

22f M255 – From Photo to Deep Fakes

George Legrady © 2022

Experimental Visualization Lab

Media Arts & Technology

University of California, Santa Barbara

Basics

Tuesdays: Lectures

Thursdays: Studio lab work and sharing

--

Previous Syllabus:

<https://www.mat.ucsb.edu/~g.legrady/academic/courses/21s255/21s255.html>

--

Student forum: <https://w2.mat.ucsb.edu/forum/viewforum.php?f=86>

--

Course focus: What is an image today?

--

Critical Questions

We visually represent the world, ideas, and explore new forms by which to visualize and new kinds of visualization

We continue to believe in the photographic image even though computation has further transformed it into a fictional document

--

Text-to-Image Synthesis, Dalle-2, MidJourney, etc. where an image is created by software based on CNN applications to an existing database of images according to semantic descriptions and metadata

Weekly Lectures on Tuesdays / Image production on Thursdays

- **Samples form the history of photography**
- **Raster digital image, the Pixel**
- **Image processing overview (noise, 2D convolution**
- **Computational Photography**
- **Computational Aesthetics (Engineering)**
- **Aesthetic components of an image (Artistic)**
- **Image Synthesis: Artificial generation of images**
- **Convolution Neural-networks, machine learning, deep learning**
- **Style Transfer, Style Gans, Generative Adversarial Networks**
- **Generative Art, human-in-the-loop computational image generation**
- **Semiotics, cultural analysis, social signification, belief in the image**
- **Critical perspectives with text-to-image software**

Image Capture: Optics, Photons, Image Recording

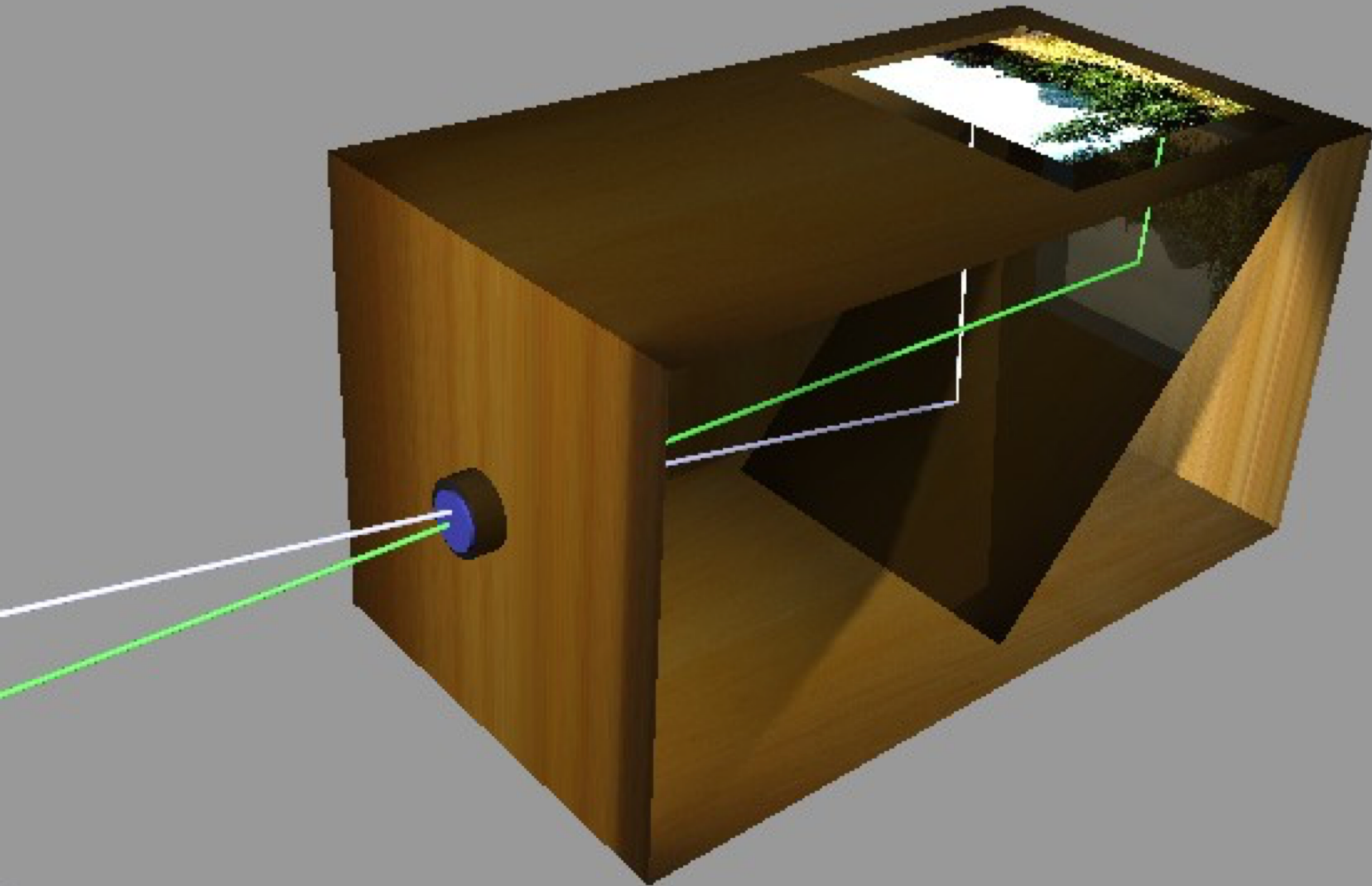
Optics: the lens through which directed light enters the darkened recording space

