Generative Adversarial Text-to-Image Synthesis

Reed et al., 2016

Explainable Machine Learning Seminar

Frank Gabel - Thursday, 5 July

Motivation

Who recognizes these celebrities ?

Motivation

Who recognizes these celebrities ?

You can't, they have been synthesized from white noise.

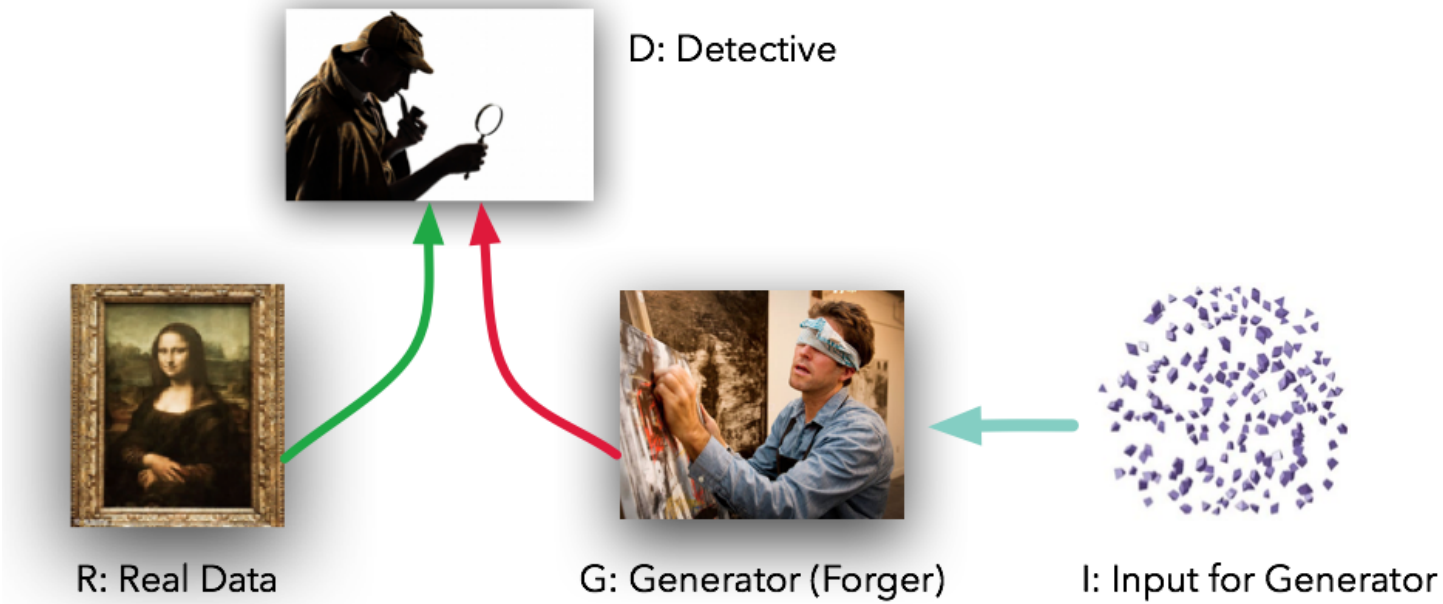
Source : Progressive Growing of GANs for Improved Quality, Stability, and Variation (2017)

