ITMO UNIVERSITY

Saint Petersburg, Russia

Специализированные технологии машинного обучения / Advanced Machine learning Technologies

Lecture 5 – ML for Modern Image Processing

Outline

ITMO UNIVERSITY

1. Digital camera: trends and challenges
2. Image signal processing stages:
   - denoising
   - demosaicing
   - superresolution
   - HDR and color processing
3. 3D data
4. Synthetic datasets

ITMOre than a UNIVERSITY

Digital cameras industry

ITMO UNIVERSITY

DSLR camera

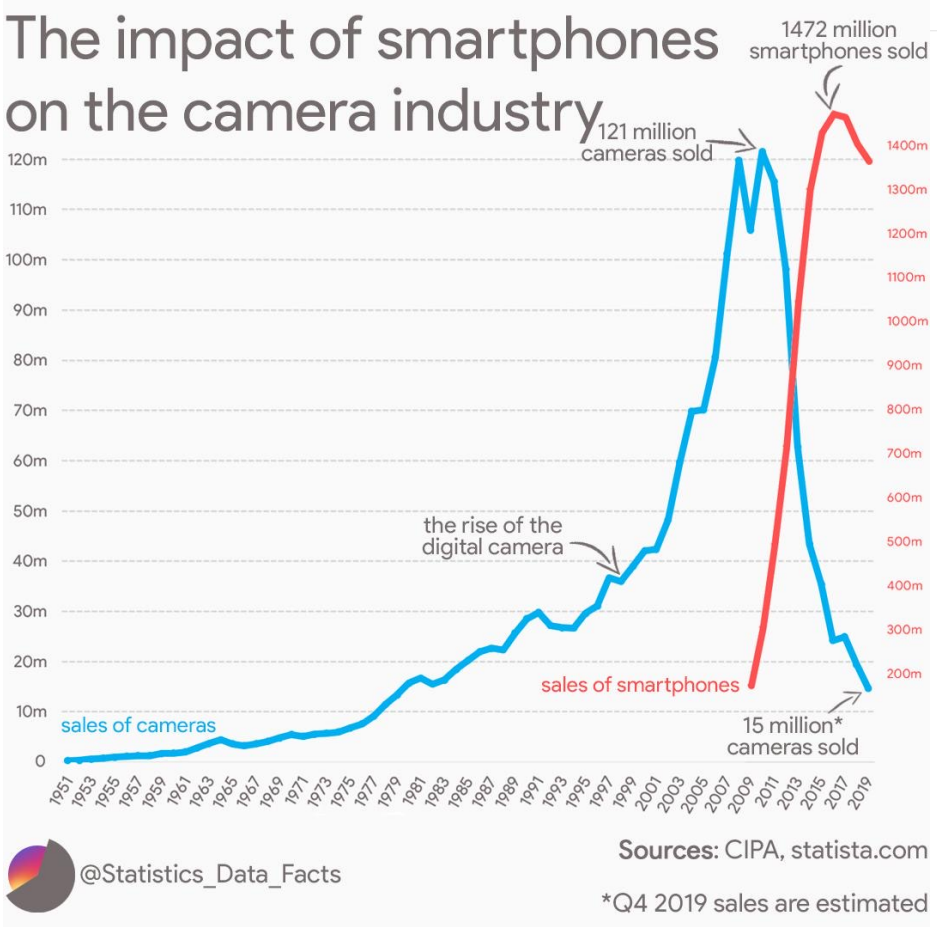
vs.

Smartphone

Increasing quality of **hardware** and **software (algorithms)** for smartphones cameras led to collapse of DSLR camera market in 2010's.

Camera is one of the 3 main criteria (together with performance and battery capacity) when choosing a new smartphone

ITMO re than a UNIVERSITY



Digital cameras applications

ITMO UNIVERSITY

Main applications:
- Mobile Cameras
- CCTV
- Autonomous Driving
- Aerospace
- Healthcare
- Robotics (industry)

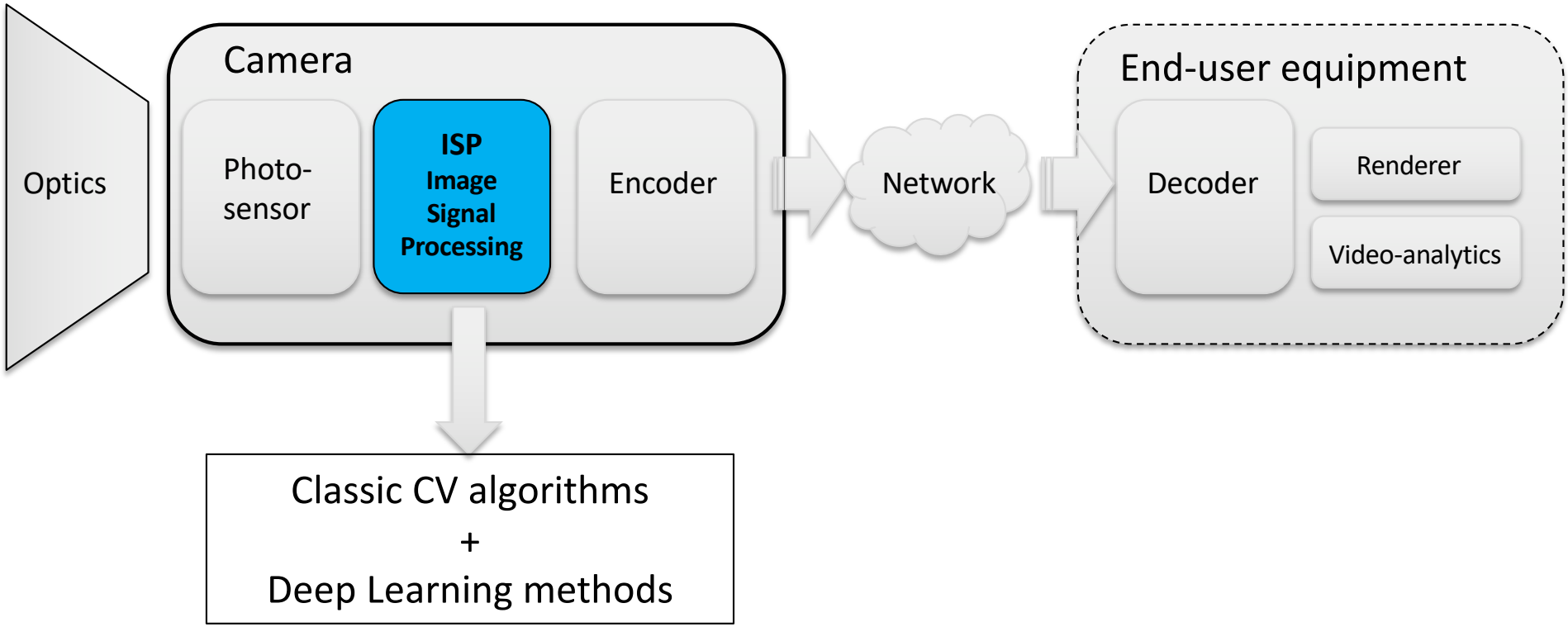
Challenging scenarios:
- Low-light
- HDR
- Harsh weather (for, snow, rain, haze etc.)

Main users and corresponding goals:
- people (humans) → perceptual quality
- machines (algorithms) → CV algorithms performance

ITMOre than a UNIVERSITY

Digital camera scheme

ITMO UNIVERSITY

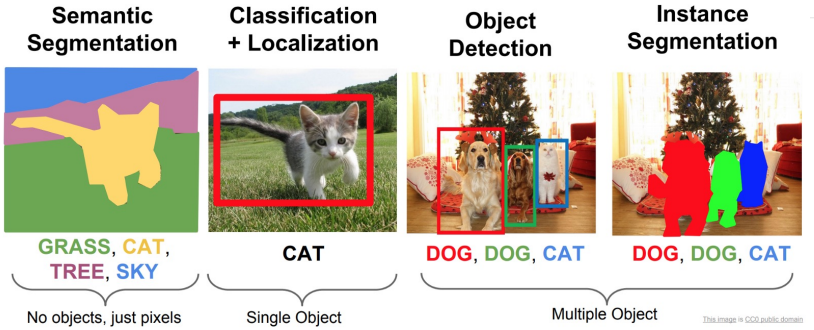


ITMO more than a UNIVERSITY

ISP task

ITMO UNIVERSITY

“Classic” supervised tasks for ML and DL in CV – object detection, recognition, semantic segmentation, object tracking etc.
-the goal – to increase an **accuracy of the model**, which **retrieve information from the image** according to its content
- metrics of quality are easy to formulate (ROC-AUC, mAP, IoU etc.)



