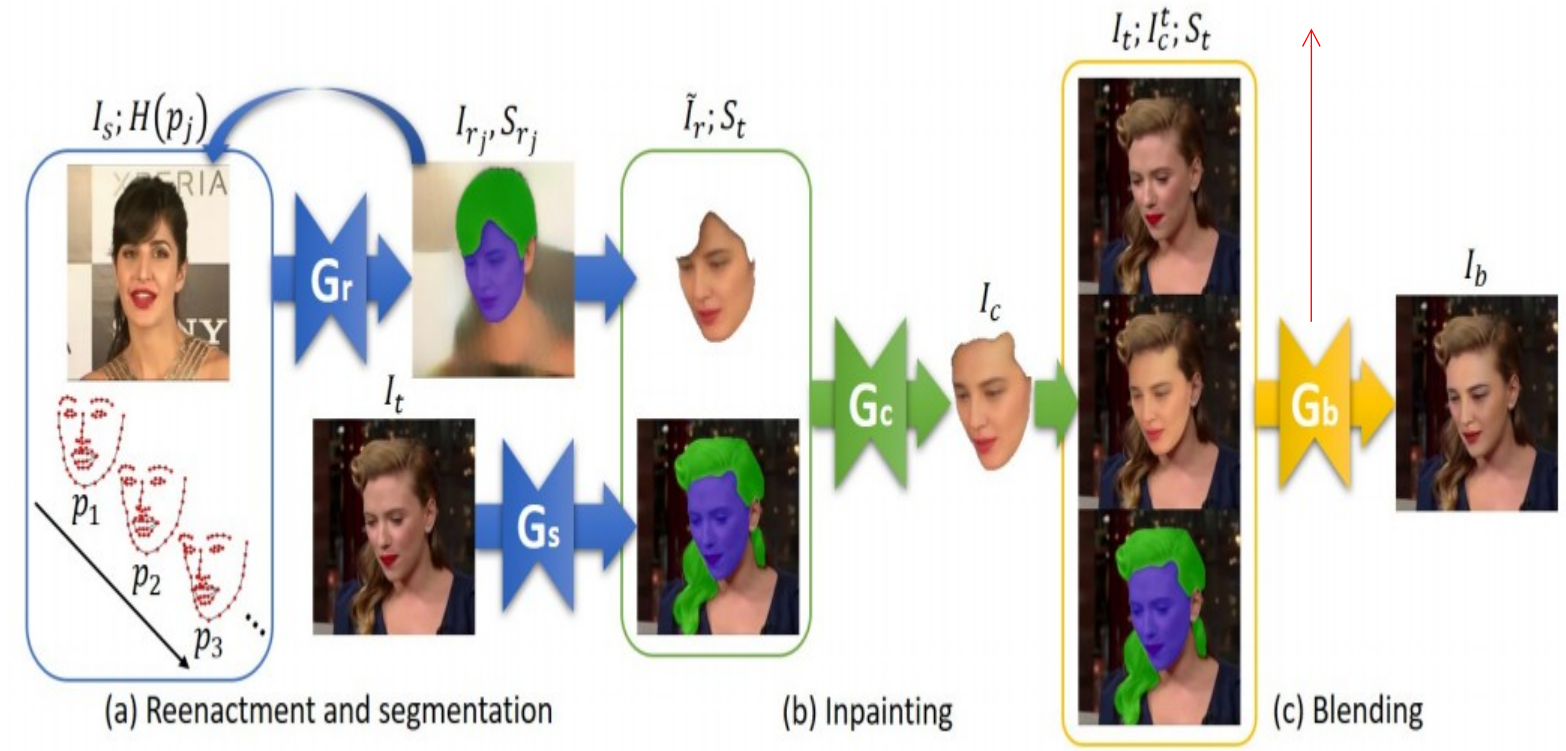
Re-enacted image might contain missing parts



Use a face inpainting network

FSGAN

Blend the completed face Fc to the target face, accounting for different skin tones and lighting conditions



CLIP

FOOD101

guacamole (90.1%) Ranked 1 out of 101 labels

✓ a photo of **guacamole**, a type of food.

× a photo of **ceviche**, a type of food.

× a photo of **edamame**, a type of food.

× a photo of **tuna tartare**, a type of food.

