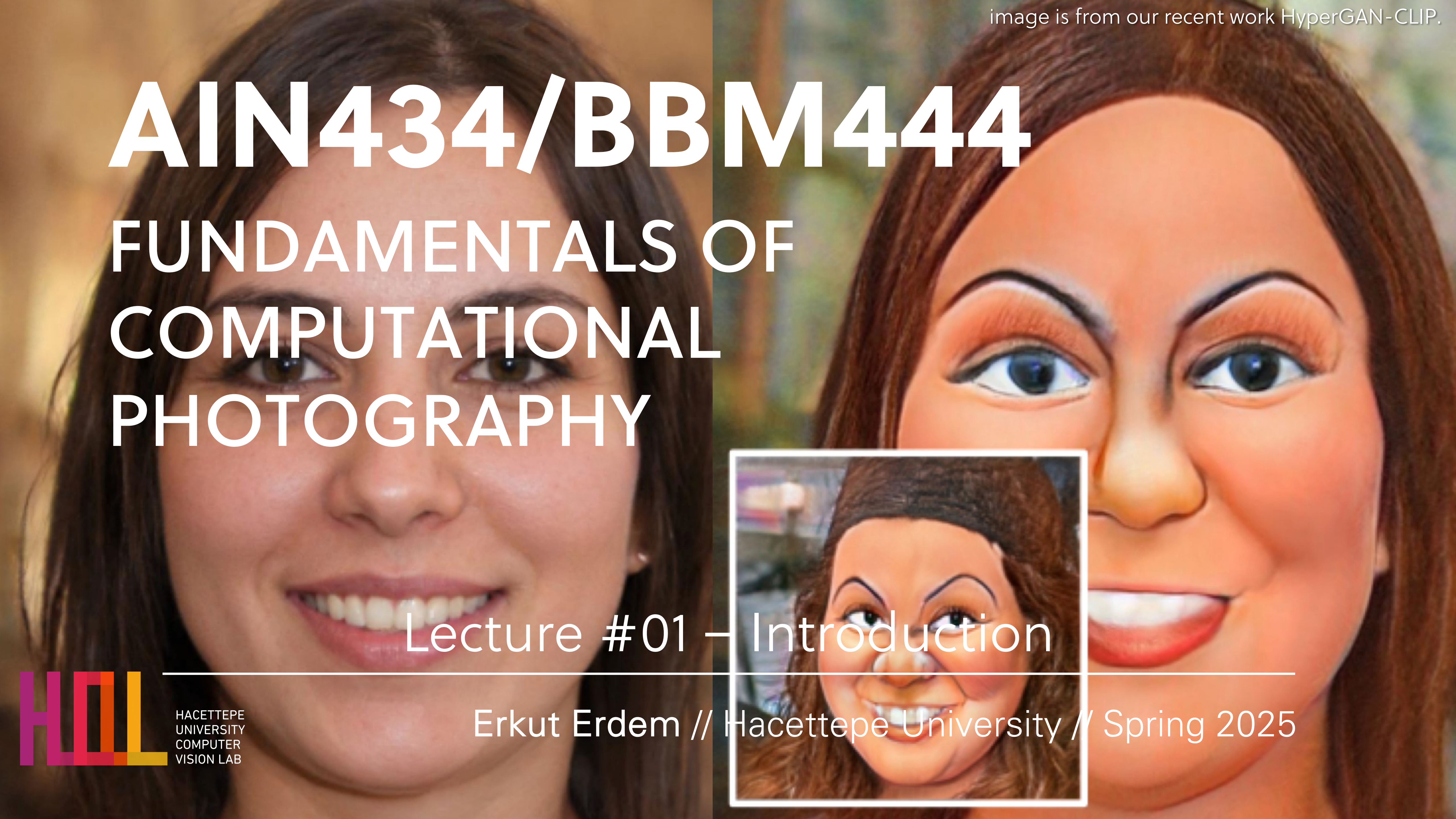


image is from our recent work HyperGAN-CLIP.

AIN434/BBM444

FUNDAMENTALS OF COMPUTATIONAL PHOTOGRAPHY

Lecture #01 – Introduction

Erkut Erdem // Hacettepe University // Spring 2025

HACETTEPE UNIVERSITY COMPUTER VISION LAB

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

Disclaimer: Some of the material and slides for this lecture were borrowed from

—Alexei Efros's CS194-26/294-26 "Intro to Computer Vision and Computational Photography" class

—Steve Marschner's CS6640 "Computational Photography" class

—Fredo Durand's slides on "The History of photography"

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

Welcome to AIN434/BBM444

- An advanced undergraduate course is about the fundamentals of computational photography
- Introduces students a number of different computational techniques to capture, manipulate and enrich visual media.

A little about me...

Koç University-İş Bank
Artificial Intelligence Center
Adjunct Faculty
2020-now

Hacettepe University
Professor
2010-now

Télécom ParisTech
Post-doctoral Researcher
2009-2010

Middle East Technical University
1997-2008
Ph.D., 2008
M.Sc., 2003
B.Sc., 2001

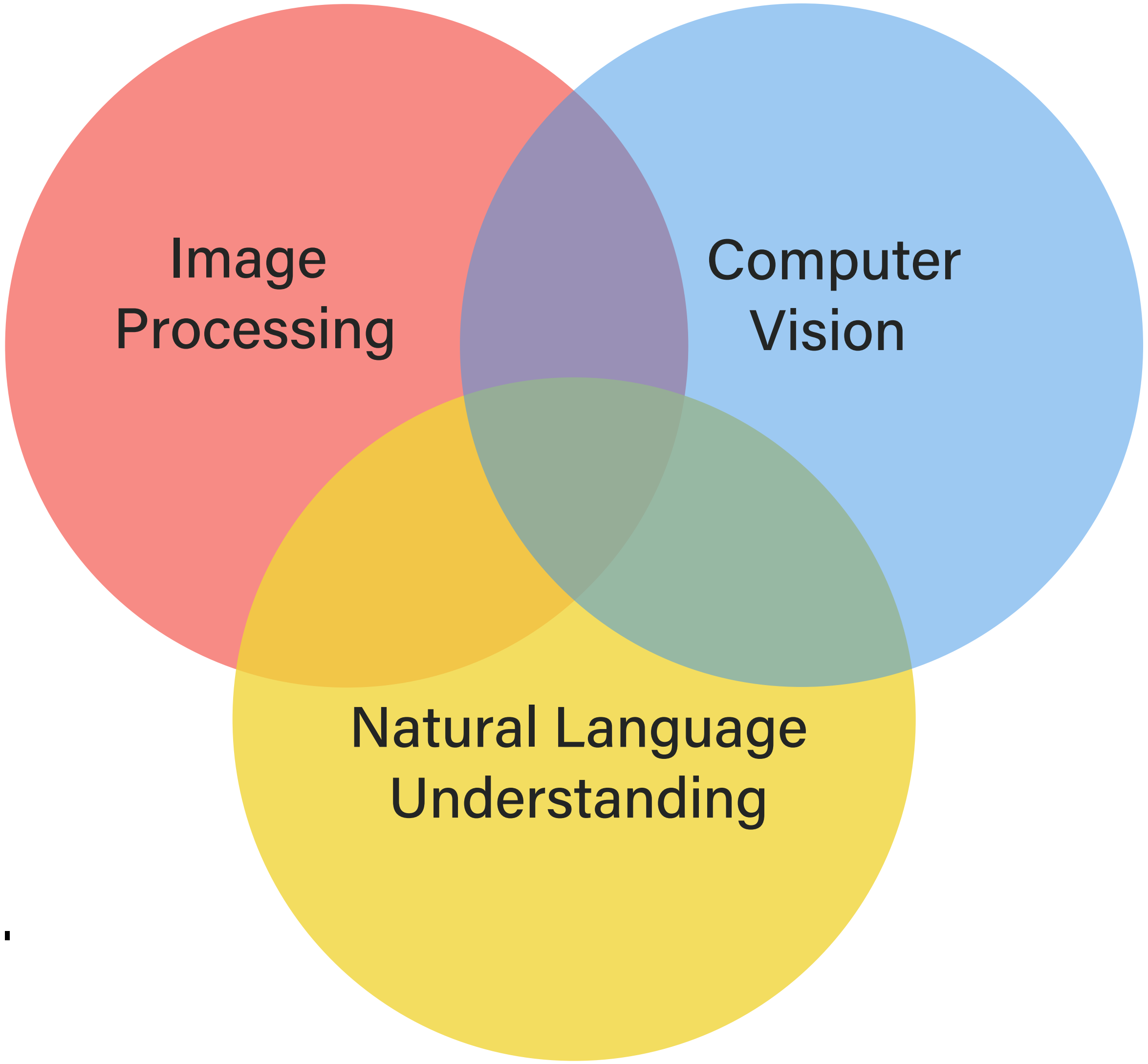
UCLA
Fall 2007
Visiting Student

VirginiaTech
Visiting Research Scholar
Summer 2006

http:// web.cs.hacettepe.edu.tr/~erkut
@erkuterdem
erkut@cs.hacettepe.edu.tr

Research Interests

- I study better ways to understand and process visual data.
- My research interests span a diverse set of topics, ranging from image editing to visual saliency estimation, and to multimodal learning for integrated vision and language.



Course Logistics

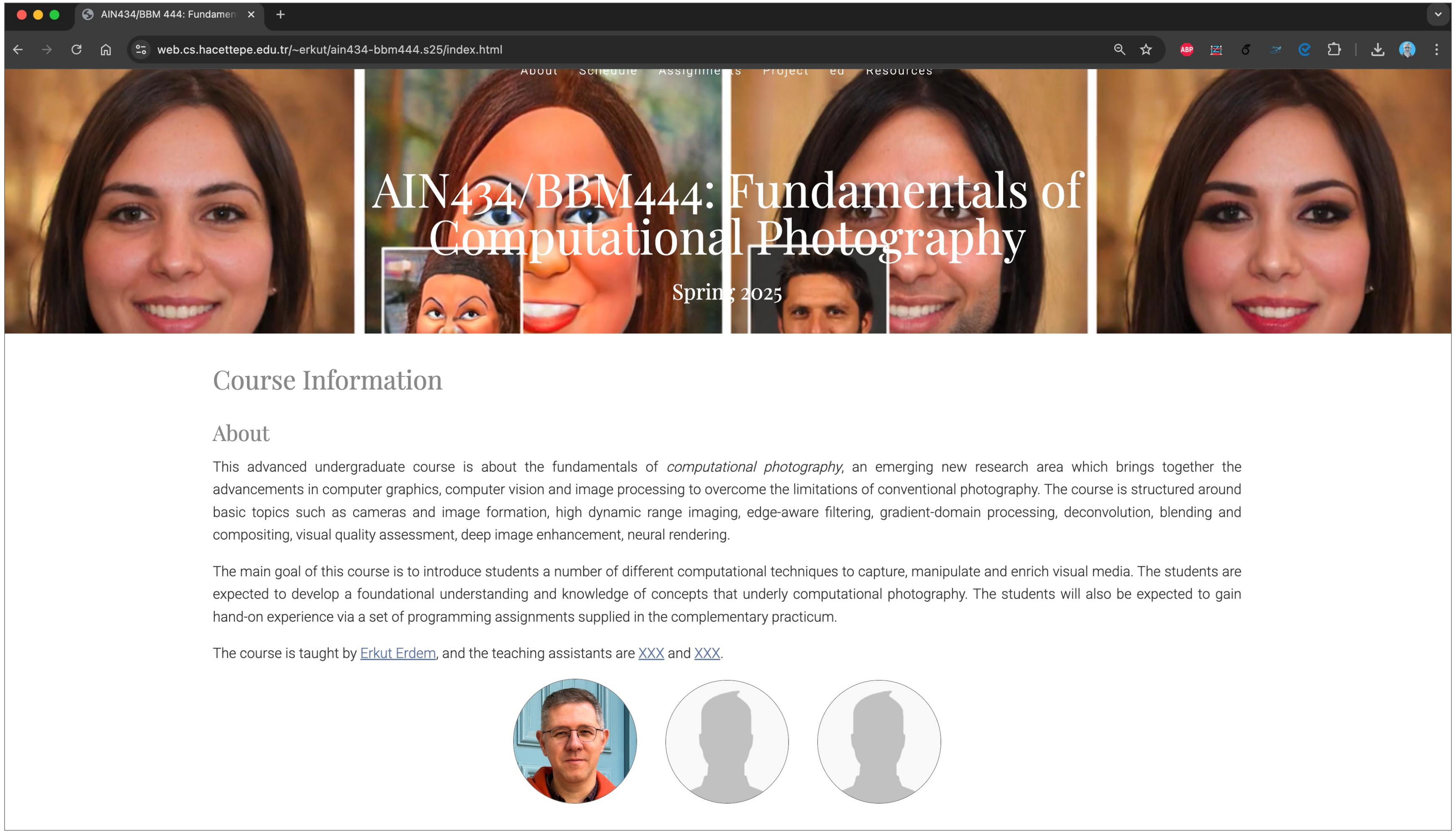
Course information

Time/Location	09:40-12:30pm Monday, D9 (AIN434/BBM444)
	16:40-17:30pm Monday, D8 (AIN435/BBM446)
Instructor	Erkut Erdem

- ed for course related announcements:
  https://edstem.org/eu/courses/2002
- Course webpage:
  https://web.cs.hacettepe.edu.tr/~erkut/ain434-bbm444.s25/index.html

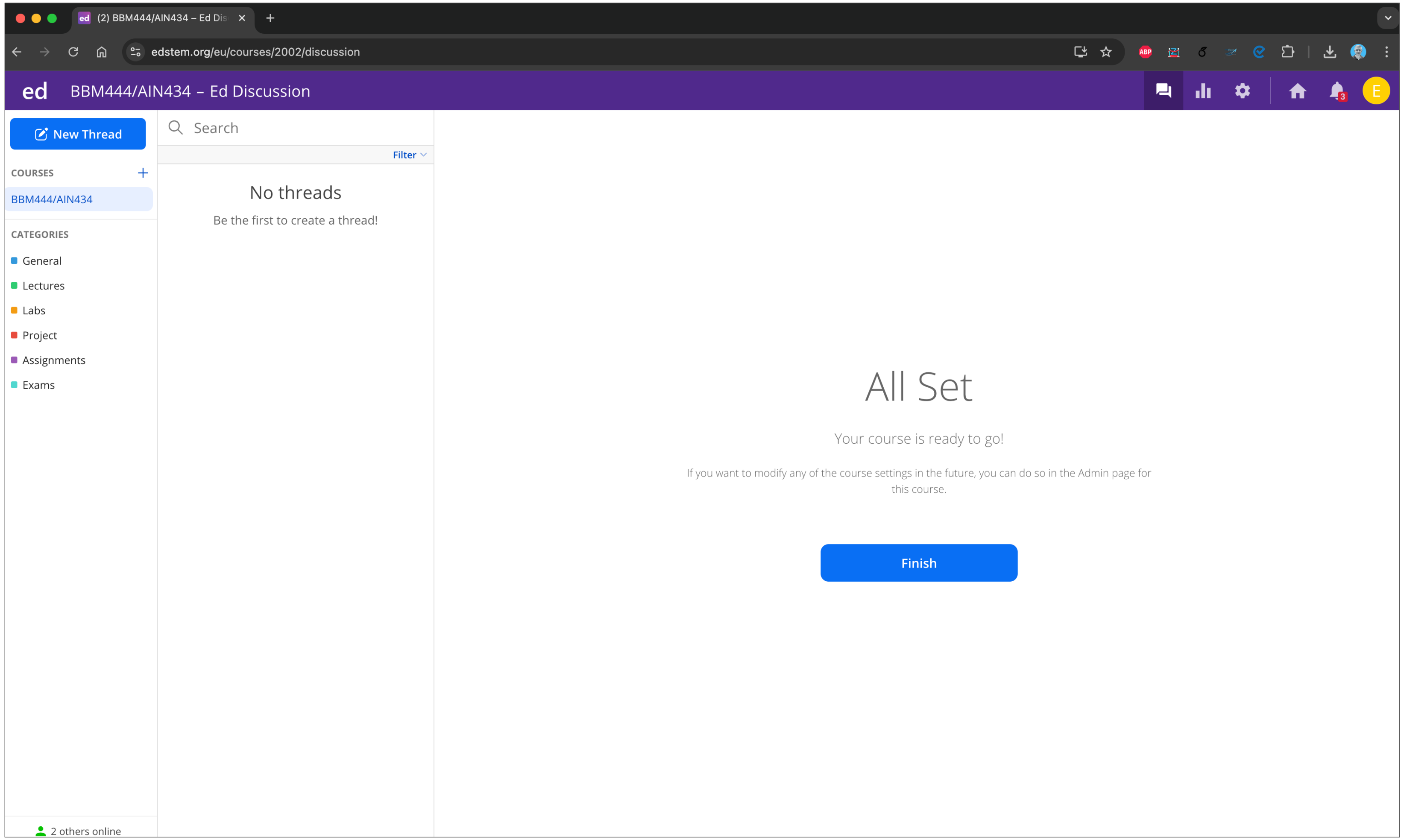
Course webpage

- will be updated regularly.
- will include lecture slides, additional reading material, course-related resources, and information about assignments and projects.

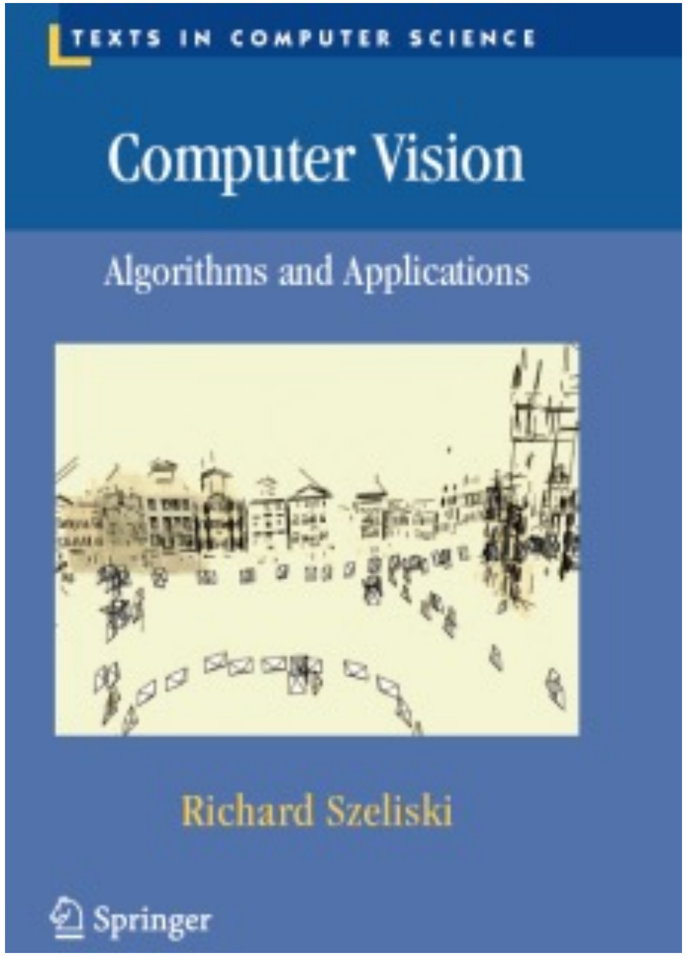


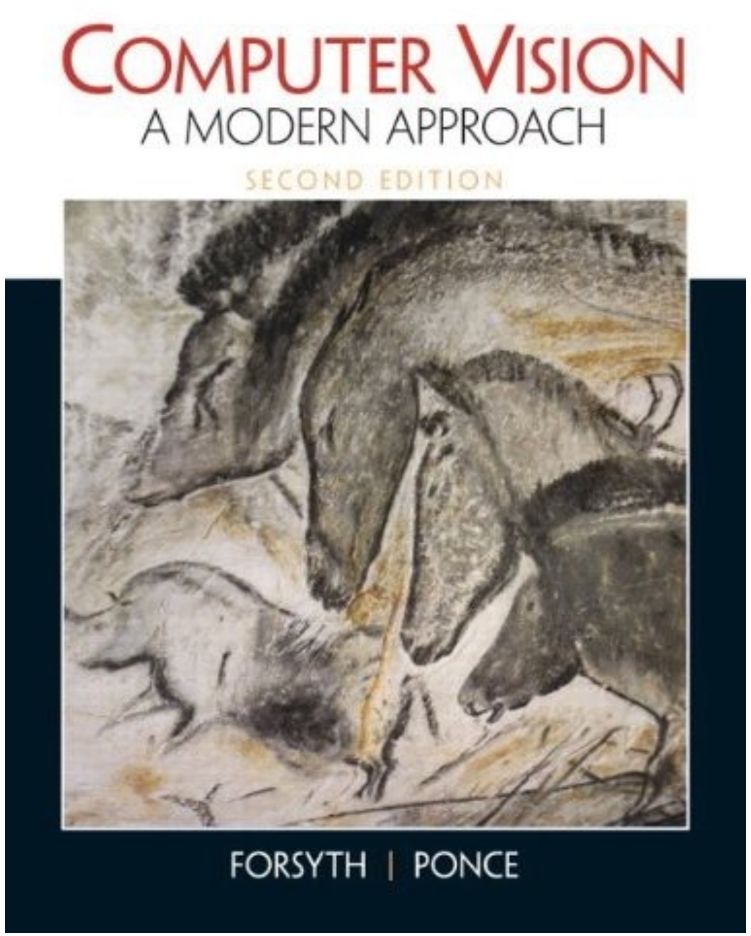
ed

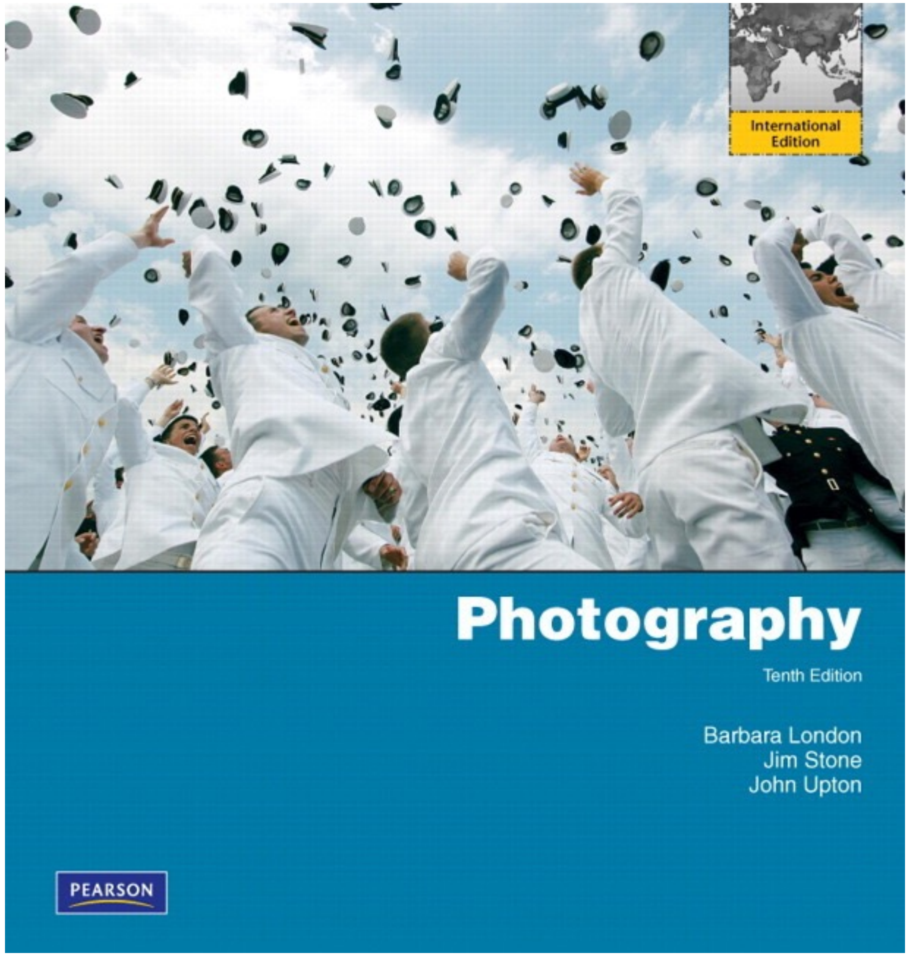
- Enrollment link is available at the course webpage.
- will be used for course-related announcements.
- similar to piazza, but with more capabilities.



Reference Books







Prerequisites

- Good math (calculus, linear algebra, statistics) and programming skills.
- An introductory course in image processing (BBM413/AIN430), and/or computer vision (BBM416/AIN431) and/or machine learning (BBM406/AIN311) is highly recommended.

Grading

- Grading for AIN434/BBM444 will be based on
  - Class participation (5%),
  - Course project (done in pairs) (30%),
  - Midterm exam (30%), and
  - Final exam (35%).

- Grading for AIN435/BBM446 will be based on
  - Four assignments (done individually) (25% each).