ISP goal – to improve the quality of the image itself (low-level features of the image)

- Not well formalized problem
- What is the **quality of the image** itself? Sometimes it is not simple to formulate in terms of metrics for training the algorithm..
- we should somehow work with perceptual quality and aesthetics of the images and videos

Challenging scenes scenarios

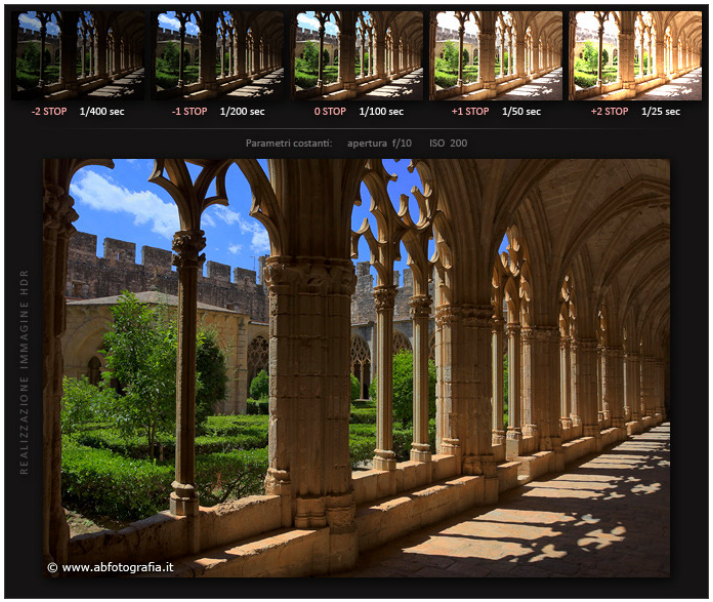
ITMO UNIVERSITY

Low-light case

Low signal-to-noise ratio (SNR)

Non-uniform illumination

Dynamic range of the scene is higher than sensor capabilities



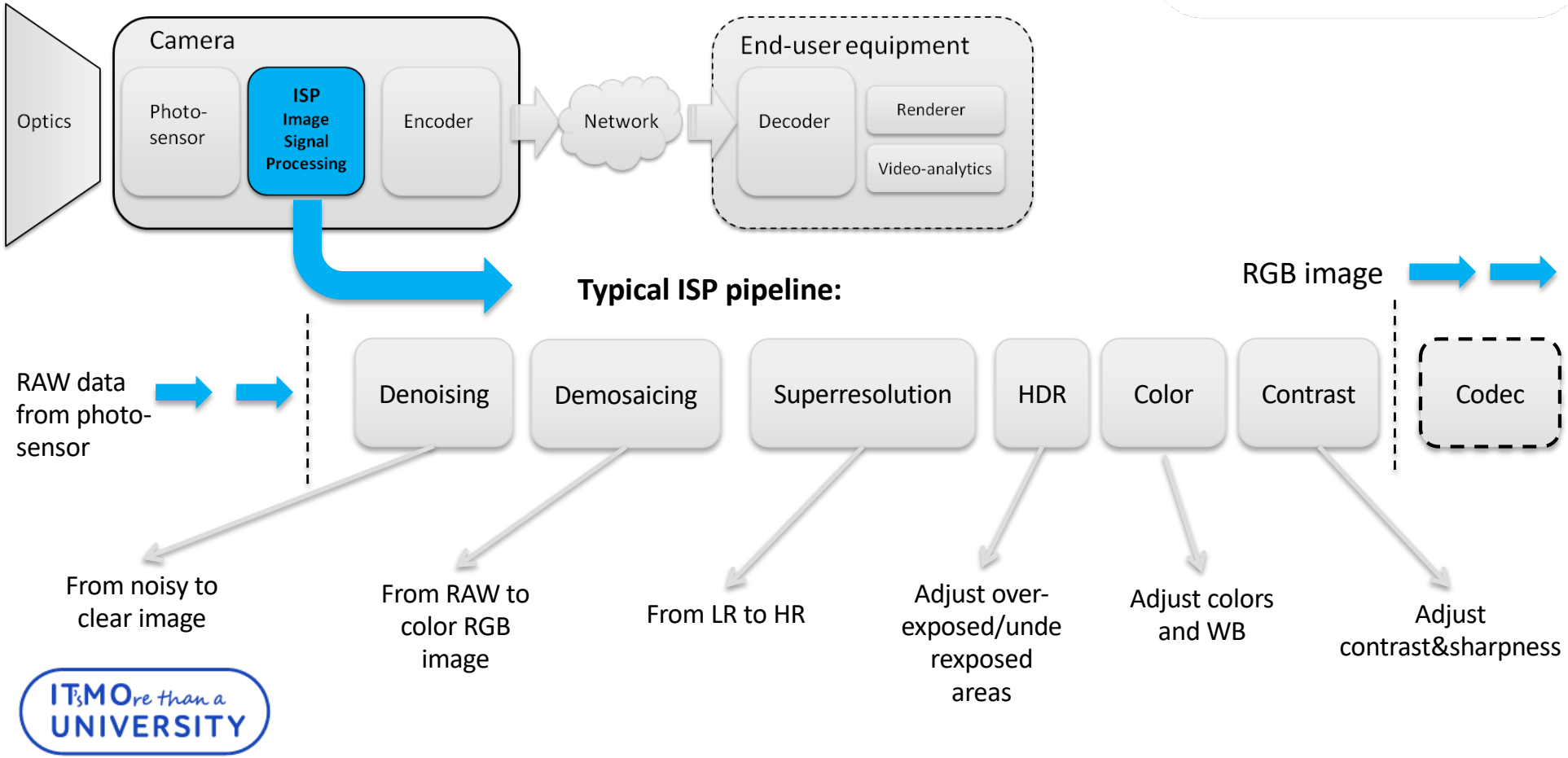
Harsh (bad) weather

Difficulties of light due to dirty objectives or windscreen, mixing with atmospheric light.

These problems are solving today by applying more efficient optics, research of new photo-sensors and hardware, using wide spectral range cameras, **developing new algorithms of ISP.**

