Course Overview

Multimedia Content Analysis

Wei-Ta Chu

Class Information

- **Time:** 9:10~12:00 Monday
- **Location:** CSIE New Building 65104
- **Lecturer:** Wei-Ta Chu (Office 65B08, CSIE Building)
- **Textbook:** Readings from books, journals and proceedings. Some references will be given after each class.
- **Course website:**
  - moodle.ncku.edu.tw
- **TA: 潘玠佑**, p76124540@gs.ncku.edu.tw

Grading (subject to change)

- Homeworks (25%)
  - Including program implementation and writing report
- Midterm (25%)
- Final Exam (20%)
- Final project (30%)
  - Including project proposal, project implementation, writing report, and oral presentation

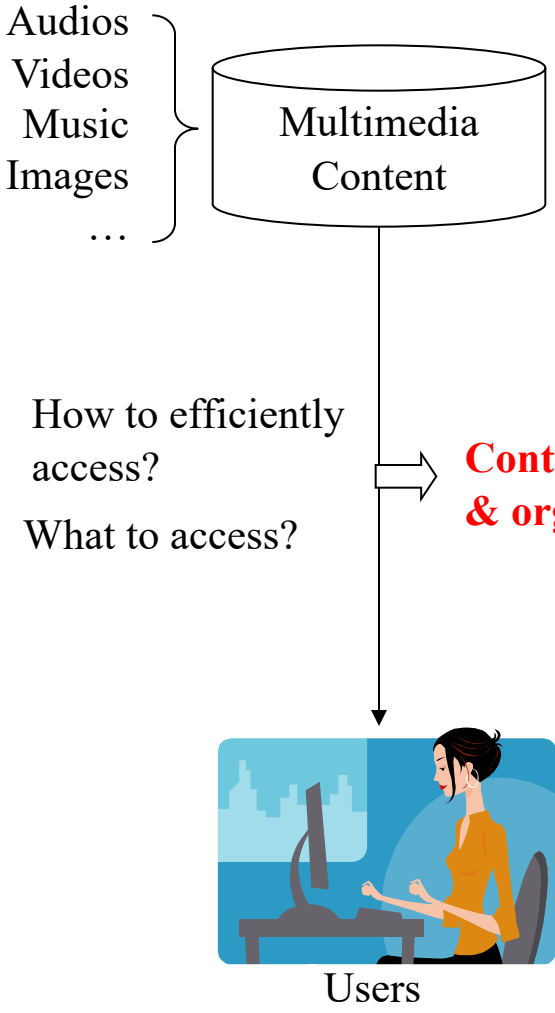
Schedule (subject to change)

日期	安排
2月19日	開課
2月26日	hw1公布 (project分組、主題方向制定）
3月4日	
3月11日	hw1繳交、hw2公布(shot change detection or CBIR)
3月18日	
3月25日	hw2繳交、hw3公布(project期中報告)
4月1日	
4月8日	
4月15日	老師出國，期中考
4月22日	hw3繳交、hw4公布 (machine learning or audio analysis)
4月29日	
5月6日	hw4繳交, hw5公布
5月13日	
5月20日	hw5繳交
5月27日	
6月3日	期末考
6月10日	端午節放假
6月17日	Final project報告&書面報告

Final Project

- You can
  - Work on a project related to your master thesis
  - Propose a new topic that has never been done
  - Participate competitions released on some public platforms, like aidea-web.tw or Kaggle
- Of course, all should be related to multimedia
- Start at the beginning of this semester

Why This Course? (1/2)



Find what we want in large volumes of content.

1. Classification
   - ·Music or video genre classification
   - ·Content-based clustering
2. Retrieval
   - ·Content-based image retrieval
   - ·Query by keywords

…

Get what we want in a specific data item.

1. Segmentation
   - ·Speech/music discrimination in audio streams
   - ·Structure analysis in sports videos
2. Adaptation
   - ·Video summarization / highlight extraction
   - ·Event detection

…

Information is of no use unless you can actually access it.