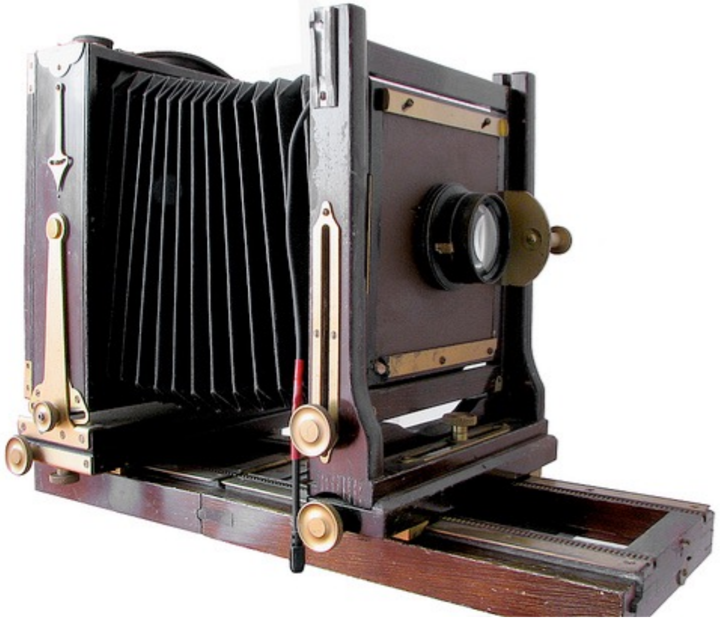
Schedule

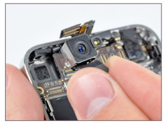
Week 1	Introduction, Digital photography
Week 2	Image formation
Week 3	Noise and Color
Week 4	Exposure and high-dynamic-range imaging
Week 5	Edge-aware filtering
Week 6	Gradient-domain image processing
Week 7	No class – National Holiday

Schedule

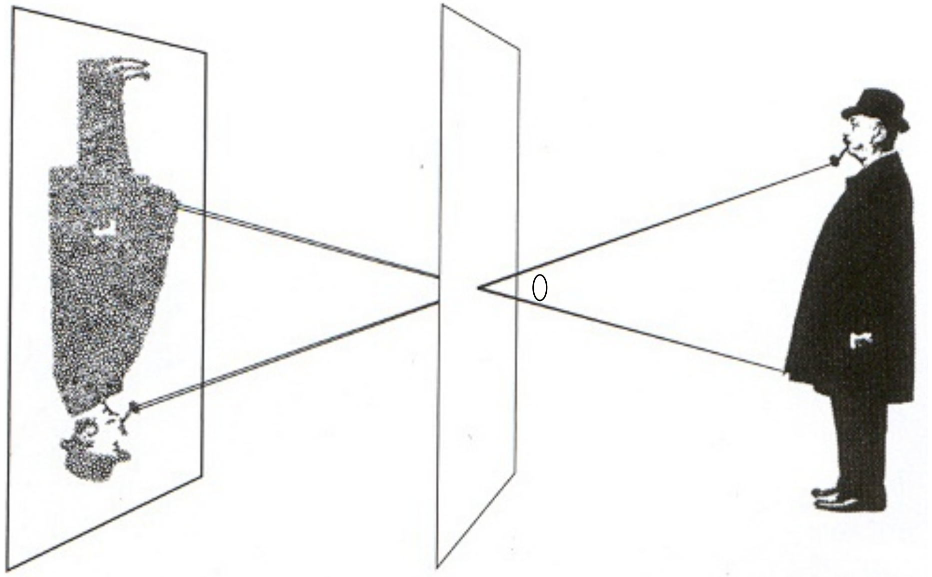
Week 8	Focal stacks and lightfields
Week 9	Midterm Exam
Week 10	Deconvolution, Coded photography
Week 11	Convolutional Neural Networks
Week 12	Deep Generative Models and their applications
Week 13	Visual quality assessment
Week 14	Project presentations, Course wrap-up

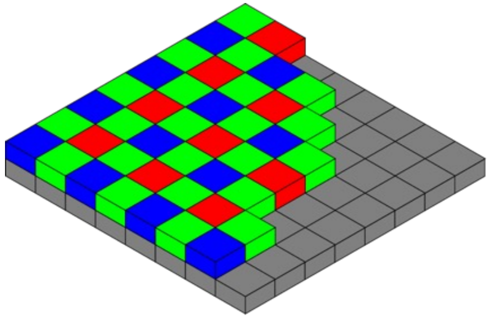
Lecture 1: Introduction to Digital photography

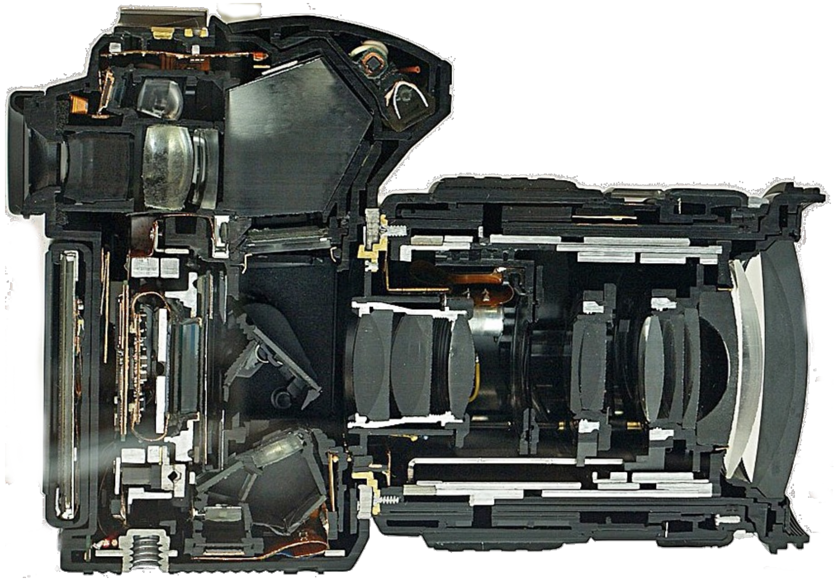




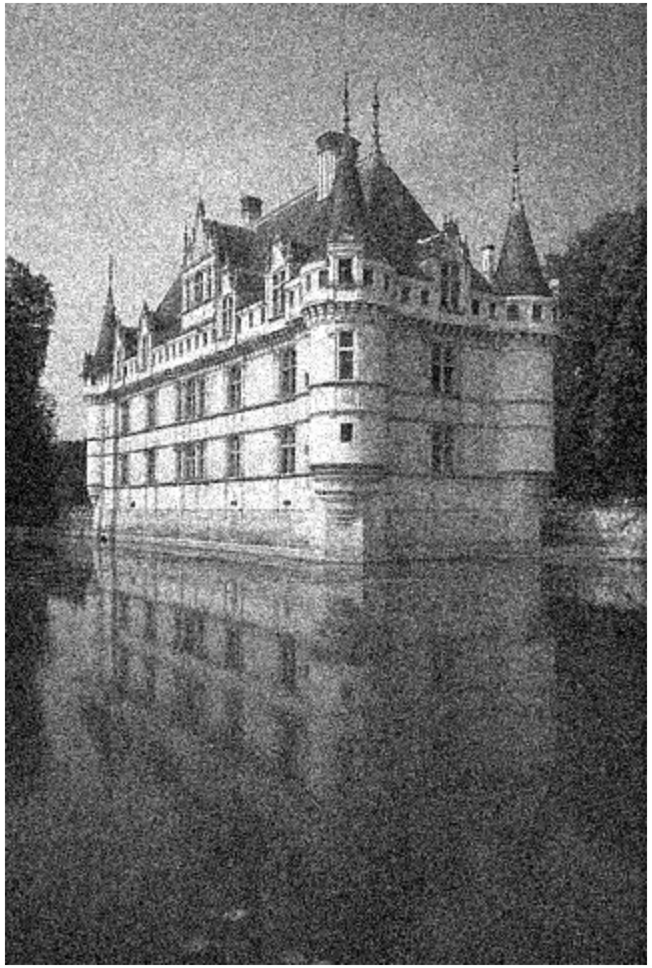
Lecture 2: Image formation

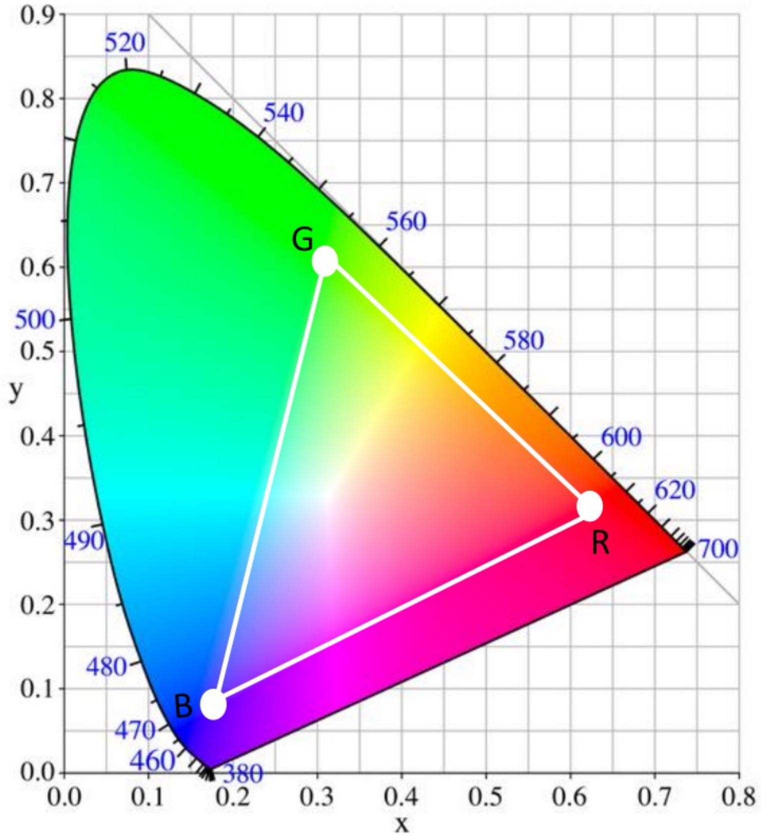






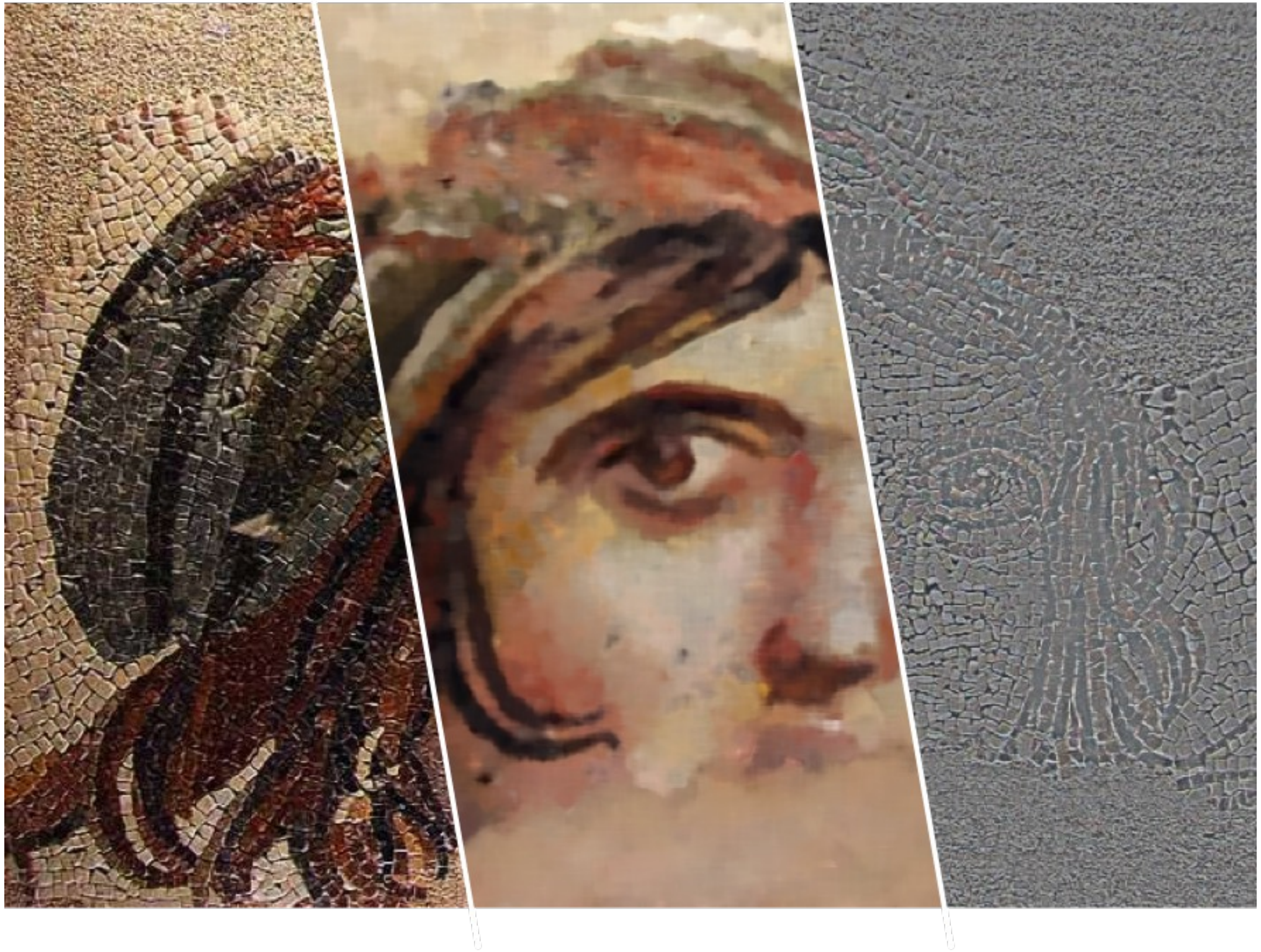
Lecture 3: Noise and Color



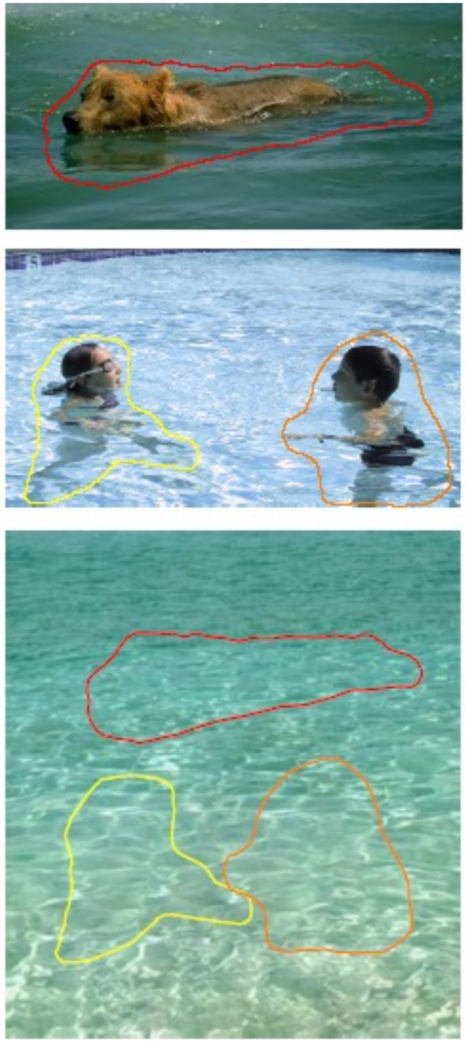


Lecture 4: Exposure and high-dynamic-range imaging

Lecture 5: Edge-aware filtering



Lecture 6: Gradient-domain image processing

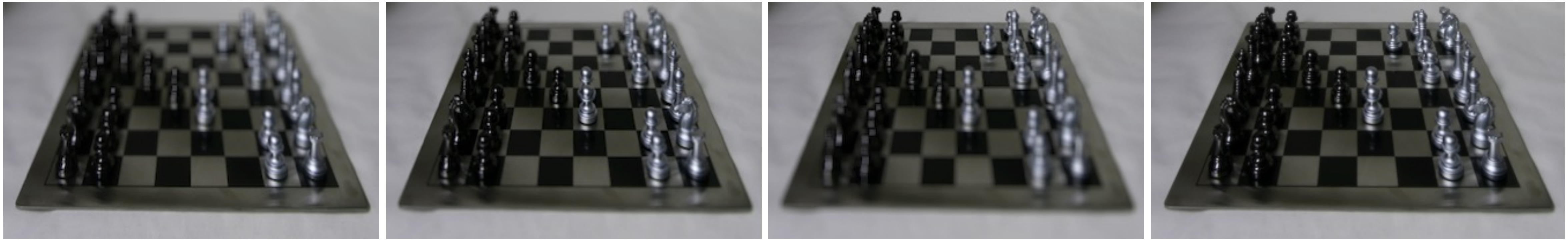


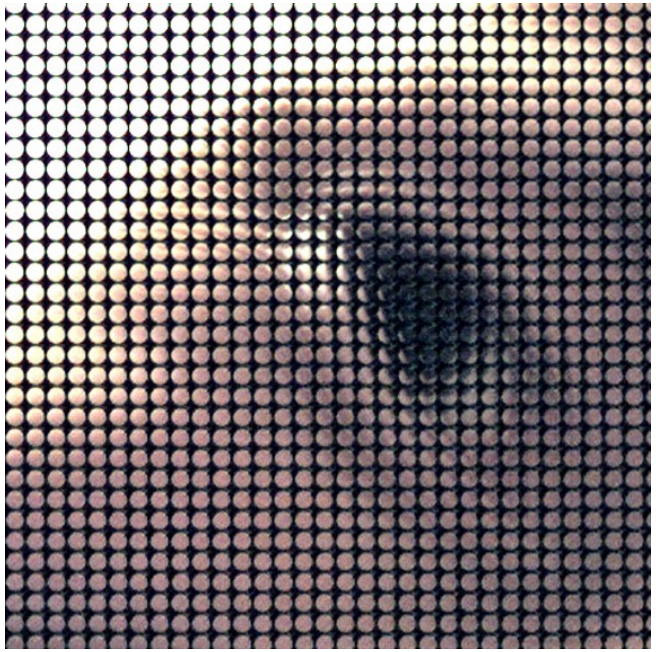
sources/destinations

cloning

seamless cloning

Lecture 7: Focal stacks and lightfields

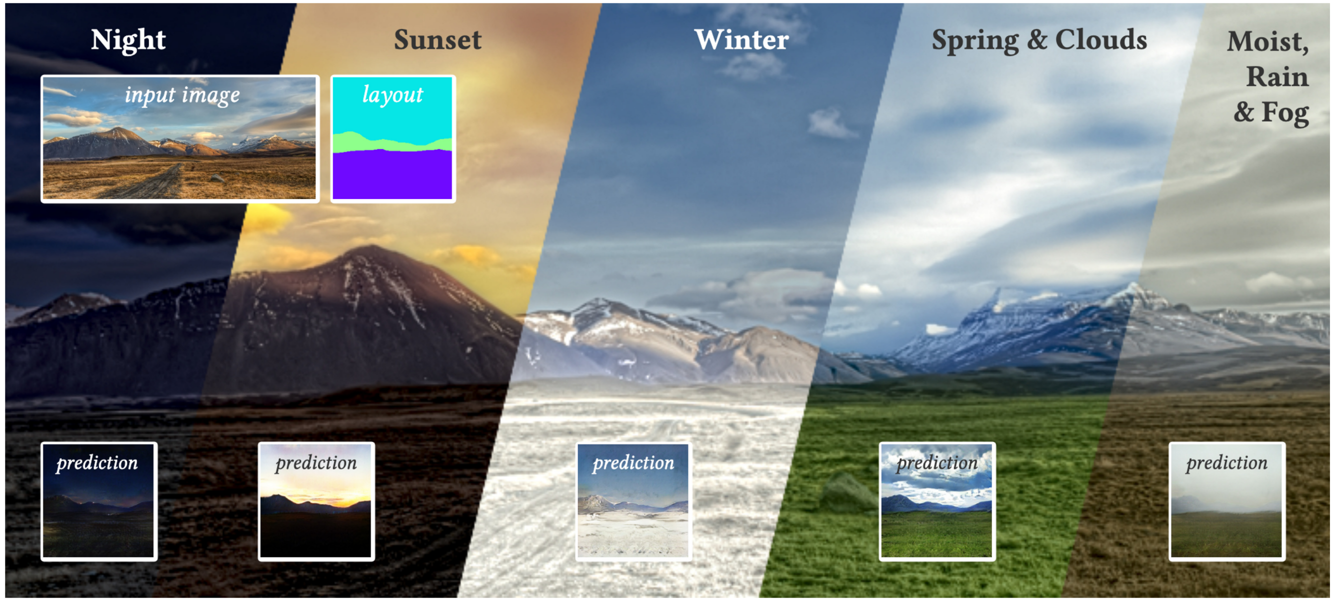




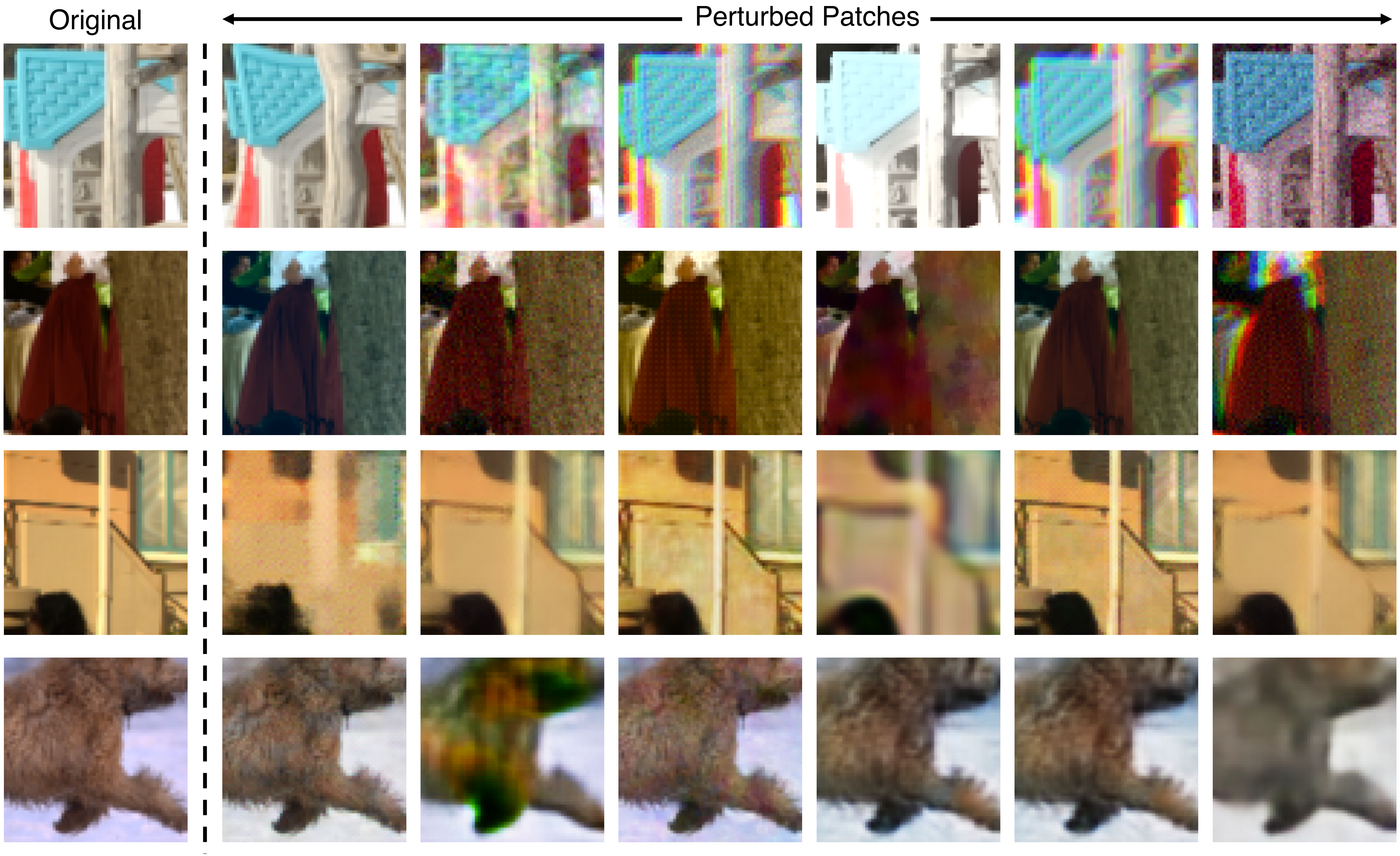
Lecture 8: Deconvolution, Coded photography

Lecture 9: Convolutional Neural Networks

Lecture 10: Deep Generative Models and their applications



Lecture 11: Visual quality assessment



Programming Assignments

- 4 programming assignments (25% each)
- Should be done individually
- Involve implementing an algorithm, carrying out a set of experiments to evaluate it, and writing up a report on the experimental results.
- **Late policy:** You have 5 slip days in the semester.

- Tentative Dates
  - Assignment 1 Out: March 3, Due: March 17
  - Assignment 2 Out: March 17, Due: March 31
  - Assignment 3 Out: April 7, Due: April 21
  - Assignment 4 Out: April 28, Due: May 12

Course project

The students who need GPU resources for the course project are advised to use Google Colab.

- The course project gives students a chance to apply the methods discussed in class to a research oriented project.
- The students can work in pairs.
- The course project may involve
  - Design of a novel approach and its experimental analysis, or
  - An extension to a recent study of non-trivial complexity and its experimental analysis.
  - A comparative analysis of methods
- Deliverables

- Proposals	Mar 24, 2025
- Project progress reports	April 28, 2025
- Final project presentations	TBA
- Final reports	May 25, 2025

Sample Course Projects – Spring 2024

Generating Hyperspectral Images From RGB Images by Utilizing Denoising Networks

Damla Akcaoğlu
n23130119@hacettepe.cs.edu.tr
Hacettepe University, Computer Engineering Department
Ankara, Turkey

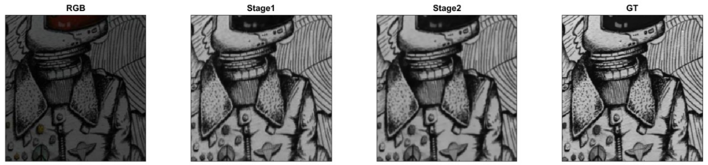


Figure 1: From left to right: RGB Image, Result of First Stage, Result of Second Stage, Ground Truth

ABSTRACT

This project explores the generation of hyperspectral images from RGB counterparts using denoising networks, presenting a two-stage training approach for spectral super-resolution. In the first stage, networks including CanNet, HSCNN+, and HSRNet are trained to generate initial hyperspectral images from RGB images. Subsequently, a spatial super-resolution network from HIR-Diff is employed in the second stage to enhance the quality of initial hyperspectral images through denoising. Experimental evaluation is conducted using the ARAD1K dataset, with performance assessed using PSNR, SSIM, SAM, and LPIPS metrics. Results indicate a decrease in overall performance when the second stage network is added, attributed to issues such as hyperparameter incompatibility and limitations of the pre-trained network. Despite these challenges, the study highlights the importance of the first stage performance and the modular nature of the proposed approach.

KEYWORDS

hyperspectral image generation, diffusion networks, denoising

ACM Reference Format:
Damla Akcaoğlu. 2018. Generating Hyperspectral Images From RGB Images by Utilizing Denoising Networks. In *Proceedings of BBM444 '23-24'*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

BBM444 '23-24', May 03–05, 2024, Ankara, Turkey

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The generation of hyperspectral images from RGB counterparts presents a cost-effective approach to acquiring rich spectral information about a scene. By leveraging advancements in deep learning and denoising techniques, this study aims to enhance the spectral resolution of RGB images through a comprehensive two-stage training approach. Hyperspectral images, with their wider range of spectral bands, offer valuable insights into the characteristics of captured scenes, making them essential for various applications in remote sensing and image analysis. However, the high cost associated with capturing hyperspectral imagery necessitates the development of efficient methods for estimating hyperspectral versions of RGB images. This paper addresses this challenge by proposing a novel methodology that combines initial hyperspectral image generation with subsequent denoising, utilizing state-of-the-art deep learning architectures. The experimental evaluation conducted on the ARAD1K dataset provides valuable insights into the performance and limitations of the proposed approach, paving the way for further advancements in hyperspectral image generation techniques.

