Introduction to Katto Lab. Multimedia Group

Katto Laboratory,
Department of Communication and Computer Engineering,
Faculty of Science and Engineering, WASEDA University

Multimedia group

- Focus on image processing research
- Main Topic
  - Image/Video Compression, Coding
  - Image Recognition
  - Image Sharpening
  - 4K, 8K, HDR
  - VR, AR
  - Biometric Sensors

- We handle a wide range of research on images/videos from image signal processing to computer vision
  - You can study anything related to images/videos

Image/Video Compression, Coding

- Coding
  - Technology that digitizes information and compresses data
  - Current methods include MPEG-2 (digital terrestrial broadcasting) and H.264/AVC (Internet video sharing service, Blu-ray Disc)

MPEG-2

746Mbps ⇒ 20Mbps

Compressed to about 1/40

Tokyo Sky Tree

H.264/AVC

746Mbps ⇒ 10Mbps

Compressed to about 1/80

Blu-ray disc recorder

Image/Video Compression, Coding

- Latest video compression method → H.265/HEVC
  - Approved as an international standard in January 2013
  - Achieves about twice the compression efficiency of H.264/AVC
  - Support compression of high-resolution moving images such as 4K and 8K

16 times the resolution of terrestrial digital broadcasting

【 H.264/AVC】

【H.265/HEVC】

Source : https://www.ntt-at.co.jp/product/rfs_hevc_sdk/detail.html

Image/Video Compression, Coding

- Katto Lab is proposing compression methods using deep learning and improving the accuracy of existing compression methods (HEVC, VVC, etc.)

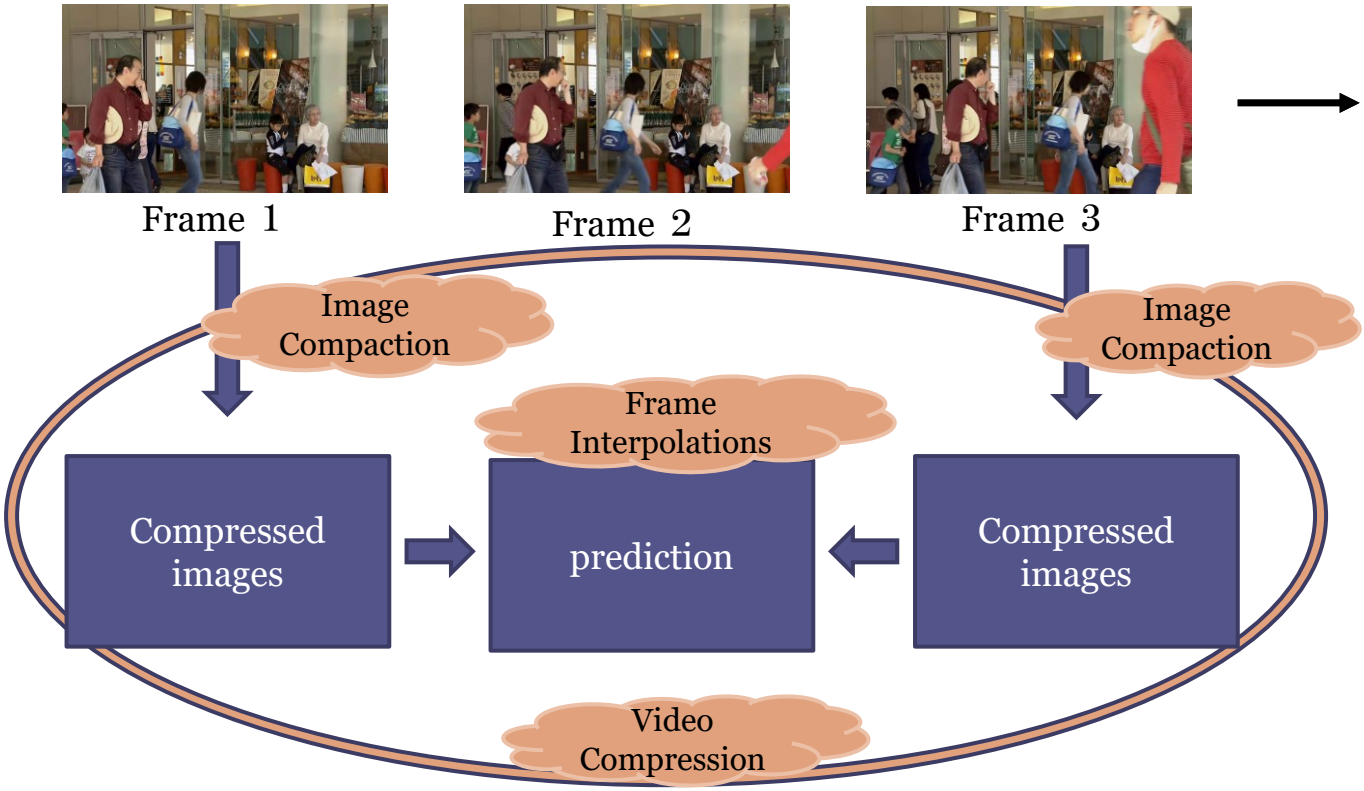
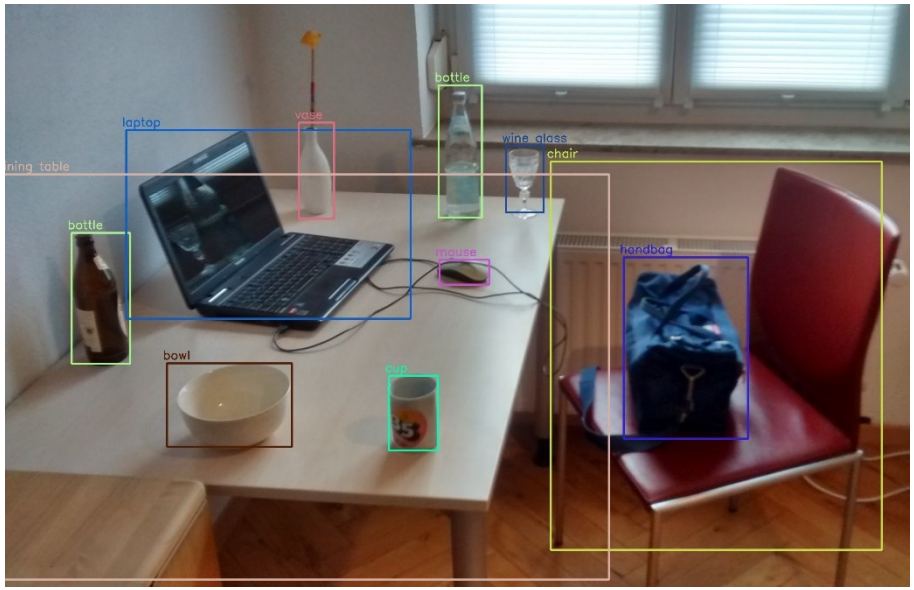


