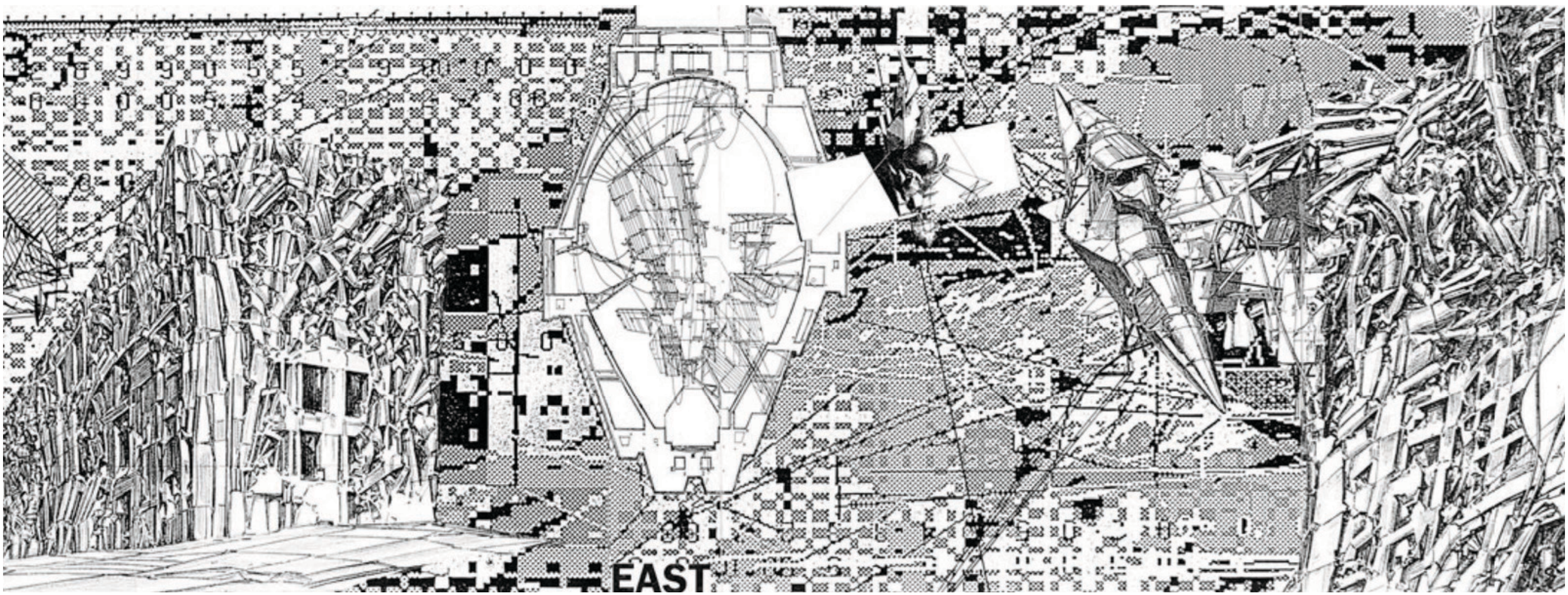
Style & Content Analysis



Style & Content Analysis

- What is content, what is style in an image?
- What are the methods of classification for each?
- With computation-generated images, how can the distinction be defined between style and content?
- What if the style is the content?

--

- Visual Modality – the ability to see: How to evaluate the degree to which pictorial expression through color, detail, depth, tonal shades are used?

Style Transfer (Gatys)

A Neural Algorithm of Artistic Style
Leon A. Gatys, Alexander S. Ecker, Matthias Bethge (2015)

In fine art, especially painting, humans have mastered the skill to create unique visual experiences through composing a complex interplay between the content and style of an image. Thus far the algorithmic basis of this process is unknown and there exists no artificial system with similar capabilities. However, in other key areas of visual perception such as object and face recognition near-human performance was recently demonstrated by a class of biologically inspired vision models called Deep Neural Networks.

Here we introduce an artificial system based on a Deep Neural Network that creates artistic images of high perceptual quality. The system uses neural representations to separate and recombine content and style of arbitrary images, providing a neural algorithm for the creation of artistic images. Moreover, in light of the striking similarities between performance-optimised artificial neural networks and biological vision, our work offers a path forward to an algorithmic understanding of how humans create and perceive artistic imagery.