Photons: Light travels from objects in the world are constrained by the optics

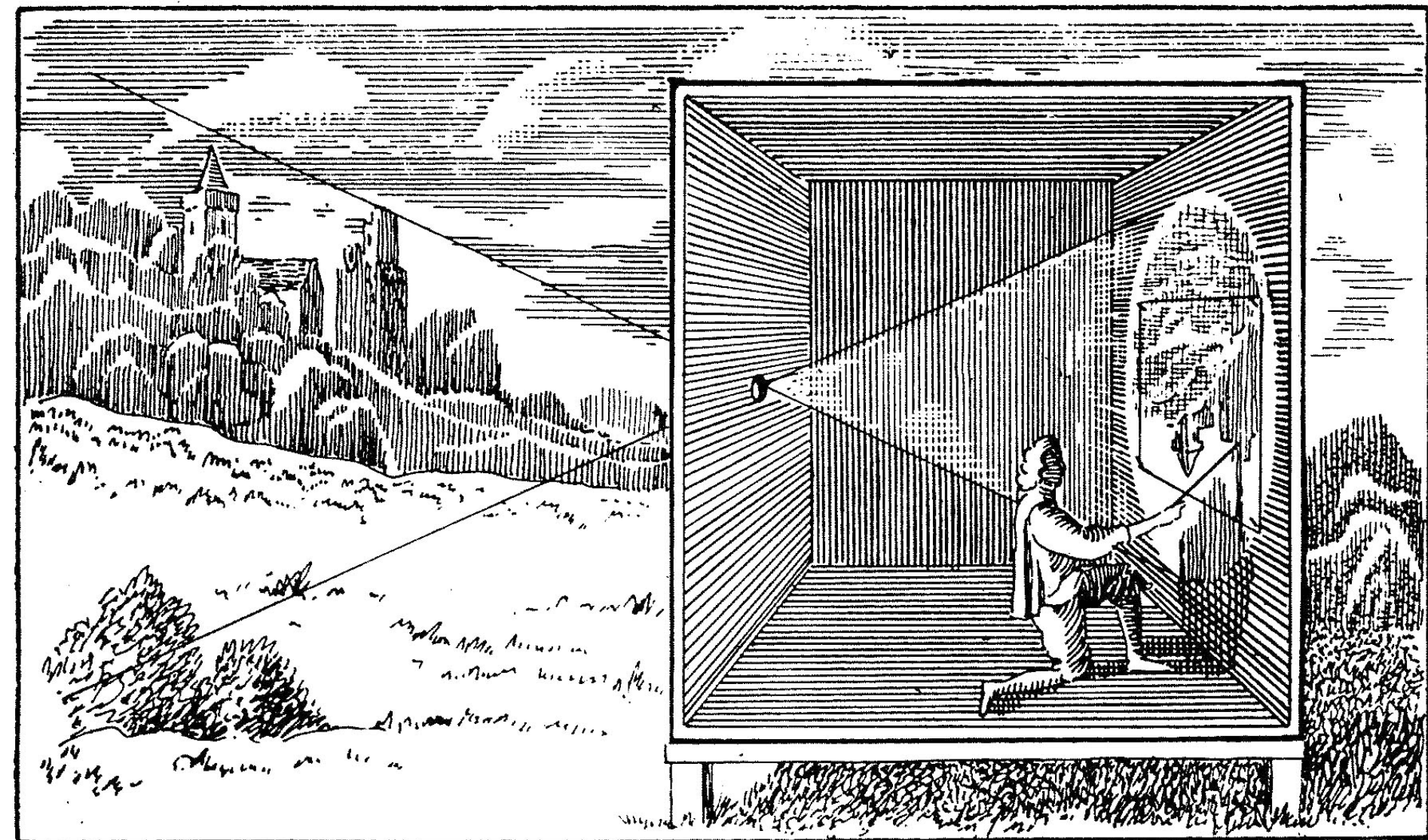
Image Recording: Chemical or electronic sensors that record the photons