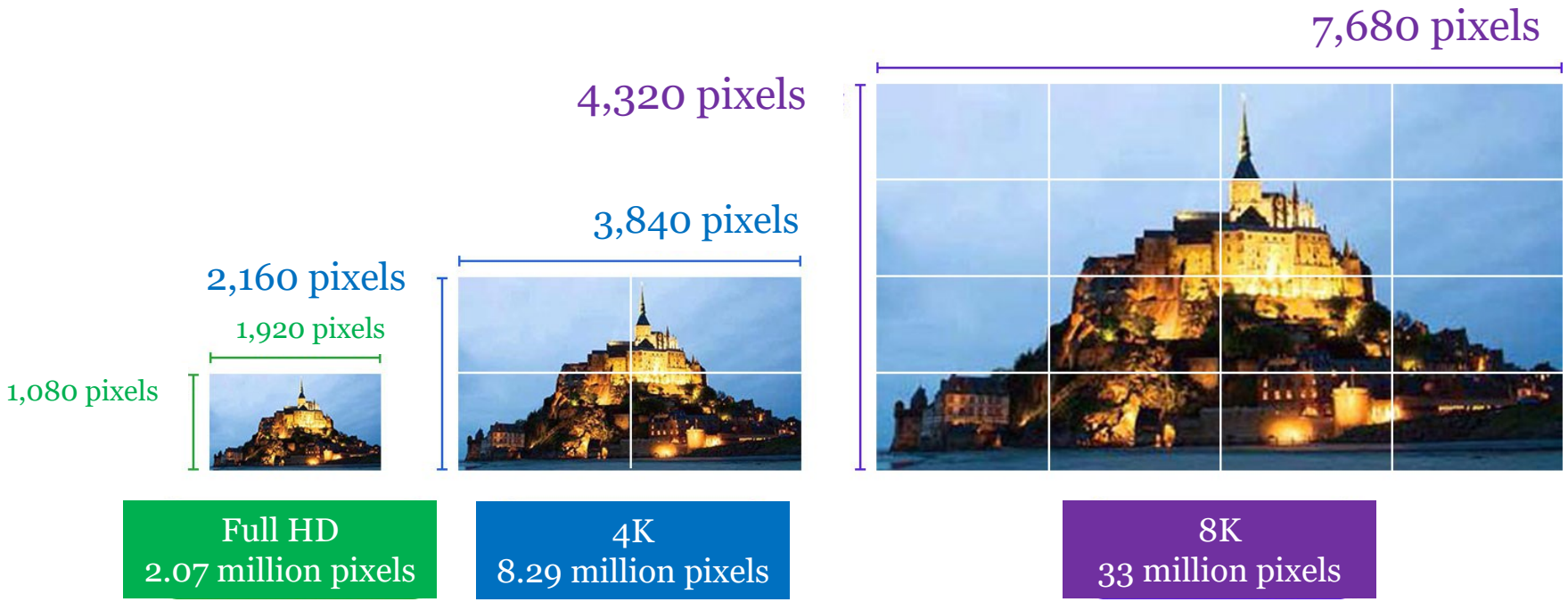
Image Recognition

- What is Image Recognition?
  - Research filed to give computers the visual functions that humans are doing
  - Recent developments in deep learning have improved the accuracy of Object Detection
- Katto Lab is applying this technology to detect abnormal persons , determine road conditions and so on



4K, 8K, HDR

- 4K, 8K
  - Next-generation ultra-high-definition video with 4 times or 16 times the resolution of terrestrial digital broadcasting
  - 4K broadcasting have already started in some areas
  - 8K main broadcast will start in 2020



4K, 8K, HDR

- HDR (High Dynamic Range)