2 PROBLEM DESCRIPTION

The project focuses on generating hyperspectral images by conditioning on their RGB counterparts. The difference between RGB images and hyperspectral images are different number of channels that covers same interval of electromagnetic spectrum. While RGB images divide the visible spectrum (380nmm-7000nm) into 3 channels, hyperspectral images divide the same range into much more bands whose number changes with the specifications of the camera. These bands carry more information than the RGB images about the captured scene, because the interaction of light with the matter changes with changing wavelength of the light. Therefore, each

A Comparative Study of Deep Learning Models for Blind Motion Deblurring on Single Image

Can Ali Ateş
canaliatess@gmail.com
Hacettepe University
Ankara, Turkey

Emre Çoban
emrecobann02@gmail.com
Hacettepe University
Ankara, Turkey

Figure 1: Example Image Deblurring Results of Restormer (Input Image vs Deblurred Image)

Abstract

Single-image blind motion deblurring is an active research area in which researchers have developed different techniques for a long time. It is widely used in different vital fields such as forensics and medical imaging, therefore it is a very important image restoration technique. In this comparative study, the recent state-of-the-art (SOTA) methods which are MAXIM, Restormer, NAFNet, and LaKDNet that bring their solutions to this problem are compared and evaluated. Each of these methods has unique characteristics and contributions, so they produce different output quality under different conditions. Accordingly, the study aims to highlight the differences between these methods to understand their advantages and disadvantages. Also, it motivates the development of a new optimal method that can be achieved by using the advantages of the methods that are compared.

1 INTRODUCTION

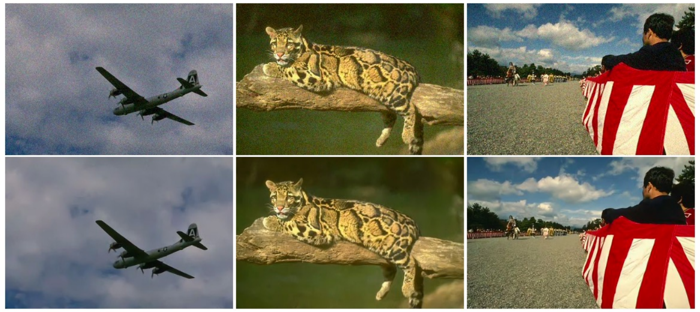
As an image restoration technique, single-image blind motion deblurring aims to produce sharper and more informative output images from the blurry input image. The "blind motion" nomenclature comes from not knowing the prior reason for the blur. This prior reason can be shaky hands, moving objects, or another factor that degrades the image with blur. In this field, different kinds of approaches that use Transformer-based models, MLP-based models, hierarchical CNNs, etc. are proposed by researchers.

In this comparative study, four different approaches that offer unique solutions for single-image blind motion deblurring are compared by using four different evaluation metrics over five different datasets. The first work is "MAXIM: Multi-Axis MLP for Image Processing" proposed by Z. Tu et al. [1], which proposes spatially-gated MLPs that enable the capture of long-range pixel interactions by making the network global and fully convolutional. The second work is "Restormer: Efficient Transformer for High-Resolution Image Restoration" proposed by Zamir et al. [2], which suggests multi-Dconv head transposed attention (MDTA) and gated-Dconv feed-forward network (GDFN) blocks for capturing both local and global pixel interactions while does not affect from the resolution of the image. The third work "Simple Baselines for Image Restoration" proposed by Chen et al. [3], offers extracting essential components by decomposing SOTA methods to form a baseline achieving better results with a lower system complexity for image restoration tasks. The fourth work "Revisiting Image Deblurring with an Efficient ConvNet" proposed by Ruan et al. [4], suggests a pure CNN block with a large kernel convolution named LaKD to explore the effect of an effective receptive field to get better performance than Transformers while has less computational costs. By comparing these models, the study intends to shed on light the development of new models that can capture the different conditions of blur effectively.

2 RELATED WORKS

There are many studies with different techniques proposed so far in the field of single-image blind motion deblurring. The early-stage techniques mainly focus on kernel estimation that has quality and computational cost inefficient shortcomings because of the iterative procedure of optimization [5, 6, 7]. These techniques have been replaced by CNN-based models with an increase in the availability of large-scale datasets. CNN-based models [8, 9, 10, 11, 12] have started to dominate state-of-the-art (SOTA) performance due to their power to learn generalizable image priors that are important for restoration tasks, but these CNN models have two main shortcomings which are local receptive field and static weights for inference.

Sample Course Projects – Spring 2023



A Comparative Study Of Image Denoising Methods

Ilayda Sahin
ilayda.sahin@hacettepe.edu.tr
Hacettepe University
Ankara, Turkey

Figure 1: From top to bottom noised images with $\sigma = 15$, denoised images with DnCNN-B from CBSD68 dataset.

ABSTRACT

The problem of image denoising is a problem that concerns every aspect of computational photography. In this regard, Adaptive Gaussian White Noise, which is synthetic, is mostly used for training and testing the denoising algorithms , and also most of the denoising algortihm's results are compared with these kind of synthtetic datas. However, the results with real-world data are not always the same. Since noise may exist in some parts while it may not exist in others, the denoising process we apply to the entire image can remove noise very well for non-textured areas, while it can also remove some details with the noise on textured areas. Therefore, in this paper, we will examine the results of FFDNet, IRCNN, and DnCNN algorithms, which are currently considered to provide very good results for denoising, by comparing them with real-world data, and we will strengthen these results with synthetic data to achieve a more powerful outcome. In addition to quantitative methods, we will also evaluate the results qualitatively, and we will reveal the differences between different performance methods.

KEYWORDS

Datasets, Convolutional Neural Networks, Denoising

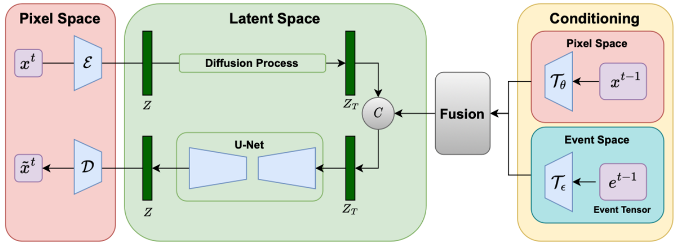
ACM Reference Format:
Ilayda Sahin. 2023. A Comparative Study Of Image Denoising Methods. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Denoising Algortihms).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Image denoising is a fundamental problem in image processing, with applications in a wide range of fields, including medical imaging, remote sensing, and computer vision. In recent years, deep learning-based denoising algorithms have emerged as a promising approach for achieving state-of-the-art performance on this task. Among these algorithms, Image Restoration Convolutional Neural Network (IRCNN)[5], Deep Convolutional Neural Network (DnCNN)[4], and Fast and Flexible Denoising Neural Network (FFDNet)[6] have shown particularly impressive results.

Denoising Algortihms, June 03–04, 2022, Ankara, TR

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX



Enhanced Frame Reconstruction from Event Data using a Latent Diffusion Model as a Post-processing Step

Canberk Sağlam*
canberksaglam1@gmail.com
Department of Computer Engineering, Hacettepe University
Ankara, Turkey

Enes Karanfil*
enkaranfiles@gmail.com
Department of Computer Engineering, Hacettepe University
Ankara, Turkey

Figure 1: High level overview of proposed post-processing model. x^{t-1} indicates the reconstructed frame from event data and x^{t-1} indicates event data. Fusion block fuses information from reconstructed frames and event data. C block where concatenation of condition and input takes place. Rest of the model is same as in [20].

ABSTRACT

Event cameras are sensors inspired by biology that differ significantly from traditional cameras, which gives them certain advantages that make them increasingly popular in many applications that prioritize low power consumption and low latency. However, event cameras produce output that is incompatible with existing computer vision algorithms, which are designed to process frame-based inputs. To address this issue, there are existing studies that converts event data into frames and videos such as E2VID [18]. The goal of this project is to enhance the quality of frames reconstructed from event data using a novel approach. We propose using latent diffusion model as a post-processing step in order to improve the quality of reconstructed frames by leveraging diffusion model's generative power.

KEYWORDS

diffusion models, event-based vision, video reconstruction

*Both authors contributed equally to this research.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Canberk Sağlam and Enes Karanfil. 2023. Enhanced Frame Reconstruction from Event Data using a Latent Diffusion Model as a Post-processing Step. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the last decade, deep learning research has greatly advanced computer vision. Advanced computer vision techniques using state-of-the-art methods have achieved remarkable outcomes in various tasks that involve digital images and videos. Nonetheless, these methods encounter difficulties with tasks in the real world that include high-speed motion and scenes with high dynamic range. The principles of conventional frame-based sensors are one of the causes of these issues.

Novel sensors called event cameras have the potential to address the aforementioned challenges. Event cameras are distinct from conventional cameras in their approach. While conventional cameras capture frames at a fixed rate measured in FPS, event cameras operate on a per-pixel basis and asynchronously measure the changes in pixel brightness, which are referred to as events. This difference in approach provides several advantages to event cameras. The quick processing of brightness changes provides high temporal resolution, which minimizes motion blur. Additionally, event cameras work with low latency and consume low power as

Sample Course Projects – Spring 2022

A Comparative Study Of Image Super-Resolution Methods

Arda Hüseyinoğlu
ardahuseyinoglu@hacettepe.edu.tr

Batuhan Orhon
b21727556@cs.hacettepe.edu.tr

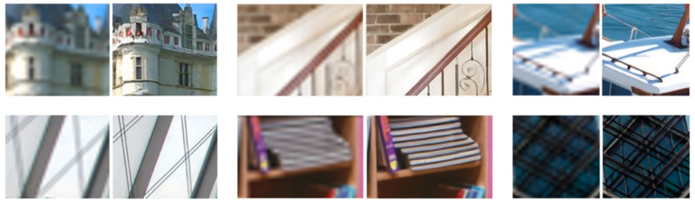


Figure 1: Low-resolution image regions, and their super-resolution correspondings which is obtained from the SPSR model[1].

ABSTRACT

Image super-resolution (SR) has been an open research area for a long time. It is an important image enhancement technique and is widely used in some areas such as satellite and medical imaging. In this paper, we aim to compare some of the recent state-of-the-art solutions to SR problems. Each of the works we compare has its own characteristics and they might provide different kinds of outputs in different conditions. We aim to highlight those differences and try to understand the reasons behind them.

1 INTRODUCTION

Image super-resolution is the process of increasing the resolution of a low-resolution image in a way that the output image is a high-quality image as possible. In recent years different kinds of approaches performed state of the art performances such as GAN architectures, Reference-based models, patch-based and cross-scale feature extractors, transformers, etc.

In this work, we compare four different approaches that were offered in recent years and all of them have different kinds of solution techniques. The first one is structure-preserving super resolution with gradient guidance proposed by Ma et al. [1], which uses image gradients that guides the generator to recover structures in images better. The other one is Deep Unfolding Network for Image Super-Resolution proposed by Zhang et al. [2], which leverages both learning based methods and model-based methods by unfolding the MAP inference via a half-quadratic splitting algorithm. The other one is the model proposed by Mei et al. [14]. They aimed to find non-local, patch-based, cross-scale pixel and patch similarities while applying super-resolution. They offer a Cross-Scale Non-Local attention module and Self Exampler Mining Cells which includes those CS-NL attention modules. In the last work we examine, Hui et al. [15] offer a lightweight information multi-distillation network by constructing the cascaded information multidistillation blocks to extract hierarchical features step-by-step and they use fusion module to aggregate them in according to their importance.

2 RELATED WORK

There are many studies on different techniques proposed so far for the image super resolution problem. Early approaches use PSNR scores as objective functions to map low-resolution images to high-resolution images [17, 18, 19]. However the models targeting the high PSNR score suffer from producing blurry images. After that the several models are proposed [20, 23, 24], which use perceptual loss to improve the visual quality of super-resolution images. One of the first model which uses perceptual loss and generates photo-realistic super-resolution images is the SRGAN proposed by Ledig et al. [21]. With their proposed model, Wang et al. [22] improve previous methods and introduce Residual-in-Residual Dense Block (RRDB) used in ESRGAN.

With the developments in the deep neural network, the state-of-art methods are proposed. Wei, Yunxuan, et al. [25] propose domain-distance aware super-resolution approach. The author says that their unsupervised approach outperforms the previous discriminatively trained, supervised, or blind SR algorithms especially in the generalization in the practical world part. The paper also includes a short part to compare the model with previous approaches. Ma, Cheng, et al. [1] propose a structure-preserving super resolution method to ease undesired structural distortions in the recovered images by generative adversarial networks (GANs) while maintaining the advantages of GAN-based methods to generate perceptual-pleasant details. Yang, Fuzhi, et al. [15] use transformer networks to take the texture information from the reference image and transfer those to the low-resolution image. Zhang, Kai, Luc Van Gool, and Radu Timofte [2] proposes an end-to-end trainable unfolding network which leverages both learning based methods and model-based methods by unfolding the MAP inference via a half-quadratic splitting algorithm.

Low Light Image Enhancement with InvertibleISP

Atakan Filgöz
atakan.filgoz@cs.hacettepe.edu.tr
Hacettepe University

ABSTRACT

Images are often captured under sub-optimal lighting conditions such as low brightness. These images exhibits characteristics such as low brightness and color distortion. Low light image enhancement (LLIE) research area has emerged to solve this issue. In this study, the InvertibleISP method [18], which was previously suggested in the literature, was used to contribute to low light image enhancement. With the InvertibleISP method, RAW images can be converted to sRGB images and vice versa. In this way, performance evaluations were made by selecting an example from the literature, a LLIE method, which takes sRGB images as input, and a LLIE method, which takes a RAW images as input. Promising results have been obtained and it has been seen that InvertibleISP can be used in the LLIE domain.

KEYWORDS

low light image enhancement, computational photography, image processing pipeline, deep neural networks

ACM Reference Format:
Atakan Filgöz. 2018. Low Light Image Enhancement with InvertibleISP. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Many learning strategies, network structures, loss functions, training data, etc. have been proposed in the literature for the low-light image enhancement problem, which aims to improve the perception or interpretability of an image taken in a poorly illuminated environment. The solutions proposed for low light image enhancement over the past decade have often been based on deep learning. These studies are divided into two branches when viewed from the inputs; the studies that take RAW images as input and the studies that take sRGB images as input. Due to the ease of data collection, the use of sRGB images as input has dominated these studies in the literature. While RAW data is limited and hard to collect, using this type of data covers a wider color gamut and higher dynamic range images. Therefore, deep models trained on RAW data often recover sharp details and high contrast, achieve vivid colors, reduce the effects of noise and artifacts, and improve the brightness of extremely low-light images.

In parallel, studies that produce RAW images from sRGB images have started in order to facilitate obtaining RAW data in the literature. One of the current studies in this area, InvertibleISP [18] aimed to produce RAW images from sRGB images and sRGB images from RAW images and achieved satisfactory results.

In this study, RAW images will be converted to sRGB images using the InvertibleISP [18] method and their performance will be measured on methods that use existing sRGB images as input. At the same time, sRGB images will be converted to RAW images and their performance will be measured on methods that use existing RAW images as input. For both image types, public datasets which previously published in the literature will be used.

2 RELATED WORK

Studies related to this study can be examined under two headings as studies that use RAW and sRGB images as inputs for the low light image enhancement task, and studies that aim conversions between RAW images and sRGB images. Although solutions are presented with different learning strategies such as reinforcement learning and unsupervised learning, supervised methods were examined due to the subject of the study.

Low Light Image Enhancement: For the low light image enhancement task, the first method using deep learning in the literature is LLNet [10]. The authors proposed a deep autoencoder-based approach to identify signal properties from low-light images and adaptively brighten images without over-amplifying and over-saturating the lighter portions in high-dynamic-range images. The results demonstrate significant reliability of the approach. In addition, this work was the pioneer of deep learning-based low light image enhancement. An end-to-end multibranch enhancement network, MBLLEN [13] improves the performance of low light image enhancement by extracting active feature representations with a feature extraction module, a development module, and a fusion module. Additionally, subnetworks such as Illumination-Net, Fusion-Net, and a Restoration-Net have been proposed [12] to improve performance. Recently, Ren et. al. [16] presented an RNN-based and an encoder-decoder-based method to solve the same problem. The EEMEFN method [22] has been proposed as a solution to the problems that existing methods cannot recover very low light or very bright areas, cannot correct the color of images exactly, and cannot focus on object edges. This method is a two-stage method using different exposure images. Lu et. al. [11] proposed TBEFN for LLIE, a multi-exposure fusion network which estimates a transfer function in two different parts that are used to obtain two enhancement results. Xu et al. [19] proposed a frequency-based

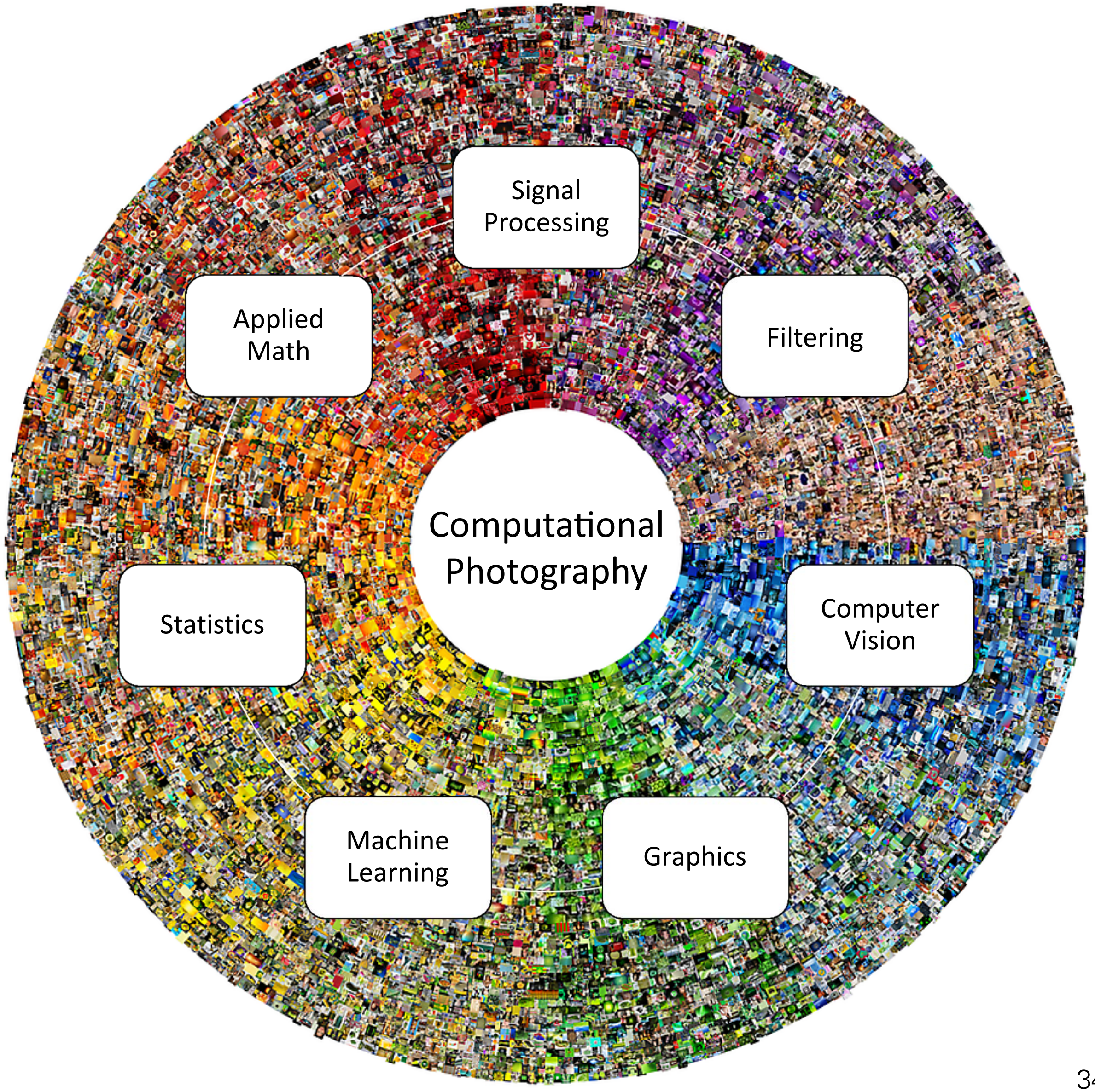
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/XXXXXXX.XXXXXXX

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

What is Computational Photography ?

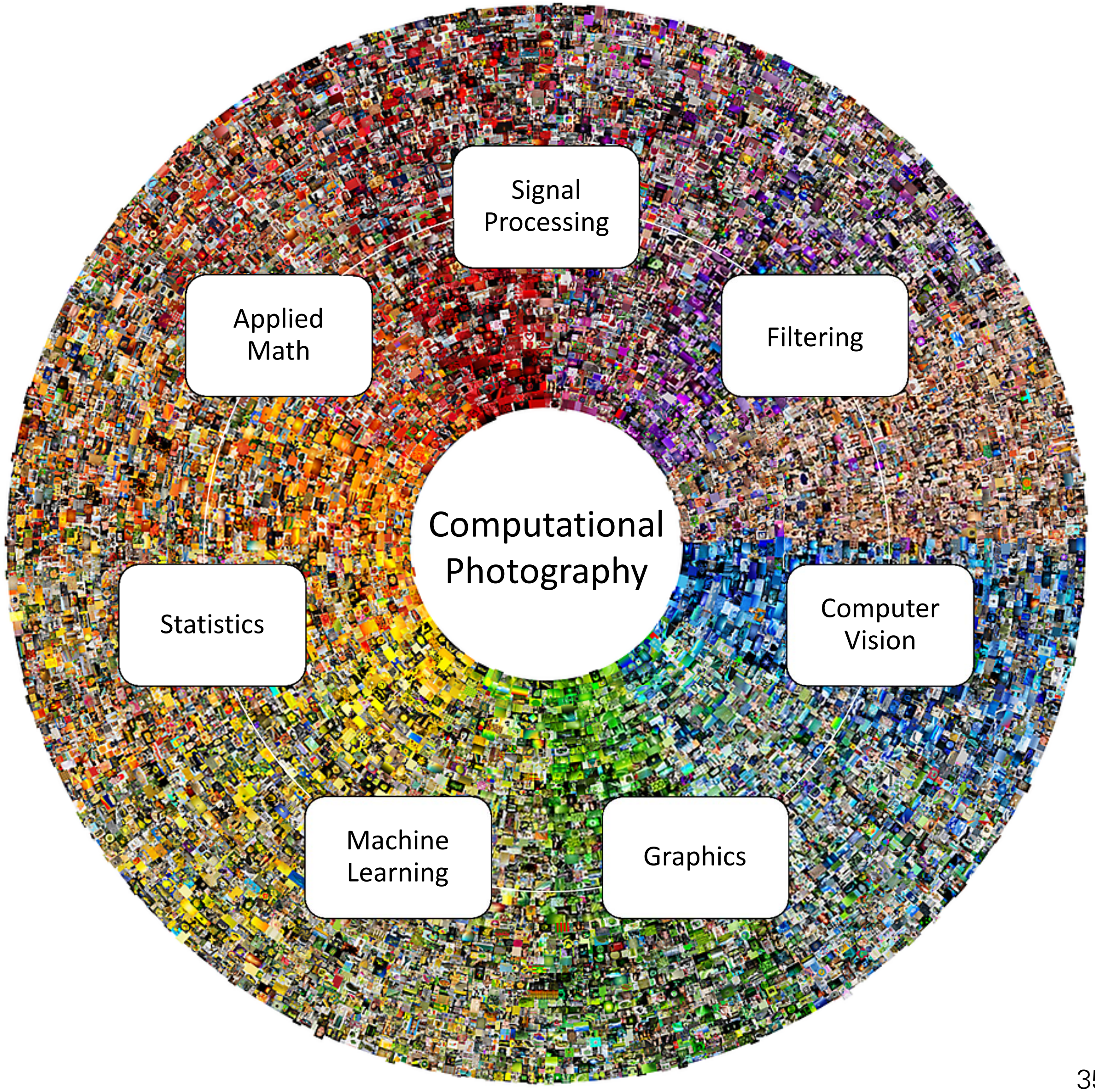
- It refers to an emerging new research area.
- It covers the set of methods used for capturing and processing digital images based on modern digital computation and algorithms instead of optical processes.
- It has changed the rules of photography, bringing to it new modes of capture, post-processing, storage, and sharing.



Slide adapted from P. Milanfar

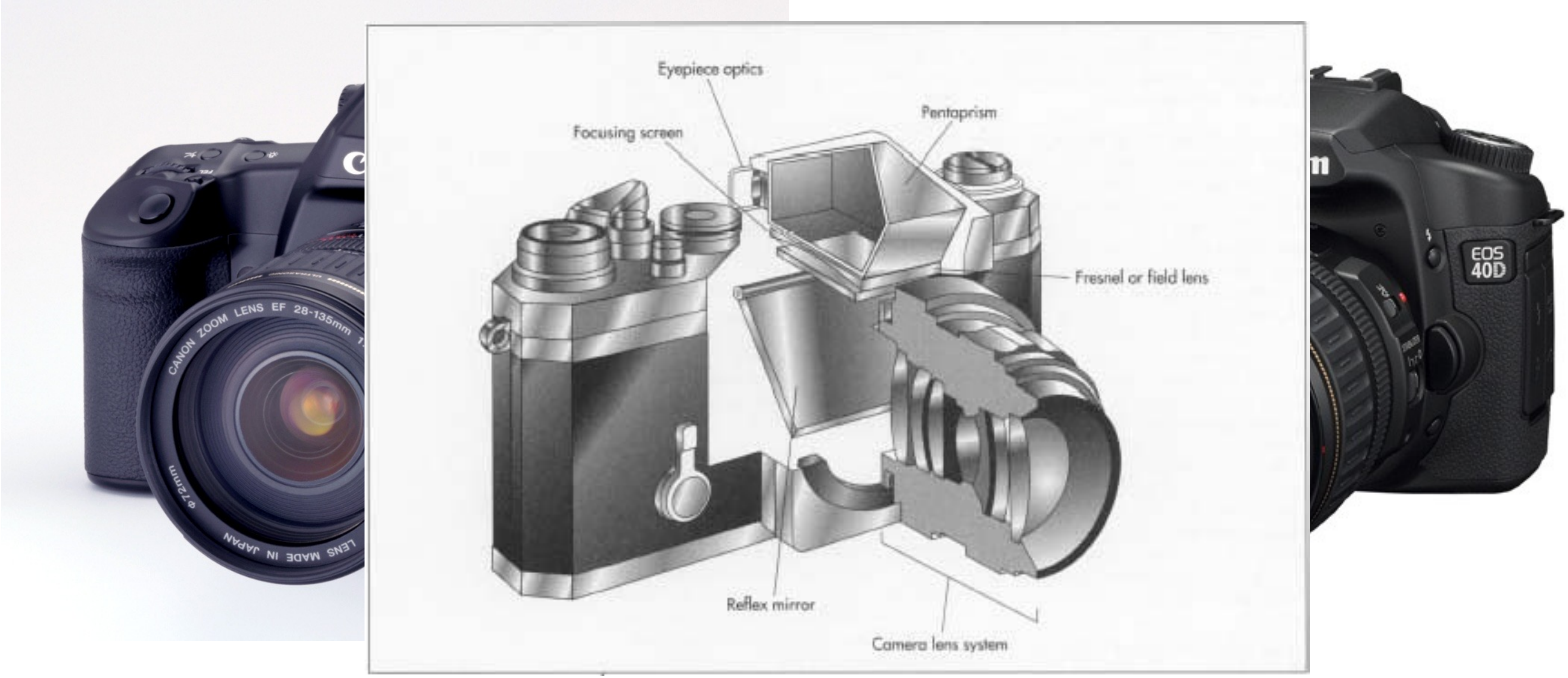
What is Computational Photography ?