Dark chamber lets photons (lightbeams) thru pinhole/lens, then captured on receiving flat surface



Tracing the projection inside a Camera Obscura



Chemical Development of a photograph

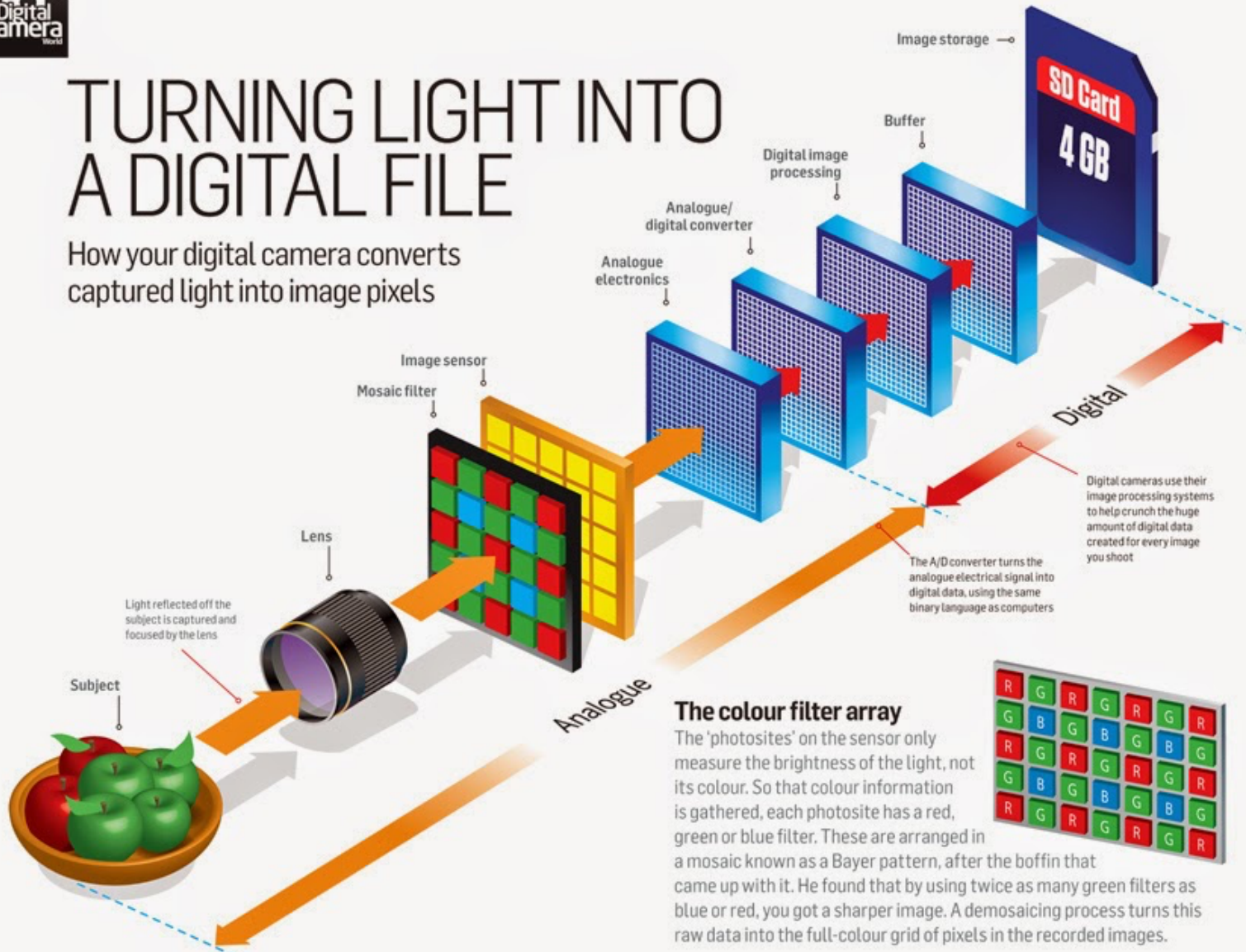


Digital Camera captures image through photo-sensors



TURNING LIGHT INTO A DIGITAL FILE

How your digital camera converts captured light into image pixels



Digital cameras use their image processing systems to help crunch the huge amount of digital data created for every image you shoot

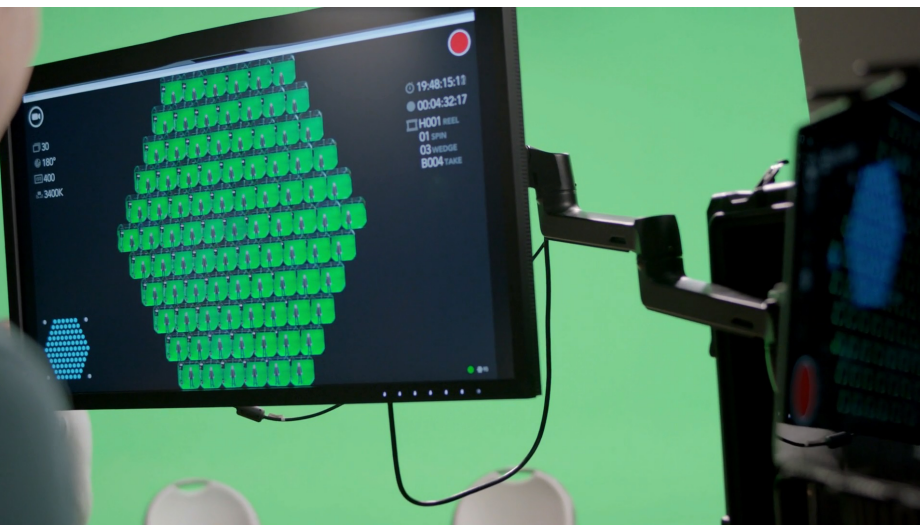
The A/D converter turns the analogue electrical signal into digital data, using the same binary language as computers

The colour filter array

The 'photosites' on the sensor only measure the brightness of the light, not its colour. So that colour information is gathered, each photosite has a red, green or blue filter. These are arranged in a mosaic known as a Bayer pattern, after the boffin that came up with it. He found that by using twice as many green filters as blue or red, you got a sharper image. A demosaicing process turns this raw data into the full-colour grid of pixels in the recorded images.



Lytro – variable depth-of-field



Digital Image Processing & Deep Fakes

Digitization also allows for various techniques such as **statistical averaging**

Deep Fakes: A blend of “Deep Learning” with “fake” – photo manipulation

A synthetic image where data from one source is integrated into another (a person is replaced with another)



People have been faking photos since long before digital tools came along.





How DALL·E 2 Works

⊕ DALL·E 2 is a system for text-to-image generation developed by my coauthors and me at OpenAI. When prompted with a caption, the system will attempt to generate a novel image from scratch that matches it. It also has additional capabilities like:

- **Inpainting:** perform edits to an image using language;
- **Variations (Figure 1):** generate new images that share the same essence as a given reference image, but differ in how the details are put together; and
- **Text diffs (Figure 4):** transform any aspect of an image using language.

The system underlying DALL·E 2, which we call unCLIP, is based on two key technologies: CLIP and diffusion. As stated in the blog, CLIP is a model that “efficiently learns visual concepts from natural language supervision”. Diffusion is a technique to train a generative model for images by learning to undo the steps of a fixed corruption process. We briefly describe both of these technologies next.