Discriminative models : $p(Y|X)$ learn decision boundaries

Generative models : $p(X, Y)$ learn distributions

Introduction

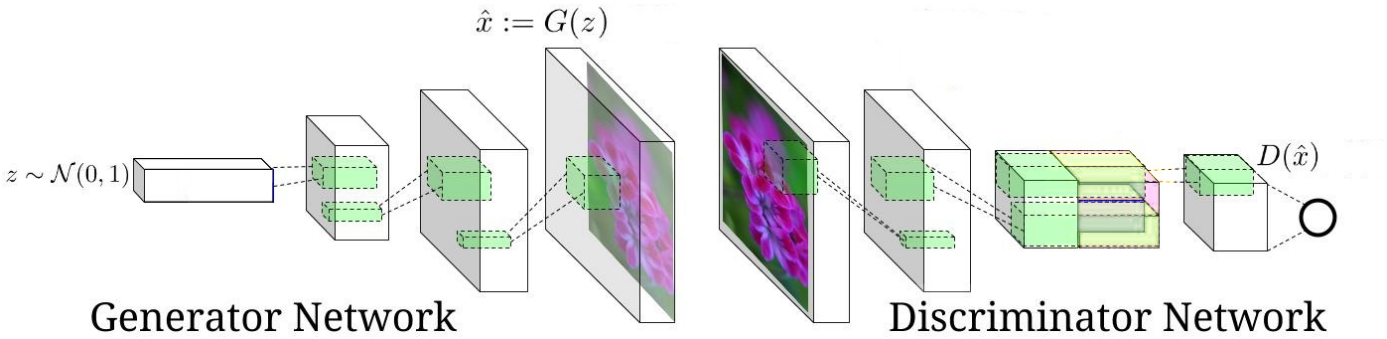
Conceptually...



Mathematically...

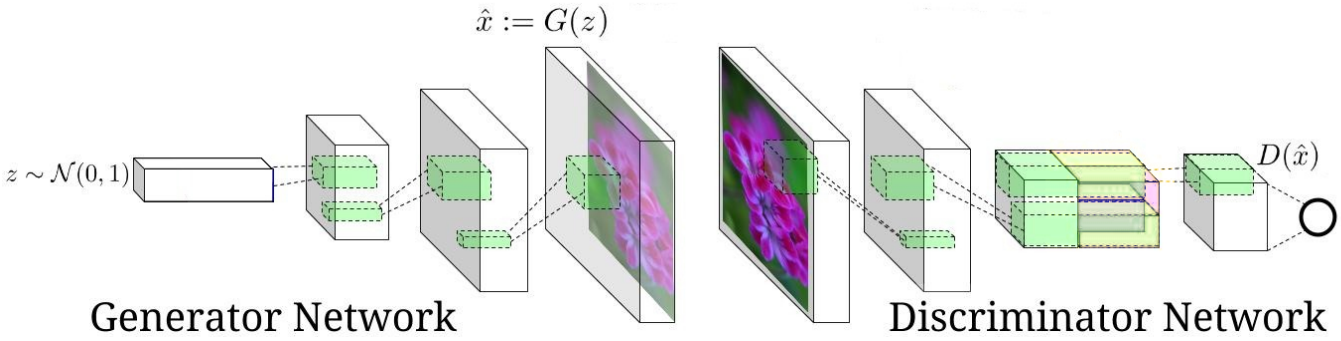
$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x\sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

Vanilla GAN



$$G : \mathbb{R}^Z \to \mathbb{R}^{D \times D} \text{ and } D : \mathbb{R}^{D \times D} \to \{0, 1\}$$