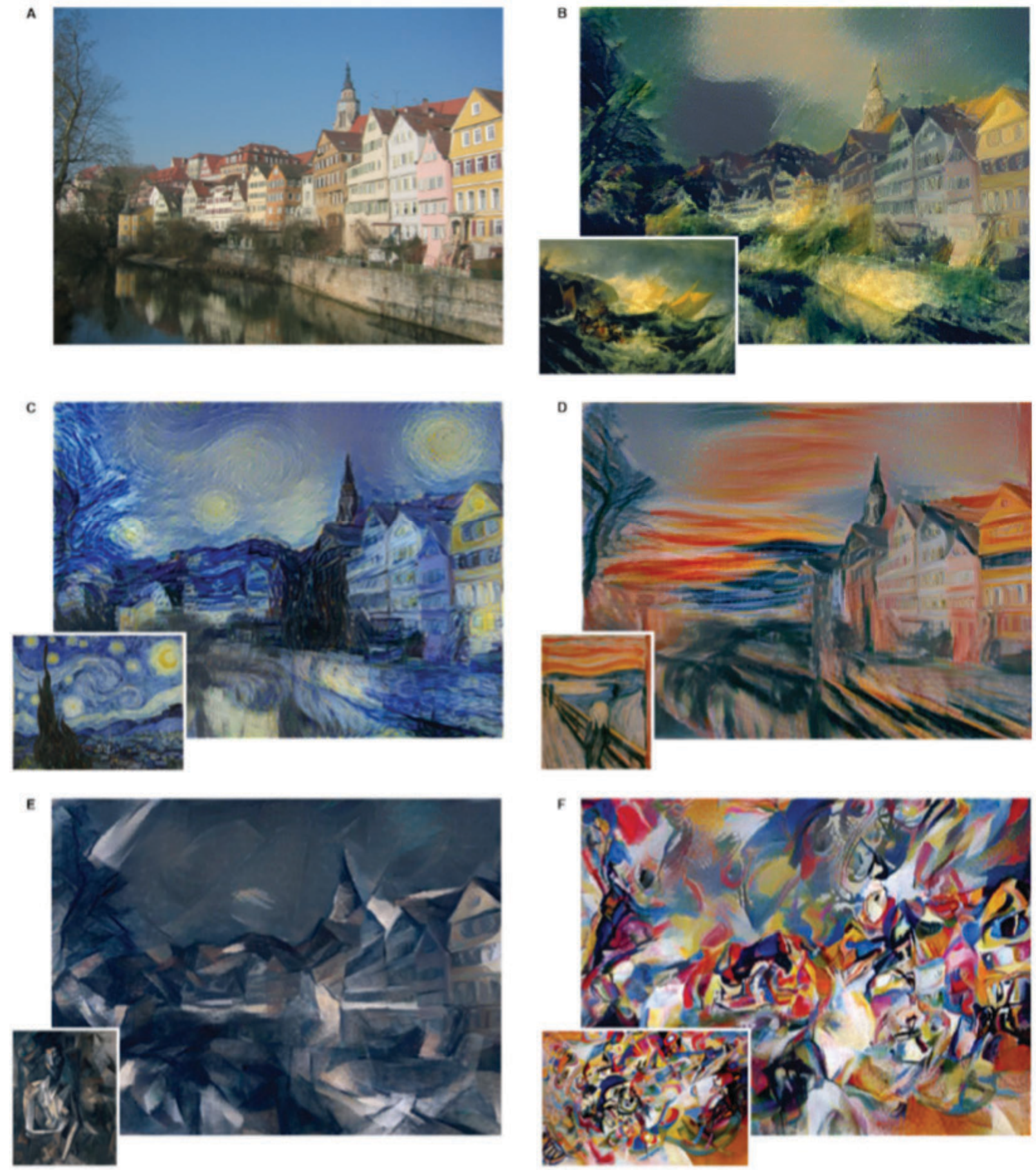


Figure 2: Images that combine the content of a photograph with the style of several well-known artworks. The images were created by finding an image that simultaneously matches the content representation of the photograph and the style representation of the artwork (see Methods). The original photograph depicting the Neckarfront in Tübingen, Germany, is shown in **A** (Photo: Andreas Praefcke). The painting that provided the style for the respective generated image is shown in the bottom left corner of each panel. **B** *The Shipwreck of the Minotaur* by J.M.W. Turner, 1805. **C** *The Starry Night* by Vincent van Gogh, 1889. **D** *Der Schrei* by Edvard Munch, 1893. **E** *Femme nue assise* by Pablo Picasso, 1910. **F** *Composition VII* by Wassily Kandinsky, 1913.

Style Transfer (Gatys)