- Digital photography:
  - Simply replaces traditional sensors and recording by digital technology
  - Involves only simple image processing
- Computational photography
  - More elaborate image manipulation, more computation
  - New types of media (panorama, 3D, etc.)
  - Camera design that take computation into account



Slide adapted from R. Fergus

Spot the difference



Film camera

Digital camera

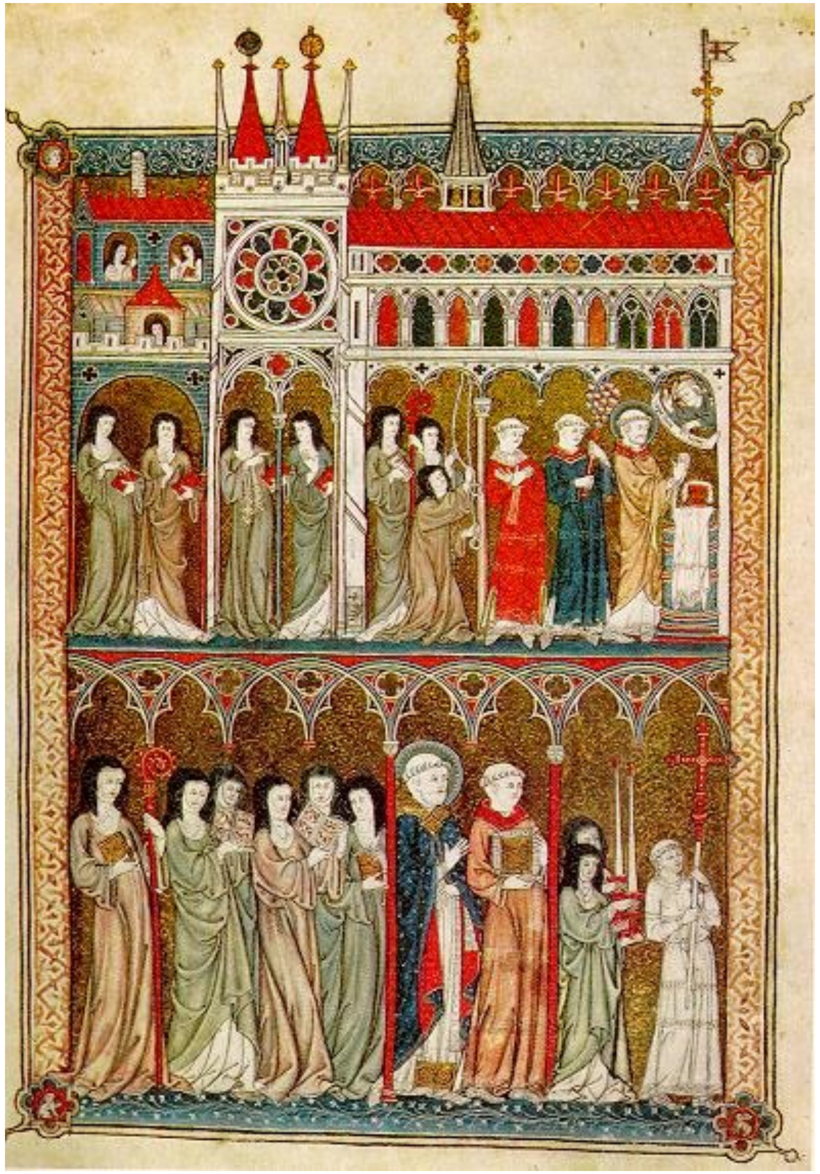
Depicting Our World: Prehistory

Prehistoric Painting, Lascaux Cave, France ~ 13,000 -- 15,000 B.C.

Depicting Our World: Middle Ages

The Empress Theodora with her court., Ravenna, St. Vitale 6th c.

Depicting Our World: Middle Ages



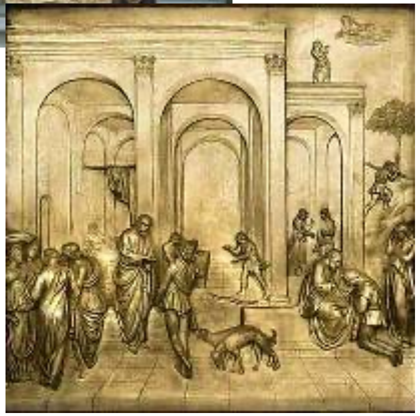
Nuns in Procession. French ms. ca. 1300.

Depicting Our World: Renaissance

North Doors (1424)

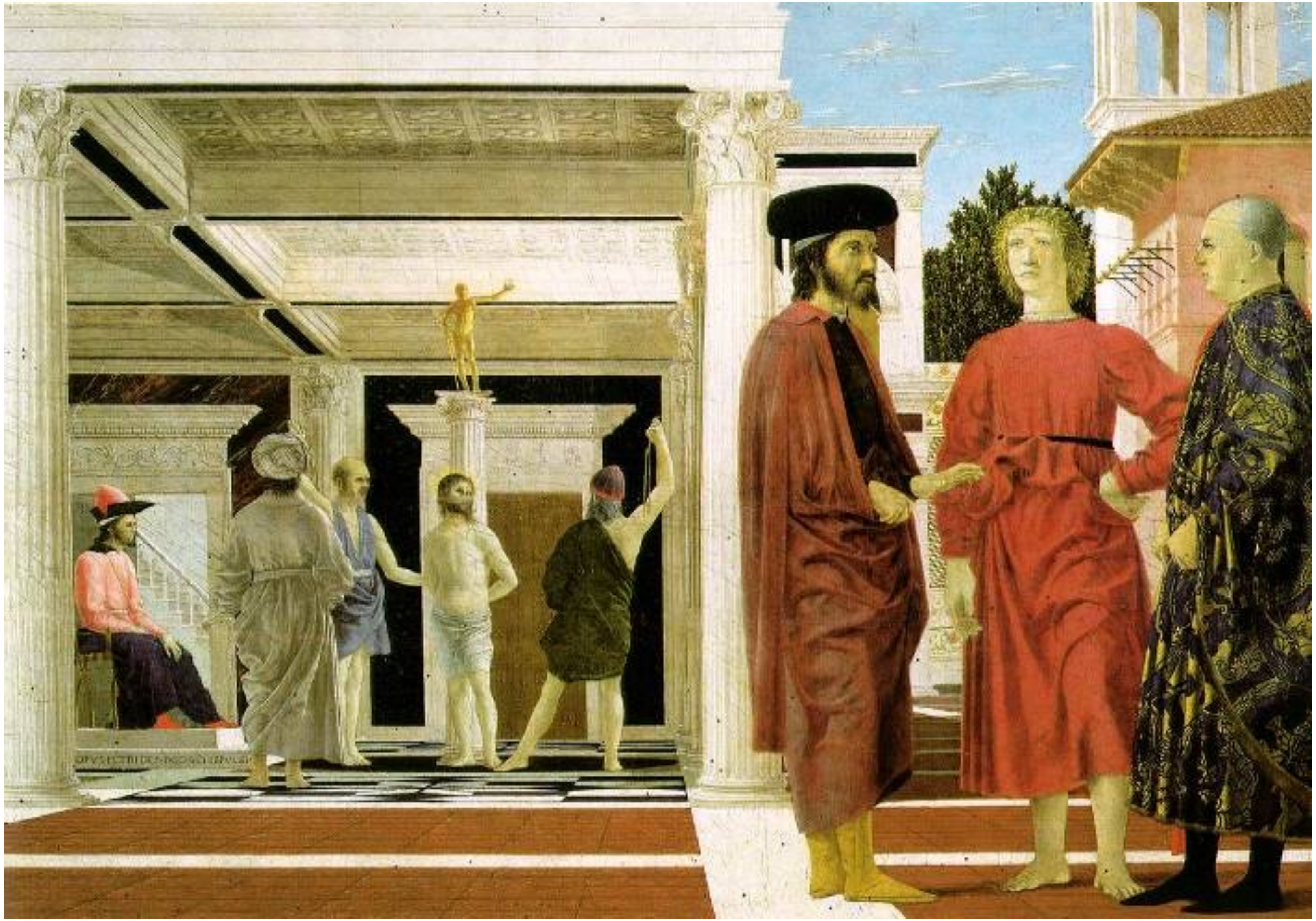
Lorenzo
Ghiberti
(1378-1455)

East Doors (1452)



Cathedral of Santa Maria del Fiore

Depicting Our World: Renaissance

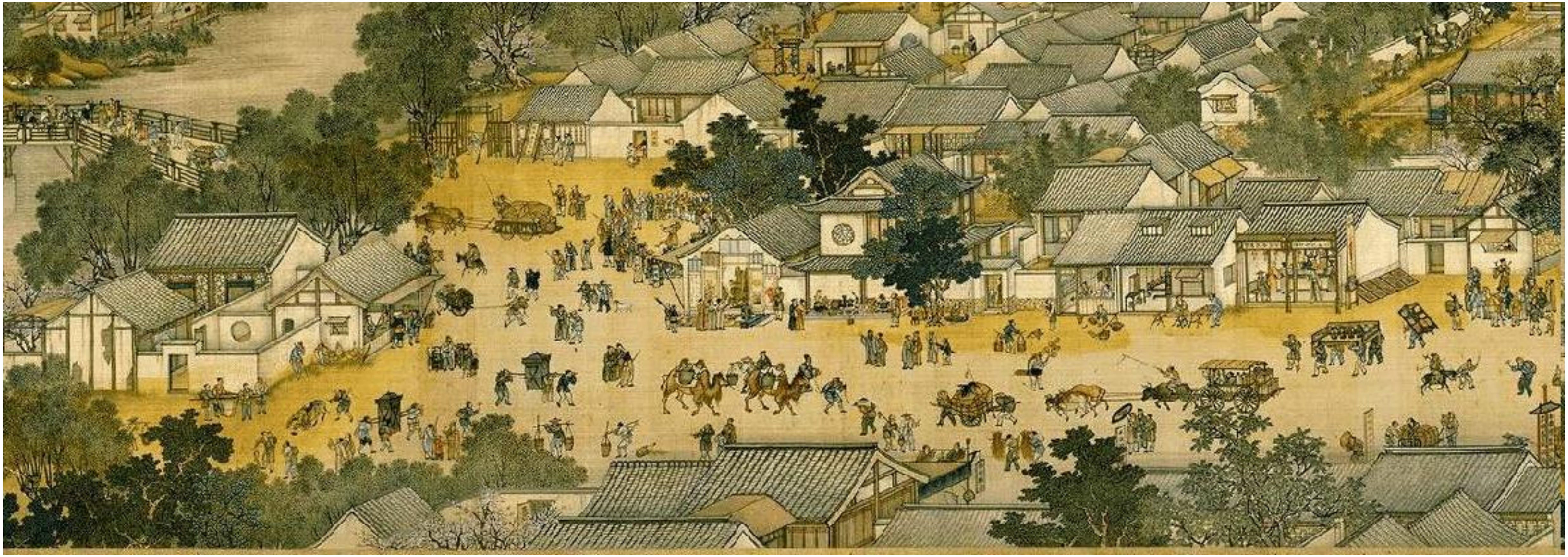


Piero della Francesca, The Flagellation (c.1469)

Depicting Our World: Renaissance

Paolo Uccello, Miracle of the Profaned Host (c.1467-9)

Depicting Our World: Song Dynasty (China)

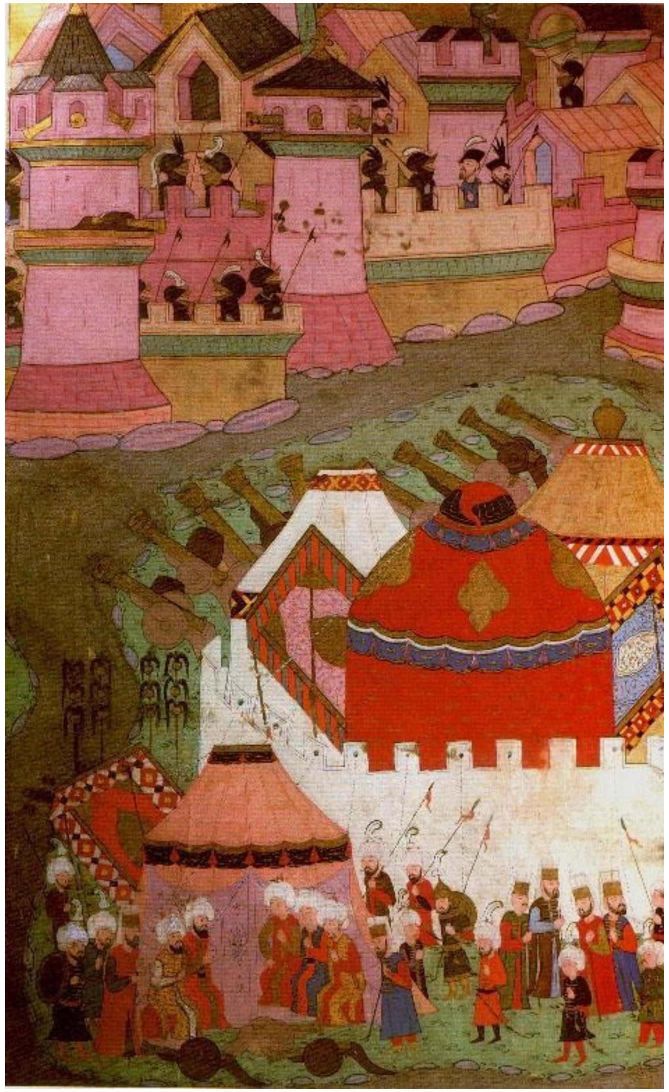


Qingming Festival by the Riverside, Zhang Zeduan ~900 AD

Depicting Our World: Edo Period (Japan)

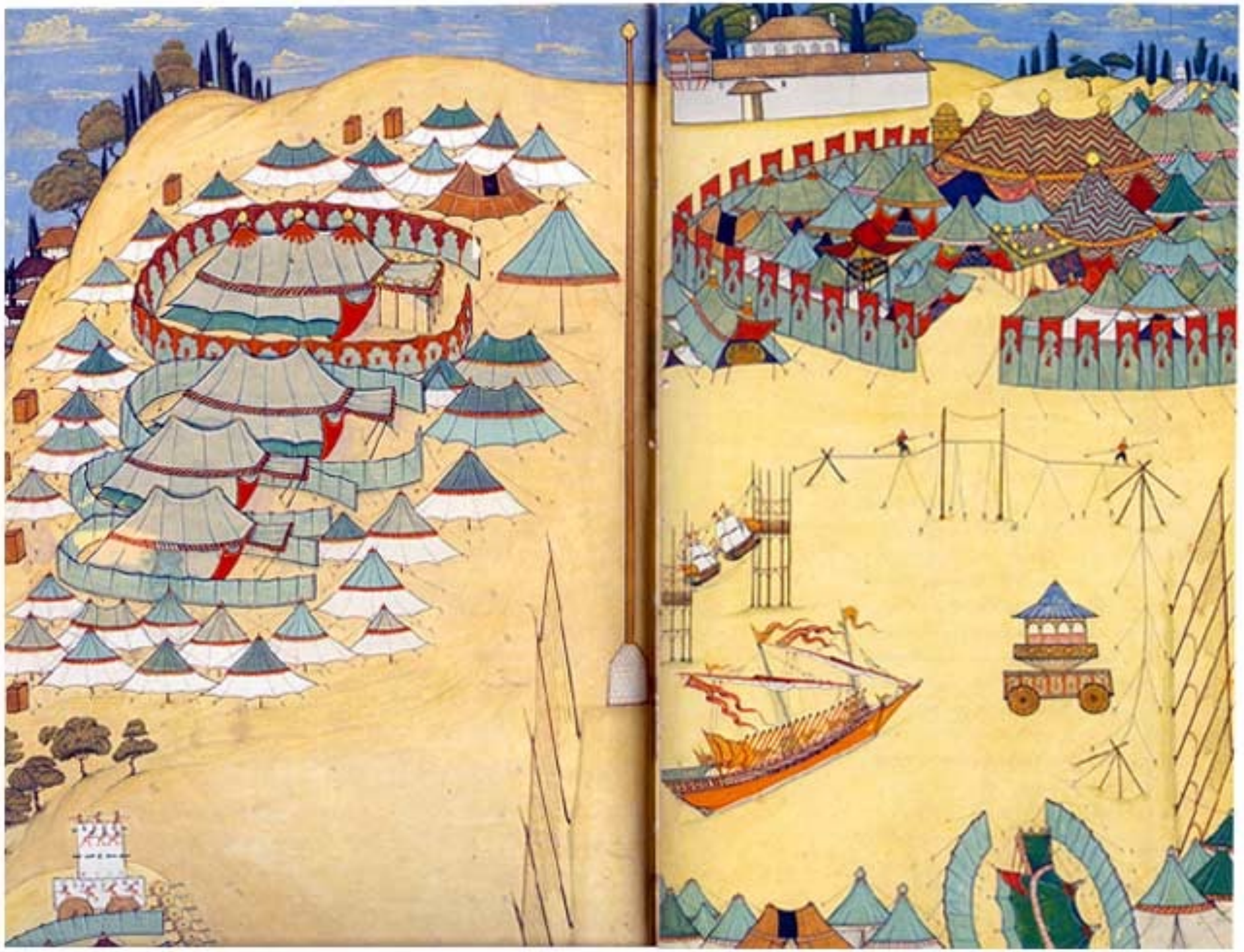
The Great Wave off Kanagawa, part of the series
Thirty-six Views of Mount Fuji, Hokusai (between 1826 and 1833)

Depicting Our World: Ottoman Miniatures



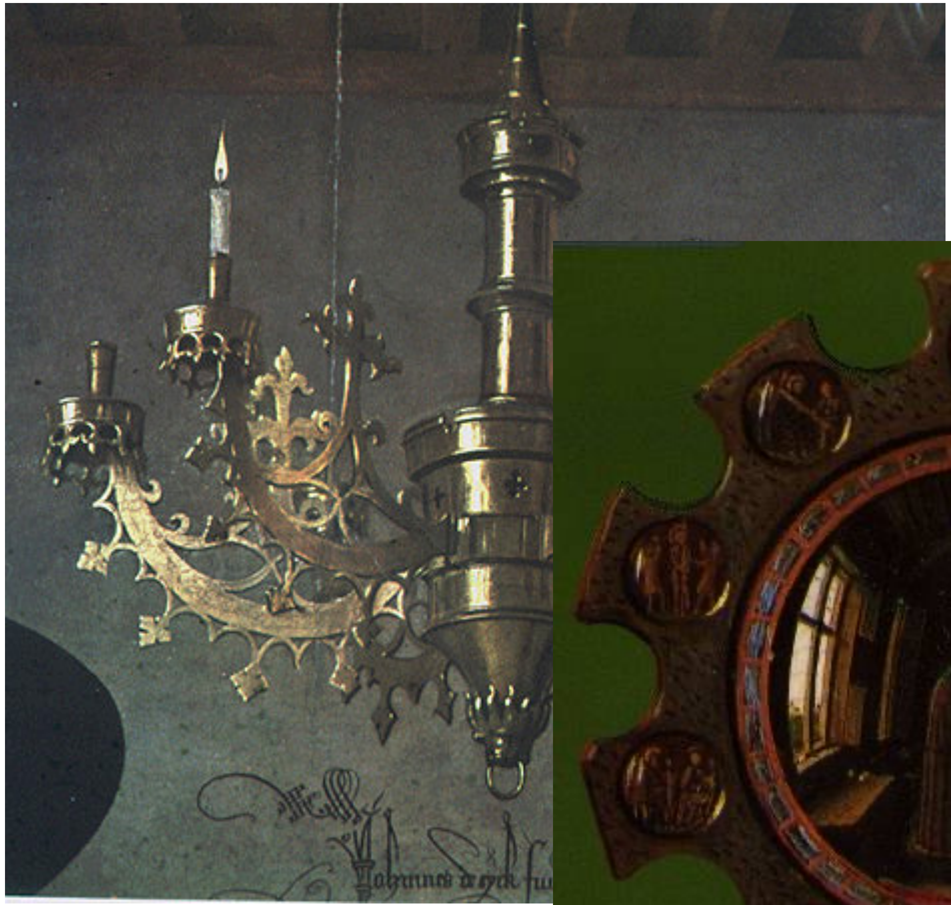
The Ottoman army besieging Vienna, from Huner-nama ('Book of Skills').
Nakkas Osman, 1588.

Depicting Our World: Ottoman Miniatures



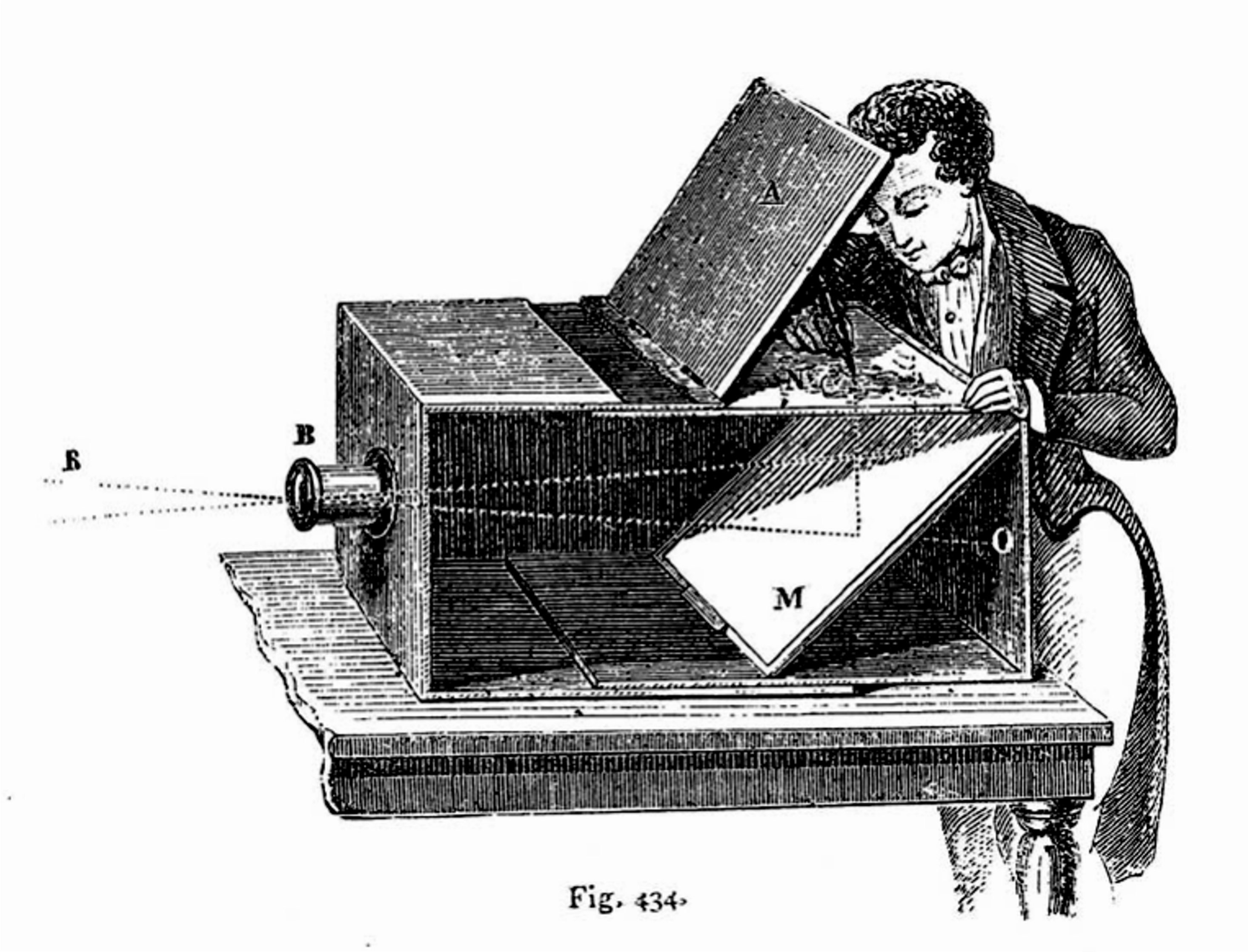
An Ottoman miniature from Surname-ı Vehbi, Abdulcelil Levni (1720)

Depicting Our World: Toward Perfection



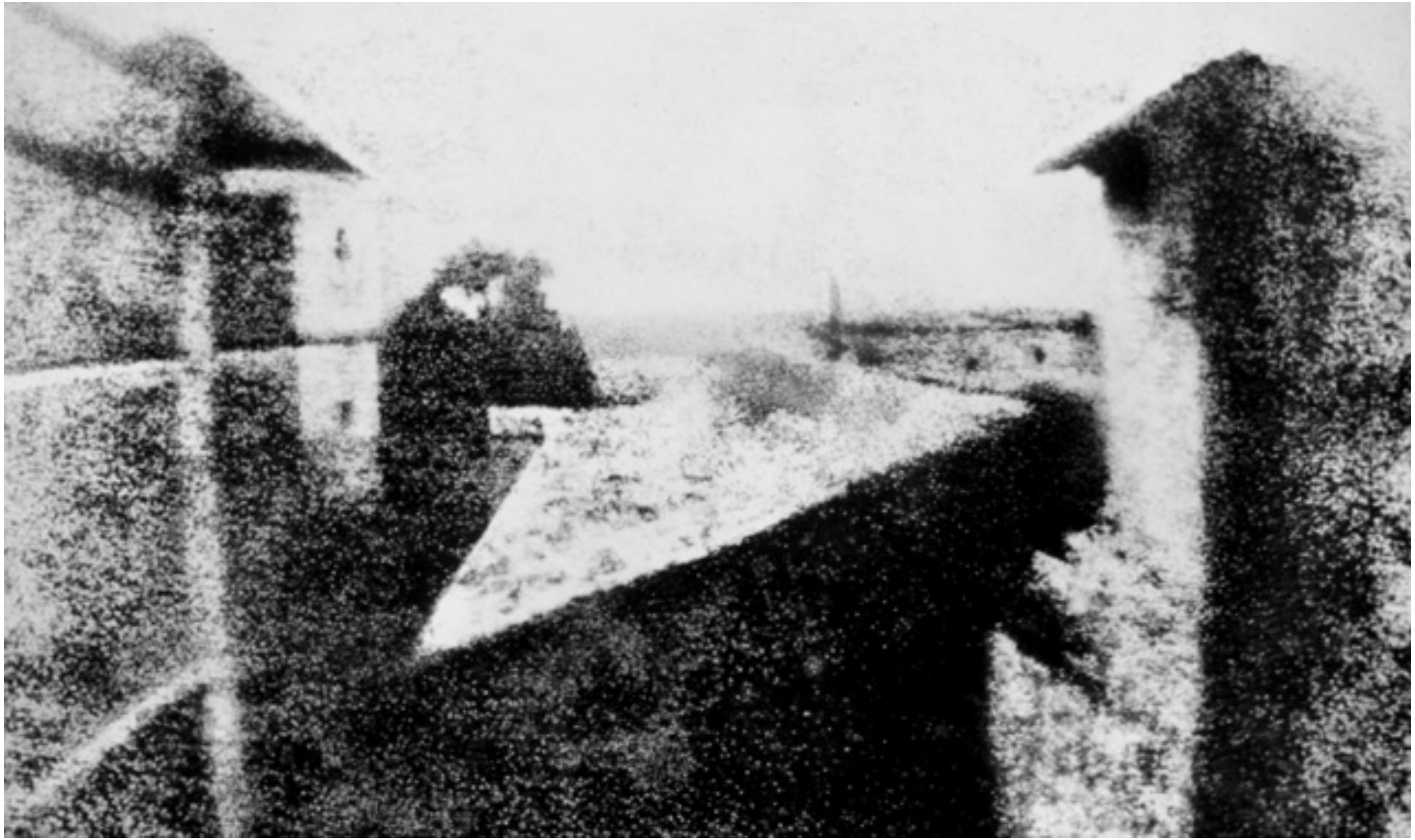
Jan van Eyck, The Arnolfini Marriage (c.1434)

Depicting Our World: Toward Perfection



Lens Based Camera Obscura, 1568

Depicting Our World: Perfection!