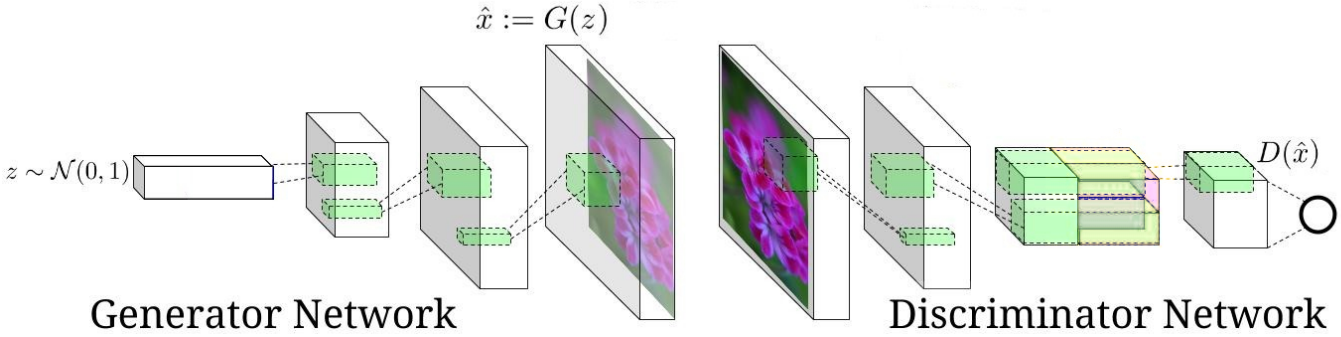
Vanilla GAN



$$G : \mathbb{R}^Z \rightarrow \mathbb{R}^{D \times D} \text{ and } D : \mathbb{R}^{D \times D} \rightarrow \{0, 1\}$$

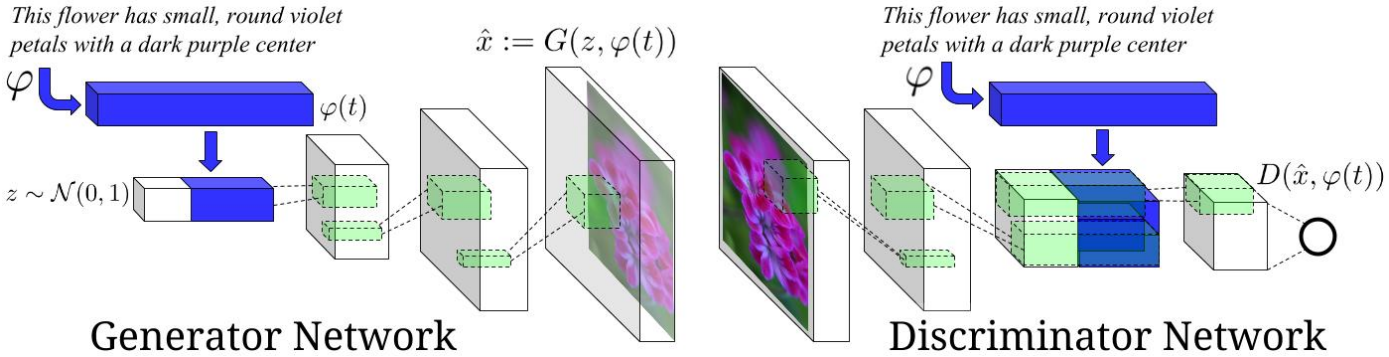
Is is possible to control the output of a GAN in a more meaningful way ?

Vanilla GAN



$$G : \mathbb{R}^Z \to \mathbb{R}^{D \times D} \text{ and } D : \mathbb{R}^{D \times D} \to \{0, 1\}$$

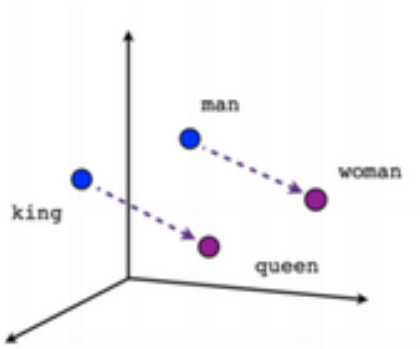
Conditional GANs



$$G : \mathbb{R}^Z \times \mathbb{R}^T \to \mathbb{R}^{D \times D} \text{ and } D : \mathbb{R}^{D \times D} \times \mathbb{R}^T \to \{0, 1\}$$

Introduction to Word Embeddings in NLP

- Map words to a high-dimensional vector space
- preserve semantic similarities :
  - president-power $\approx$ prime minister
  - king-man+woman $\approx$ queen.