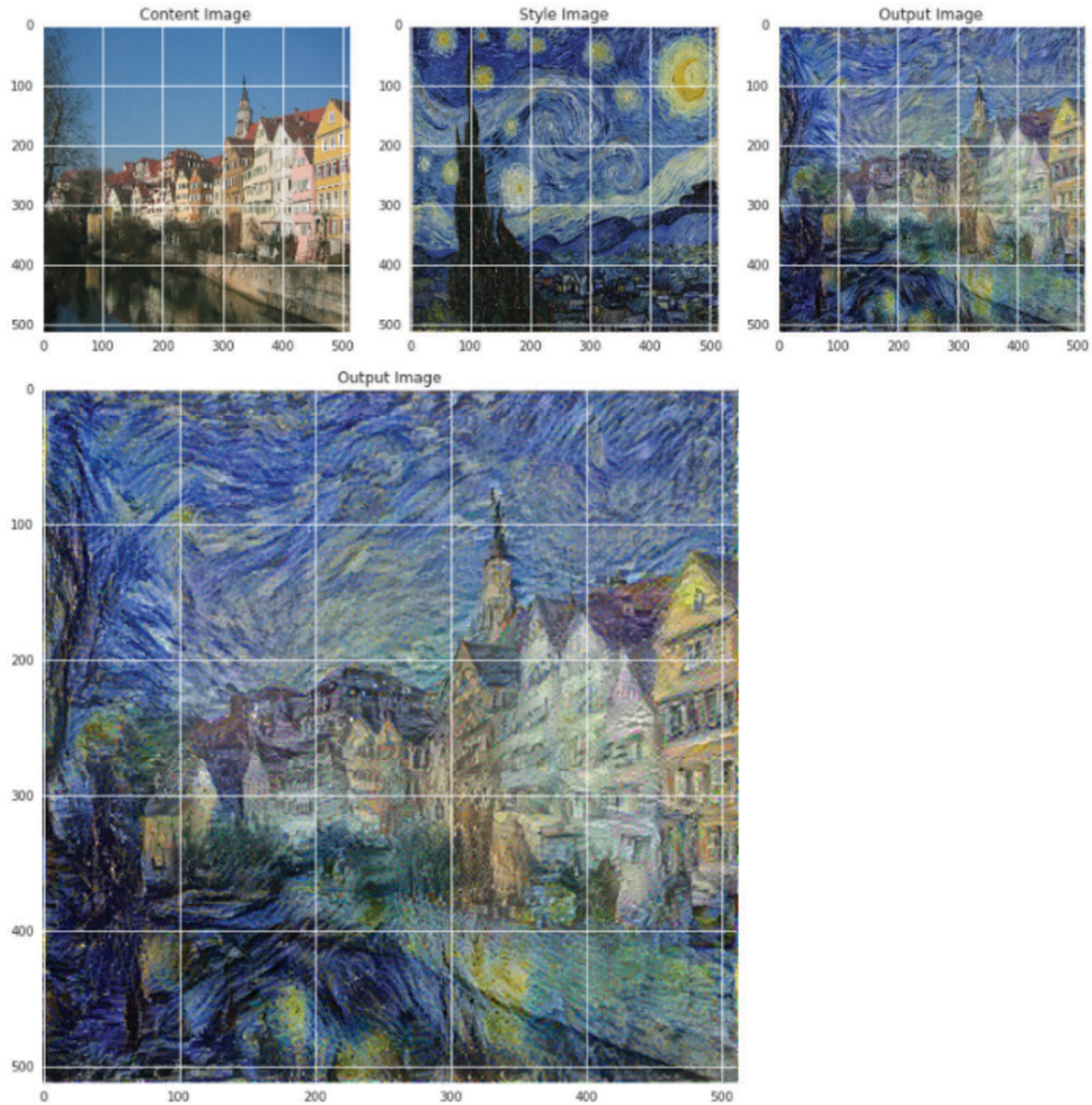
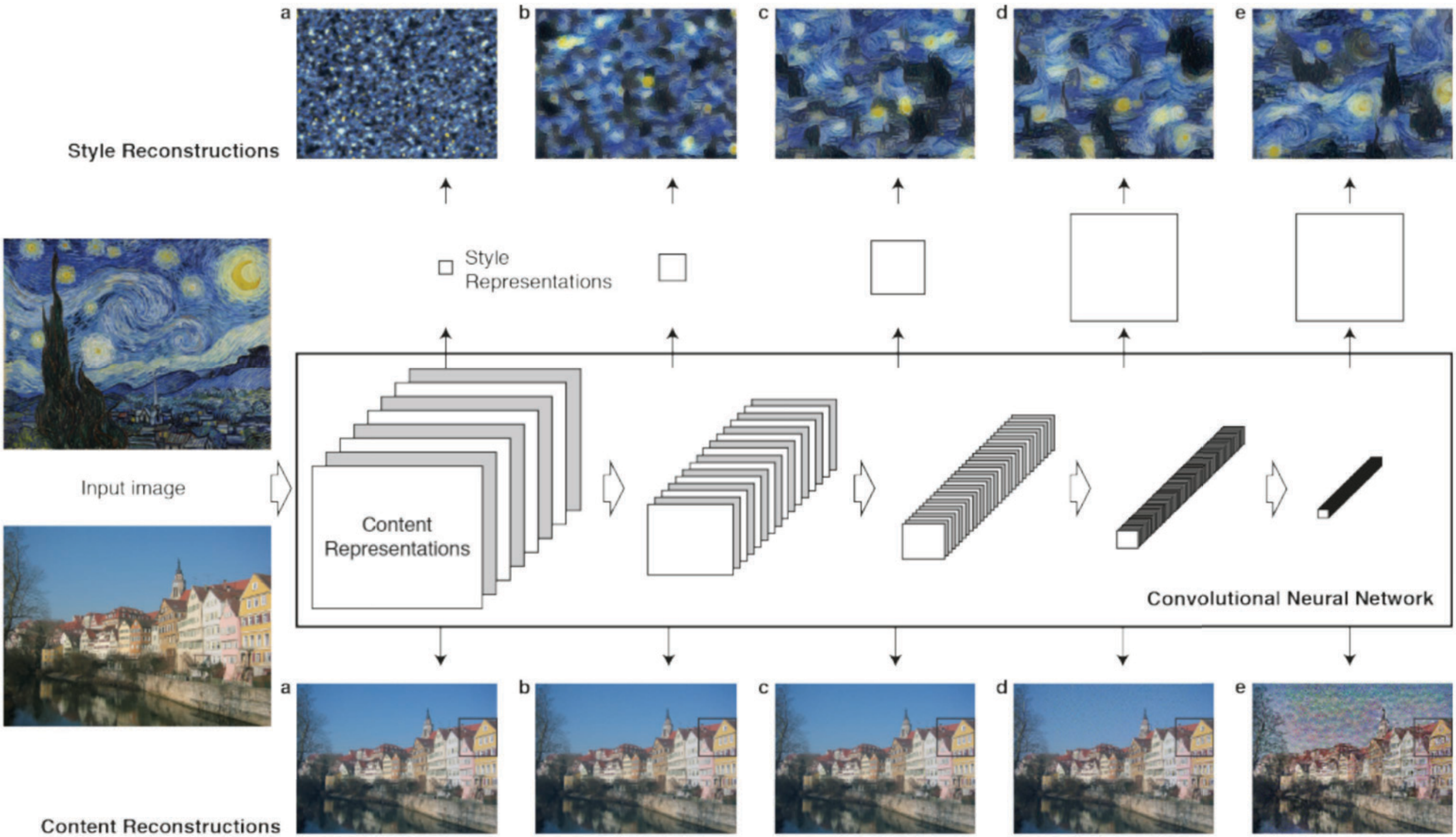


Image of Tuebingen — Photo By: Andreas Praefcke [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)], from Wikimedia Commons and Image of Starry Night by Vincent van Gogh Public domain

Style Transfer

- **Style transfer** – Transferring the style of the source image, while preserving the content of the target image
- Style specific to *Neural Style Transfer* is defined as the Texture of an image - Texture of an image captures the geometric shapes, patterns and transitions
- **Goal**: To synthesize a tecture from a source image while constraining the texture synthesis in order to preserve the semantic content of the target image