Image signal processing (ISP) pipeline

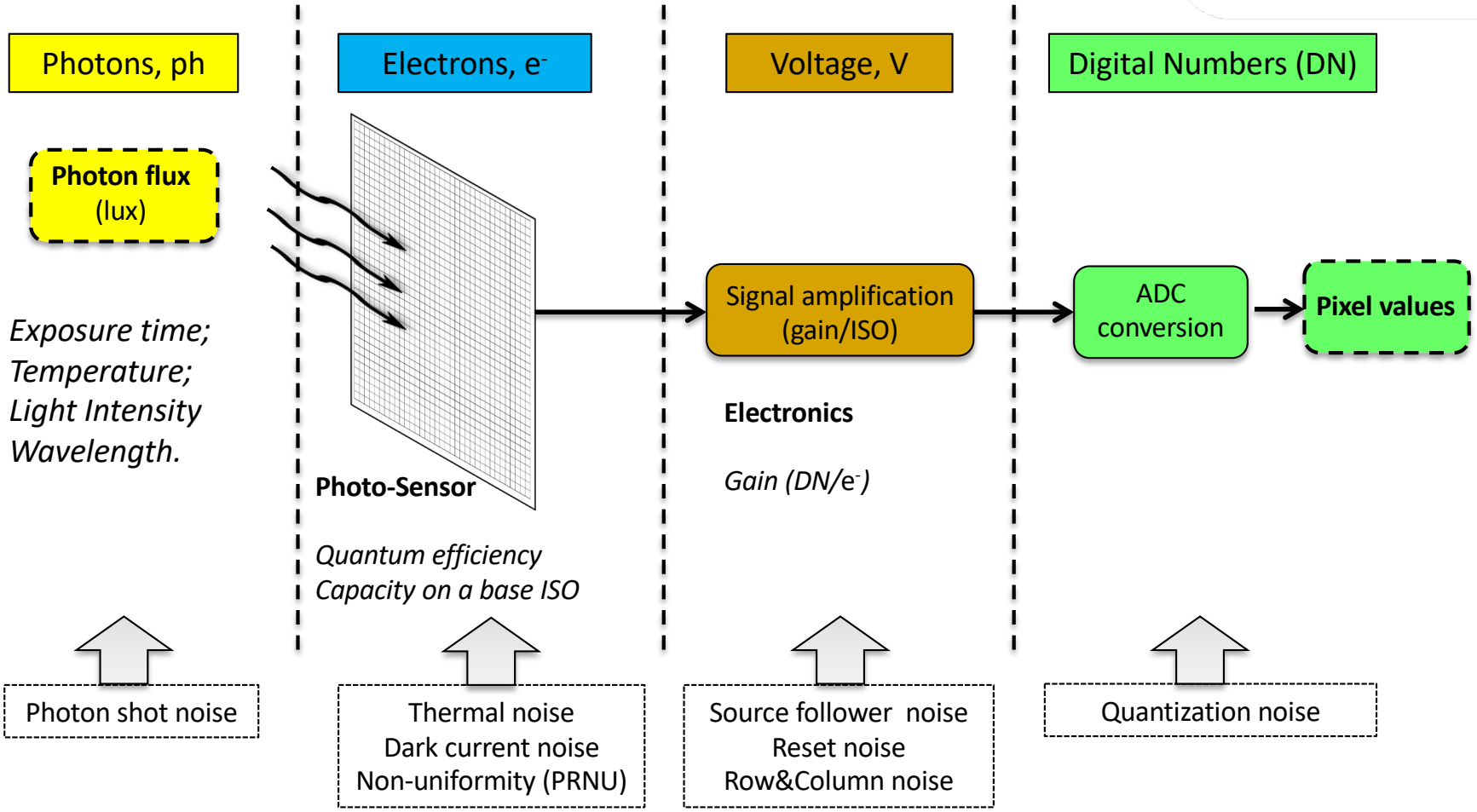
ITMO UNIVERSITY



ITMO More than a UNIVERSITY

Analog to Digital Conversion (ADC) scheme

ITMO UNIVERSITY



ITMO UNIVERSITY

Image Signal Processing (ISP) pipeline

ITMOre than a UNIVERSITY

ISP as optimization task

ITMO UNIVERSITY

Given input image (frame) and a desired reference frame at each stage of ISP pipeline there can be formulated ML supervised learning problem which is usually solved as optimization of neural networks parameters with respect to its parameters θ:

$$\hat{\theta} = \arg\min L(\hat{I}, I) + \lambda\Phi(\theta)$$

where $L(\hat{I}, I)$ - loss function and $\Phi(\theta)$ - regularization term.

Loss functions should reflect the difference between the reference (ideal) image and the current output of the model. This difference can be formulated as pixel-wise (as L1, L2 norm), as feature-based metric according to the convolutional layers outputs of the network or according to the share of style or content preserving.
If the model is GAN – loss is Generator and Discriminator losses *(see Generative models lecture)*

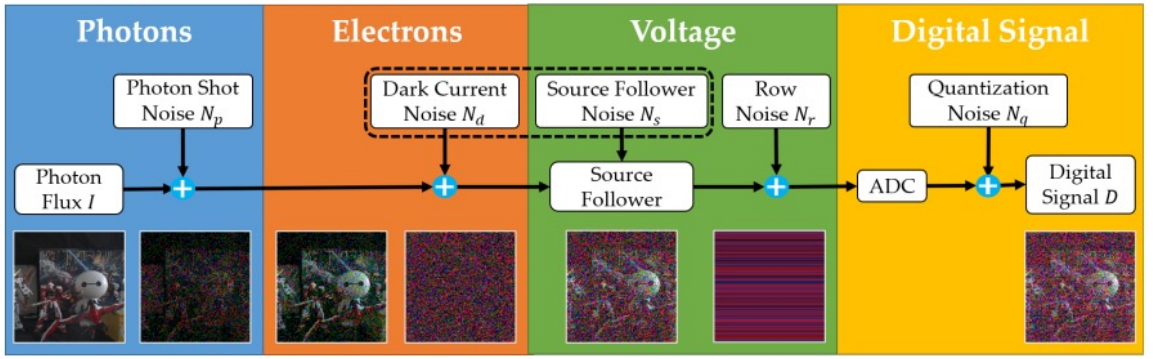
ITMO more than a UNIVERSITY

Denoising

ITMO UNIVERSITY

During image capturing and processing through ISP pipeline a noise of a different nature is added to the signal:

- **photon shot noise** – natural unremovable noise;
- **dark current noise** – charge generated is common through the sensors types.
- **pattern noise** – caused by small variations in the responses of different pixels;
- **read noise** – generated by electronics
- quantization noise – electrons to BT transformation
- **cross-talk** – charge spread between the pixels



Source: K.Wei et al., Physics-based Noise Modeling for Extreme Low-light Photography, 2021

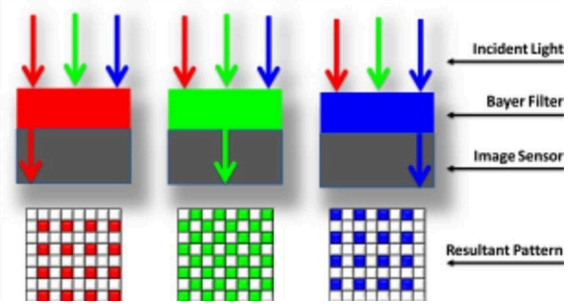
Input | Gaussian | Gaussian + Poisson | Paired Data | Paper

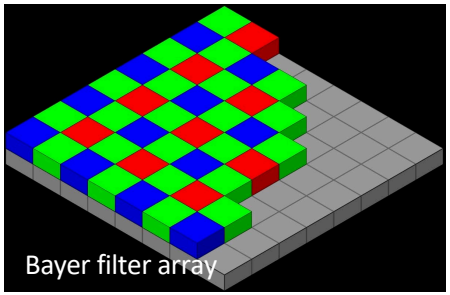
Research of noise statistical distribution according to its physical nature is crucial for denoising methods.

Complex models of noise generation (with calibration of noise parameters) can improve the quality of denoising algorithms via more efficient training pairs.

ITMOre than a UNIVERSITY

Color filter array (CFA) and color perception





Each pixel from the photo-matrix is filtered to record only one of 3 colors. To obtain a full-color image various **demosaicing algorithms** can be used to interpolate red, green and blue values for each pixel. These algorithms make use of the surrounding pixels of the corresponding colors to estimate the values for a particular pixel.

Human's cones of 3 types (S, M and L) have different spectral range sensitivities. Visual range is approx 400-700 nm.