  - In addition to ultra-high definition, the colors that can be expressed, that is, the dynamic range, have also expanded
  - Enables video expression closer to the real world
- The Katto Lab is investigating a method to make existing SDR (Standard Dynamic Range) images equivalent to HDR images.

Conventional (SDR) video

HDR video

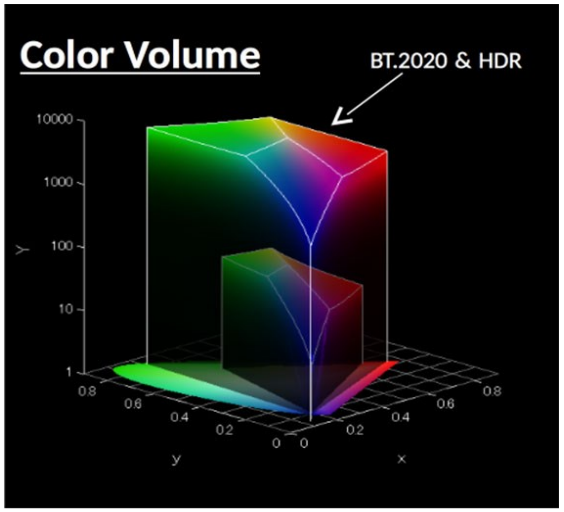
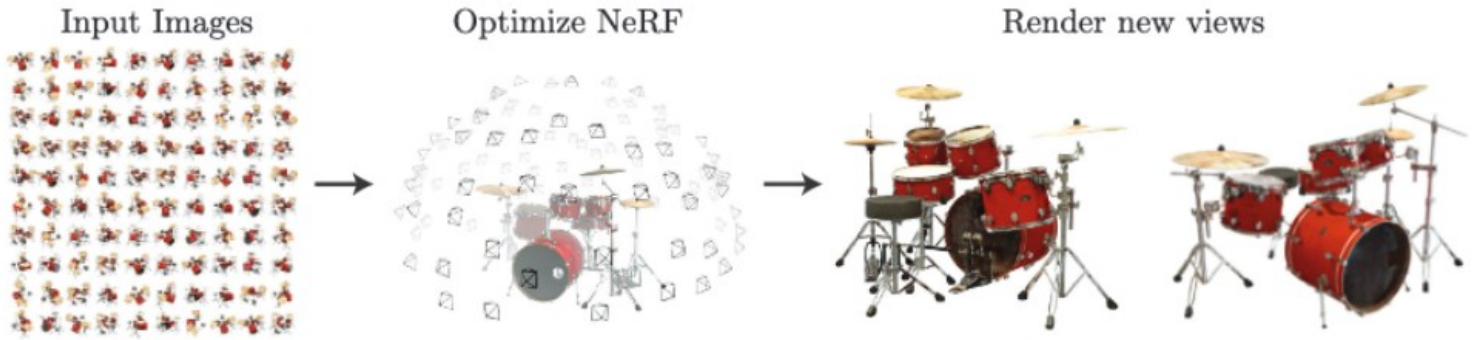


Image source：http://www.apab.or.jp/4k-8k/appeal/

Ex1 : Neural Radiance Fields(NeRF)

synthesize 3D scenes from a collection of images using deep learning



Neural Network

Input : position + view direction (x, y, z, θ, φ).
Output : color + density (r, g, b, σ).