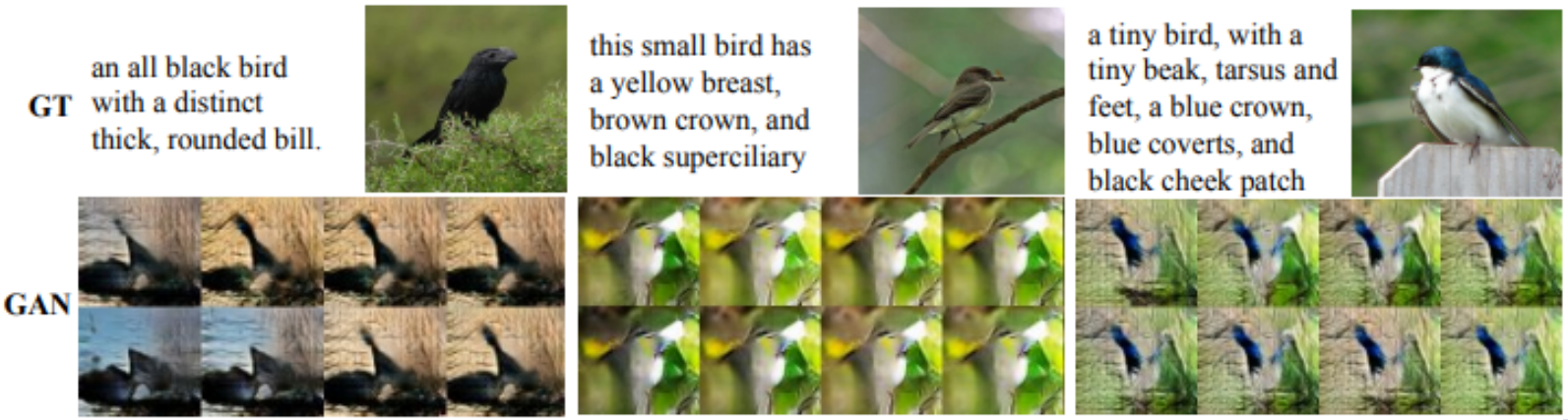
Male-Female

Verb tense

Country-Capital

- Technique for embedding descriptions : **deep symmetrical structural joint embedding** (Reed et al., 2016) $\Rightarrow$ idea is the same : preserve semantic similarities of sentences in the embedding space