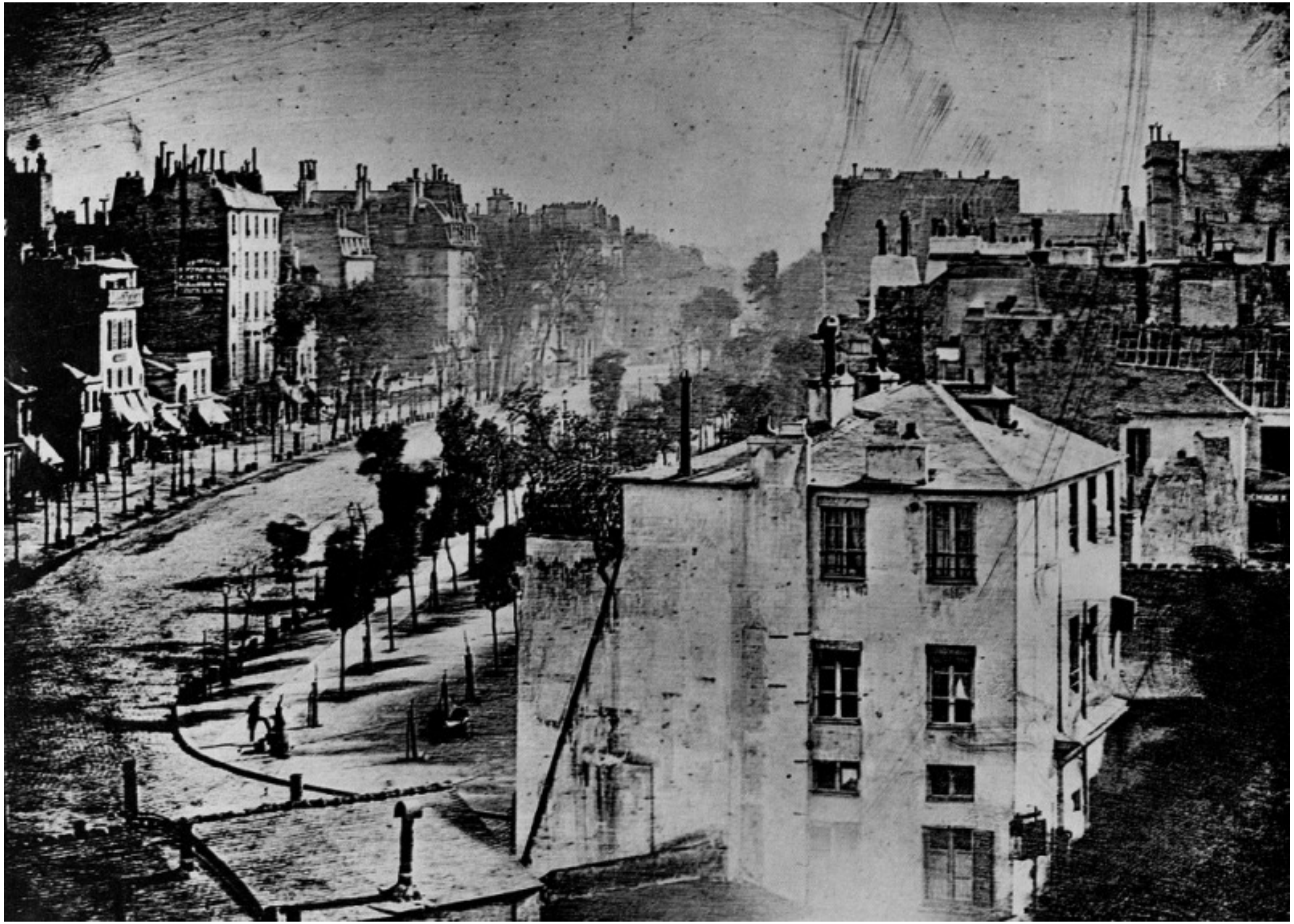


View from the Window at Le Gras, Joseph Nicéphore Niépce (1826)

Depicting Our World: Perfection!

Still Life, Louis Jaques Mande Daguerre, 1837

Depicting Our World: Perfection!



Boulevard du Temple, Louis Daguerre, 1838

After realism…

Monet, La rue Montorgueil

Depicting Our World: Ongoing Quest

Pablo Picasso

David Hockney

Depicting Our World: Ongoing Quest

David Hockney, Place Furstenberg, (1985)

Which one is right?

Multiple viewpoints

David Hockney,
Place Furstenberg, 1985

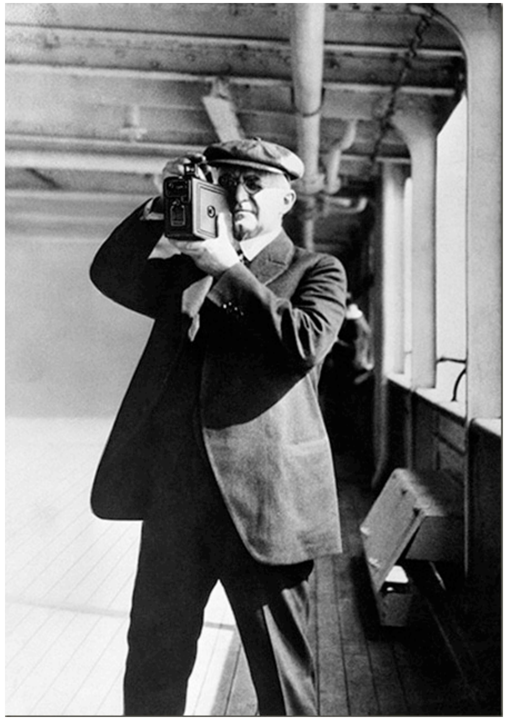
Single viewpoint

Alyosha Efros
Place Furstenberg, 2009

Recording images automatically

- Silver halide (AgCl, AgBr, AgI) salts are light sensitive
  - absorbed photons in halide ions cause free electrons
  - electrons combine with Ag+, producing metallic silver
- Daguerre: first practical and permanent photographic plate
  - Hg vapor (yikes!) combines with Ag to produce reflective amalgam Daguerrotypes were widely popular
- Indirect negative-plate processes
  - negative images on paper, glass allowed multiple copies to be printed
- Roll film: silver halide grains in gelatin on celluloid
  - introduced by Eastman in 1880s
  - portable, convenient, practical
  - sensitive (“fast”) enough for moving subjects in daylight


Daguerrotype (1839)


George Eastman with his Kodak camera

Slide adapted from P. Milanfar

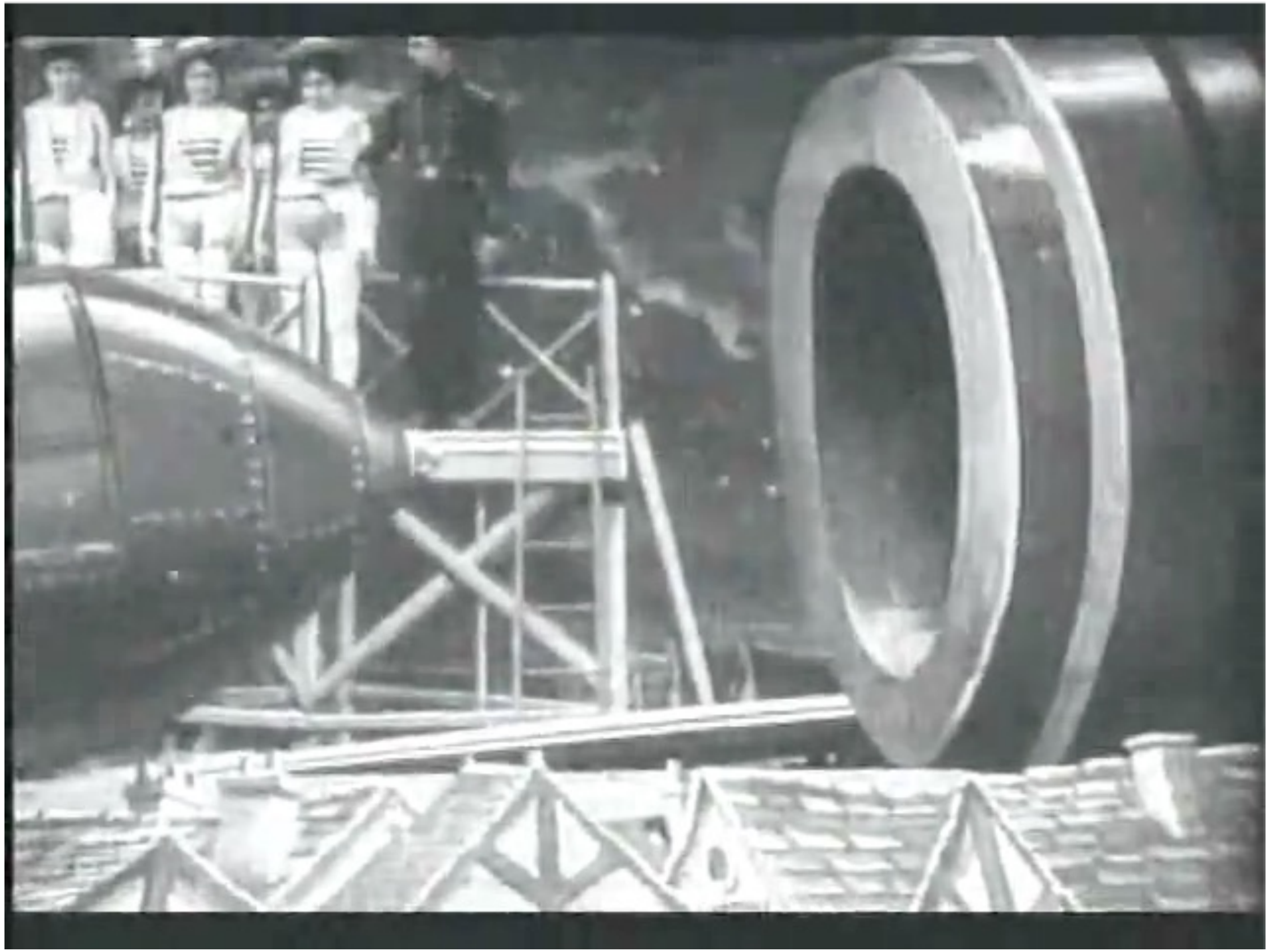
Motion pictures

- Sensitive roll film enables sampling in time
- 1890s - several cameras
  - Lumière brothers' Cinematographe
  - Edison's Kinescope

Cinématographe [Wikimedia commons]

George Eastman and Thomas Edison in 1928

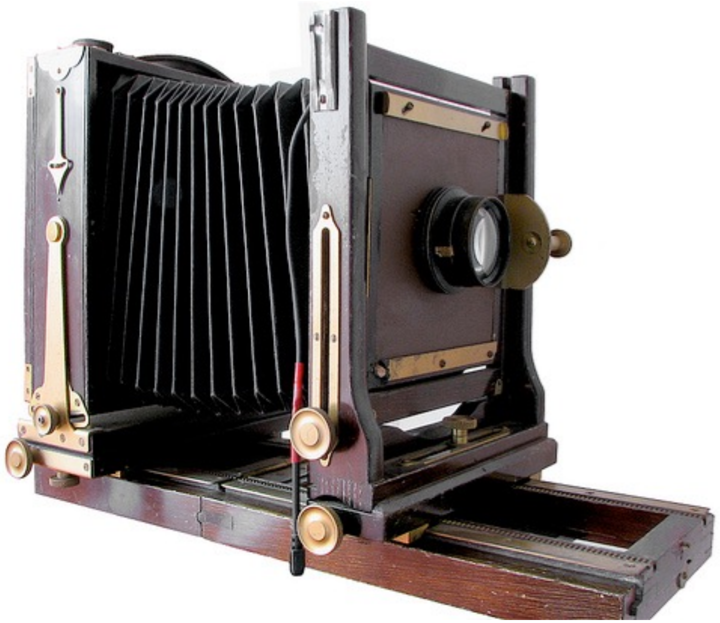
George Méliès



Georges Méliès, A Trip to the Moon, 1902

Improvements in cameras

- Size and portability
- Ease of use
- Automation



Improvements in film

- Sensitivity
  - enables photographs of faster subjects—“faster” film
- Dynamic range
  - higher quality images with detail in highlights and shadows
  - expanded “latitude” to mess up the exposure
- Resolution
  - enables smaller format cameras

Television

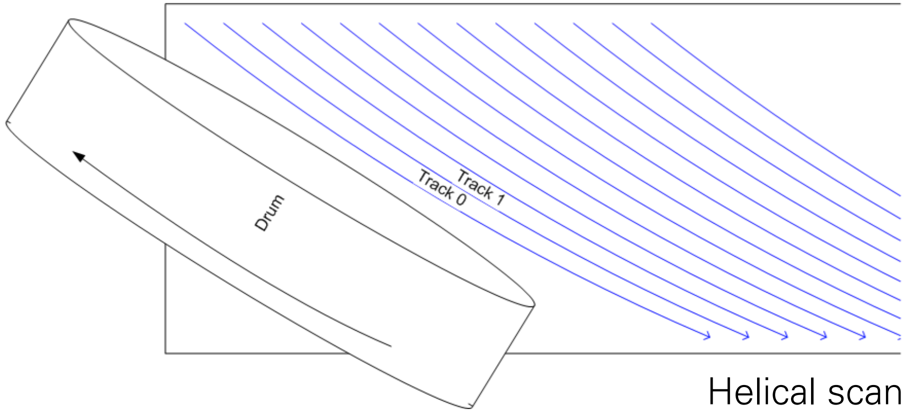
- Practical around 1927 (Farnsworth)
- Camera basically the same
  - imaging lens plus planar image sensor
- Recording is electronic
  - various early schemes
  - early winner: CRT image sensors (Orthicon, Vidicon, ...)
- Initially seems quite different from photography/cinematography
  - ephemeral output signal - live viewing only
  - low resolution, low dynamic range images

Farnsworth Archives

Philo Farnsworth, c. 1935

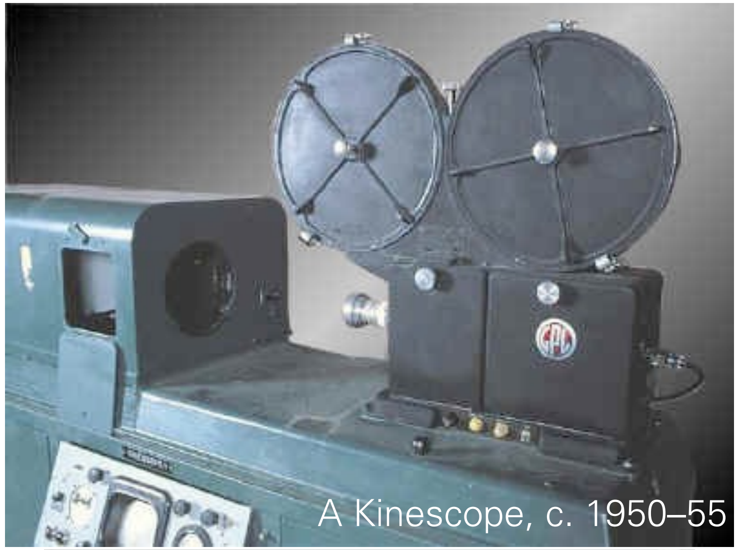
Recording video signals

- Kinescope (1940s)
  - photograph onto motion picture film re-photograph the film for replay
- Videotape (1956)
  - record signal on magnetic tape
  - very high head velocities required - transverse or helical scanning



Wikipedia

Helical scan



A Kinescope, c. 1950–55

Peter Lindell, Canada Science and Technology Museum

A 2-inch video tape recorder c.1970

Wikipedia

Imaging around 1950s–70s

- Technology improves incrementally
  - Film emulsions improve; very high quality attainable in large formats
  - Video technology improves; but standards keep resolution fixed
  - Lens designs improve, cameras become much more usable
- Usage is refined
  - Photography an established art form, widespread hobby
  - Cinematography develops as a storytelling medium
  - Television becomes dominant mass communication medium

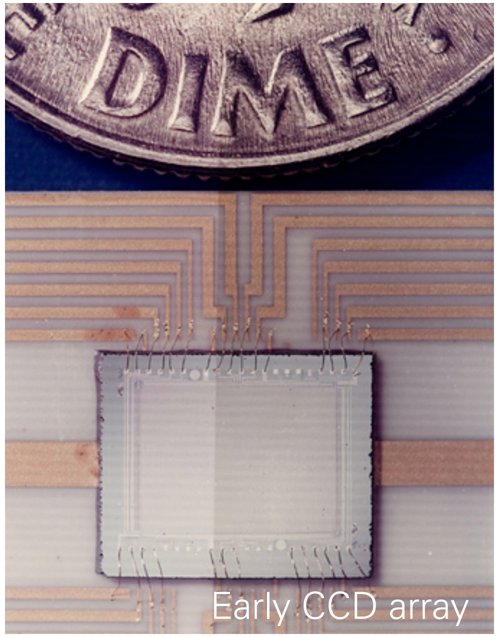
Meanwhile…

- Invention of CCD (1969)
  - solid-state, fundamentally discrete image sensor
  - quickly established in astronomy, space
  - by mid-80s, displaces tubes in video cameras (as drop-in replacement)
- Computing and computer graphics
  - sufficient memory to store images becomes available
  - first framebuffers developed 1972–74
- Digital signal transmission and processing
  - used for audio and telephone
- These set the stage for the next revolution

Photo: Alcatel-Lucent/Bell Labs

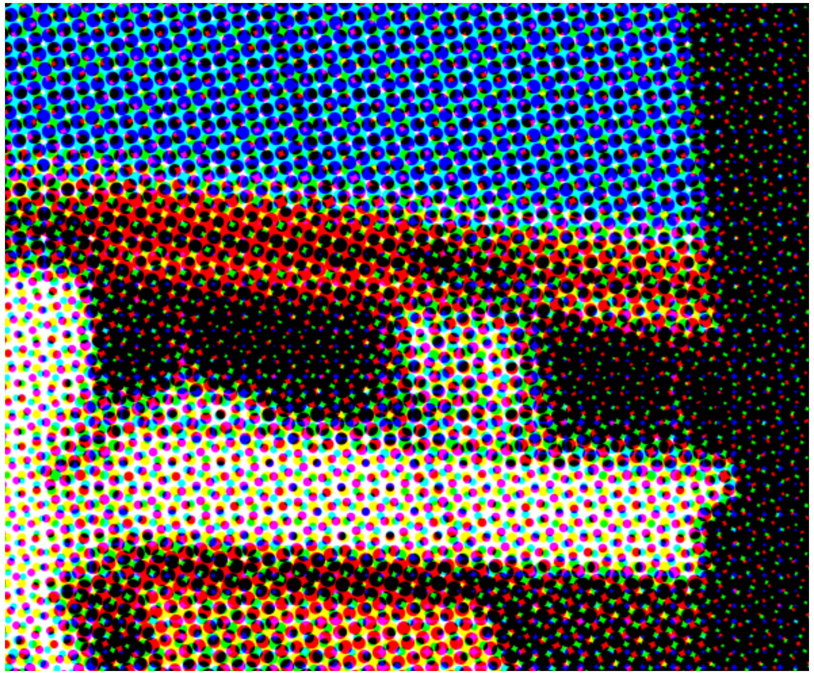

George Smith and Willard Boyle in 1970

Photo: Alcatel-Lucent/Bell Labs



Early CCD array

Digital imaging



- Halftone printing of images
  - halftone process around for a while
  - complex, delicate optical procedure
  - moving images from place to place requires moving film or paper
- Digital imaging
  - scan images from film or paper
  - transmit images by phone
  - do processing (e.g. halftone separation) by computing print images using laser printer or laser film recorder
- Image editing
  - 1990—Adobe Photoshop 1.0
- Image compression algorithms
  - make image storage, transmission more practical

Photoshop toolbar from version 1.07

Digital photography

- Digital images are established
    - people can make use of them directly
- CCD sensors improve
    - Moore’s law makes pixels smaller
    - video cameras already recording images electronically
    - digital image capture used in scientific applications
- Analog electronic still camera (aka. still video camera)
    - is just a video camera that takes one frame at a time
    - several manufacturers made them
    - but high image quality expectations for stills delays acceptance

First microprocessor in a camera, Canon AE-1976

Canon RC-701 still video camera, 1986

Early digital cameras

- Important limitations
  - low image quality (relative to film)
  - slow camera performance
  - large, heavy, clunky
  - limited, expensive image storage
- Important advantages
  - immediate availability of images
  - zero (well...) marginal cost per exposure
- First adopters: photojournalists
- Kodak DCS series
  - based on film camera bodies
  - early commercial success
  - storage: PCMCIA hard disks (mid 90s)

digicamhistory.com

Kodak DCS-100, 1991

Kodak DCS-100, 1991

Digital rivals film

- Key improvements
  - cameras become more compact
  - resolution and dynamic range improve
  - LCD displays for immediate image review
  - costs drop
- Meanwhile
  - computers with high-quality color displays become pervasive
- User experience
  - image review is a big change for users
  - sharing of digital images suddenly becomes easier than prints

Digital video

- Initially: improved recording medium
  - record the same old signal, but digitally best-quality medium for professional use
- Improvements
  - storage and bandwidth improve by orders of magnitude
  - video compression algorithms advance
  - digital formats become simpler/better than analog-derived
  - flexibility finally unlocks video resolution
- Digital recording becomes standard for video
  - basic experience similar
  - cost and quality greatly improved

Digital displaces film and video

- Move from convenience vs. quality to convenience and quality
- Digital slowly takes over for basically all users
  - advances in storage/transmission and compression algorithms
  - ecosystem for online sharing of photos, videos
  - declining use of printed images
- Last bastion: cinematography
  - delay: quality standards plus tradition
  - first took over low end because of film costs
  - now taking over high end because of superior quality/usability

Digital cameras today

- Digital SLRs
  - high-end product for professionals and enthusiasts
- Digital cinema
  - high-resolution cameras for big-budget film production
- HD video
  - medium resolution for low-end film and high-end TV production
- Mirrorless system cameras
  - smaller high-end cameras with electronic viewfinding
- Compact still cameras
  - inexpensive, auto-everything for day-to-day usage
- Tiny cameras in all cell phones
  - "The best camera is the camera that is with you"

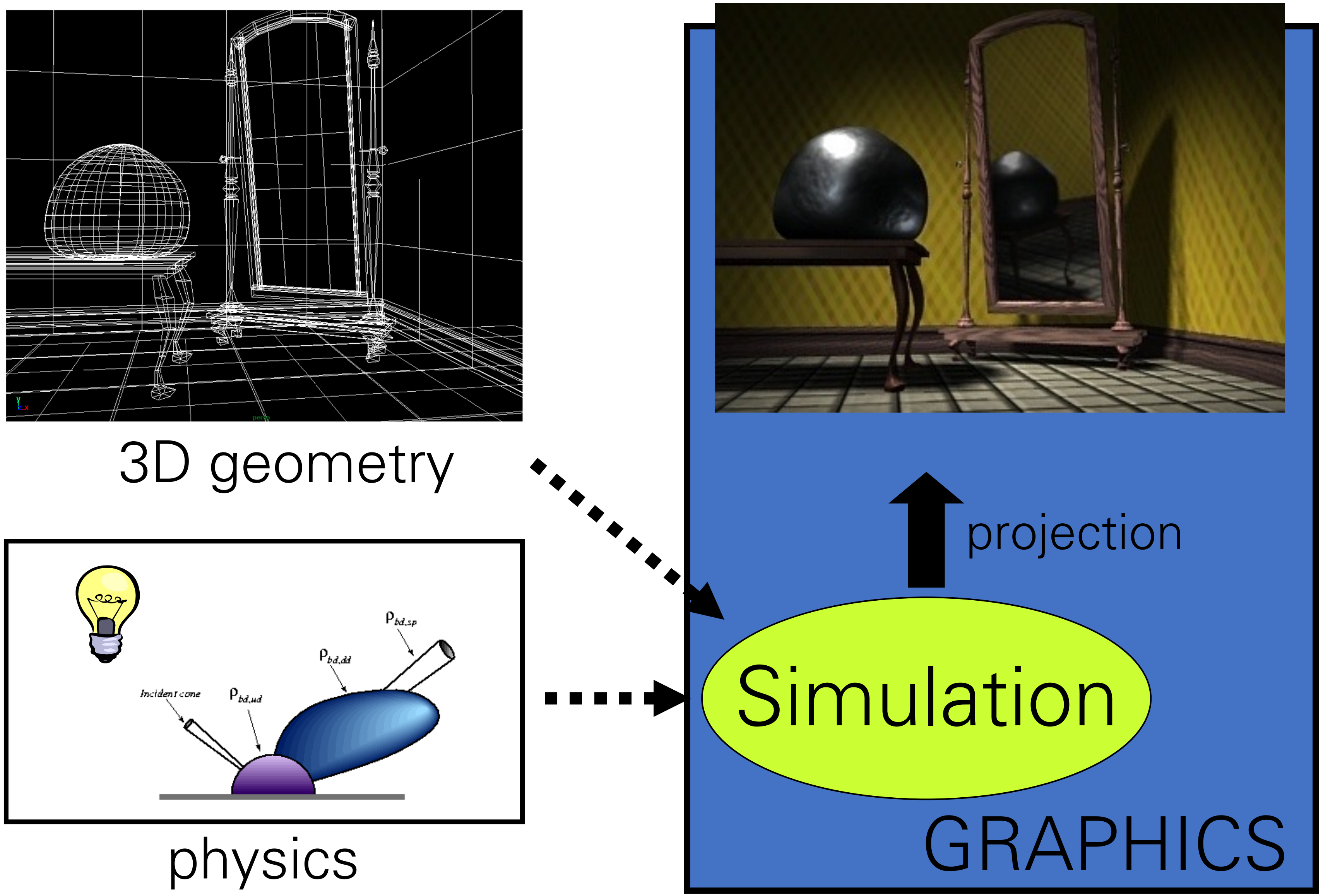

Digital photography today

- Video, photography, and cinema have converged
  - all using the same basic technology
  - all modern still cameras do video too (and many vice versa)
- Cameras becoming completely pervasive
  - film-equivalent quality possible in <1 cm^3
  - mobile applications driving much sensor/lens development
  - mobile cameras had eaten compact digicam market
- Computing power still rapidly advancing
  - more and more computation being done on images

Computer Graphics?