Subsequent research

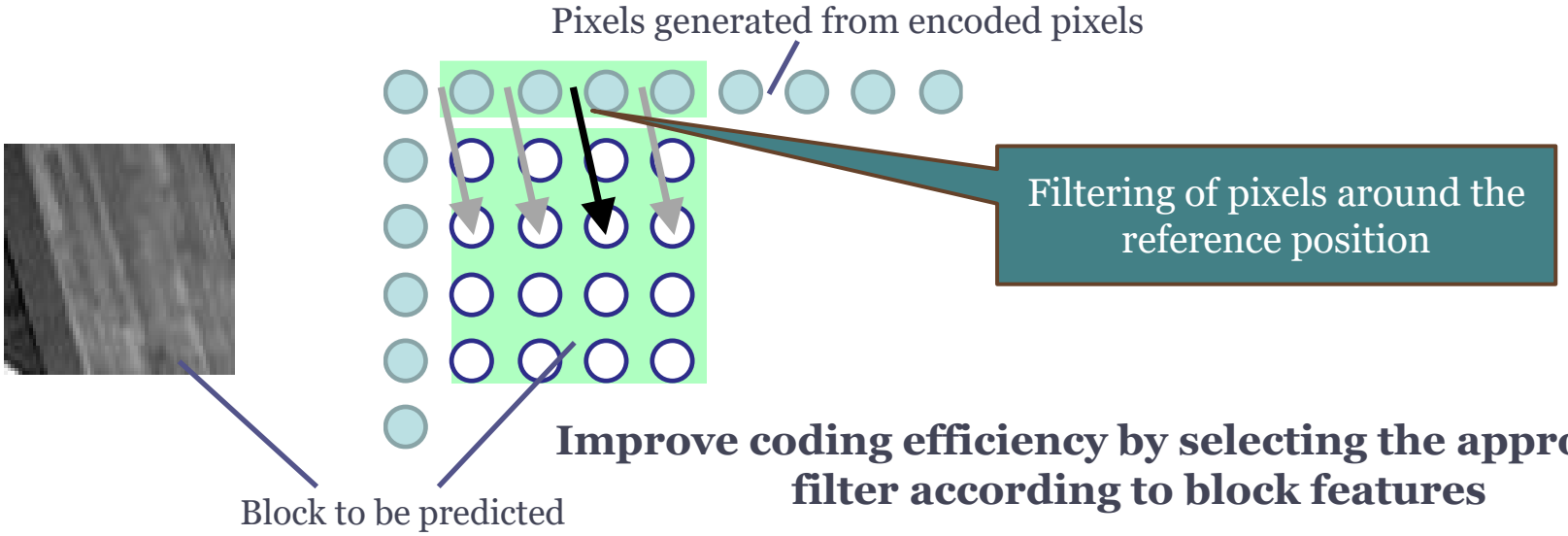
- Accelerating NeRF learning.
- Building models with unorganized image capture conditions.

Ex2 : A Study on Reference Pixel Interpolation Filters in Intra-Prediction

Intra-forecasting allows...
- Divide the images into blocks
- Use the encoded blocks to predict other blocks

Directional Prediction

One of the intra-prediction modes
Predicts the target pixel using pixels in a specific direction from each pixel



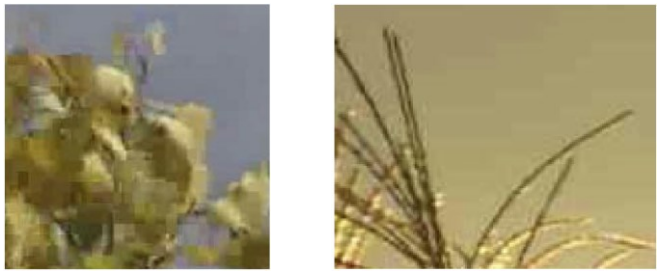
Improve coding efficiency by selecting the appropriate filter according to block features

Smoothing filter or sharpening filter, etc.