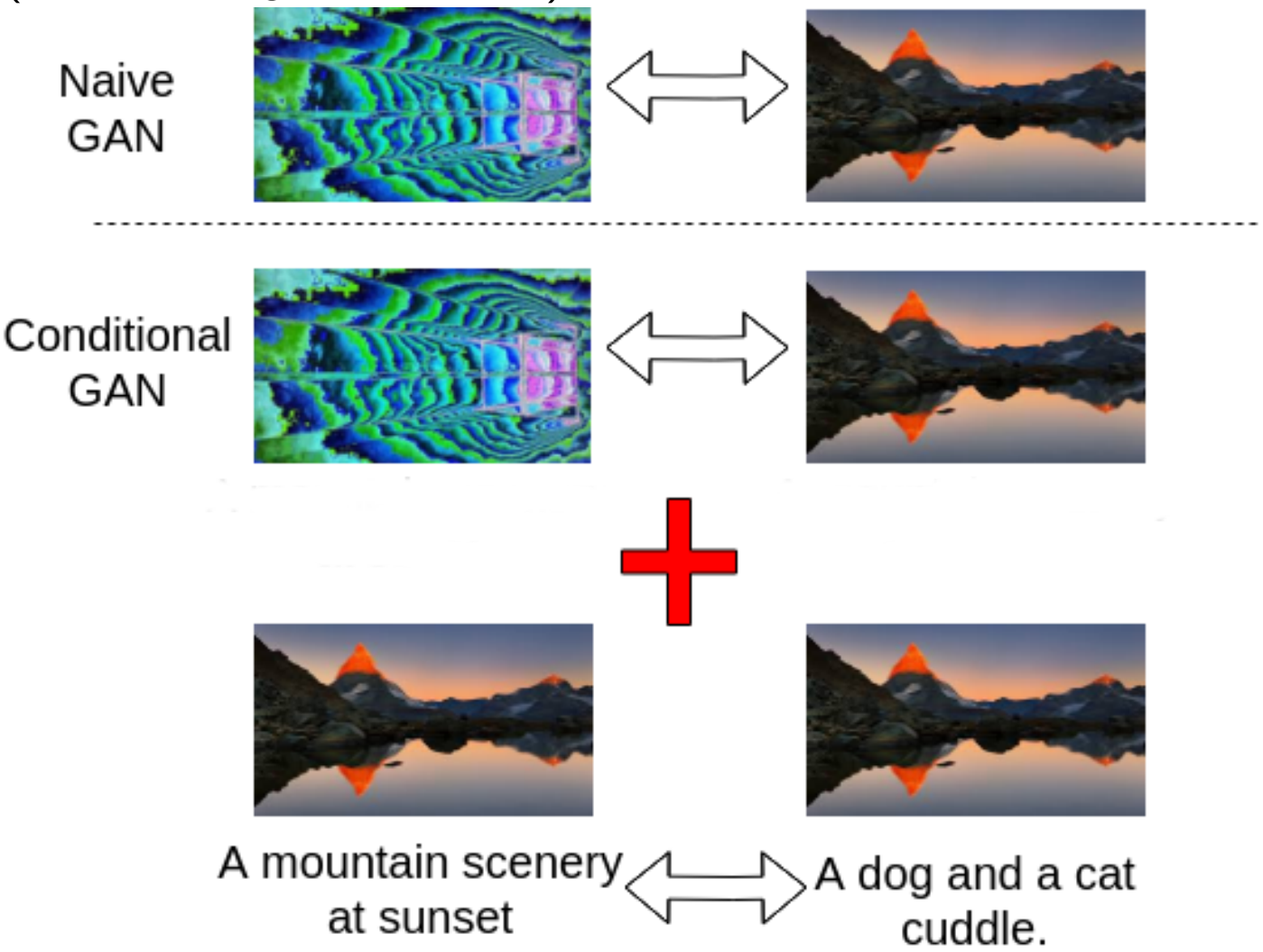
Naive result



This does not work really well.

Conditional GAN training dynamics

Conditional GANs have an additional error source (unmatching description)



Extensions and tricks

Matching-aware discriminator (GAN-CLS)

so far :

- (real image, correct description) - pairs $\rightarrow$ classified as 1
- (fake image, correct description) - pairs $\rightarrow$ classified as 0

Extensions and tricks

Matching-aware discriminator (GAN-CLS)

so far :

- (real image, correct description) - pairs
  $\rightarrow$ classified as 1
- (fake image, correct description) - pairs
  $\rightarrow$ classified as 0

now :

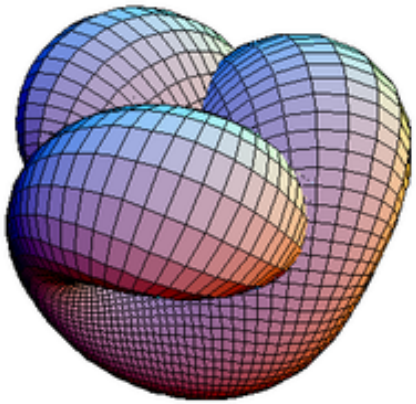
$\Rightarrow$ add

- (real image, false description) - pairs
  $\rightarrow$ classified as 0

Extensions and tricks

Learning with manifold interpolation (GAN-INT)

interpolate between embedding pairs



Build the arithmetic mean between embeddings from the training set : $t_{new} = \beta t_1 + (1 - \beta) t_2$

$\rightarrow$ free training data

Extensions and tricks

Inverting the generator for style transfer

inject style information by training the z vector (background color, lighting etc.)