- Computers to create image
- Sketchpad, 1961, Ivan Sutherland’s MIT PhD thesis

Traditional Computer Graphics



State of the Art

- Amazingly real
- But so sterile, lifeless, futuristic (why?)

The richness of our everyday world

Photo by Svetlana Lazebnik

Beauty in complexity

University Parks, Oxford

Which parts are hard to model?

Photo by Svetlana Lazebnik

People

From “Final Fantasy”

On the Tube, London

GenAI - Generative AI

- refers to the set of recent techniques (mostly based on deep learning) which employs existing content (like text, images, videos, speech, codes, etc.) to generate new plausible content.

- Many interesting applications, and application domains.

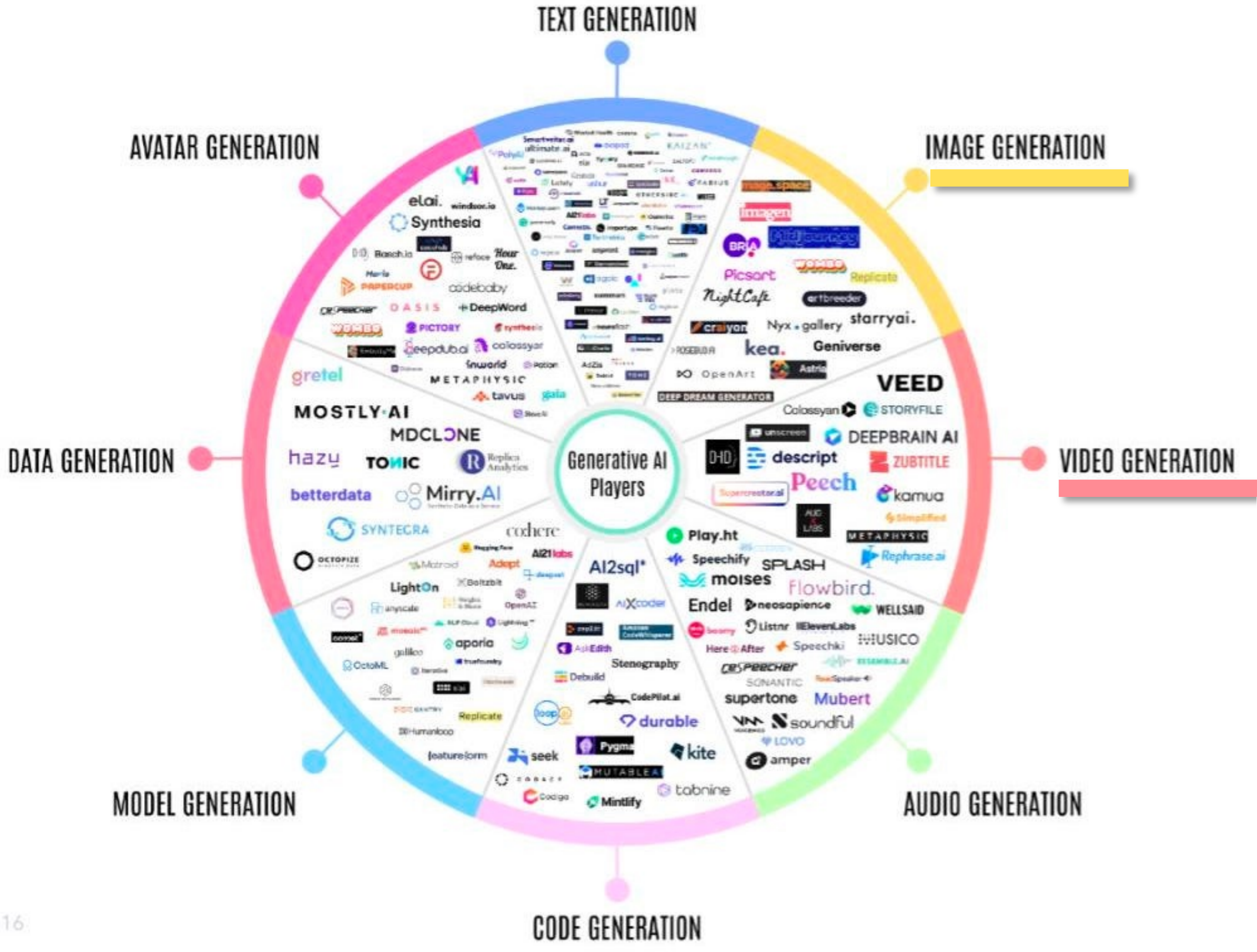


Image credit: B. Kaplan

GenAI - Generative AI

Image credit: B. Kaplan

GenAI - Generative AI

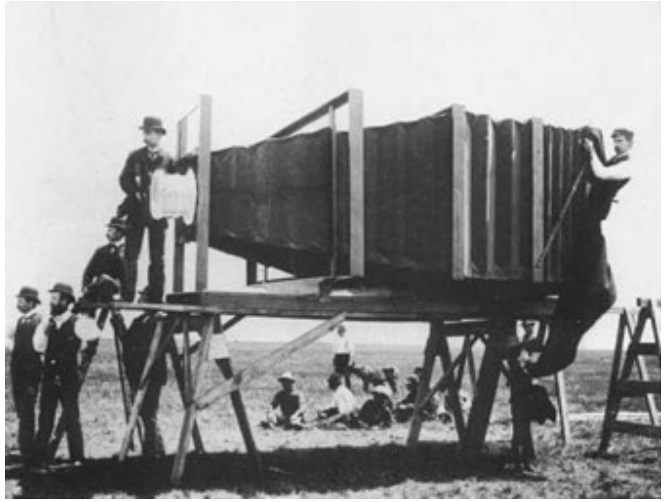
Image credit: Rombach et al., 2022

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

The unfinished revolution

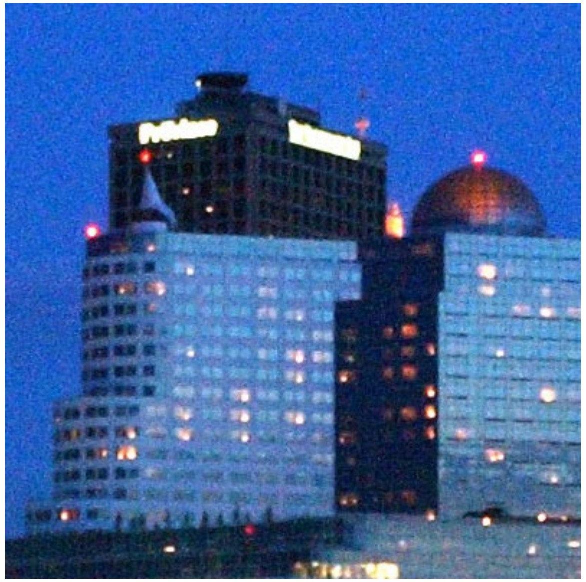
- Traditional photography:
  - optics focuses optical array onto sensor
  - chemistry records final image
- Digital photography
  - optics focuses optical array onto sensor
  - digital sensor records final image



Slide credit: F. Durand

Limitations of traditional photography

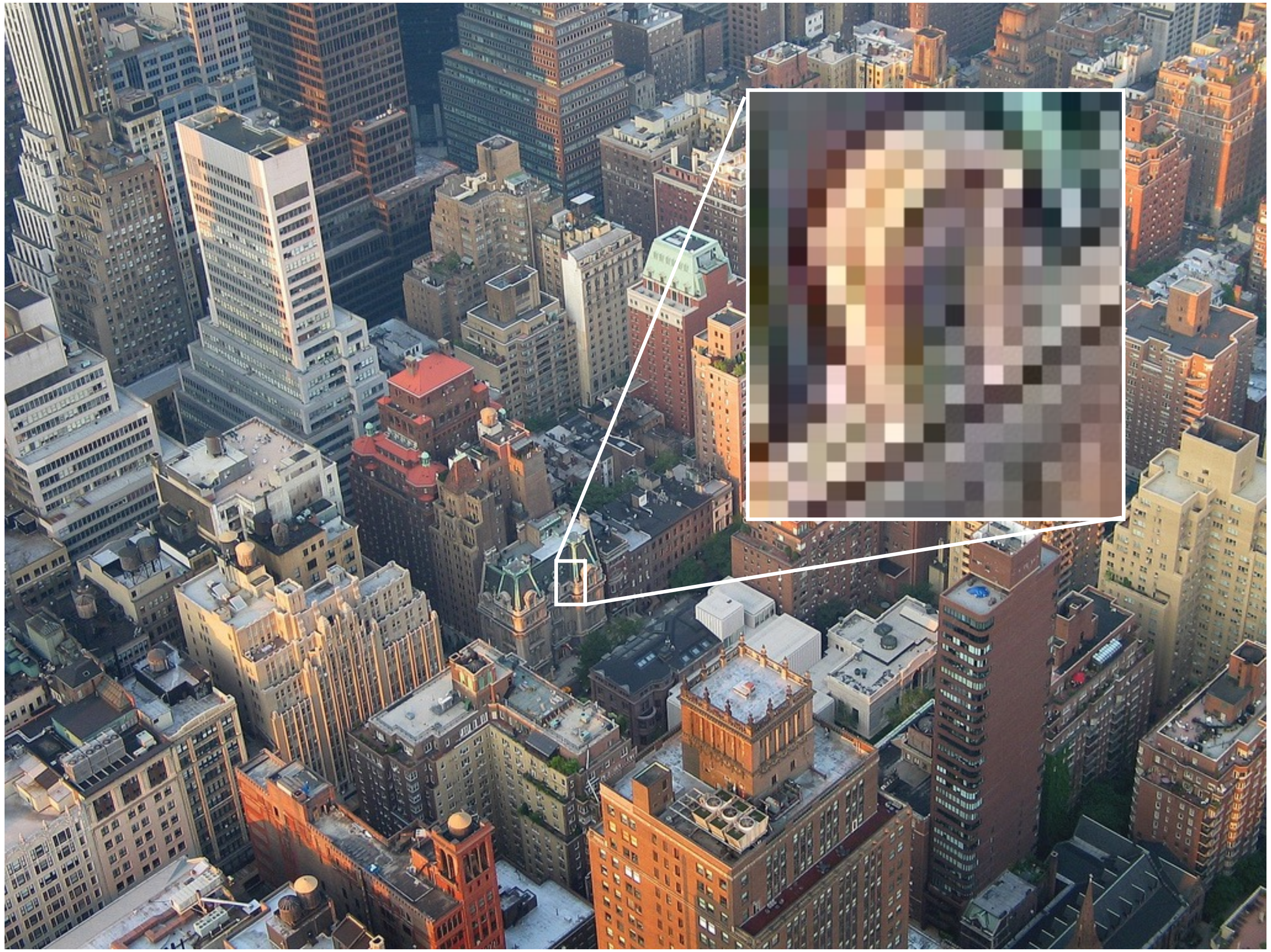
- Blur, camera shake, noise, damage



Slide credit: S. Lazebnik

Limitations of traditional photography

- Limited resolution



Slide credit: S. Lazebnik

Limitations of traditional photography

- Bad color / no color

Slide credit: S. Lazebnik

Limitations of traditional photography

- Unwanted objects

Slide credit: S. Lazebnik

Limitations of traditional photography

- Unfortunate expressions

Slide credit: S. Lazebnik

Limitations of traditional photography

- Limited dynamic range

Slide credit: S. Lazebnik

Limitations of traditional photography

- Single viewpoint, static 2D picture

Slide credit: S. Lazebnik

Limitations of traditional photography

- Single depth of focus

Slide credit: S. Lazebnik

Creating Realistic Imagery

Computer Graphics

+ great creative possibilities

+ easy to manipulate objects or viewpoint

- tremendous expertise and effort to obtain realism

Computational Photography

Realism
Manipulation
Ease of capture

Photography

+ instantly realistic

+ easy to acquire

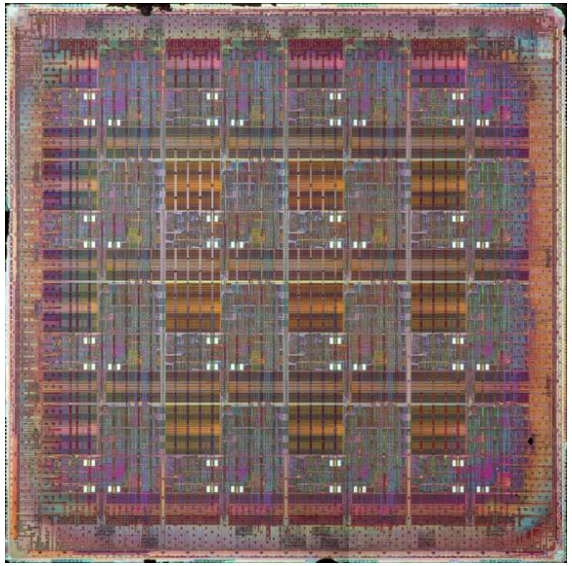
- very hard to manipulate objects or viewpoint

Slide credit: S. Lazebnik

Computational Photography

- Arbitrary computation between the optical array and the final image
- Data recorded by sensor is not the final image

Generalized imaging



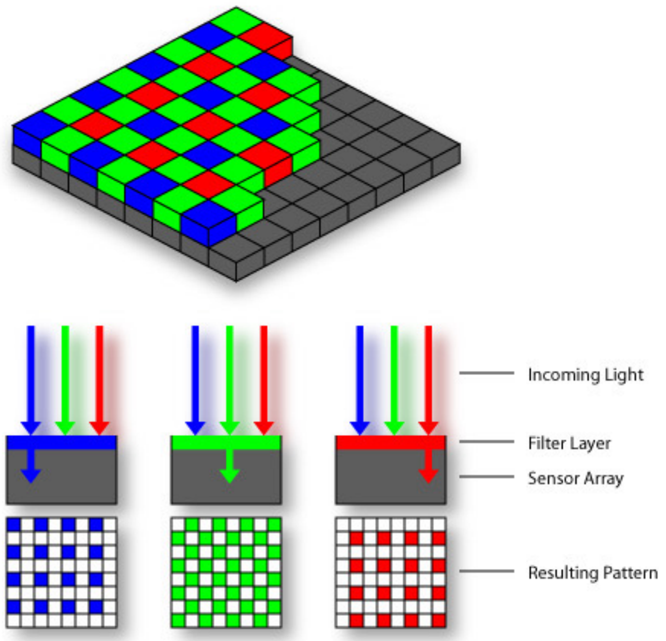
Lots of computation

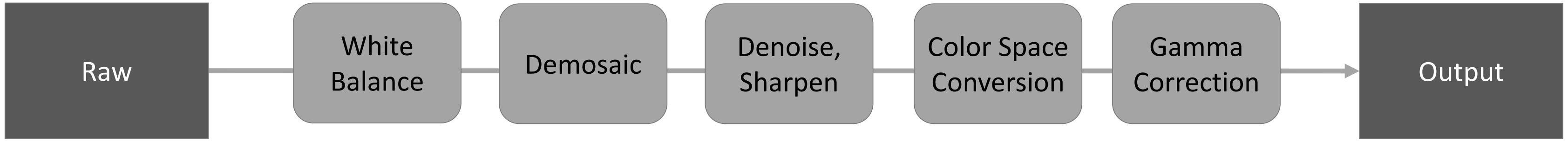
Final image

Slide credit: F. Durand

Computational Photography

- Arbitrary computation between the optical array and the final image
- Data recorded by sensor is not the final image





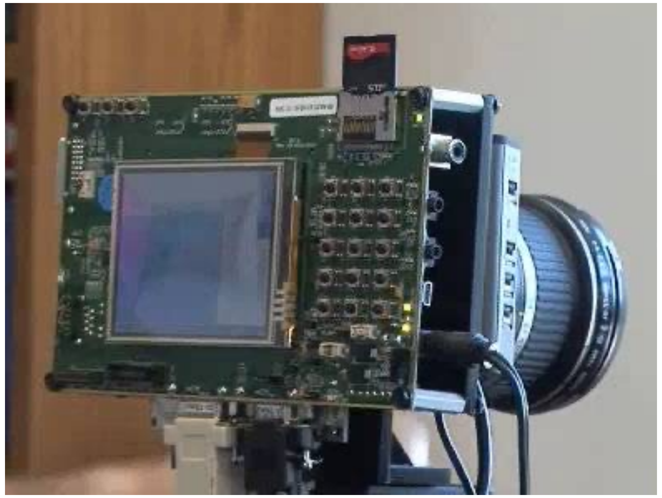
Slide adapterd from F. Durand

Computational Photography

- Arbitrary computation between the optical array and the final image
- Post-process after traditional imaging
  - a.k.a. image processing (maybe more interactive)
  - But also combine multiple images to overcome limits of traditional imaging (HDR, panorama)
- Design imaging architecture together with computation
  - Computational cameras, computational illumination, coded imaging, data-rich imaging
- Extract more than just 2D images
- New media (panorama, photo tourism)

Slide credit: F. Durand

Computational Photography



- How can I use computational techniques to capture light in new ways?
- How can I use computational techniques to breathe new life into the photograph?
- How can I use computational techniques to synthesize and organize photo collections?

Slide credit: F. Durand

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

Photo Style Transfer

Image credit: Ledig et al., 2017

Photo Style Transfer

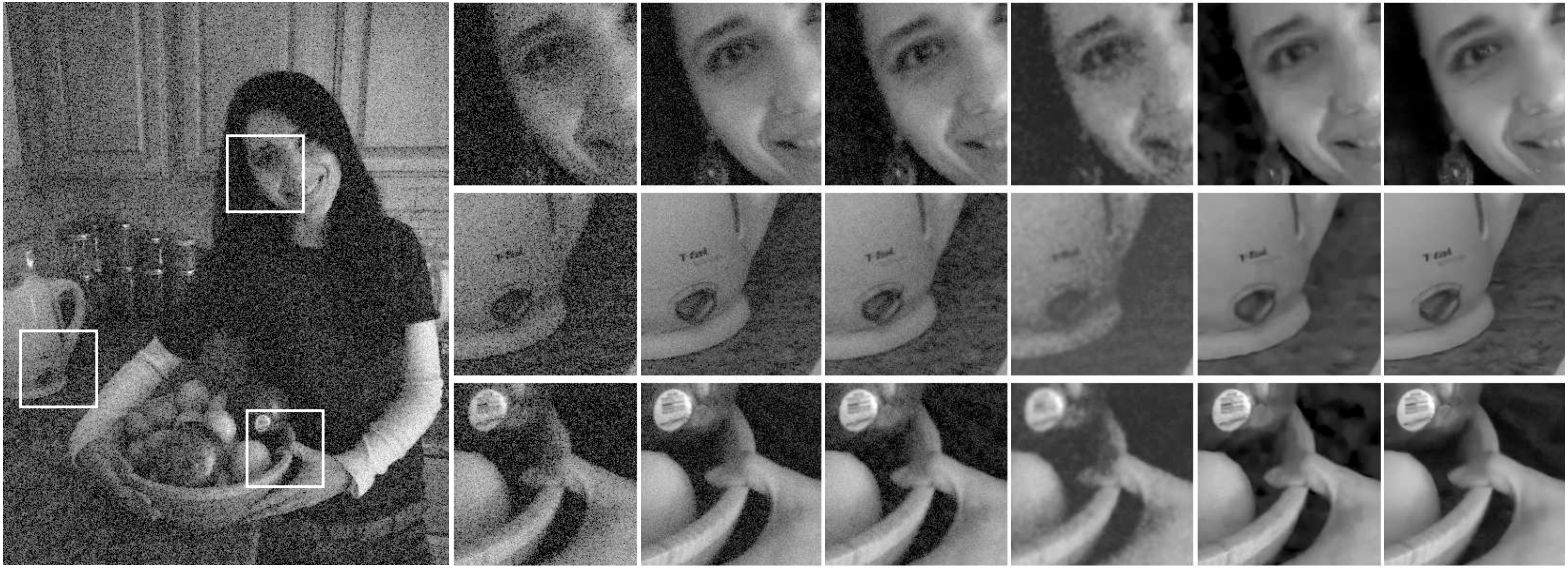
Image credit: Ledig et al., 2017

Image Relighting


(a) Input image and estimated lighting (b) Rendered images from our method under three novel illuminations

Image credit: Sun et al., 2019

Image Denoising



Reference frame (a) Reference (b) Average (c) HDR+ (d) NLM (e) VBM4D (f) Ours (KPN)

Image credit: Mildenhall et al., 2018

Image Super Resolution

bicubic (21.59dB/0.6423) | SRResNet (23.53dB/0.7832) | SRGAN (21.15dB/0.6868) | original

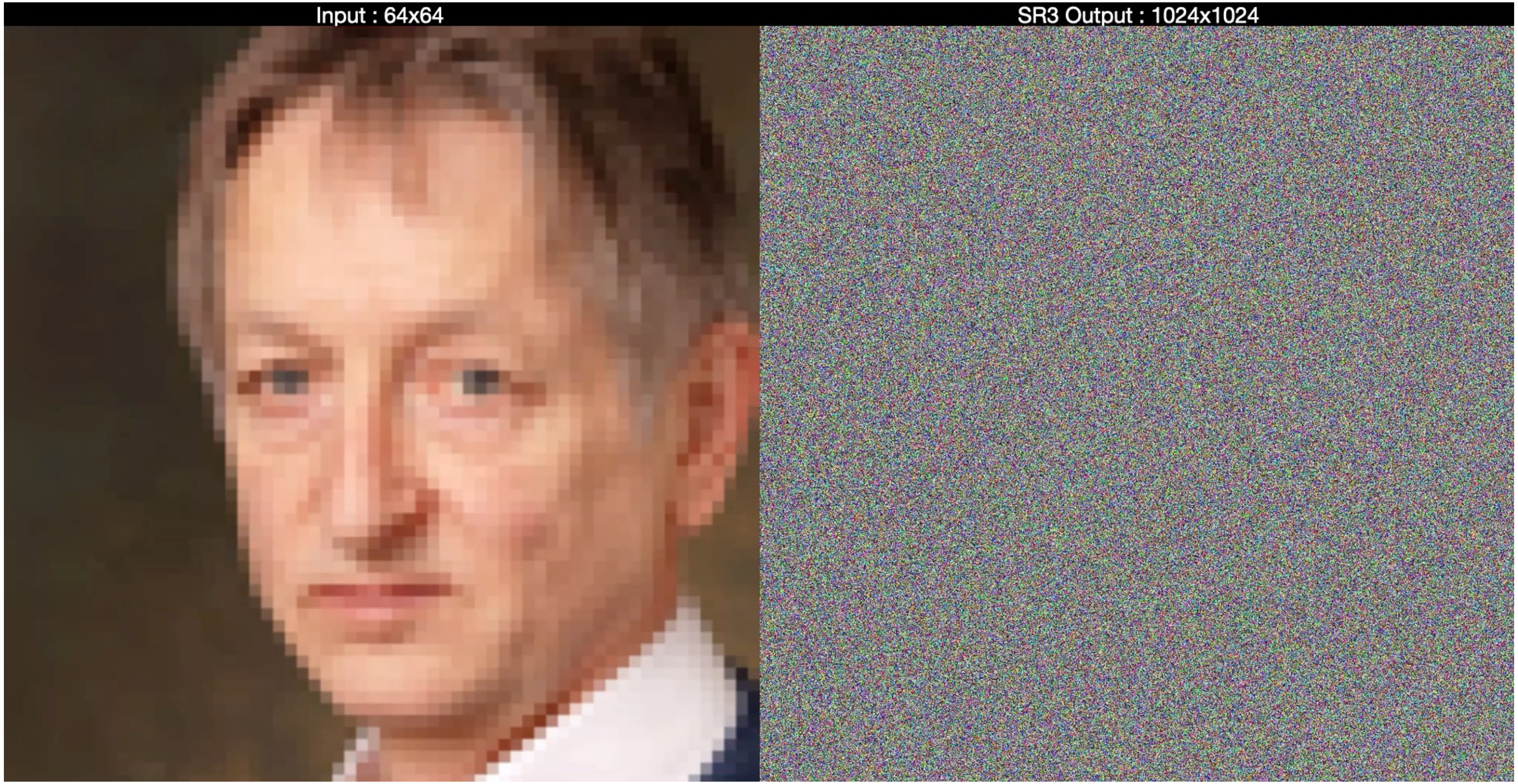
Image credit: Ledig et al., 2017

Image Super Resolution

LR ($\times 8$) | RRDB | ESRGAN | SRFlow, $\tau = 0.8$ | **HCFlow**, $\tau = 0.8$ | **HCFlow++**, $\tau = 0.8$ | Ground Truth

Image credit: Liang et al., 2021

Image Super Resolution



Results of a SR3 model (64×64 → 512×512), trained on FFHQ, and applied to images outside of the training set.