Ex3 : Reduction of Coding Distortion Using Super-Resolution Network

Coding distortion

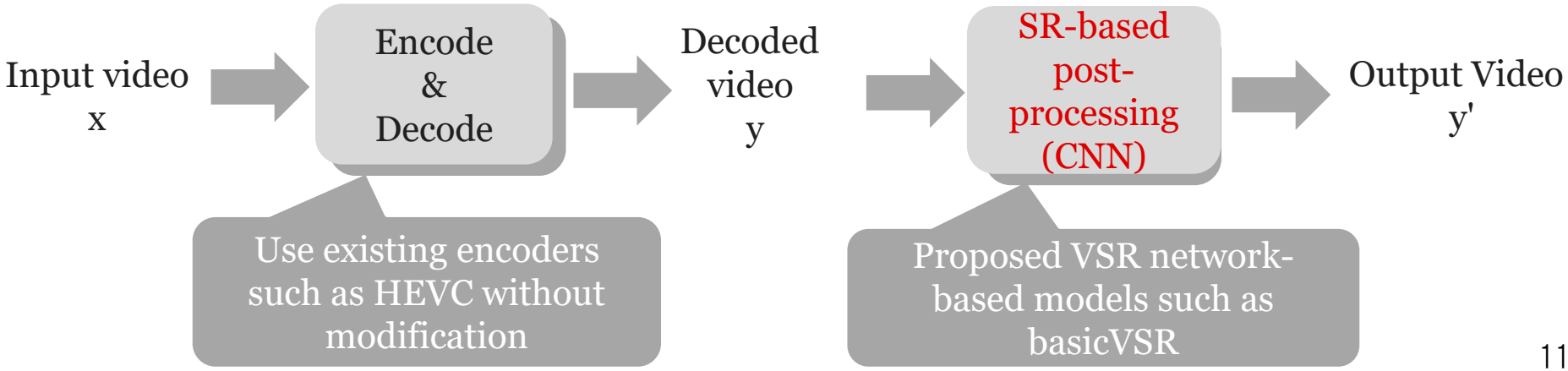
Distortions that occur during lossy encoding
Block noise originating from the coding block that is the unit of compression,
mosquito noise caused by censoring of high-frequency components, etc



Block noise　　Mosquito noise

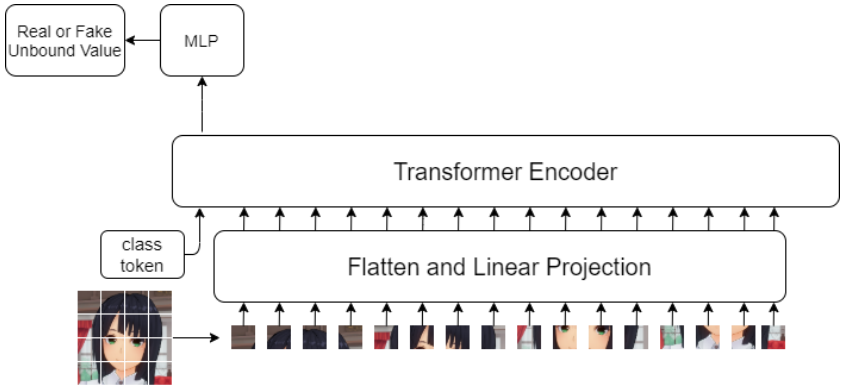
Image source：http://www.ieice-hbkb.org/files/02/02gun_05hen_09.pdf

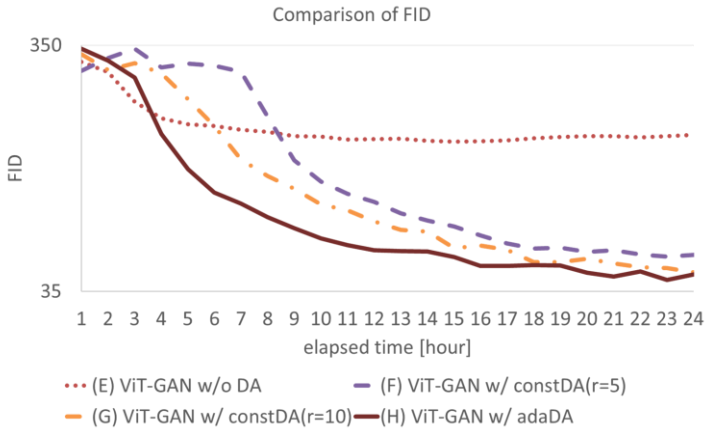
- Reduces compression artifacts by applying a super resolution (SR) network to the decoded video



Ex4 : Research on image generation methods

- Using Vision Transformer as a GAN discriminator





- Real/fake decision by making the output of MLP's one dimensional
- Reduce the number of parameters by sharing the Transformer encoder
- Change the strength of data augmentation based on loss to increase generalization

Reference: Shota HIROSE, Naoki WADA, Jiro KATTO, Heming Sun:" ViT-GAN: Using Vision Transformer as Discriminator with Adaptive Data Augmentation", ICCCI 2021