Train a style encoder network S : $s \leftarrow S(x),\ \hat{x} \leftarrow G(s, \phi(t))$

Extensions and tricks

Matching-aware discriminator (GAN-CLS) introduce new types of samples

Learning with manifold interpolation (GAN-INT) interpolate between embedding pairs

Inverting the generator for style transfer inject style information (background color, pose etc.)

Results - birds dataset

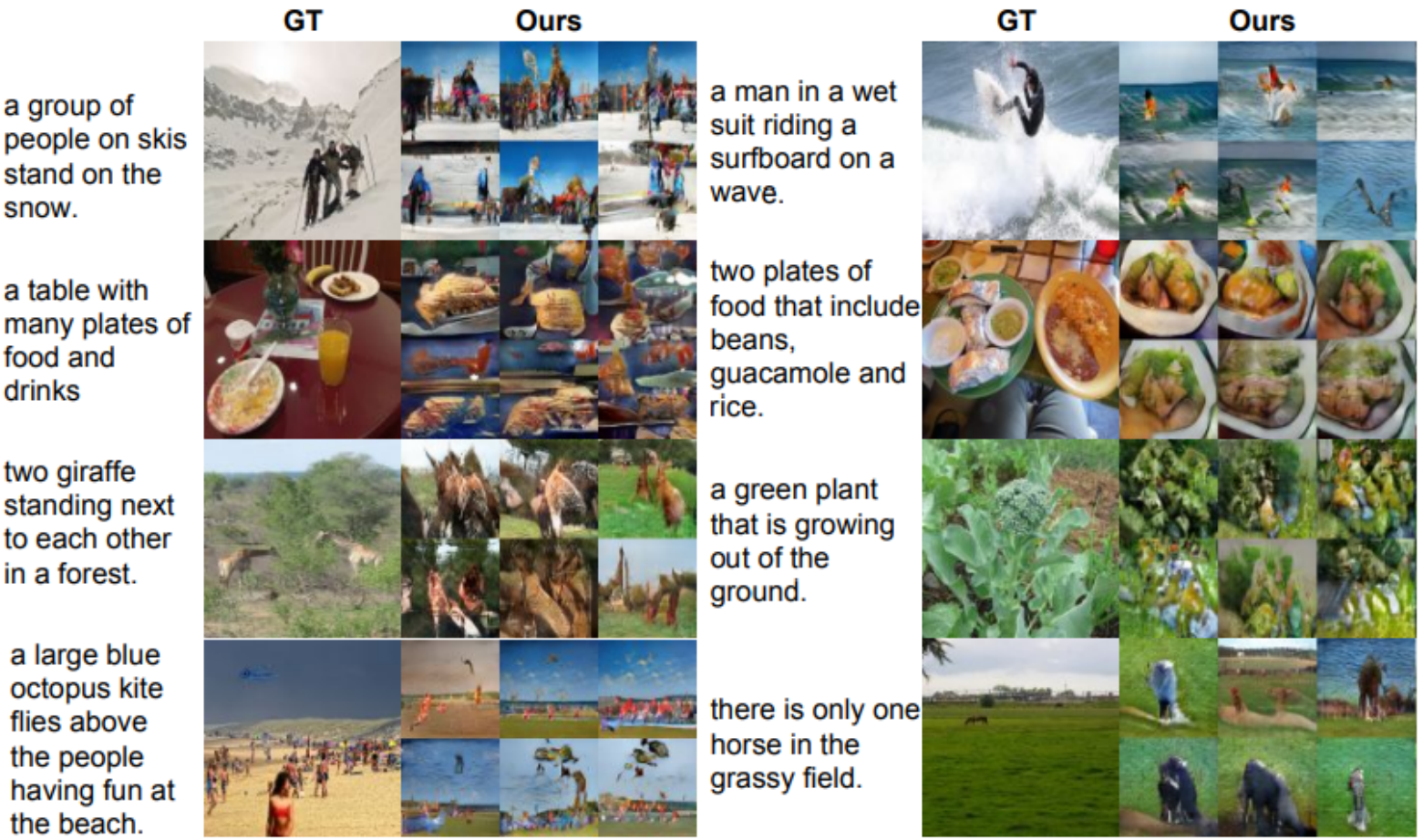
Results - birds dataset - GAN-INT

‘Blue bird with black beak’ →
‘Red bird with black beak’