Image credit: Saharia et al., 2021

Image Deblurring

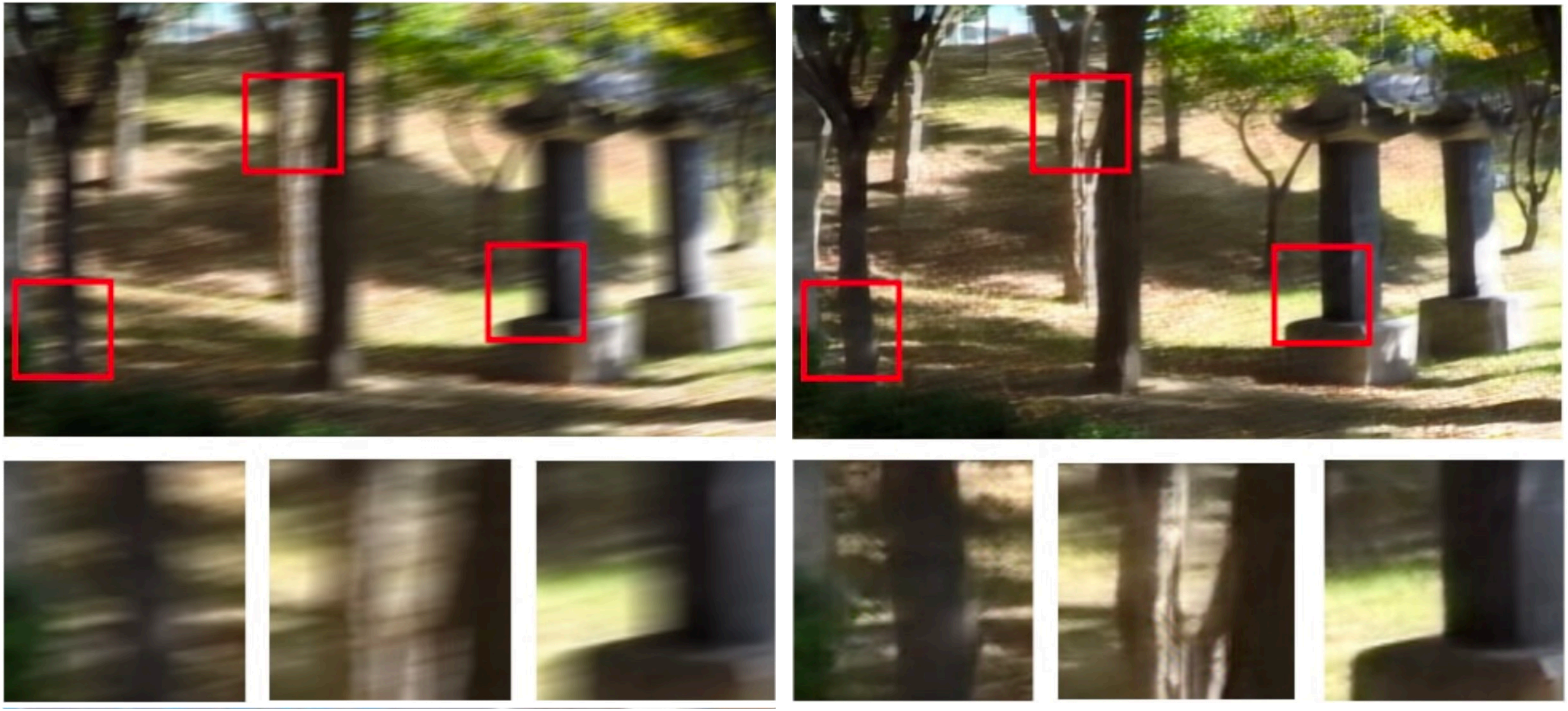


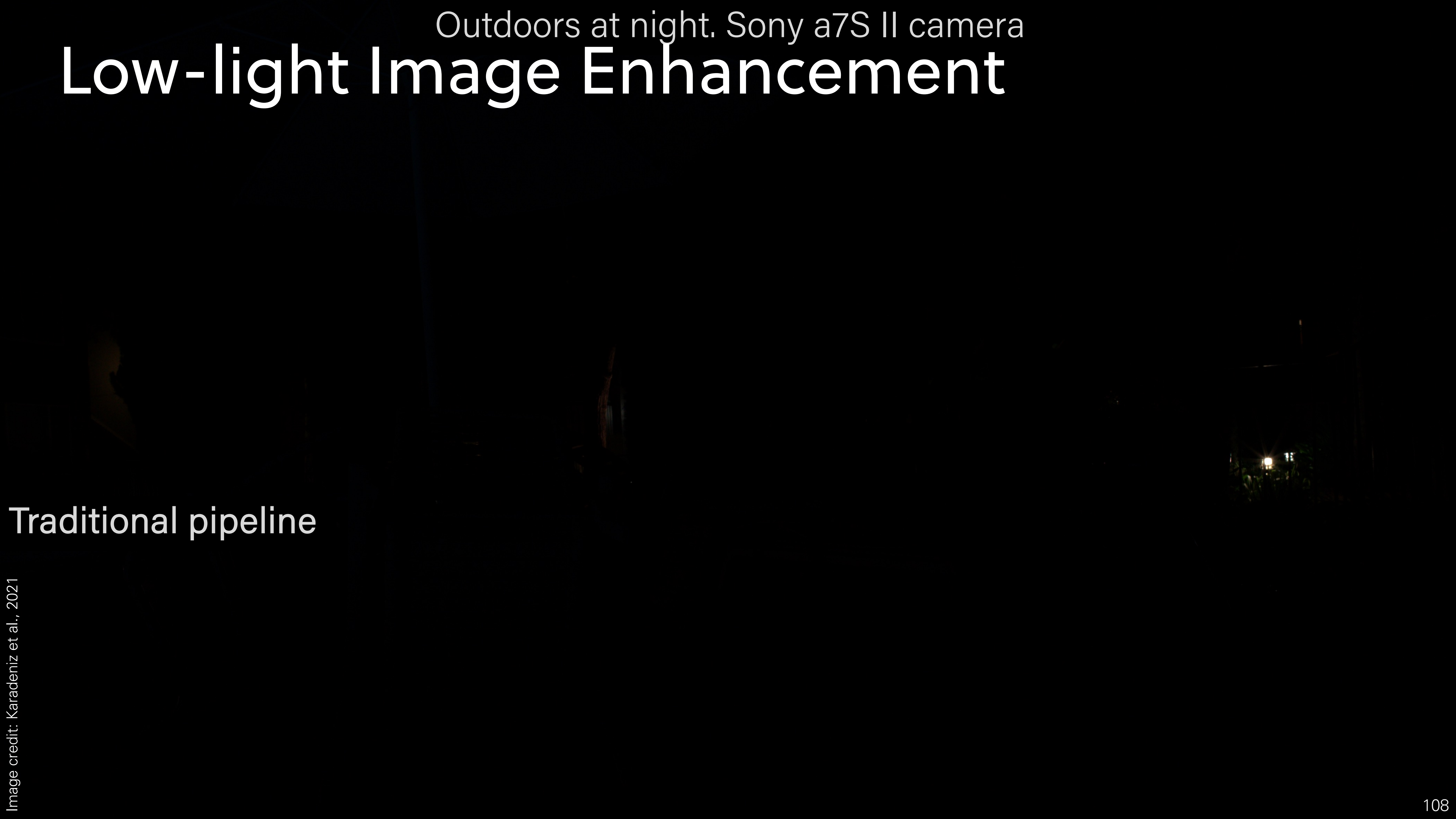
Image credit: Kupyn et al., 2018

Image Deblurring

Image credit: Tu et al., 2022

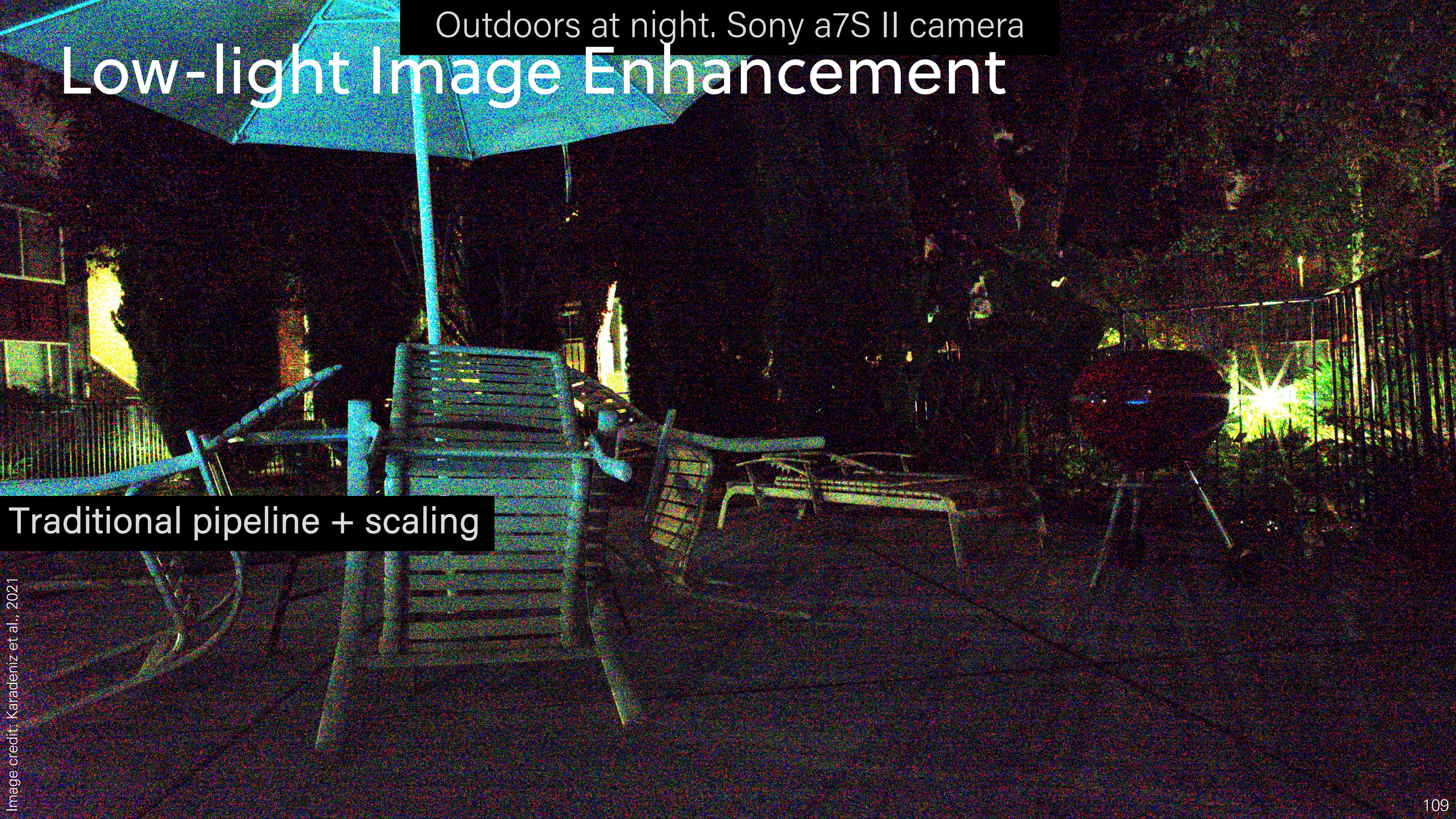
Low-light Image Enhancement

Outdoors at night. Sony a7S II camera



Traditional pipeline

Image credit: Karadeniz et al., 2021


Outdoors at night. Sony a7S II camera
Low-light Image Enhancement
Traditional pipeline + scaling
Image credit: Karadeniz et al., 2021

Low-light Image Enhancement

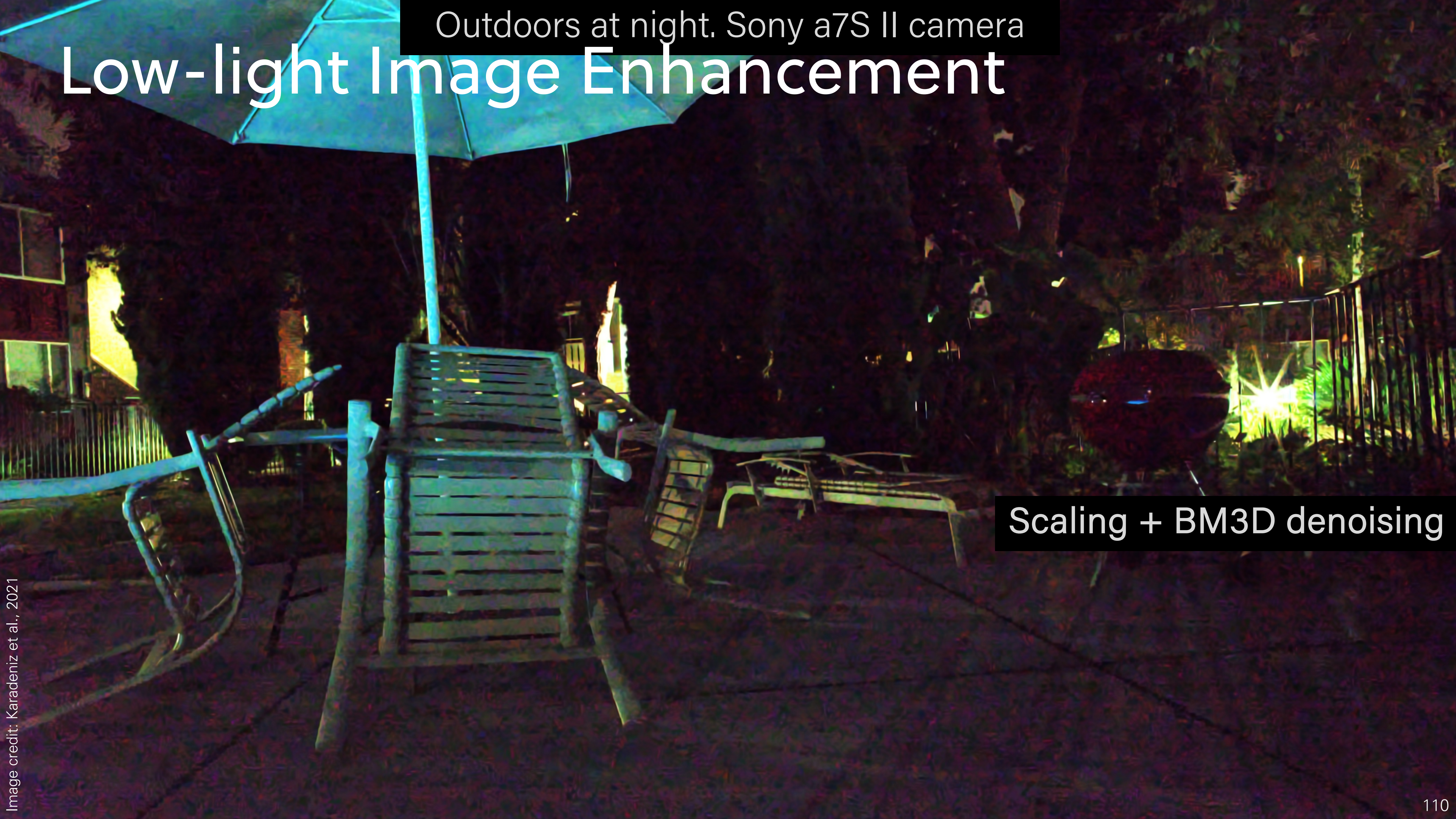


Image credit: Karadeniz et al., 2021

Low-light Image Enhancement

Low-light Image Enhancement

Image credit: Karadeniz et al., 2021

Generating Synthetic Images

Image credit: Brock et al., 2018

Generating Synthetic Images

Image credit: Rombach et al., 2022

SDXL

Image credit: Podell et al., 2023

SDXL

Image credit: Podell et al., 2023

Emu

Image credit: Sun et al., 2023

Emu

Utensils, a bottle, and a glass positioned behind a stove

A decadent chocolate treat adorned with decorative sugar art

A beaver dressed in a vest, wearing glasses and a vibrant necktie, in a library

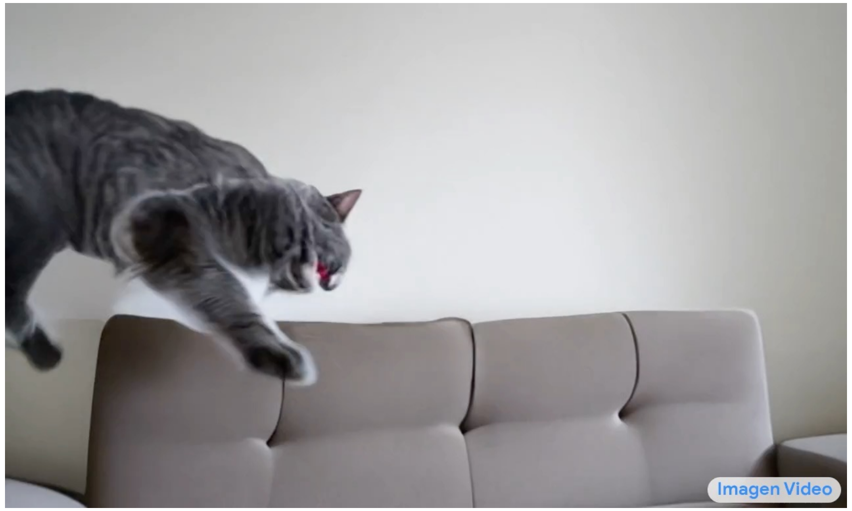
a cow eating a green leafy plant

An emu wearing sunglasses and chilling on a beach

Image credit: Sun et al., 2023

Generating Videos from Text - Imagen Video

A teddy bear
running in New York City

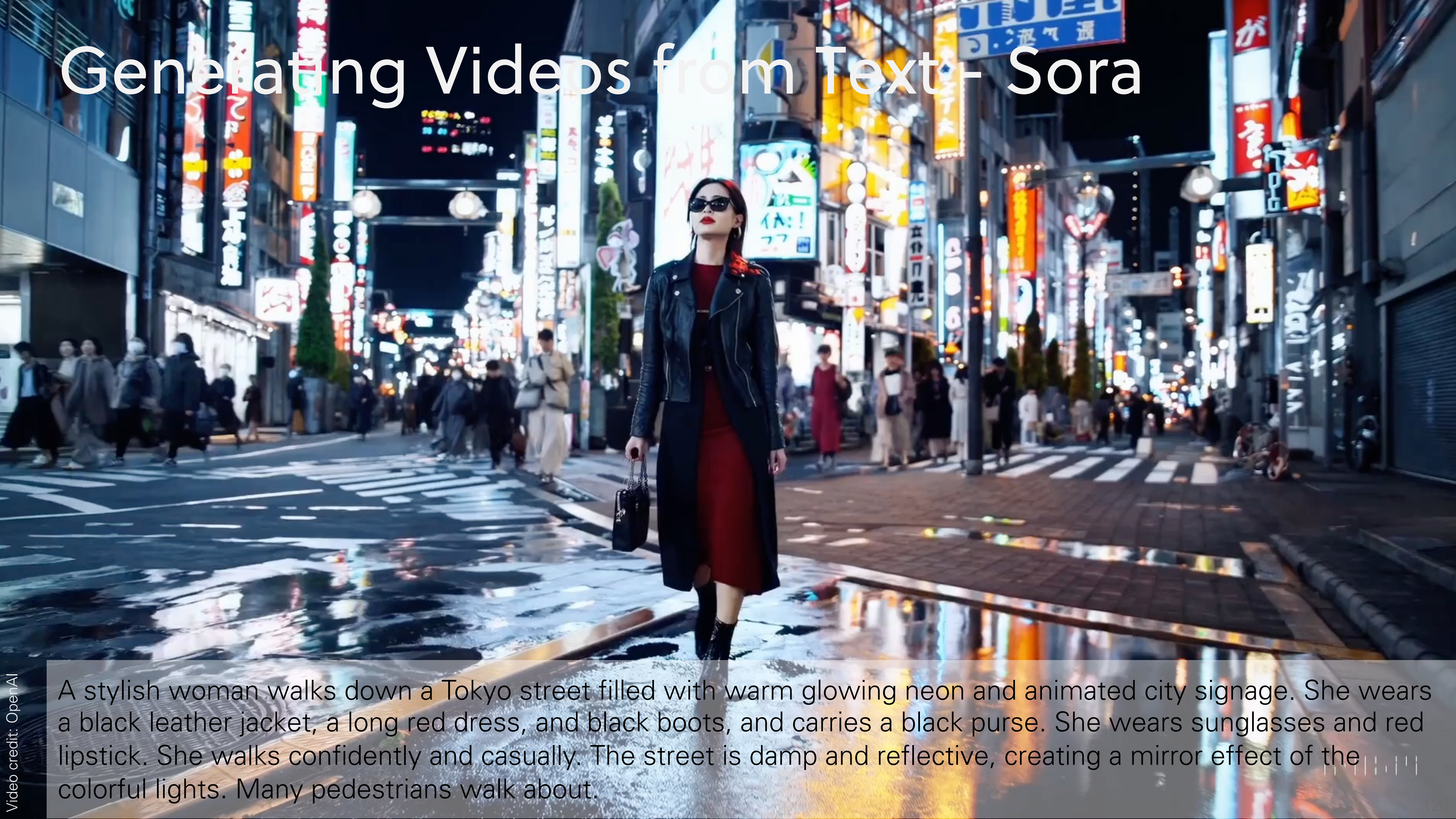


A british shorthair
jumping over a coach

A swarm of bees
flying around their hive

Video credit: Ho et al., 2022

Generating Videos from Text - Sora



A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

Video credit: OpenAI

Generating Videos from Text - Sora

A young man at his 20s is sitting on a piece of cloud in the sky, reading a book.

Video credit: OpenAI

Generating Videos from Text - Sora

The camera directly faces colorful buildings in burano italy. An adorable dalmation looks through a window on a building on the ground floor. Many people are walking and cycling along the canal streets in front of the buildings.

Video credit: OpenAI

Time-travel Rephotography

Image credit: Luo et al. 2021

Image Inpainting

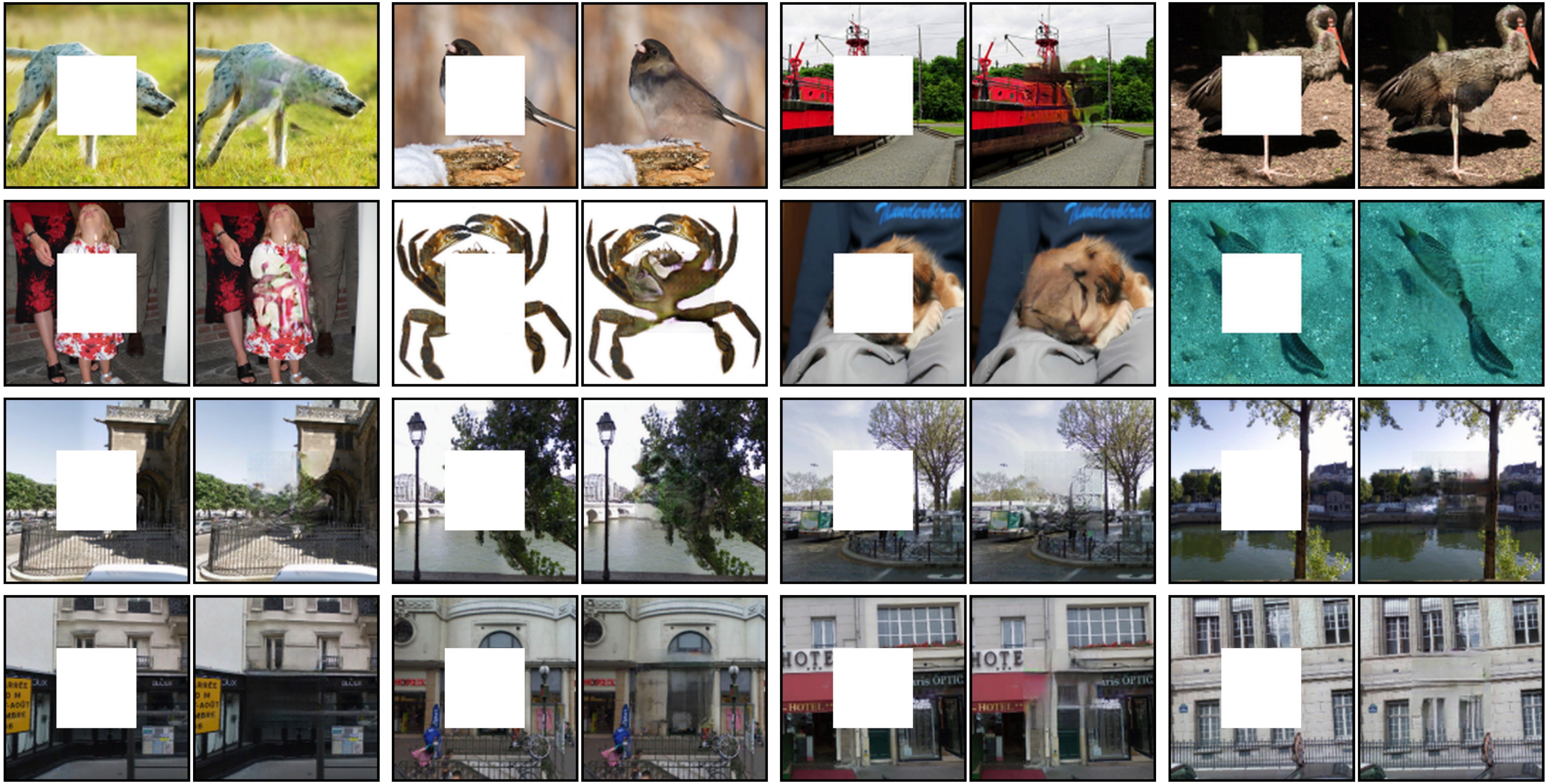


Image credit: Pathak et al., 2016

Image Inpainting

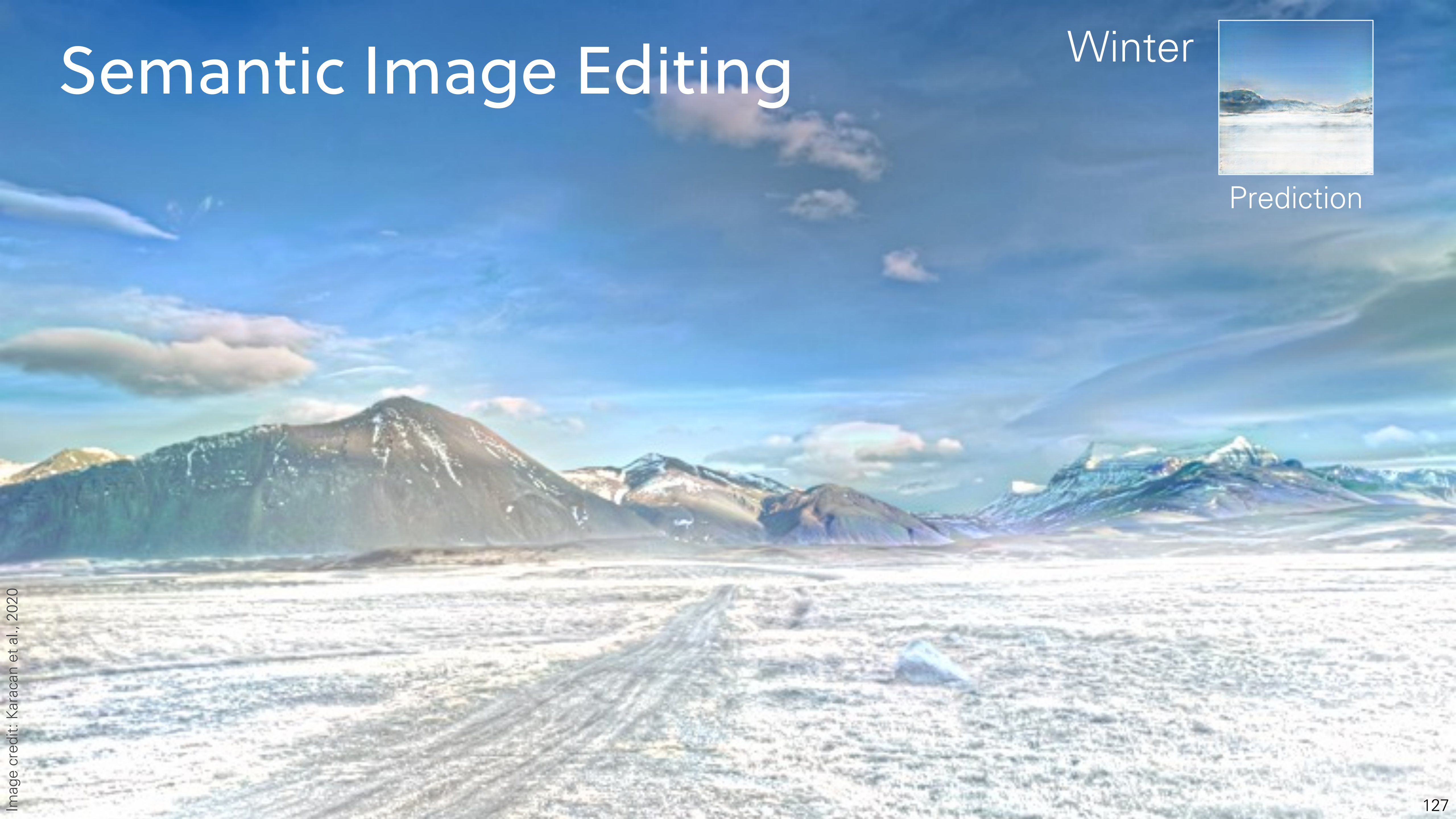
Image credit: Rombach et al., 2022

Semantic Image Editing

Semantic Layout

Image credit: Karacan et al., 2020

Semantic Image Editing



Winter

Prediction

Image credit: Karacan et al., 2020

Semantic Image Editing

Spring
+
Clouds

Prediction

Image credit: Karacan et al., 2020

Semantic Image Editing

Input image | Input mask | "beach" | "big mountain" | "The Great Pyramid of Giza"