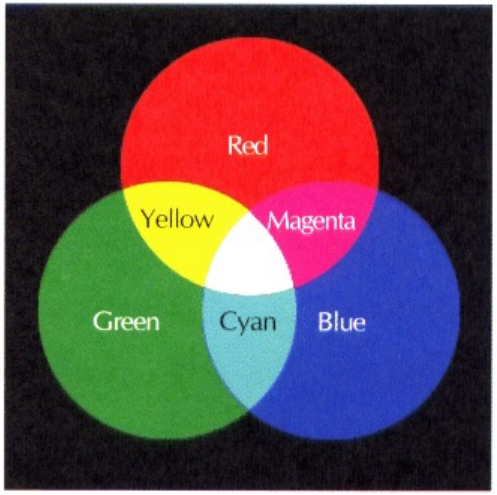
Why This Course? (2/2)

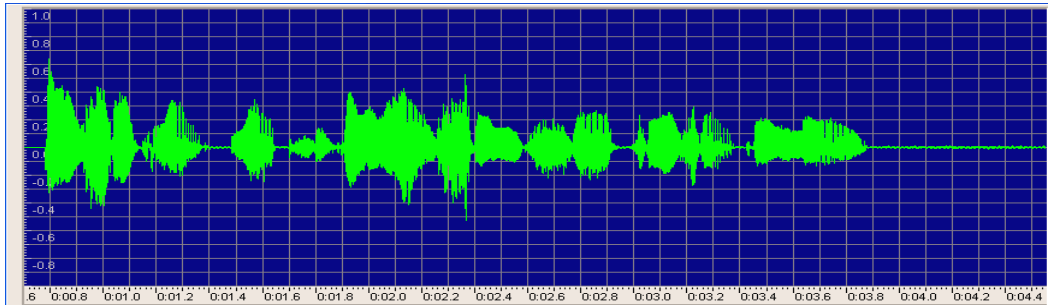
- The goals of this course are to
  - Introduce basic ideas, techniques, and tools for analyzing multimedia content
  - Inspire students to propose and develop new multimedia-related techniques/systems
- Enhance the efficiency and effectiveness of multimedia information access.

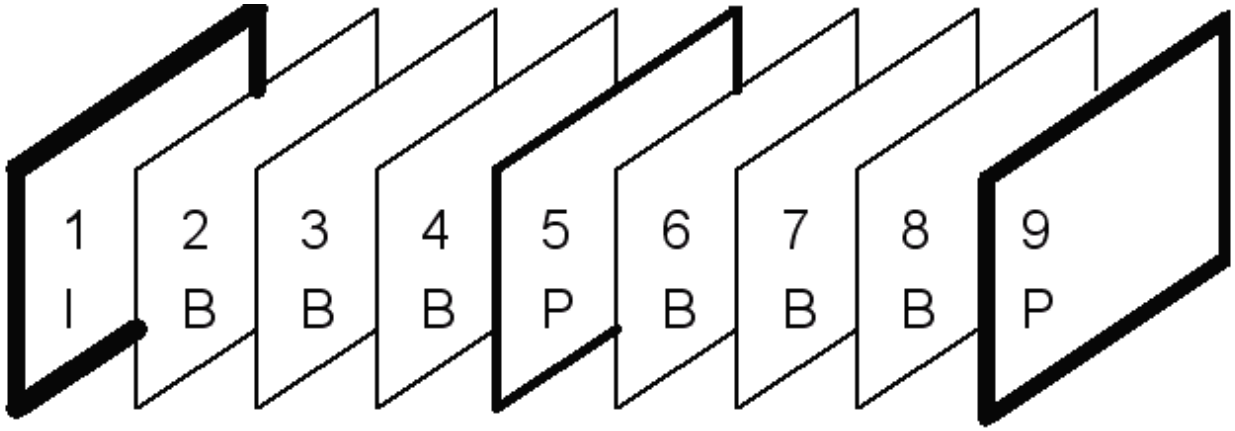
Syllabus

- PART 1: Conventional multimedia analysis
  - Essence of Image, Video, and Audio
  - Video/Audio Content Analysis
  - Machine Learning Techniques
- PART 2: Deep Learning and Recent Works
  - CNN, RNN, GAN, transformer, object detection, semantic segmentation, image translation