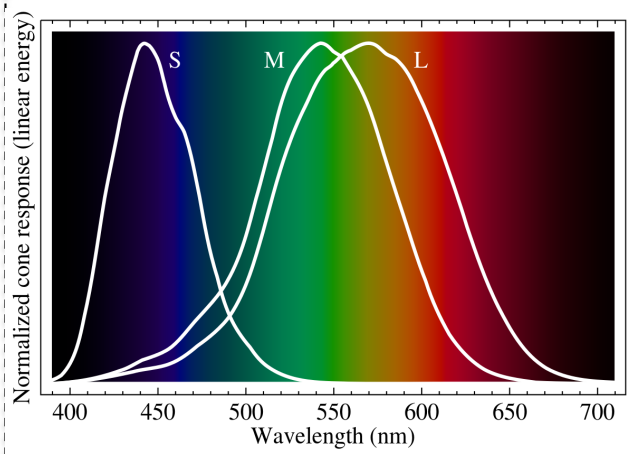
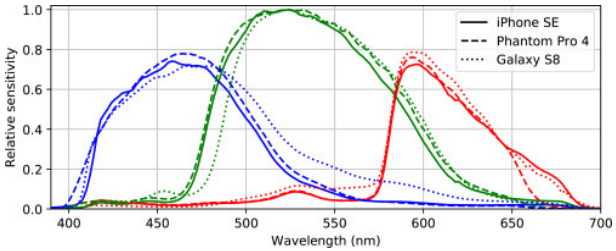
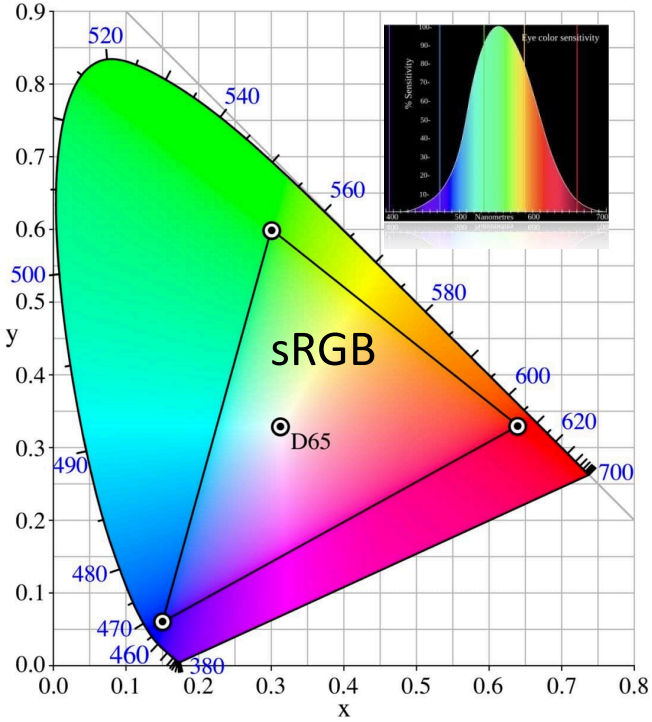


Photo-sensors have different quantum efficiency depending on the specific spectral range



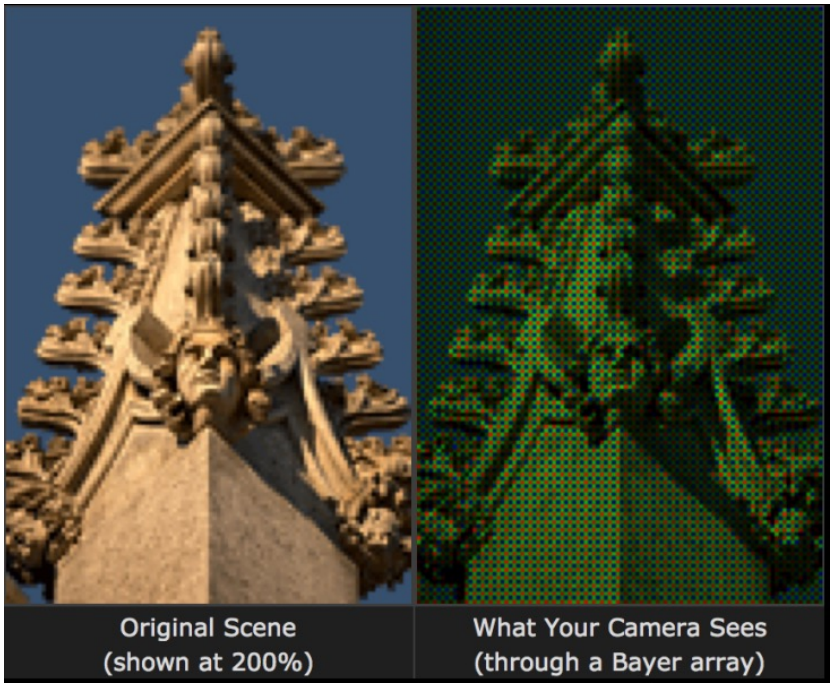
sRGB is a standard RGB (red-green-blue) color space that is used today in displays, printers and Web.

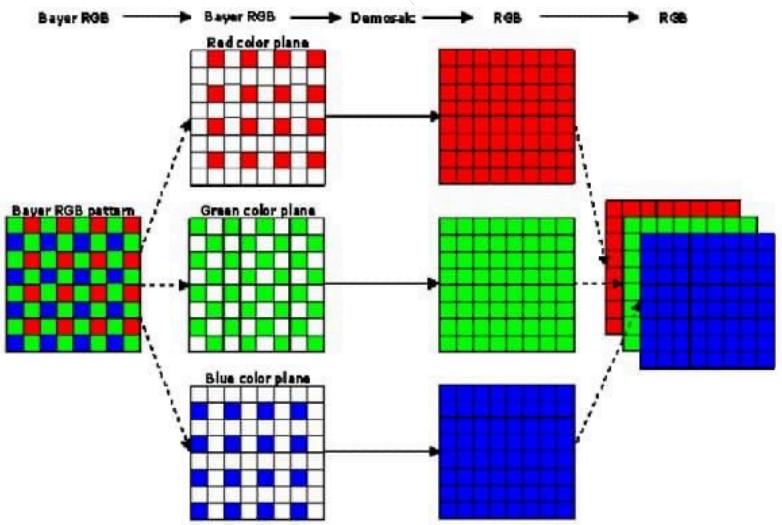


Demosaicing

ITMO UNIVERSITY

Demosaicing algorithms reconstruct a full color image (i.e., a full set of color triples, usually RGB) from the spatially undersampled color channels output from the color filter array (usually Bayer filter).





Trends&challenges:

1. Recently the most efficient models for this task imply deep learning architectures to represent adequate colors of the images ;
2. There exist as well classic algorithms (like adaptive demosaicing with bilateral filtering (Ramanath, Snyder, 2003) and using specific kernels for high quality linear interpolation, taking into account edges and luminance changes (Malvar et. al., 2004)
3. Modern approaches usually combine DM and DN into a single model (Liu et. Al., 2020, Sharif, 2021, etc.) which is called Joint DD approach.

Superresolution

ITMO UNIVERSITY

SR goal – increase the resolution of low-res images (video-frames), enhance them, add more details, but with saving the fidelity.

Models:

- Classic approaches (bilinear or bicubic interpolation) – are stable but add a lot of blur and have low perceptual quality;

- DL models – trained on LR-HR paires – generally good, but decrease its quality on the real world scenes with unknown downsampling kernels and noise models; **GANs** – usually generate artefacts

Trends and challenges:

- Data for training – not many realistic datasets with high diversity. **Downsampling kernels?**
- Low fidelity scores for GAN-based models
- Challenging tasks with high magnification factors (x4, x8) especially for >1K-resolution images
- Very new idea – SR3 model from Google Brain – based on denoising diffusion probabilistic model https://iterative-refinement.github.io/ - “SR via iterative refinement”