‘Small blue bird with black wings’ →
‘Small yellow bird with black wings’

‘This bird is bright.’ → ‘This bird is dark.’

Results - more general dataset



Results - style transfer on birds dataset

Text descriptions (content)

Images (style)

The bird has a **yellow breast** with **grey** features and a small beak.

This is a large **white** bird with **black wings** and a **red head**.

A small bird with a **black head and wings** and features grey wings.

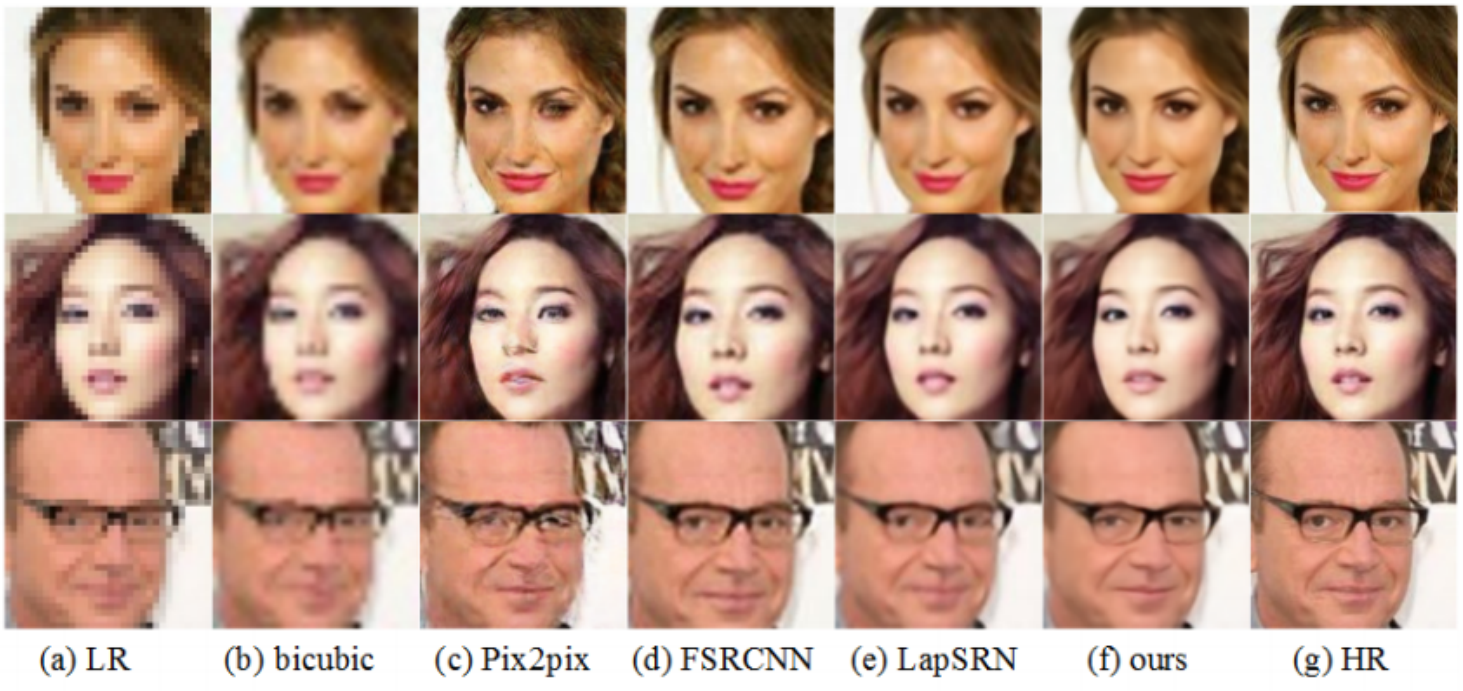
This bird has a **white breast**, brown and white coloring on its head and wings, and a thin pointy beak.