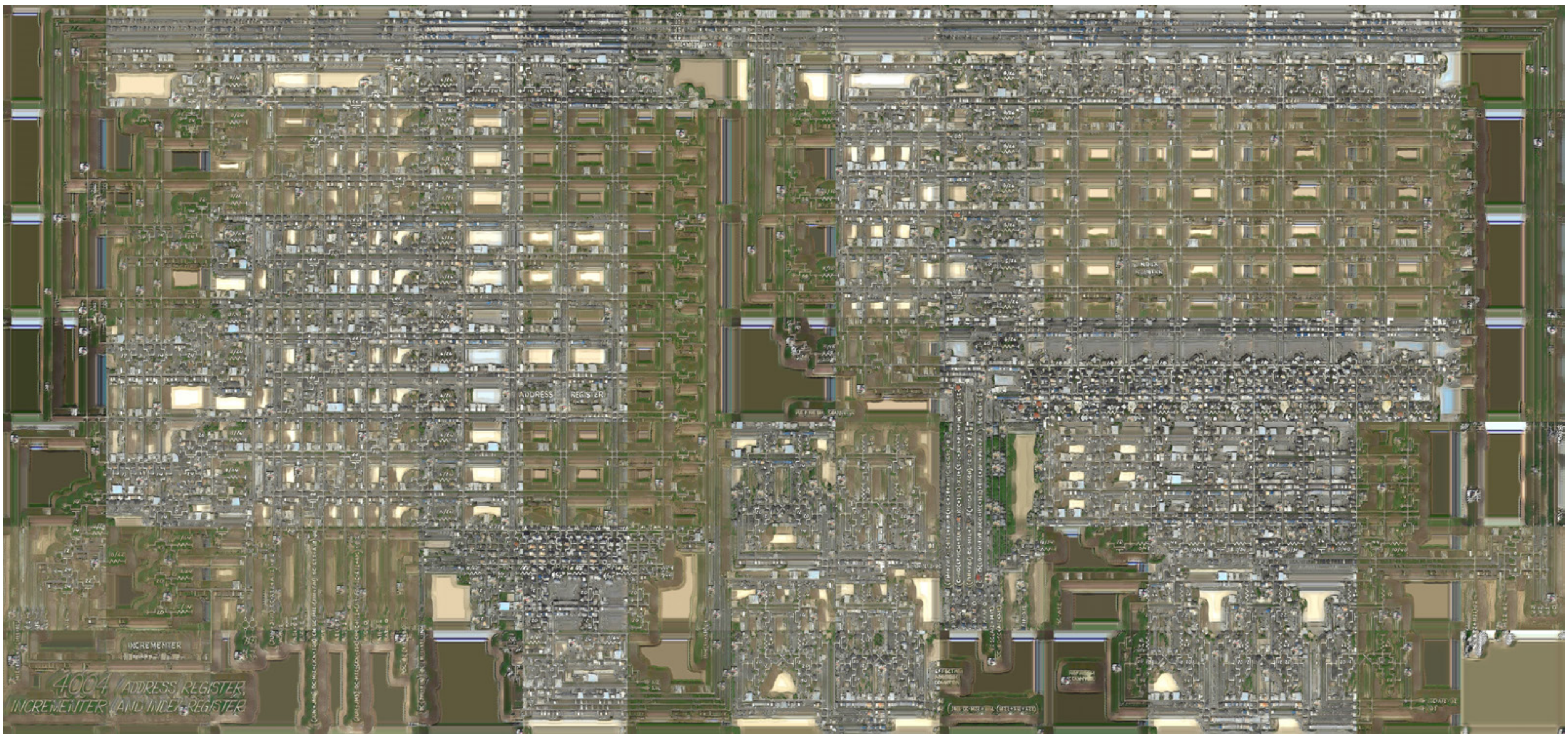
Style Transfer (Gatys)



Style Transfer (***Intel 4004 processor into city map*** by Fabian Offert)

Created by training a pix2pix GAN (implementations are freely available on GitHub, originally proposed in this paper: https://arxiv.org/abs/1611.07004 on pairs of map data and corresponding satellite images, scraped from Google Maps with the help of a script that I wrote.

Video by Fabian translating Stockhausen's scores into cities: https://vimeo.com/235286541



Computational Decomposition of Style for Controllable and Enhanced Style Transfer

Minchao Li Shikui Tu Lei Xu
Shanghai Jiao Tong University
{marshal_lmc,tushikui,leixu}@sjtu.edu.cn

Abstract

Neural style transfer has been demonstrated to be powerful in creating artistic image with help of Convolutional Neural Networks (CNN). However, there is still lack of computational analysis of perceptual components of the artistic style. Different from some early attempts which studied the style by some pre-processing or post-processing techniques, we investigate the characteristics of the style systematically based on feature map produced by CNN. First, we computationally decompose the style into basic elements using not only spectrum based methods including Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) but also latent variable models such Principal Component Analysis (PCA), Independent Component Analysis (ICA). Then, the decomposition of style induces various ways of controlling the style elements which could be embedded as modules in state-of-the-art style transfer algorithms. Such decomposition of style brings several advantages. It enables the computational coding of different artistic styles by our style basis with similar styles clustering together, and thus it facilitates the mixing or intervention of styles based on the style basis from more than one styles so that compound style or new style could be generated to produce styled images. Experiments demonstrate the effectiveness of our method on not only painting style transfer but also sketch style transfer which indicates possible applications on picture-to-sketch problems.

1. Introduction

Painting art, like Vincent van Gogh's "The Starry Night", have attracted people for many years. It is one of the most popular art forms for creative expression of the conceptual intention of the practitioner. Since 1990's, researches have been made by computer scientists on the artistic work, in order to understand art from the view of computer or to turn a camera photo into an artistic image automatically. One early attempt is Non-photorealistic rendering (NPR)[18], an area of computer graphics, which focuses on enabling artistic styles such as oil painting and drawing for digital images. However, NPR is usually limited to specific styles and hard to generalize to produce styled images for any other artistic styles.

One significant advancement was made by Gatys *et al*. in 2015 [7], called neural style transfer, which could separate the representations of the image content and style learned by deep CNN and then recombine the image content from one and the image style from another to obtain styled images. During this neural style transfer process, fantastic stylized images were produced with the appearance similar to a given real artistic work, such as Vincent van Gogh's "The Starrry Night". The success of the style transfer indicates that artistic styles are computable and are able to be migrated from one image to another. Thus, we could learn to draw like some artists apparently without being trained for years.