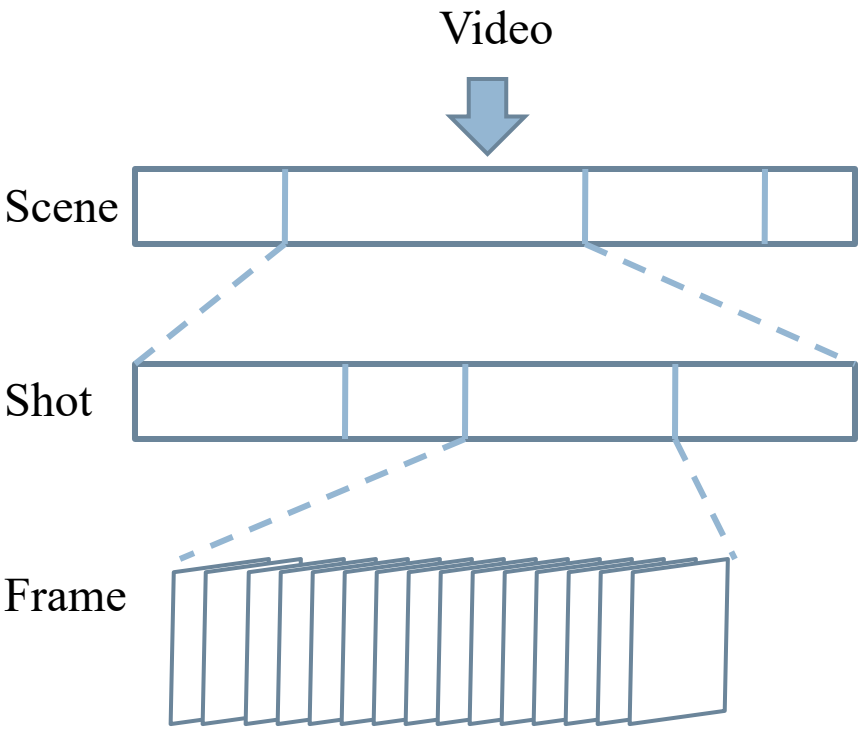
Essence of Image, Video, and Audio

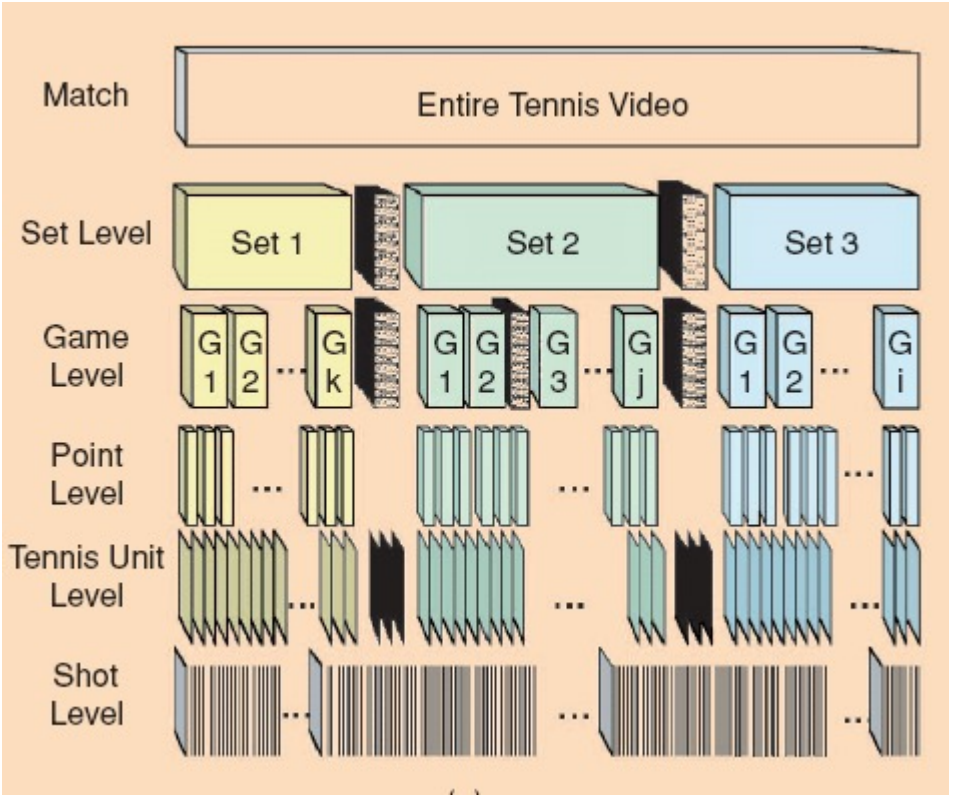






Video Structure Analysis

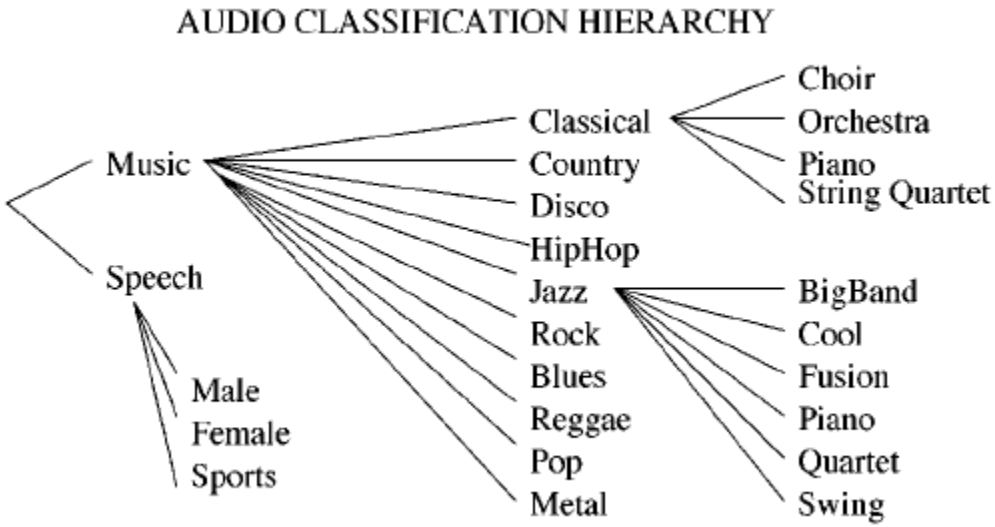




A. Kokaram, et al., "Browsing Sports Video",2006

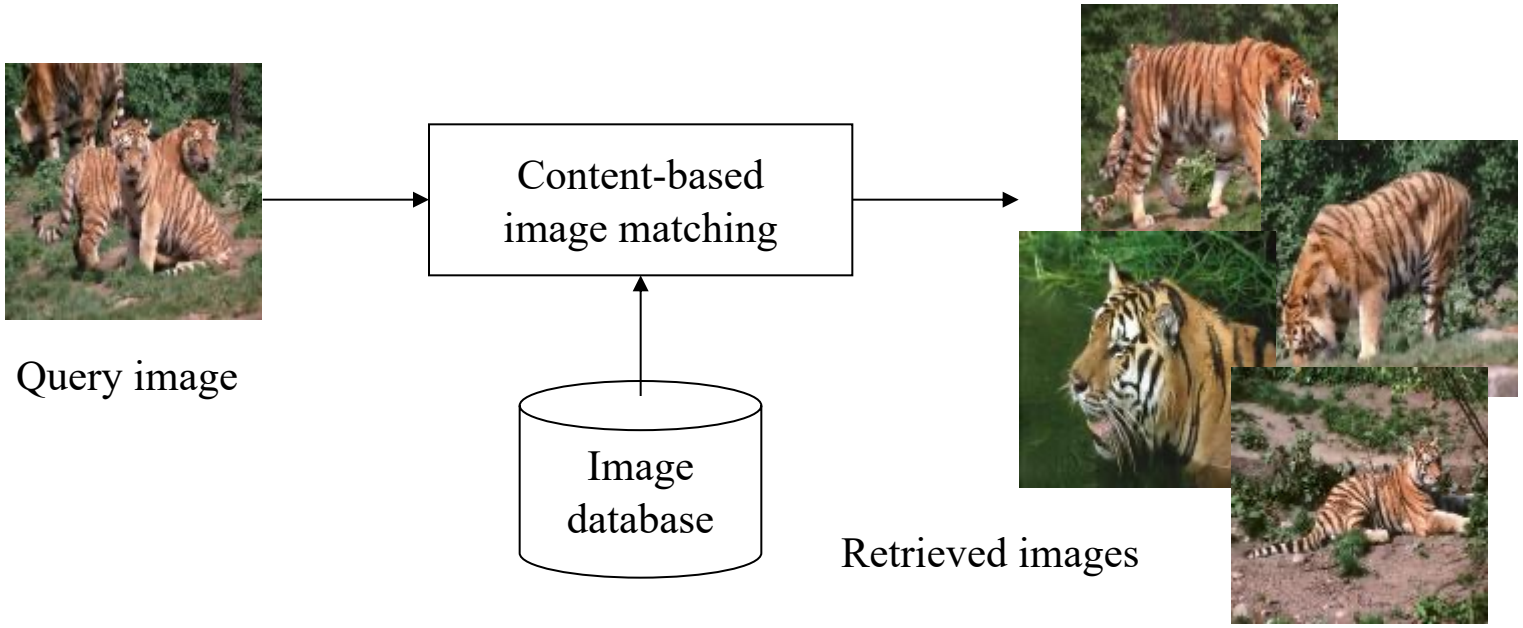