A small bird with **white base** and **black stripes** throughout its belly, head, and feathers.

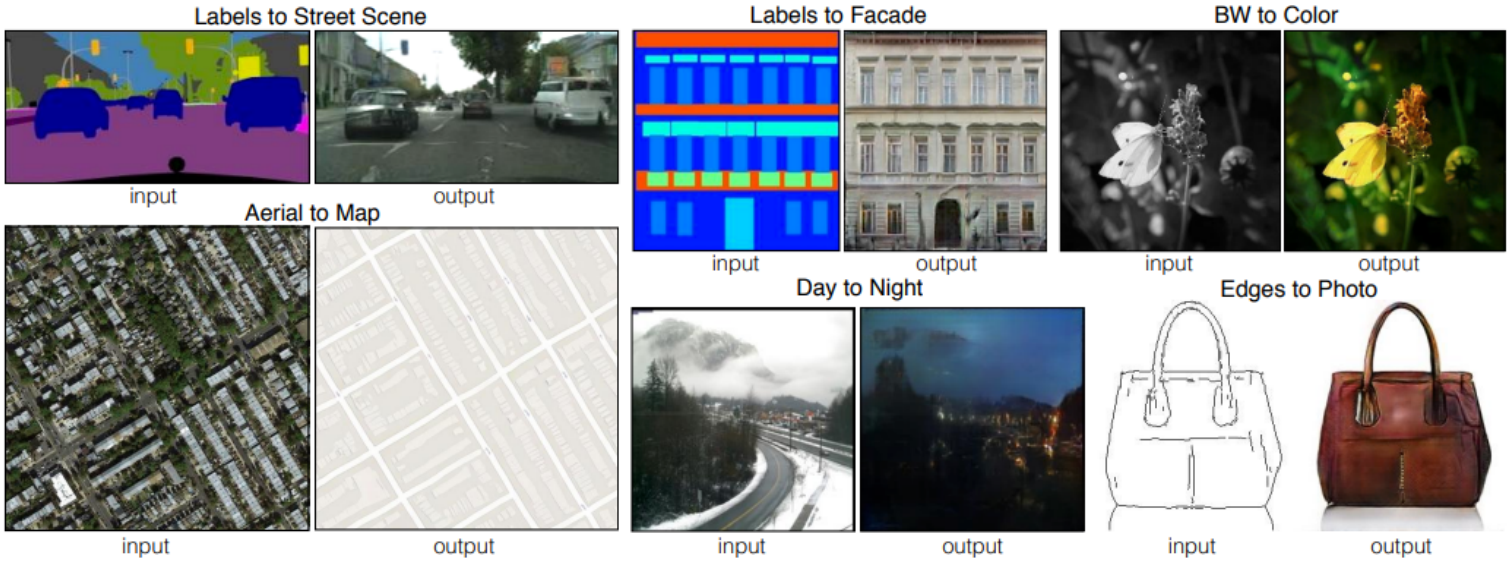
A small sized bird that has a cream belly and a short pointed bill.

This bird is **completely red**.

Recent application examples



Source : High-Quality Face Image Super-Resolution Using Conditional Generative Adversarial Networks (2017)



Source : Image-to-Image Translation with Conditional Adversarial Networks (2017)

"In my opinion, among many interesting recent developments in deep learning, adversarial training is the most important one."

Yann LeCun, Director of Facebook AI