ITMOre than a UNIVERSITY

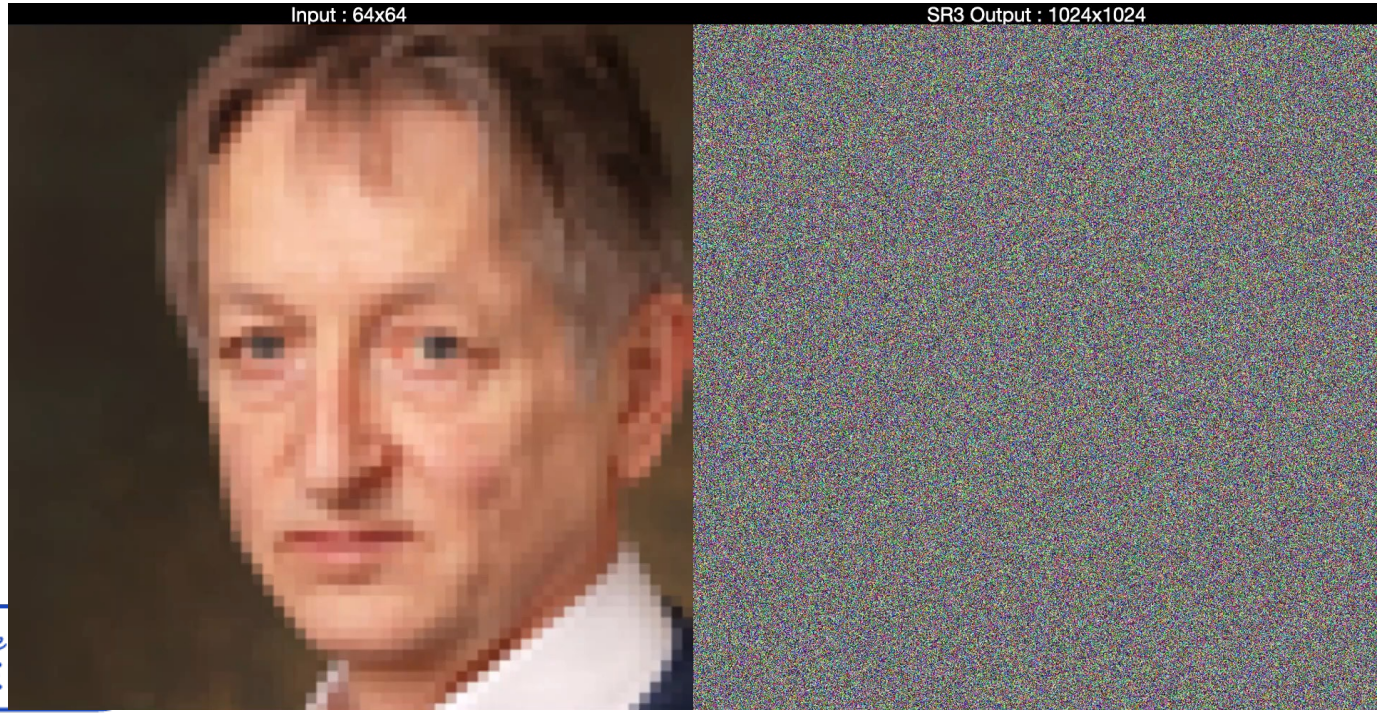
Superresolution

ITMO UNIVERSITY

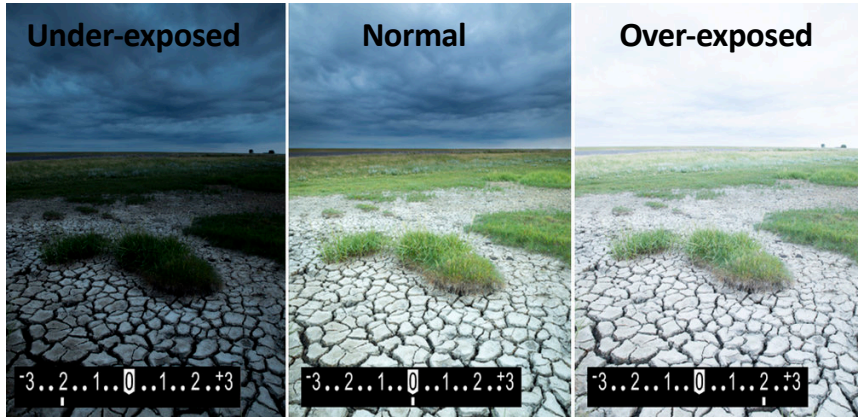
- **SR3 model** from Google Brain – based on denoising diffusion probabilistic model

https://iterative-refinement.github.io/ - *"SR via iterative refinement" Saharia et.al., (2021)*

- Inference starts with pure Gaussian noise and iteratively refines the noisy output using a U-Net model trained on denoising at various noise levels.



HDR – high dynamic range processing



Dynamic Range processing – set of techniques used to reproduce a greater range of luminosity that is possible with standard photographic techniques.

Main tasks:
-Enhance dynamic range of the image optimizing its perceptual quality;
-Color tone mapping of the image preserving the same bitrate (dynamic range compressing)
- color temperature adjustment (automatic White Balance, AWB)
-adjusting colors for specific displays.

Trends and challenges:
1) Data generation – not many realistic datasets with high diversity. For Video HDR – there only 2 available works with creating such datasets.
2) One of the latest improvements – use of attention mechanism and dilated convolutions approach.
3) For Video HDR – using Deformable Convolutional Networks (DCN) help significantly improve the results (according to the experiments)

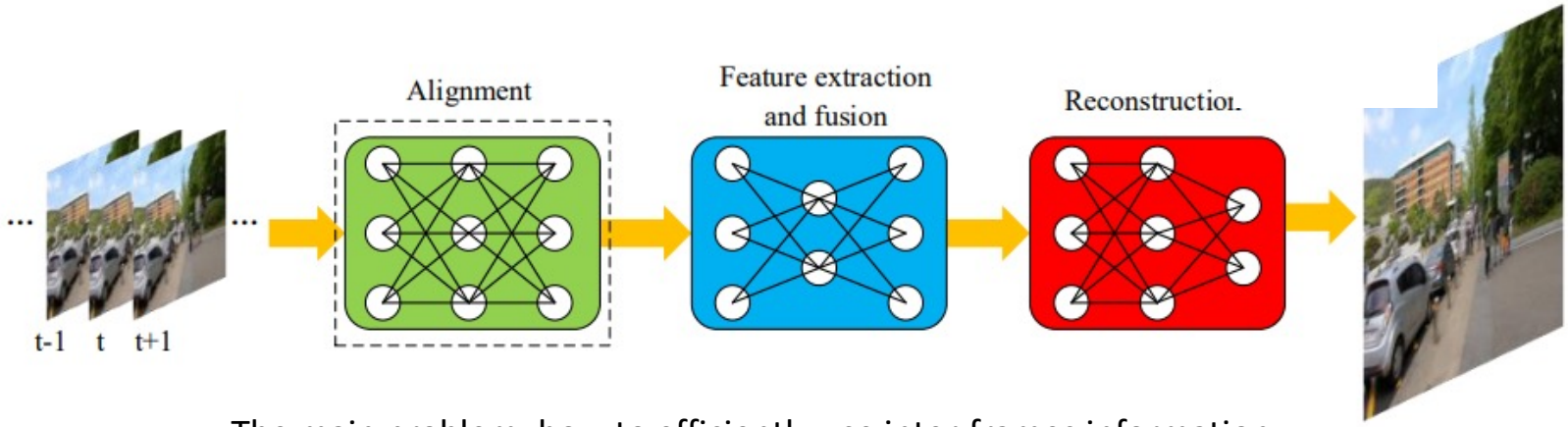
https://data.csail.mit.edu/graphics/fivek/

Video signal processing

ITMO UNIVERSITY

Two types of video signal processing:
- sequence of frames – each frame is analyzed independently → comes to **ISP**
- using temporal information between the neighboring frames → **VSP**

The same tasks as for ISP
+ frame alignment;
+ fusion the information from the neighboring frames;



The main problem: how to efficiently use inter-frames information for more robust output of SR, HDR and other ISP algorithms