⊕ CLIP consists of two neural networks – a text encoder and an image encoder – that are trained on a large, diverse collection of image-text pairs. Each encoder maps its input to a point on a globe (known as an *embedding*) that functions as a “concept space” shared by both modalities. During each step of training, CLIP receives a list of images and a corresponding list of captions that describe them. Using this data, we can form two types of image-text pairs: a *matching* pair, in which an image is paired up with its corresponding caption, and a *mismatching* pair, in which an image is paired up with any other caption. The encoders are trained to

CLIP-CLOP: CLIP-Guided Collage and Photomontage

Piotr Mirowski, Dylan Banarse, Mateusz Malinowski, Simon Osindero, Chrisantha Fernando

DeepMind, London, UK

{piotrmirowski, dylski, mateuszm, osindero, chrisantha} @ deepmind.com

Abstract

The unabated mystique of large-scale neural networks, such as the CLIP dual image-and-text encoder, popularized automatically generated art. Increasingly more sophisticated generators enhanced the artworks' realism and visual appearance, and creative prompt engineering enabled stylistic expression. Guided by an artist-in-the-loop ideal, we design a gradient-based generator to produce collages. It requires the human artist to curate libraries of image patches and to describe (with prompts) the whole image composition, with the option to manually adjust the patches' positions during generation, thereby allowing humans to reclaim some control of the process and achieve greater creative freedom. We explore the aesthetic potentials of high-resolution collages, and provide an open-source Google Colab as an artistic tool.

Introduction

A *collage*, from the French *coller*, is “a composite image made by sticking newspaper cuttings, photographs, and other printed images onto a flat surface, often combined with paint” (Zaczek and Actor 2008). *Photomontage* extends collage by manipulating and compositing photographs (Ades 1976). The origins of collage can be traced to the invention of paper in China, and *photo-collage* was a social pastime for the Victorian upper-class (of Scotland 2019), before Cubists Pablo Picasso and Georges Braque made collage into an art form (Zaczek and Actor 2008; Greenberg 1958).

In this paper, we formalize collage as a picture produced by optimizing affine spatial and color transformations of



Figure 1: *The Fall of the Damned after Rubens and Eaton*. High-resolution collage of image patches of animals (Fig.7), optimized hierarchically with 3x3 overlapping CLIP critics.

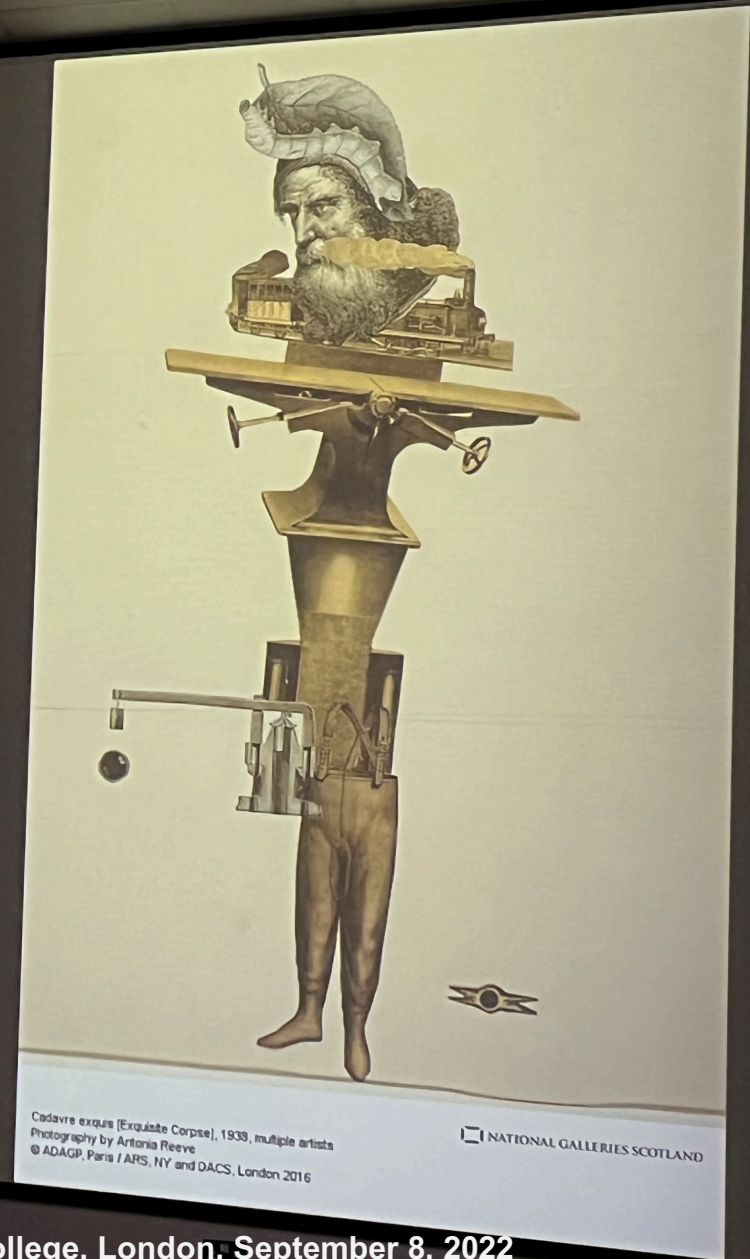
prompt and the resulting *collage image*. Therefore, it acts as an AI-based *Critic* assessing the “quality” of the artwork given its description. Large-scale dual encoders exhibit some degree of semantic compositionality, as they allow novel combinations of phrases or images, and handle such visual concepts as color, texture, shape, object rela-