Following Gatys *et al*.'s pioneering work, a lot of efforts have been made to improve or extend the neural style transfer algorithm.[25] considered the semantic content and introduced the semantic style transfer network.[15] combined the discriminatively trained CNN with the classical Markov Random Field (MRF) based texture synthesis for better mesostructure preservation in synthesized images. Semantic annotations were introduced by [1] to achieve semantic transfer. To imporve the efficiency, [14] as well as [22] introduced a fast neural style transfer method, which is a feed-forward network to deal with a large set of images per training. With help of an adversarial training network, results were further improved in [16]. For a systematic review on neural style transfer, please refer to [13].

The success of recent progress on style transfer relies on the separable representation learned by deep CNN, in which the layers of convolutional filters automatically learns low-level or abstract representations in a more expressive feature space than the raw pixel-based images. However, it is still challenging to use CNN representations for style transfer due to their uncontrollable behavior as a black-box, and thus it is still difficult to select appropriate composition of

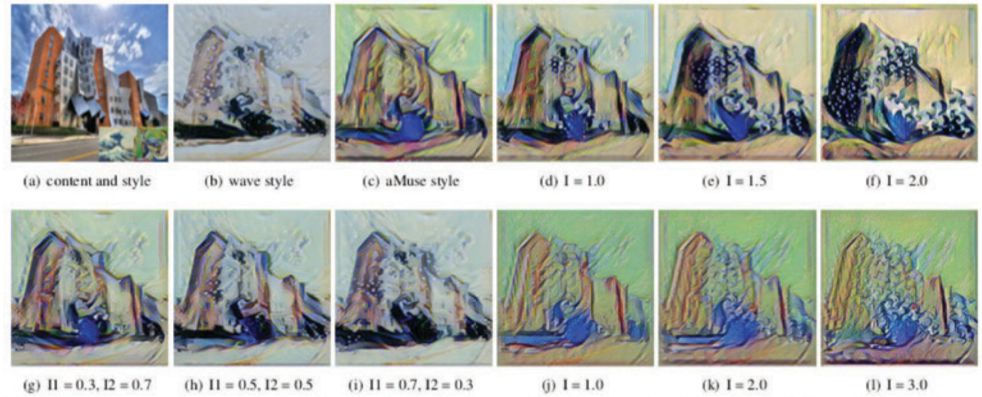


Figure 6. (a) the content image and two style images; (b)(c) styled image of single style using traditional methods [7]; (g-i) interpolation mixing where I1 and I2 are the weights of 'wave' and 'aMuse' in interpolation; (d-f,j-l) results of mixing the color of 'aMuse' and the stroke of 'wave' where I is the intervention to the stroke of 'wave'. Specifically, (d-f) use FFT; (j-l) use ICA.

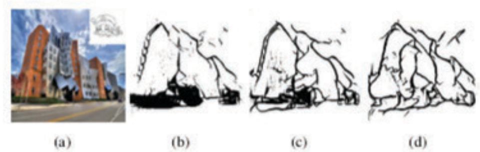


Figure 7. Picture-to-sketch using style transfer and binarization. (a) content image and style image; (b-d) styled images. From (b) to (d), the number of stroke increases as more details of the content image are restored.

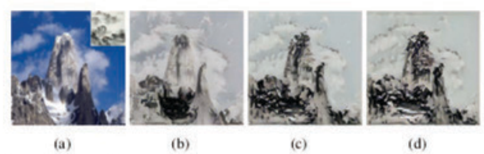


Figure 8. Neural style transfer of Chinese painting with stroke controlled. (a) content image and style image (by Zaixin Miao); (b-d) styled images. From (b) to (d), the strokes are getting more detailed which gradually turns freehand style into finebrush style.

4.5. Sketch style transfer

Picture-to-sketch problem challenges how computer can understand and represent the concept of objects both abstractly and semantically. State-of-the-art methods [2, 20] use variance model of genarative adversary network (GAN) via both supervised and un-supervised methods. One obstacle mentioned by [20] is that using supervised learning only may result in unstablity due to the noise in the dataset which is caused by variant sketch styles for the same data sample. Controllable neural style transfer proposed by us tackles the above obstacle because inconsistent styles are no more burdens, but can in turn increase the style diversity of output images. Moreover, as is shown in Figure 7, our method can control the abstract level by reserving major semantic details and minor ones automatically. In addition, our method does not require vector sketch dataset, but as the tradeoff, we cannot generate sketch stroke by stroke like [2, 20].

4.6. Chinese painting style transfer

Chinese painting is an exclusive artistic style which does not have plentiful color like the Western painting, but mostly represents the artistic conception by strokes. Taking advantage of effective controls over stroke via our methods, the Chinese painting styled image can be either misty-like (Figure 8(b)) which can be called as freehand-brush or meticulous representation (Figure 8(d)) which is called as fine-brush.

5. Conclusions

Artistic styles are made of basic elements, each with distinct characteristics and functionality. Developing such a style decomposition method facilitate the quantitative control of the styles in one or more images to be transfer to another natural image, while still keeping the basic content of natural image. In this paper, we proposed a novel computational decomposition method, and demonstrated its strengths via extensive experiments. To our best knowledge, it is the first such study, which could serve as a com-

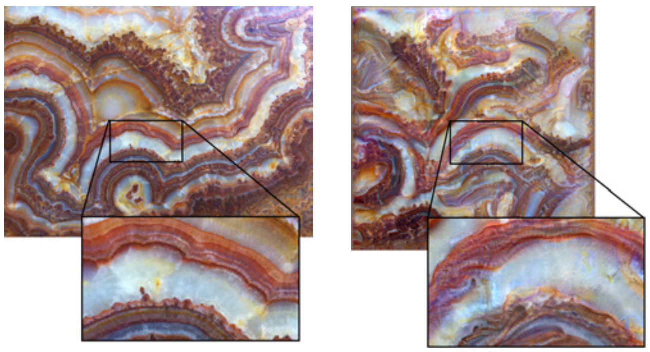
Image Style Transfer

- Find image representations that independently model variations in the semantic image content and the style which it is presented
- This calls for a need to develop a method for disentangling style and content in images
- Tutorial: https://pytorch.org/tutorials/advanced/neural_style_tutorial.html

High-Resolution Multi-Scale Neural Texture Synthesis

High-Resolution Multi-Scale Neural Texture Synthesis

Xavier Snelgrove
Toronto, Ontario, Canada
xs@wxs.ca



(a) Source texture (b) Synthesized image

Figure 1: High resolution texture synthesis matching CNN texture statistics at 5 image scales

ABSTRACT

We introduce a novel multi-scale approach for synthesizing high-resolution natural textures using convolutional neural networks trained on image classification tasks. Previous breakthroughs were based on the observation that correlations between features at intermediate layers of the network are a powerful texture representation, however the fixed receptive field of network neurons limits the maximum size of texture features that can be synthesized.