ITMO UNIVERSITY

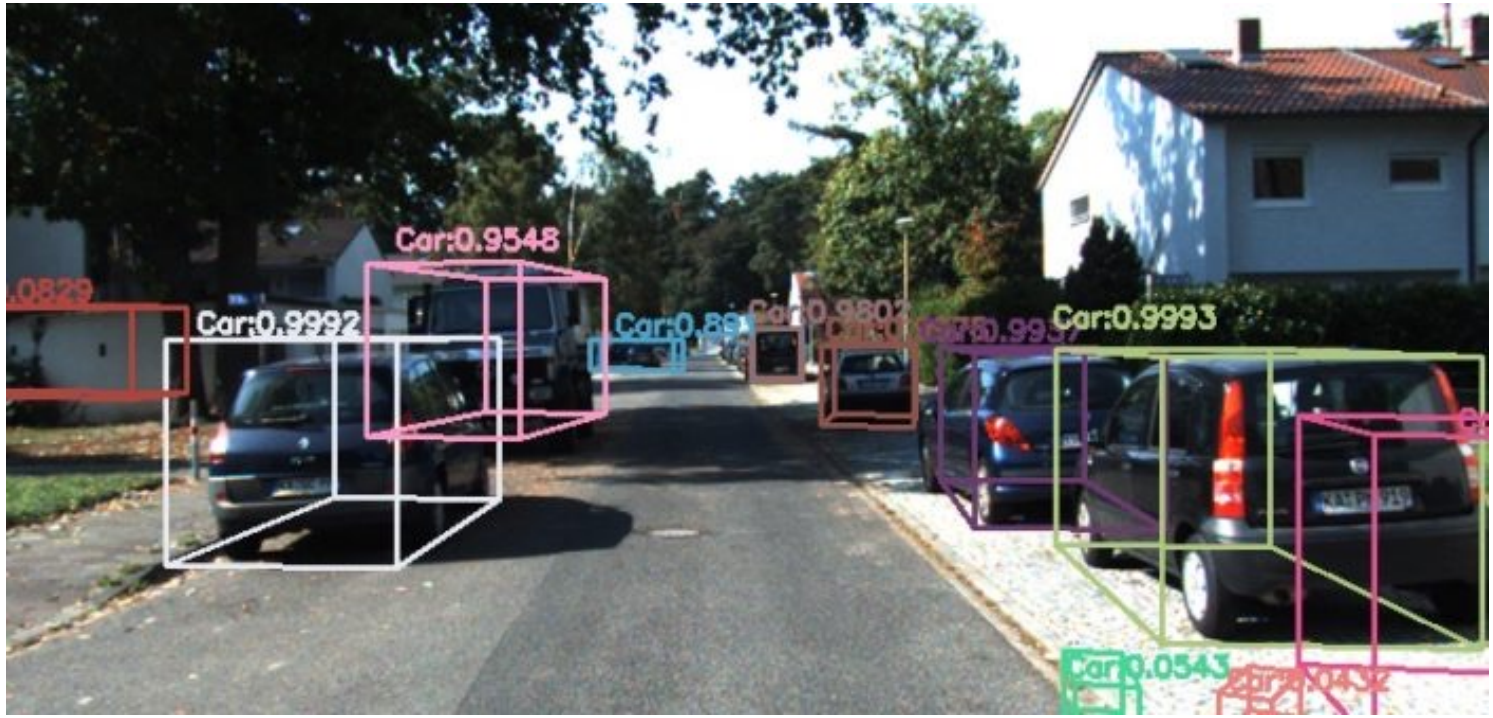
3-D

ITMOre than a UNIVERSITY

3D Object Recognition

ITMO UNIVERSITY

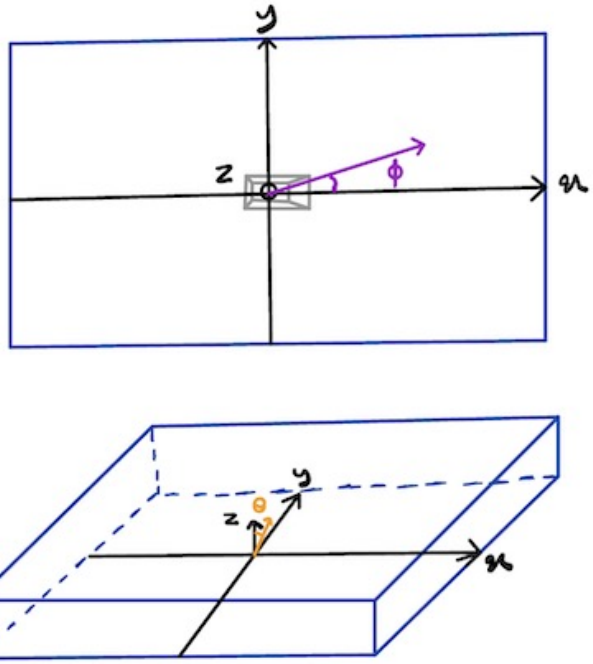
Improving quality of images and video for 3D object detection and segmentation is crucially important for **autonomous driving** where shapes and 3d spatial positions of different objects (cars, pedestrians etc.) varies very much but play a significant role as well.



2D detection:
- x direction;
- y direction;
- xy-plane rotation;

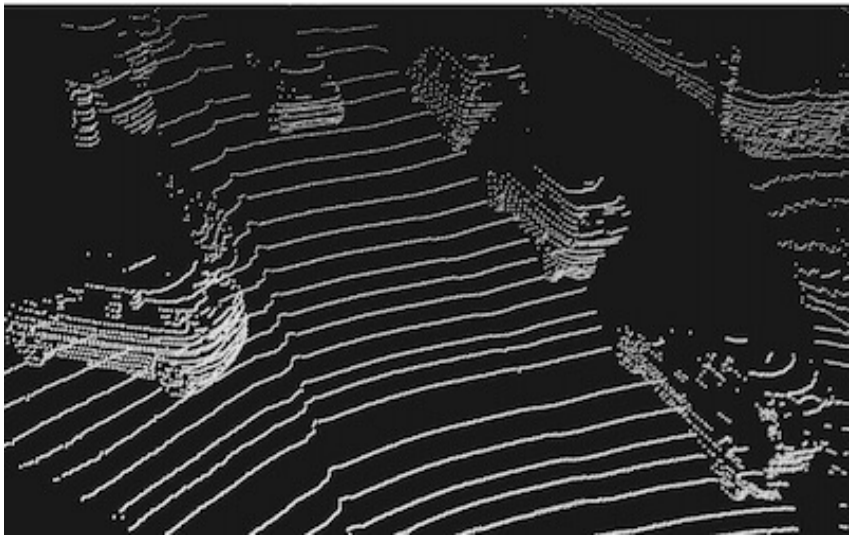
3D detection:
- x direction;
- y direction;
- z direction;
- xy-plane rotation;
- xz-plane rotation;
- yz-plane rotation.

Lidar Coordinate Frame



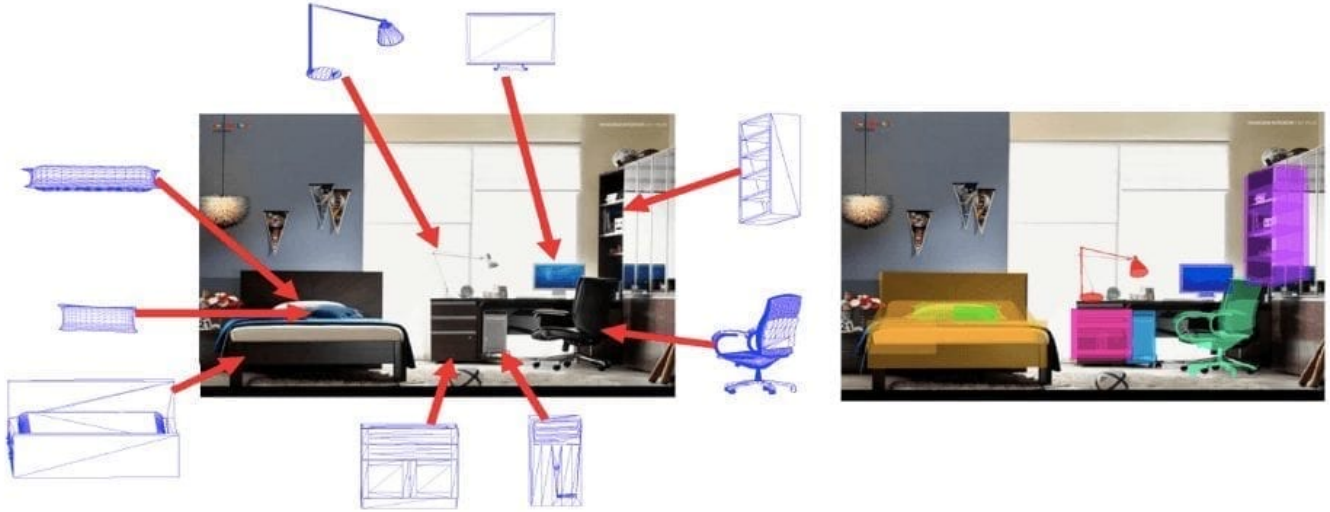
$$r = \sqrt{x^2 + y^2 + z^2}.$$
$$\phi = \arctan \frac{y}{x}.$$
$$\theta = \arccos \frac{z}{\sqrt{x^2+y^2+z^2}}.$$

Cartesian coordinates ->
Spherical coordinates
+ intensity =
4d point space

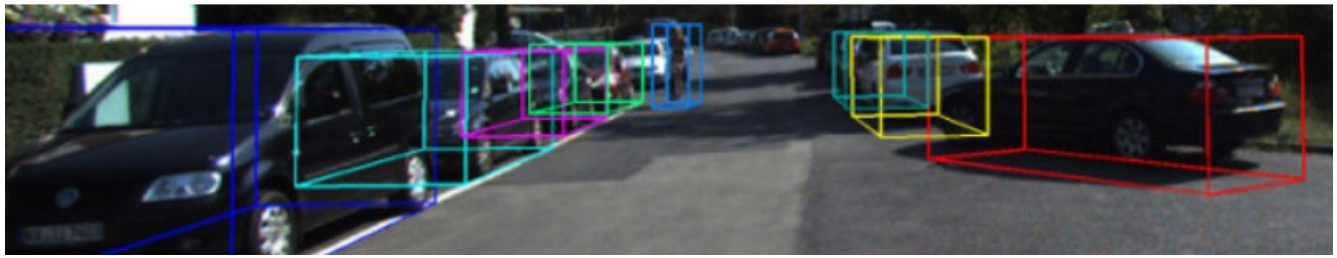


3D Object Recognition Datasets

ITMO UNIVERSITY



ObjectNet3D:
- 90k images;
- 200k objects;
- 100 categories;
- 44k 3D shapes.