Embracing the absurd

Exquisite corpses

Game played by the **Surrealists**, ca 1925
(André Breton, Yves Tanguy,
Marcel Duchamp, Jacques Prévert, etc...)

Le
Le cadavre
cadavre exquis
exquis boira
boira le
le vin
vin nouveau



Cadavre exquis (Exquisite Corpse), 1938, multiple artists
Photography by Antonia Reeve
© ADAGP, Paris / ARS, NY and DACS, London 2016

NATIONAL GALLERIES SCOTLAND

“Ricking the Reed”, Platinum print, Peter Henry Emerson (1886)



Lev Manovich (midJourney)



Lev Manovich (midJourney)



Lev Manovich (midJourney)



Lev Manovich (midJourney)



Lev Manovich (midJourney)



Lev Manovich (midJourney)



Lev Manovich



Stable Diffusion (AI Image Synthesis tool) has a parameter called Cfg scale. It adjusts how much the image will be like your prompt. Higher values keep your image closer to your prompt" (from documentation.) The scale values range from 0 to 20. In this experiment I used the same prompt and changed this parameter from 1 to 10.

The prompt text specified a still-life with a water bottle and a wine glass by 17th century Dutch artist Willem Claesz. Heda (1593 - 1680). He was indeed famous for his still-lives, but I did not try to match them, and I did not use any of his paintings images as references. The still-life the AI generated was sufficient. What I wanted to see is how it will change with different values of this parameter.

Lev Manovich

Lev Manovich is an author of books on digital culture and new media, and professor of Computer Science at the Graduate Center, City University of New York. Manovich's current research and teaching focuses on digital humanities, social computing, new media art and theory, and software studies.^[3]

Manovich is also the founder and director of the Cultural Analytics Lab (called Software Studies Initiative 2007-2016),^[4] which was described in an associated press release as computational analysis of massive collections of images and video (cultural analytics).^[5] His lab was commissioned to create visualizations of cultural datasets for Google,^[6] New York Public Library,^[7] and New York's Museum of Modern Art (MoMA).^[8]

One of his books, *The Language of New Media*, has been translated into thirteen languages.^[9] Manovich's latest academic book *Cultural Analytics* was published in 2020 by the MIT Press.^[10]