We show that rather than matching statistical properties at many layers of the CNN, better results can be achieved by matching a small number of network layers but across many scales of a Gaussian pyramid. This leads to qualitatively superior synthesized high-resolution textures.

CCS CONCEPTS

• **Computing methodologies → Texturing**;

KEYWORDS

texture synthesis, neural networks, Gaussian pyramid

SA '17 Technical Briefs, November 27–30, 2017, Bangkok, Thailand

ACM Reference Format:
Xavier Snelgrove. 2017. High-Resolution Multi-Scale Neural Texture Synthesis. In *SA '17 Technical Briefs: SIGGRAPH Asia 2017 Technical Briefs, November 27–30, 2017, Bangkok, Thailand.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3145749.3149449

1 INTRODUCTION

There have been recent significant improvements in the quality of example-based texture synthesis techniques by taking advantage of intermediate representations in a convolutional neural network (CNN) trained to classify images [Gatys et al. 2015b].

Correlations between features maps at these intermediate layers, represented by a Gram matrix, turn out to be a powerful representation of texture. By synthesizing new images whose Gram matrix is close to that of an exemplar image (for instance via gradient descent), we get images with similar texture.

However, this only works well when the semantically significant features in the image are at the correct scale for the network, and in practice the receptive field of a feature at an intermediate layer for common CNN architectures is relatively small [Luo et al. 2016]. The popular VGG architectures from Simonyan and Zisserman [2014], used by Gatys et al. and others, are trained on 224×224 pixel images, in which relevant features will be quite a bit smaller.

So, given a high-resolution source image, for optimal results an artist must scale down that image until the pixel-scale of the features of interest match the receptive field of the appropriate semantic layer of the network. This limits the resolution of the rendered image, and further breaks down for source images with

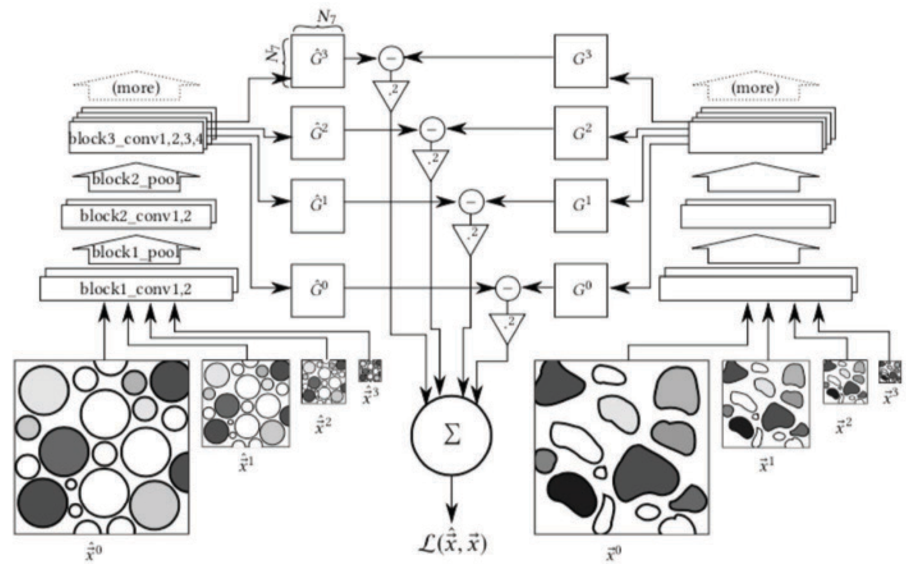


Figure 2: Block diagram of our system. The left-hand-side shows the source image, with Gram matrices of feature correlations being extracted for VGG-19 layer `block3_conv2` across 4 distinct spatial scales. On the right is the current state of the optimization $\vec{x}$, whose feature correlation matrices are extracted in the same fashion. The sum of the squared difference of these matrices corresponds to the loss function $\mathcal{L}$ which is minimized by gradient descent on $\vec{x}$.

and `block3_conv2` (the 3rd and 8th layers). Using both a low level layer and an intermediate layer sometimes allows the optimization to escape local optima that it would get stuck in using only the higher level layer.

Figure 3 shows an example result. The source texture in 3a shows texture at multiple scales. At the macro scale it is dominated by regions of blue paint and rusty-red wall, with red streaks in the blue regions. At the micro-scale we see that the red region has a rough texture, whereas the blue region has a smooth texture, and a distinct rounded "paint-chipping" shape at the border between the regions. All of these effects are captured in the synthesized image 3b. Using Gatys et al. we find that some of the micro-scale properties are captured, but the macro-scale red/blue alternation is almost completely missing. Further we can see strong artefacts around the bright blue point in the corner. Refer to Figure 4 for many more results.

Experiments were performed using the Keras [Chollet et al. 2015] neural networks framework. We used the Scipy [Jones et al. 2001] implementation of L-BFGS-B [Zhu et al. 1997]. The code for these experiments is available online[2].

5 DISCUSSION

This work has shown the power of using multi-scale representations in conjunction with neural texture synthesis. Gatys et al. can be seen as a special-case of this work for a single scale octave $S = 1$. As such, much of work building on their approach would also be applicable to ours.