Input image | Input mask | "gravestone" | "toy truck" | "snake"

Image credit: Avrahami et al., 2022

Target: This person has mustache

Image credit: Baykal et al., 2023

Target: A cat with ginger hair

Image credit: Baykal et al., 2023


Image credit: Baykal et al., 2023

Target: This bird has wings that are blue and has a white belly

Source image

Target domain

Image credit: Anees et al., 2024

Source image

Resulting image

Image credit: Anees et al., 2024

Source image

Target image

Image credit: Anees et al., 2024

Source image

Resulting image

Image credit: Anees et al., 2024

Source image

Target image

Image credit: Anees et al., 2024

Source image

Resulting image

Image credit: Anees et al., 2024

Source image

Target image

Image credit: Anees et al., 2024

Source image

Resulting image

Image credit: Anees et al., 2024

Source image

Target: This young person has black hair and bangs

Image credit

Source image

Target: Elsa from Frozen

Image credit: Anees et al., 2024

Instruction-Based Object Removal

Image credit: Yildirim et al., 2023

Instruction-Based Object Removal

remove the gray kite at the left

remove the street light at the left

remove the man at the right of the man

remove the red car at the left of the tall ladder

remove the colorful train at the right

remove the boat at the right of the small boat

Image credit: Yildirim et al., 2023

Visual Quality Assessment

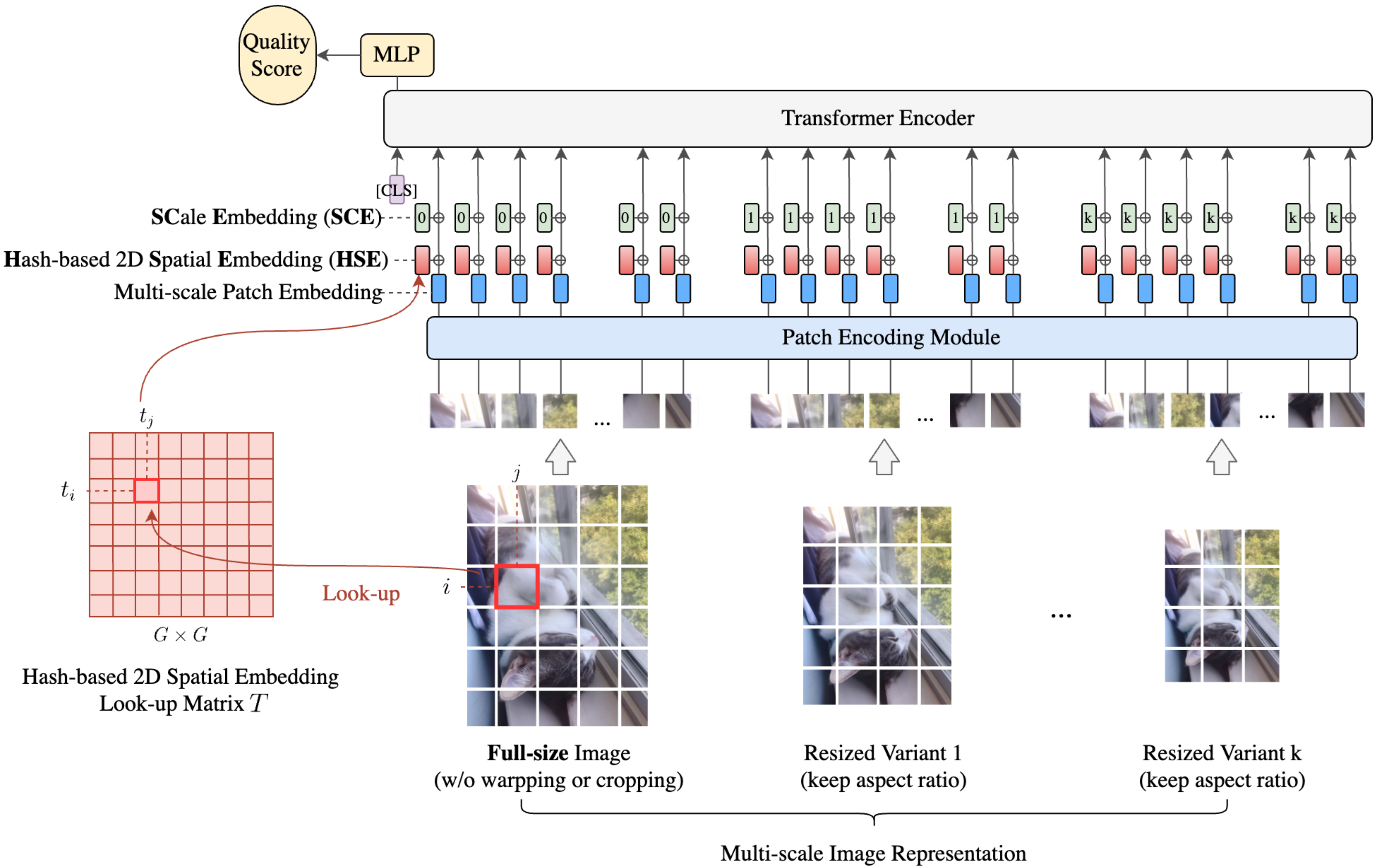


Image credit: Ke et al., 2021

Visual Quality Assessment of 360° images

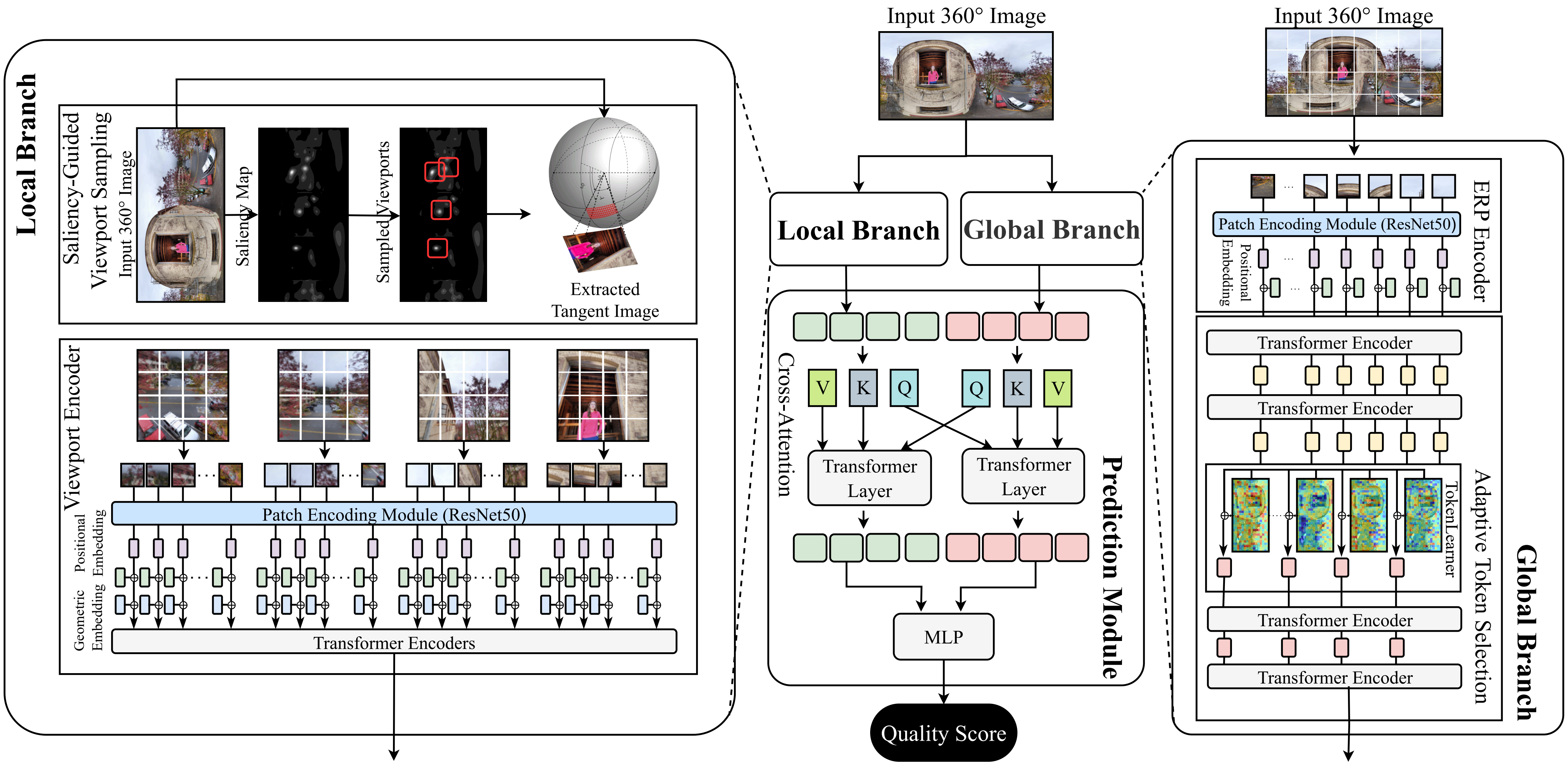
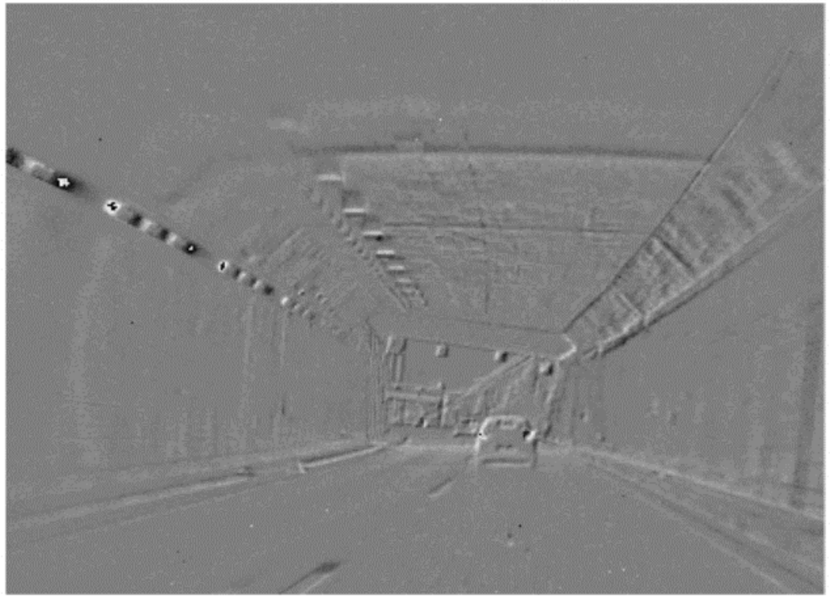
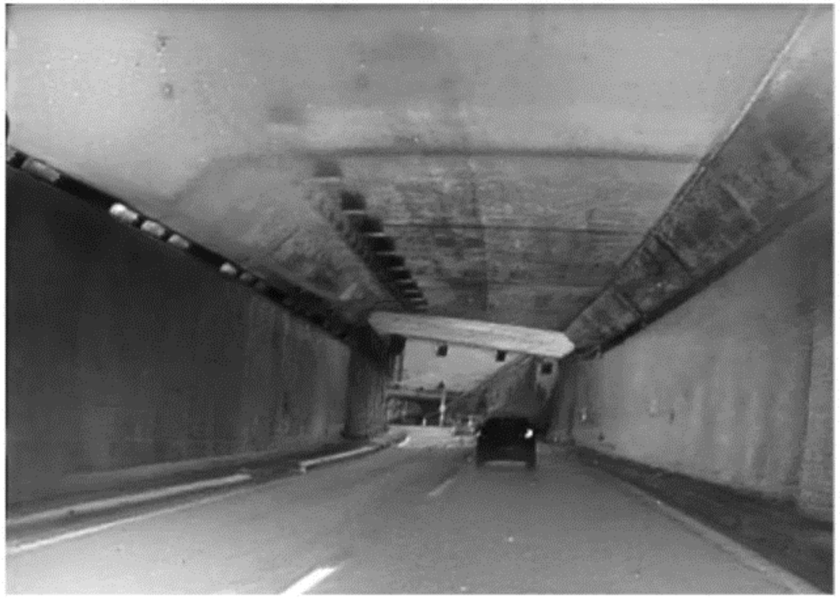


Image credit: Jabbari Tofighi et al., 2024

Video Generation from Events



Events



Our reconstruction

Phone camera

Video credit: Rebecq et al., 2019

Video Generation from Events

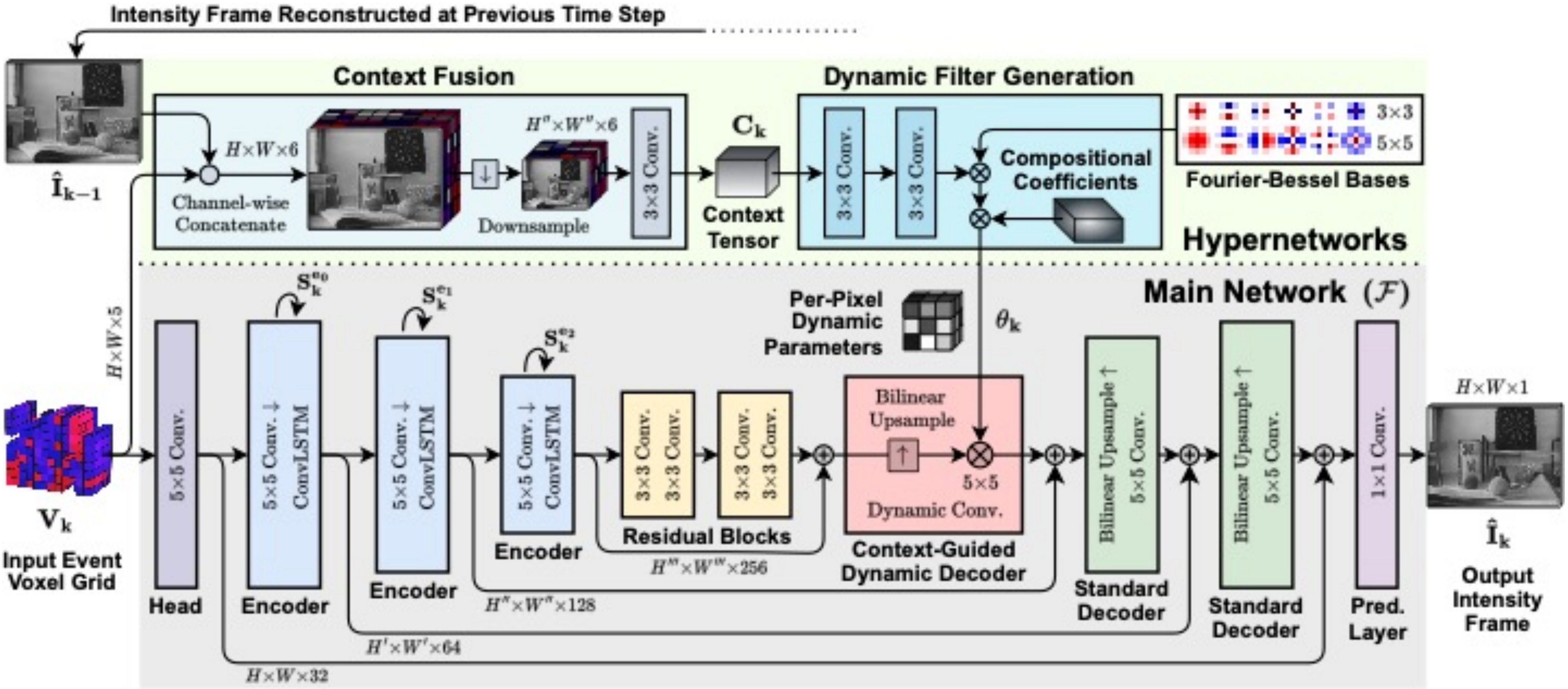


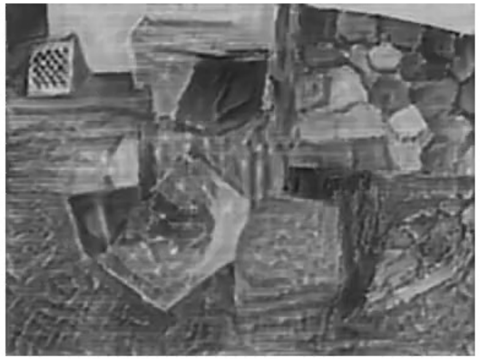
Image credit: Ercan et al., 2024

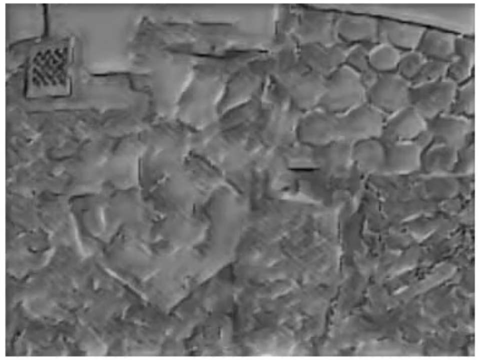
Video Generation from Events


E2VID


FireNet


SSL_E2VID


E2VID+


FireNet+


SPADE_E2VID

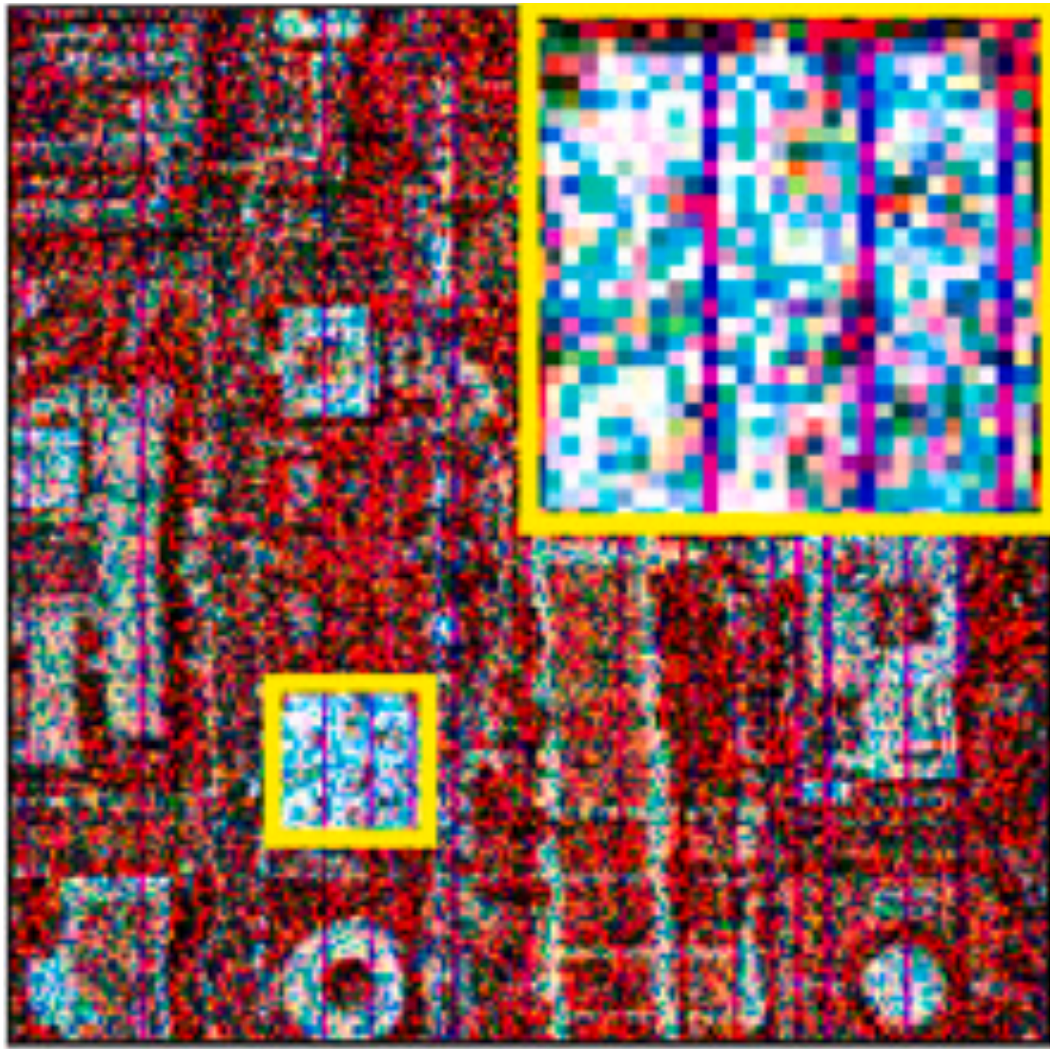

ET-Net


HyperE2VID


Ground Truth

Video credit: Ercan et al., 2024

Hyperspectral Image Denoising



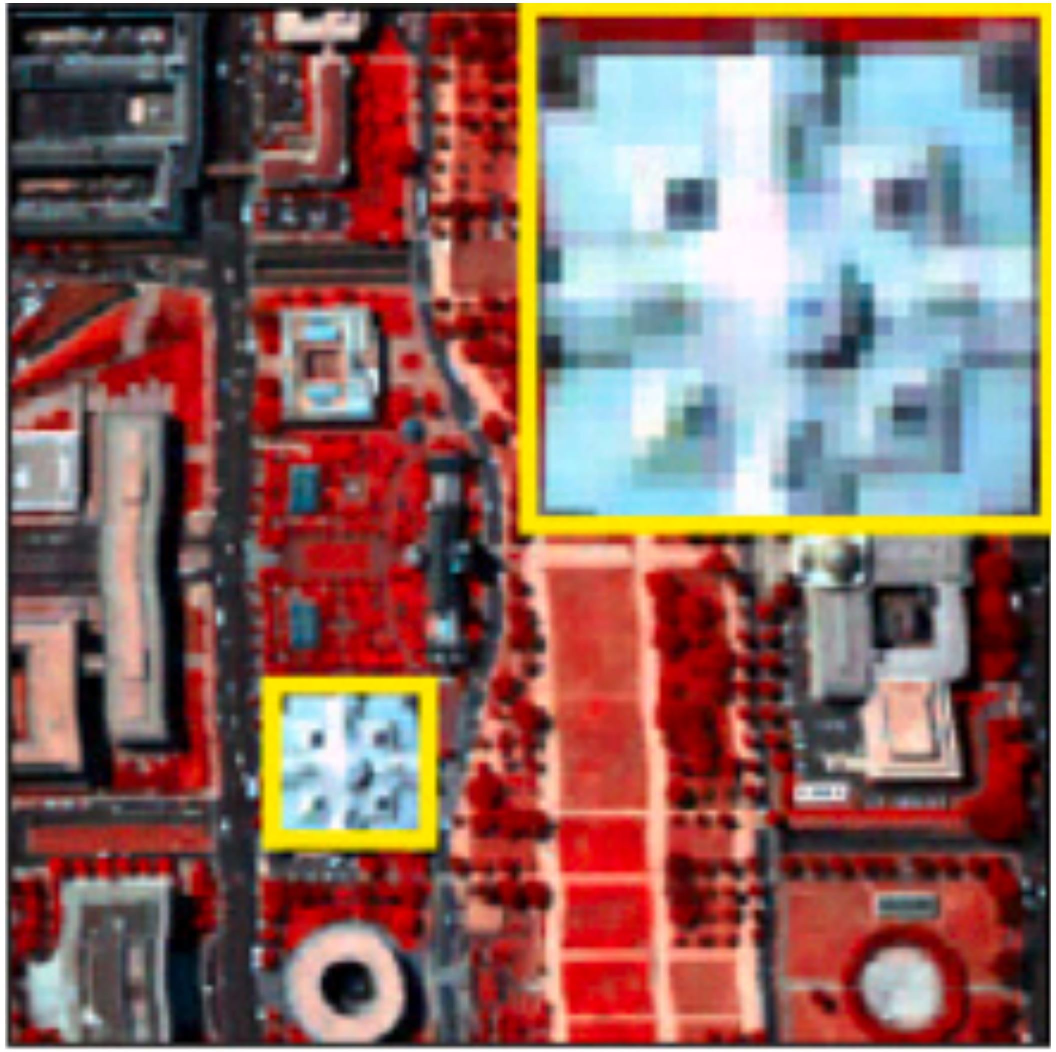


Image credit: Torun et al., 2024

Today's Lecture

- Course info
- History of photography
- Limitations of traditional photography
- Recent accomplishments

Reading Assignments

- Brian Hayes, Computational Photography, American Scientist 96, 94-99, 2008
- Michael Johnston, Your Camera Roll Contains A Masterpiece, New Yorker, March 31, 2022

Next Lecture:
Image formation