Audio Content Analysis

- Audio classification
  - Speech, music, environment sound, …
- Music genre classification
  - Jazz, pop, classical, …



G. Tzanetakis, et al., "Musical genre classification of audio signals", 2002

Content-based Image Retrieval

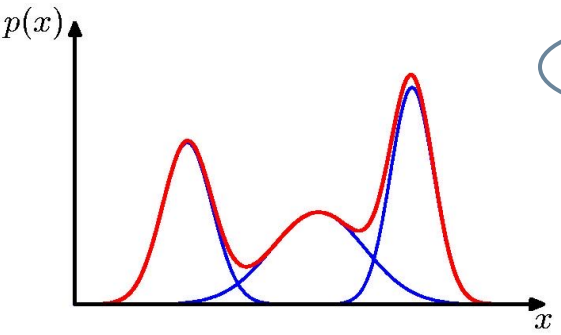


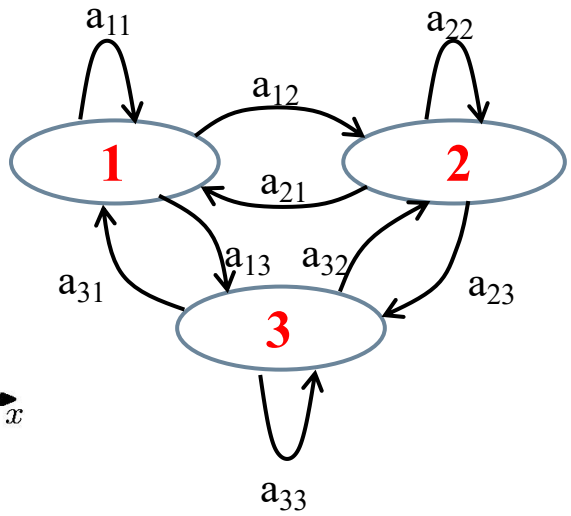
Find images/video shots that have similar features.
E.g. color, texture, edge, motion, …

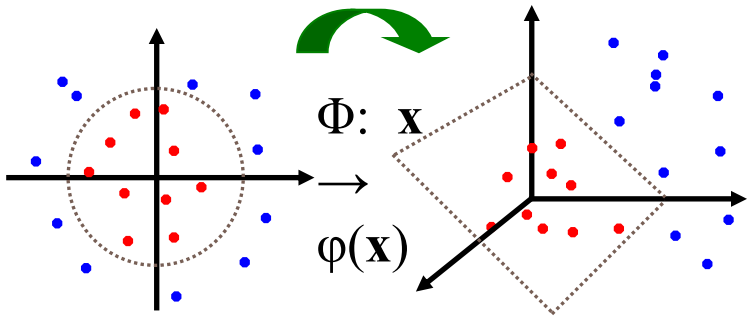
CIRES: Content-Based Image Retrieval System, http://amazon.ece.utexas.edu/~qasim/research.htm