Course Work

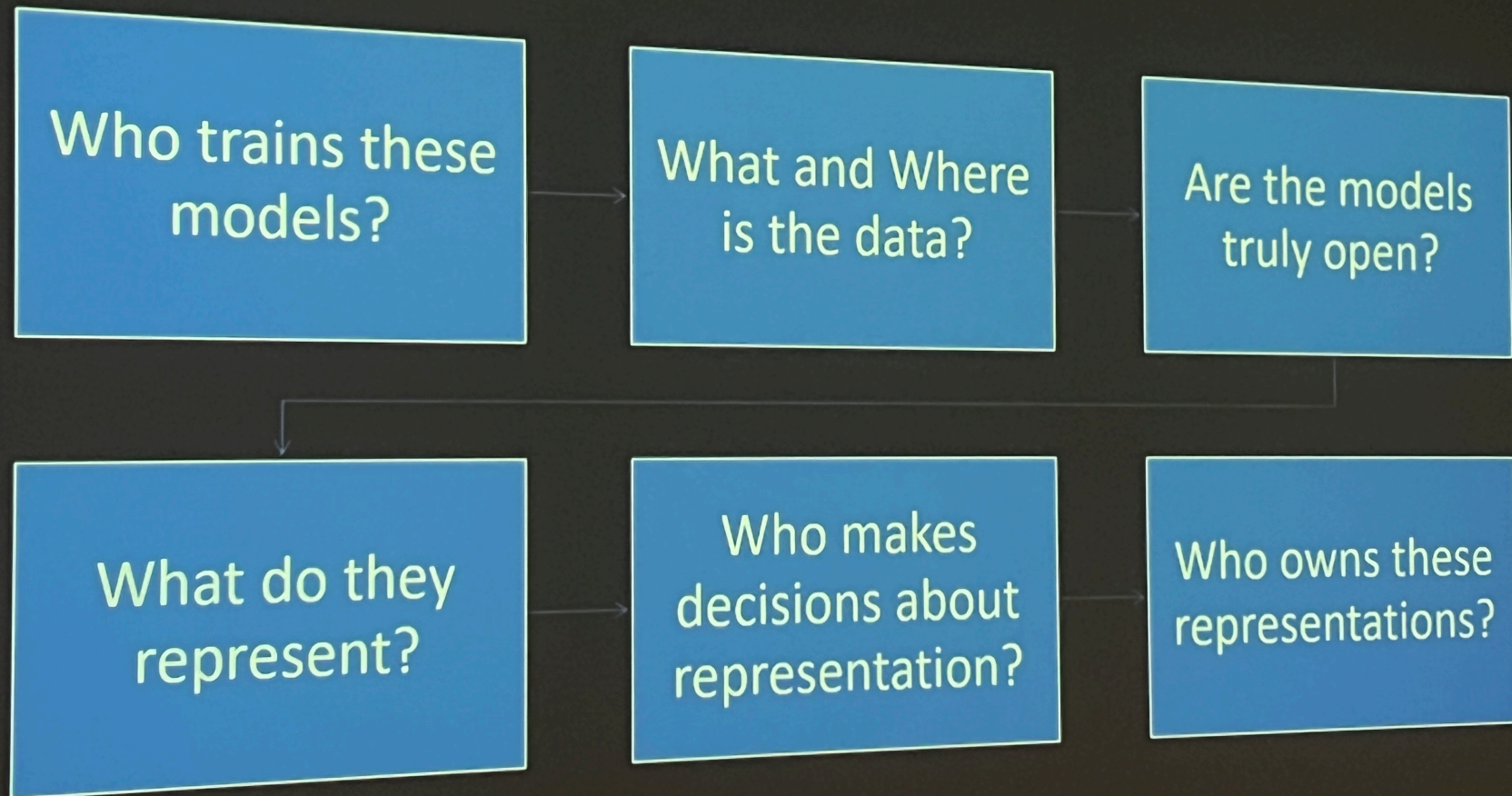
- **Weekly Image production with text-to-image software**
- **Short weekly reports on Tuesday Lectures:** Thoughts you may have about any of the materials covered in the weekly presentations
- **Final Report:** final report that analyses the image production you have achieved through the course

Analytical & Research Approach

- **Practice:** The course will involve the production of images through text-to-image software such as midjourney
- **Engineering:** What are the operations by which to generate computational – based images
- **Artistic:** What are the aesthetic aspects and history of the production of images
- **Humanities:** Analyses of how we construct images, and how we make sense of images
- **Software Development Issues:** Any software is defined by the knowledge range of the developer

Transformers

Pre-trained transformers



Digital Cultures

<https://www.youtube.com/watch?v=ButDfjQohB0>

1:19 is it like the invention of photography behnaz farahi

1:41 Discussion - Aaron Hertzmann – Dalle-statistical data

To be continued...