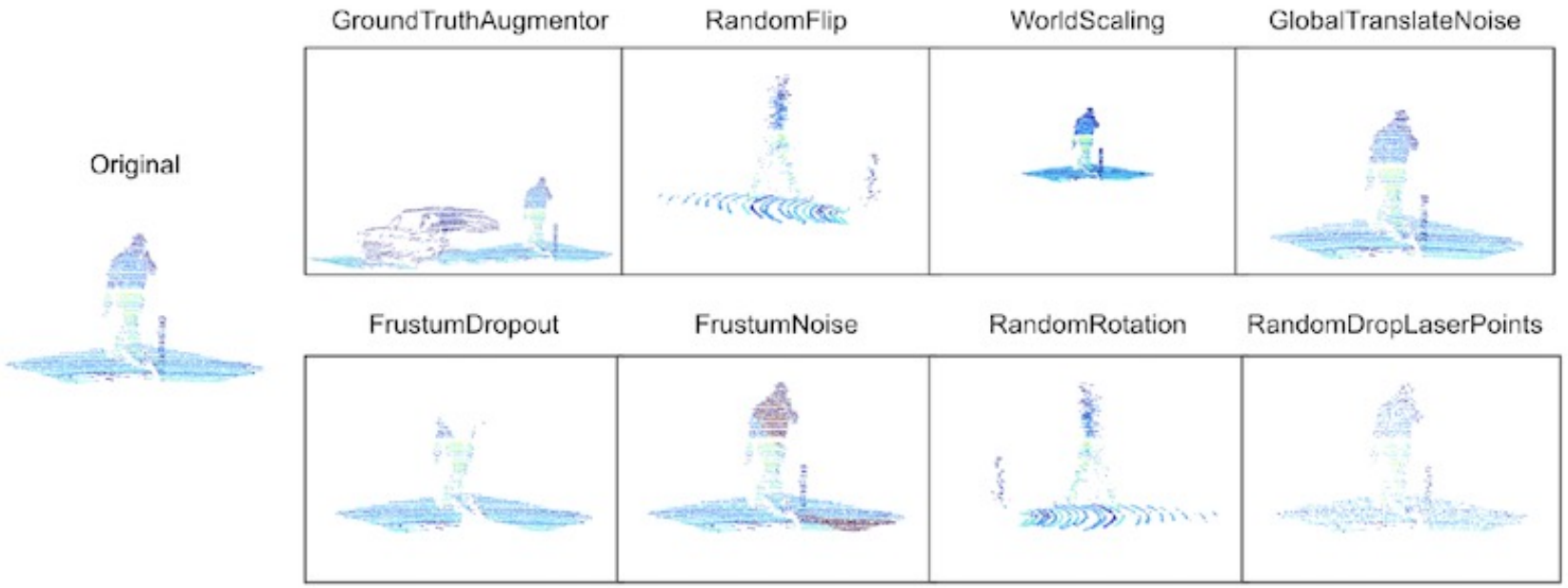


KITTI:
- Lidar Point Cloud;
- 15k images;
- 80k objects;
- Categories: cars, pedestrians, cyclists.

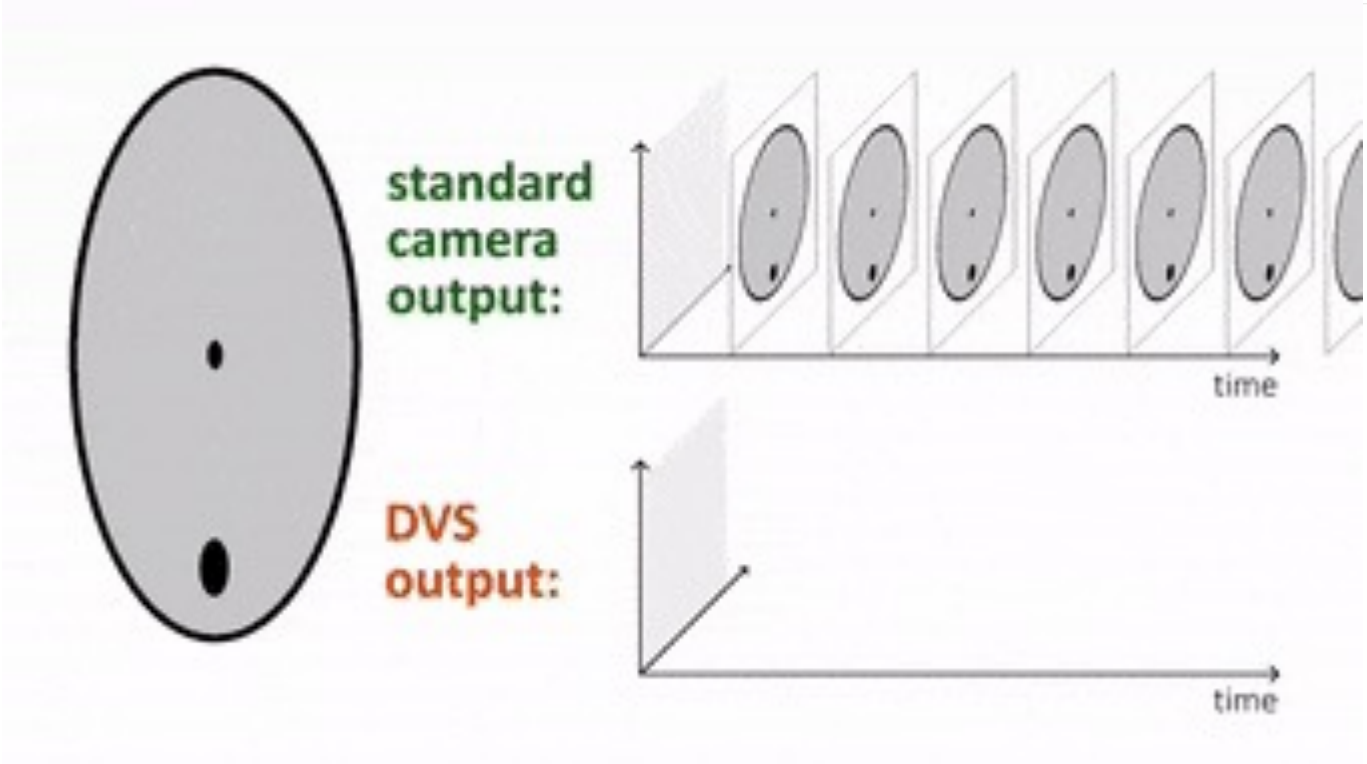
Data Augmentation

ITMO UNIVERSITY

- The data augmentation methods applied to lidar point clouds are inspired by the augmentation methods originally designed for images.
- Point cloud data augmentation methods must be compatible with the physical laws governing the propagation of laser beams.



