Title: Machine Learning.....

First Name Last Name[*]   First Name Last Name[*]   First Name Last Name
Department of Mathematics, Northeastern University

November 28, 2021

Abstract

Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn ...

1   Introduction

We can imagine all this despite never having seen a side by side example of a Monet painting next to a photo of the scene he painted. Instead, we have knowledge of the set of Monet paintings and of the set of landscape photographs. We can reason about the stylistic differences between these two sets, and thereby imagine what a scene might look like if we were to "translate" it from one set into the other.

In this paper, we present a method that can learn to do the same: 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 examples.

...

2   Related work

Image-to-Image Translation The idea of image-to-image translation goes back at least to Hertzmann et al.'s Image Analogies [?], who employ a non-parametric texture model [?] on a single input-output training image pair. More recent approaches use a *dataset* of input-output examples to learn a parametric translation function using CNNs (e.g., [?]).

Our approach builds on the "pix2pix" framework of Isola et al. [?], which uses a conditional generative adversarial network [?] to learn a mapping from input to output images. Similar ideas have been applied to various tasks such as generating photographs from sketches [?] or from attribute and
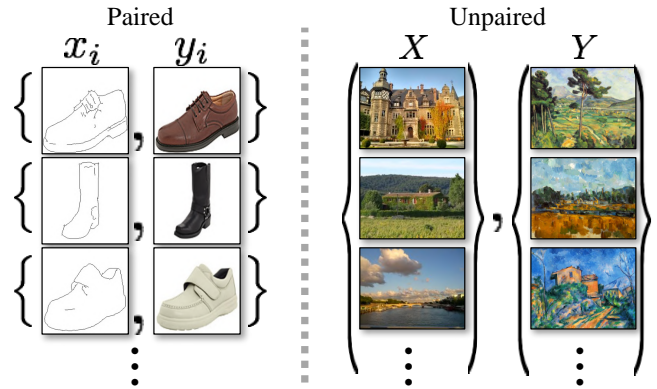

Figure 1: *Paired* training data (left) ...

semantic layouts [?]. However, unlike the above prior work, we learn the mapping without paired training examples.

....

3   Formulation

Our goal is to learn mapping functions between two domains X and Y given training samples $\{x_i\}_{i=1}^N$ where $x_i \in X$ and $\{y_j\}_{j=1}^M$

3.1   Adversarial Loss

We apply adversarial losses [?] to both mapping functions. ...

4   Implementation

Network Architecture   We adopt the architecture for our generative networks from Johnson et al. [?] who have shown impressive results for neural style transfer and super-resolution. This network contains three convolutions, ...

5   Results

We first compare our approach against recent methods for unpaired image-to-image translation on paired datasets

where ground truth input-output pairs are

5.0.1 Comparison against baselines

As can be seen in ? and ?, we were unable to achieve compelling results with any of the baselines. Our method, on the other hand, can produce ...

...

6 Acknowledgement

We would like to thank ... for providing the data. ...

References

[1] Y. Aytar, L. Castrejon, C. Vondrick, H. Pirsiavash, and A. Torralba. Cross-modal scene networks. *PAMI*, 2016.

[2] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *CVPR*, 2017.

[3] ...