A natural extension would be to attempt multi-scale style transfer, following [Gatys et al. 2015a] but using our objective function to represent the "style loss".

We use a simple optimization process to synthesize our images. Other work has shown that feed-forward neural networks can be trained to approximate this optimization [Johnson et al. 2016] so it would be interesting to attempt those methods with our multi-scale objective. It is likely that some similar multi-scale approach would be required for the architecture of that network as otherwise the same receptive-field issues would apply for that feed-forward network.

Finally, combing our work with other approaches for finding larger structures in images such as by finding symmetries as in [Berger and Memisevic 2016; Sendik and Cohen-Or 2017] would also be interesting.

[2] http://github.com/wxs/subjective-functions

Gram Matrices (Snelgrove)

High-Resolution Multi-Scale Neural Texture Synthesis

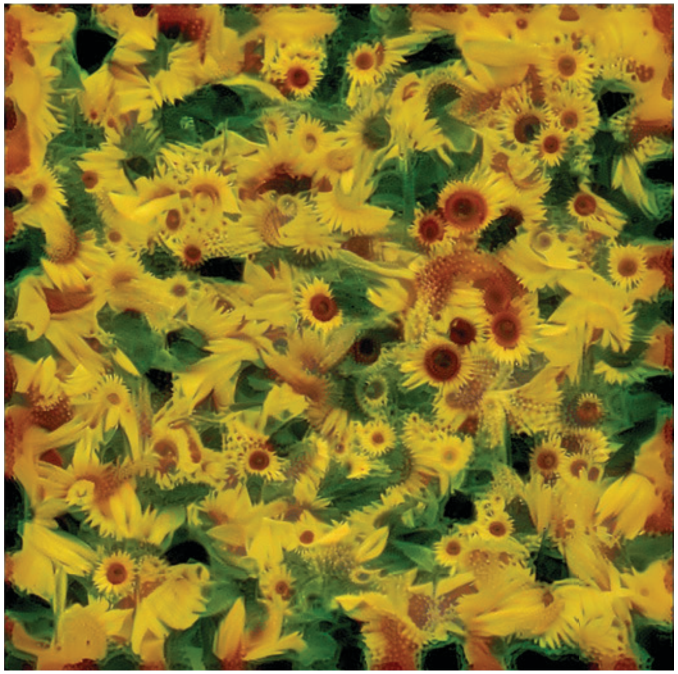
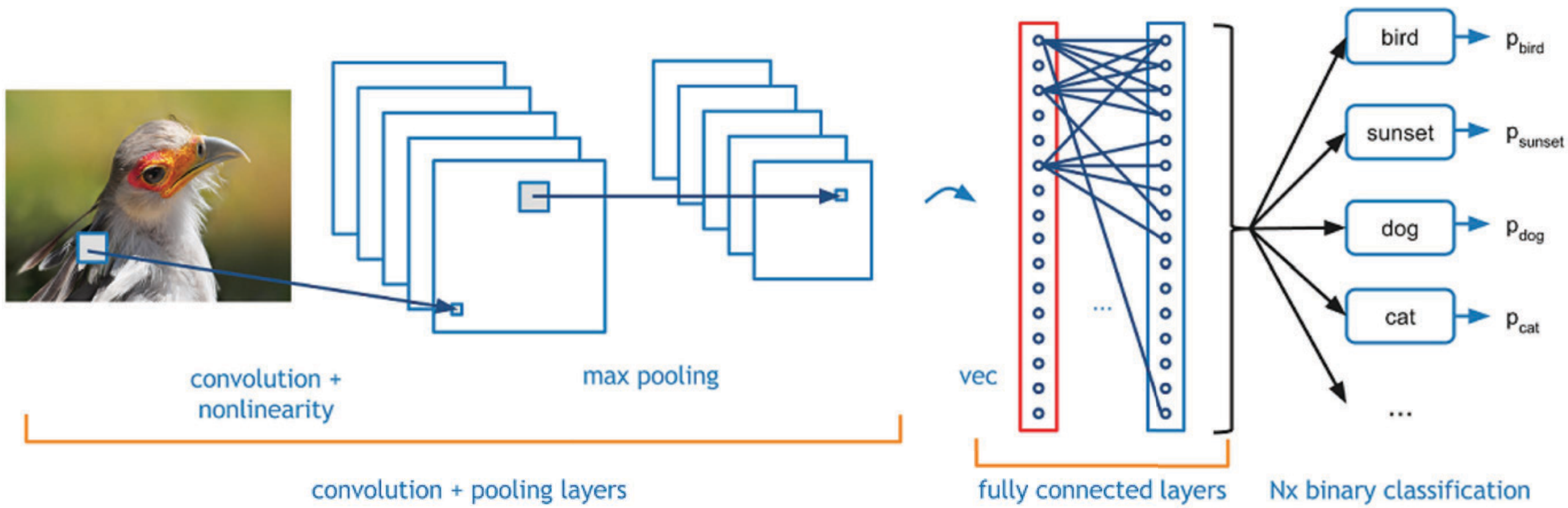


Image Analysis & Convolutional Neural Networks

- **Image analysis** is the extraction of meaningful information from images; mainly from digital images by means of digital image processing techniques
- Processes include feature extraction, pattern recognition, image classification, etc. through Computer Vision fundamentals used in Convolutional Neural Networks
- Convolutional Neural Networks uses in image classification



https://heartbeat.fritz.ai/the-5-computer-vision-techniques-that-will-change-how-you-see-the-world-1ee19334354b

Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros

Berkeley AI Research (BAIR) Laboratory, UC Berkeley

{isola,junyanz,tinghuiz,efros}@eecs.berkeley.edu

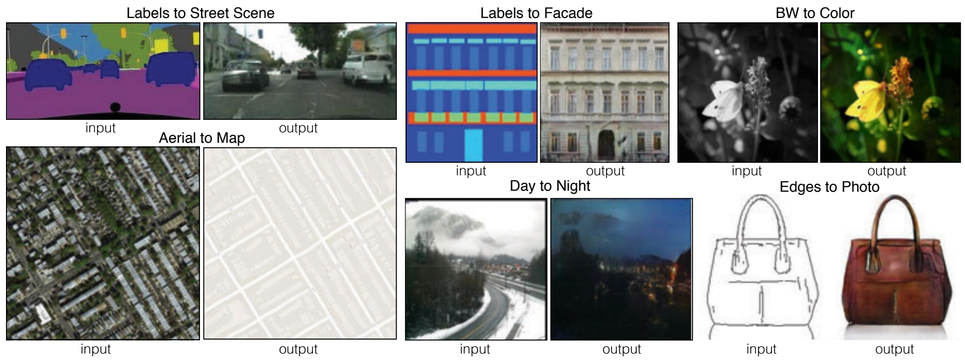


Figure 1: Many problems in image processing, graphics, and vision involve translating an input image into a corresponding output image. These problems are often treated with application-specific algorithms, even though the setting is always the same: map pixels to pixels. Conditional adversarial nets are a general-purpose solution that appears to work well on a wide variety of these problems. Here we show results of the method on several. In each case we use the same architecture and objective, and simply train on different data.

Abstract

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. Indeed, since the release of the pix2pix *software associated with this paper, a large number of internet users (many of them artists) have posted their own experiments with our system, further demonstrating its wide applicability and ease of adoption without the need for parameter tweaking. As a community, we no longer hand-engineer our mapping functions, and this work suggests we can achieve reasonable results without hand-engineering our loss functions either.*

1. Introduction

Many problems in image processing, computer graphics, and computer vision can be posed as "translating" an input image into a corresponding output image. Just as a concept may be expressed in either English or French, a scene may be rendered as an RGB image, a gradient field, an edge map, a semantic label map, etc. In analogy to automatic language translation, we define automatic *image-to-image translation* as the task of translating one possible representation of a scene into another, given sufficient training data (see Figure 1). Traditionally, each of these tasks has been tackled with separate, special-purpose machinery (e.g., [16, 25, 20, 9, 11, 53, 33, 39, 18, 58, 62]), despite the fact that the setting is always the same: predict pixels from pixels. Our goal in this paper is to develop a common framework for all these problems.

The community has already taken significant steps in this direction, with convolutional neural nets (CNNs) becoming the common workhorse behind a wide variety of image prediction problems. CNNs learn to minimize a loss function – an objective that scores the quality of results – and although the learning process is automatic, a lot of manual effort still http://www.memo.tv/portfolio/learning-to-see/

Learning to see (2017)

An artificial neural network looking out onto the world, trying to make sense of what it sees, in context of what it's seen before.

It can only see what it already knows, just like us.

"Learning To See" is an ongoing series of works that use state-of-the-art Machine Learning algorithms as a means of reflecting on ourselves and how we make sense of the world. The picture we see in our conscious mind is not a mirror image of the outside world, but is a reconstruction based on our expectations and prior beliefs. In "Learning To See", an artificial neural network loosely inspired by our own visual cortex, looks through cameras and tries to make sense of what it sees. Of course it can only see what it already knows. Just like us.

The work is part of a broader line of inquiry about self affirming cognitive biases, our inability to see the world from others' point of view, and the resulting social polarization.

The series consists of a number of studies, each motivated by related but different ideas.

Related work

Learning to See: Hello, World!

Learning to Dream: Supergan

Dirty Data

FIGHT!

All watched over by machines of loving grace. Deepdream edition

Keeper of our collective consciousness

Related texts

All watched over by machines of loving grace

A digital god for a digital culture. Resonate 2016

Deepdream is blowing my mind

Background

Originally loosely inspired by the neural networks of our own brain, Deep Learning Artificial Intelligence algorithms have been around for decades, but they are recently seeing a huge rise in popularity. This is often attributed to recent increases in computing power and the availability of extensive training data. However, progress is undeniably fueled by multi-billion dollar investments from the purveyors of mass surveillance – technology companies whose business models rely on targeted, psychographic advertising, and government organizations focussed on the War on Terror. Their aim is the automation of Understanding Big Data, i.e. understanding text, images and sounds. But what does it mean to 'understand'? What does it mean to 'learn' or to 'see'?

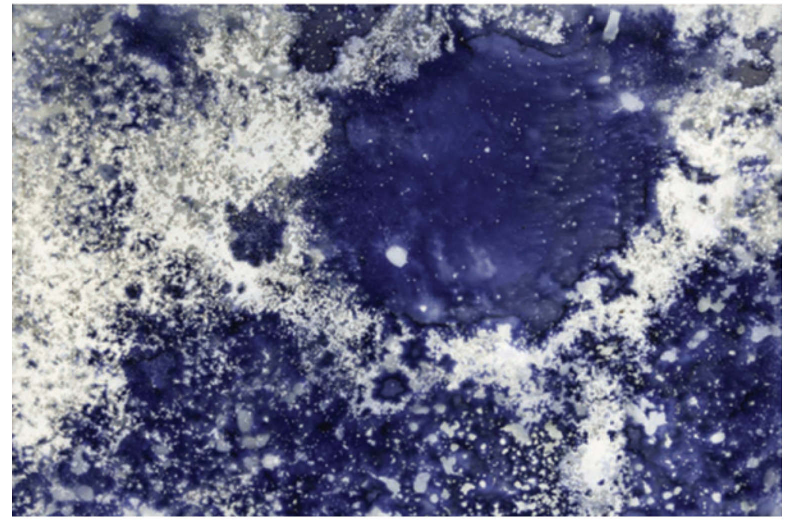
Datasets

These images are **not** using 'style transfer'. In style transfer, the network is generally run on, and contains information on, a *single image*. These networks contain *knowledge of the entire dataset*,

New abstract photography– How to describe the style and content?

Marco Breuer

Carolyn Marks Blackwood



Meghann Riepenhoff

Joachim Schulz

Ellen Carey

Mariah Robertson

Matthew Brandt