Event Cameras



ITMO UNIVERSITY

Synthetic Data for CV

ITMOre than a UNIVERSITY

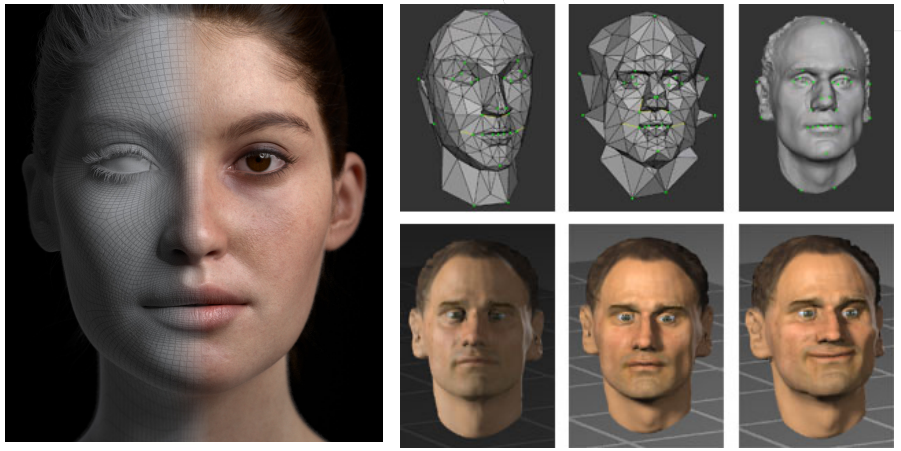
Synthetic Data

ITMO UNIVERSITY

Problem:

➢ Collecting high-quality and high-diversity datasets in Computer Vision is very challenging and time/resources consuming task

➢ With the improving of the quality of the rendering engines (UE4, Blender3D, Unity3D etc.) and generative modeling there is an opportunity to collect HQ syntetic data with desirable properties.



Advantages:

- perfect labeling (for example, pixel-perfect segmentati masks compared with its correspondent RGB image;
- actually unlimited sizes of datasets;
- great economy of resources and time.

Disadvantages:

-It's synthetic… Models trained on synthetic data not always perform as good on the real data

Synthetic Datasets

ITMO UNIVERSITY

How to organize synthetic dataset?

- **for acceleration of rendering** – one can construct images by pasting 3D models of various objects to real scenes (not to generate background);
- **balanced sampling**: place various augmented objects to the same background; every object is placed on equal number of backgrounds and every background is used with the same number of objects;
- use **superimposed objects** as hard cases for the object recognition models
- add **noise** for hard examples;
- **augmentation**, especially, for object detection really matters;
- **smart (balanced) augmentation** – that fill in the gaps in real-world distribution;

ITMOre than a UNIVERSITY

Learning From Context-Agnostic Synthetic Data, 2020

Synthetic Datasets + Tools

ITMO UNIVERSITY

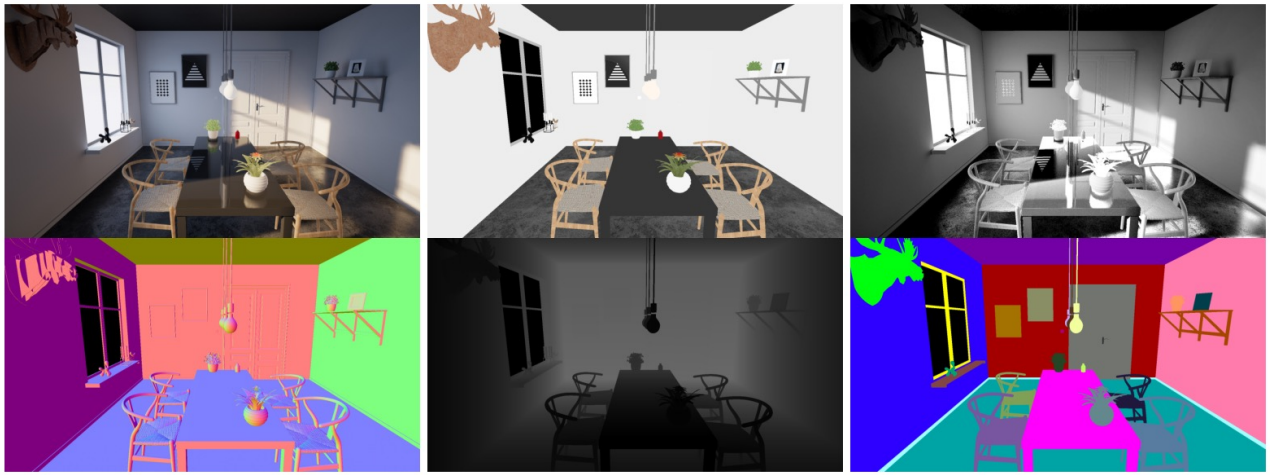
Examples of synthetic datasets:

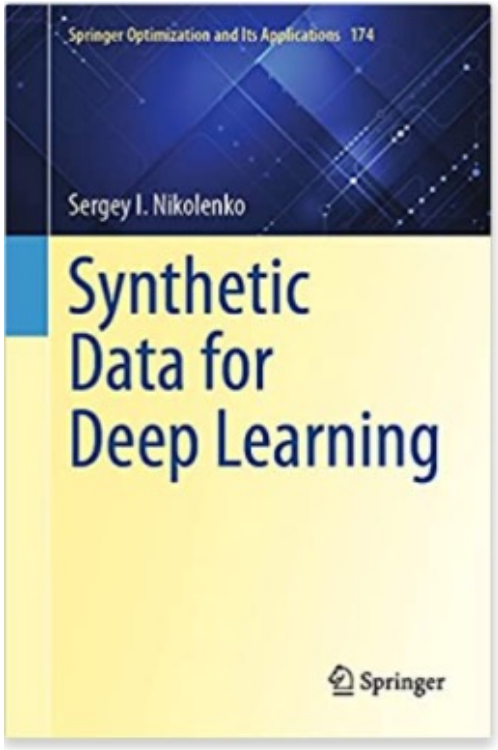
-**SynScapes** – 25000 8-bit physically based unique images of street scenes https://7dlabs.com/synscapes-overview
- ProcSy – synthetic dataset for autonomous driving https://uwaterloo.ca/waterloo-intelligent-systems-engineering-lab/procsy
- **SynSign** – augmented synthetic traffic signs against real background (small resolution)
- **AADS** – Autonomous Driving Simulation Dataset (rendered cars against real-world scenes (mostly roads)
- **FAT** – Falling Things Dataset

ITMOre than a UNIVERSITY

Tolls for generation:

- **Blender 3D**
- **NDDS** – NVIDIA Deep Learning Dataset Synthetiser – UE4 plugin - https://github.com/NVIDIA/Dataset_Synthesizer
- **UnrealROX+** - An improved Toll for Acquiring Synthetic Data from Virtual 3D Environments – UE4 plugin https://github.com/3dperceptionlab/unrealrox-plus





This is the first book on synthetic data for deep learning, and its breadth of coverage may render this book as the default reference on synthetic data for years to come. The book can also serve as an introduction to several other important subfields of machine learning that are seldom touched upon in other books. Machine learning as a discipline would not be possible without the inner workings of optimization at hand. The book includes the necessary sinews of optimization though the crux of the discussion centers on the increasingly popular tool for training deep learning models, namely synthetic data. It is expected that the field of synthetic data will undergo exponential growth in the near future. This book serves as a comprehensive survey of the field.

In the simplest case, synthetic data refers to computer-generated graphics used to train computer vision models. There are many more facets of synthetic data to consider. In the section on basic computer vision, the book discusses fundamental computer vision problems, both low-level (e.g., optical flow estimation) and high-level (e.g., object detection and semantic segmentation), synthetic environments and datasets for outdoor and urban scenes (autonomous driving), indoor scenes (indoor navigation), aerial navigation, and simulation environments for robotics. Additionally, it touches upon applications of synthetic data outside computer vision (in neural programming, bioinformatics, NLP, and more). It also surveys the work on improving synthetic data development and alternative ways to produce it such as GANs.
The book introduces and reviews several different approaches to synthetic data in various domains of machine learning, most notably the following fields: domain adaptation for making synthetic data more realistic and/or adapting the models to be trained on synthetic data and differential privacy for generating synthetic data with privacy guarantees. This discussion is accompanied by an introduction into generative adversarial networks (GAN) and an introduction to differential privacy.

Quality?

ITMO UNIVERSITY

Image quality attributes

- Sharpness
- Noise
- Dynamic range
- Tone reproduction
- Contrast
- Color accuracy
- Distortion
- Vignetting,
- Exposure accuracy
- Lateral chromatic aberration (LCA)
- Lens flare
- Color moiré
- Artifacts

Visual Perception?

ITMO more than a UNIVERSITY


ITMO UNIVERSITY
Saint Petersburg, Russia
Thank you!