× a photo of **hummus**, a type of food.

SUN397

television studio (90.2%) Ranked 1 out of 397

✓ a photo of a **television studio**.

× a photo of a **podium indoor**.

× a photo of a **conference room**.

× a photo of a **lecture room**.

× a photo of a **control room**.

YOUTUBE-BB

airplane, person (89.0%) Ranked 1 out of 23

✓ a photo of a **airplane**.

× a photo of a **bird**.

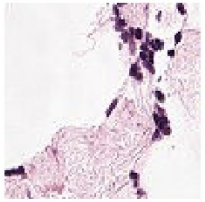
× a photo of a **bear**.

× a photo of a **giraffe**.

× a photo of a **car**.

PATCHCAMELYON (PCAM)

healthy lymph node tissue (22.8%) Ranked 2 out of 2



× this is a photo of **lymph node tumor tissue**

✓ this is a photo of **healthy lymph node tissue**

Learning Transferable Visual Models From Natural Language Supervision

CLIP

https://arxiv.org/pdf/2103.00020.pdf

CLIP

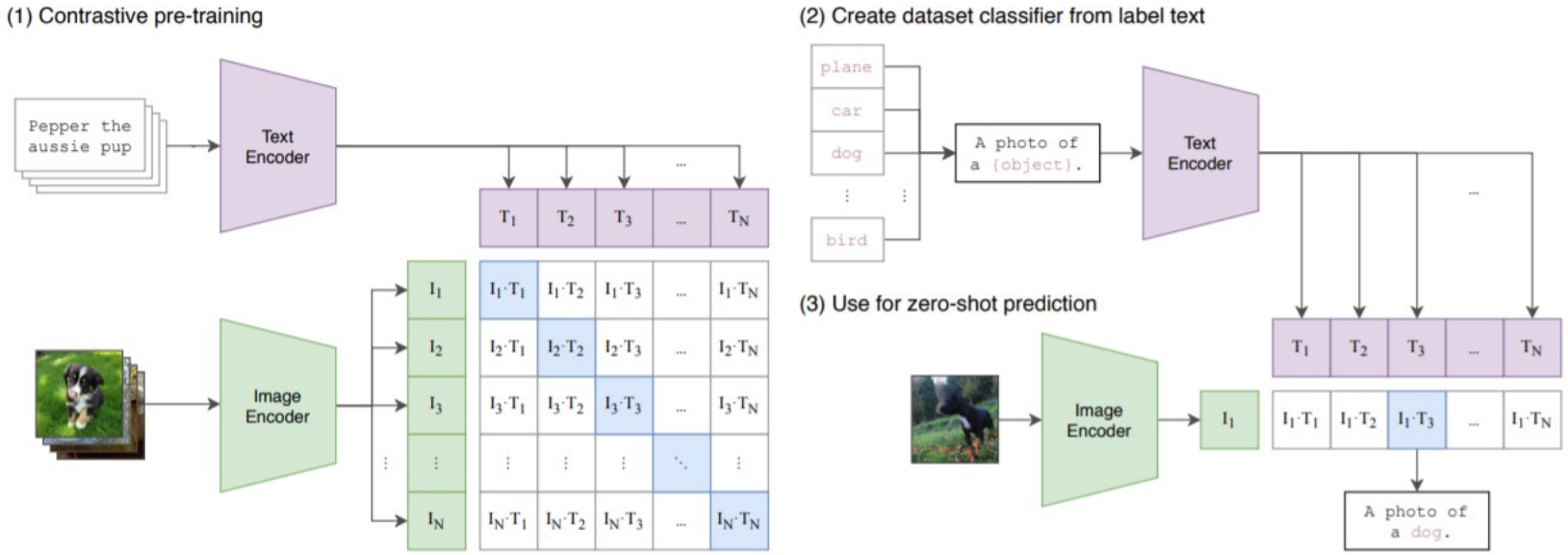


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

https://www.youtube.com/watch?v=u0HG77RNhPE

Zero-Shot Text-to-Image Generation

DALL-E

https://arxiv.org/pdf/2102.12092.pdf
https://openai.com/blog/dall-e/

Dall-E

https://www.youtube.com/watch?v=C7D5EzkhT6A

Deep fake detection

https://www.youtube.com/watch?v=poSd2CyDpyA