Machine Learning Techniques

- Some machine learning techniques
  - Gaussian Mixture Model (GMM)
  - Hidden Markov Model (HMM)
  - Support Vector Machine (SVM)







Deep Learning

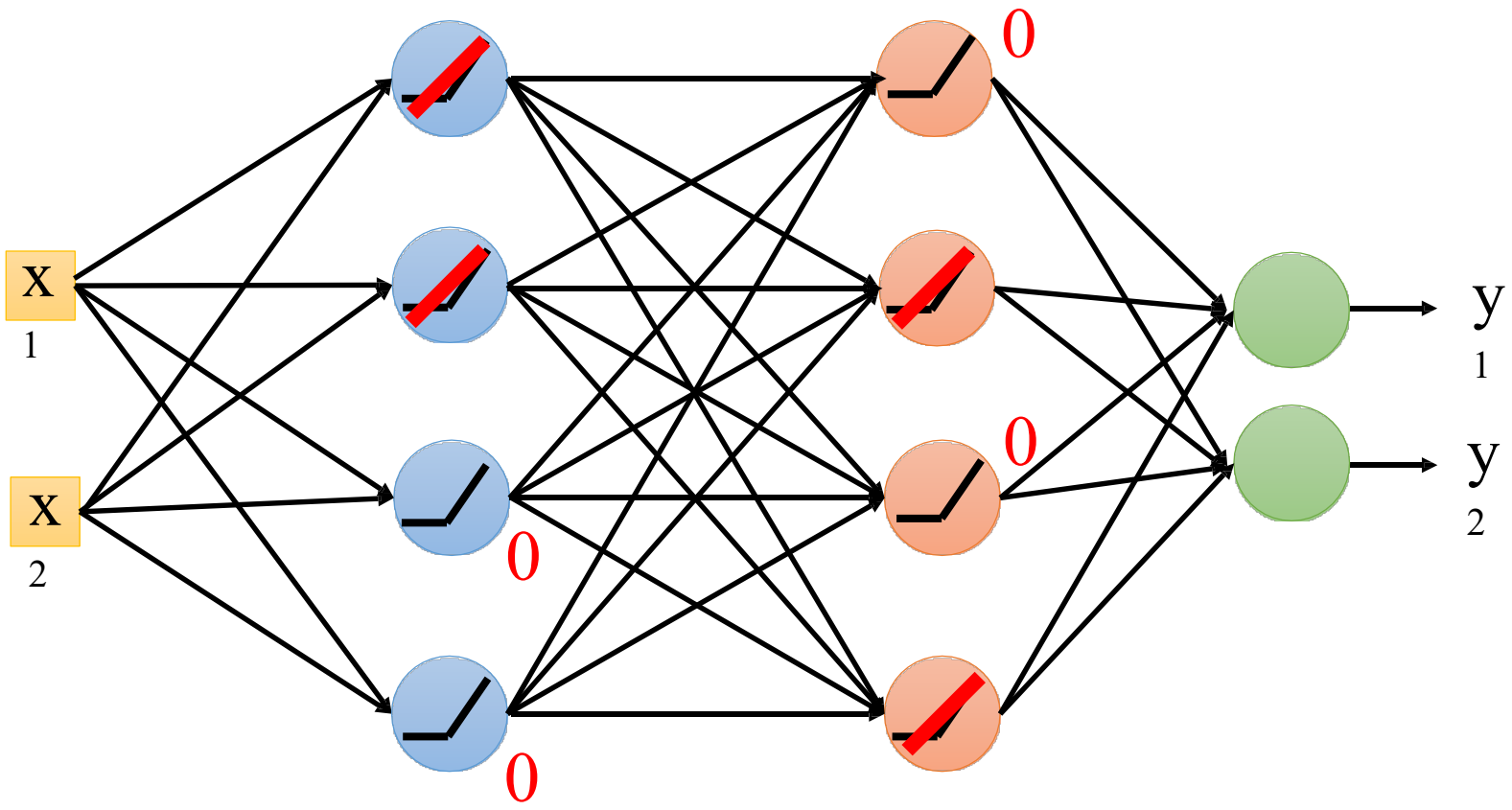


Image Segmentation

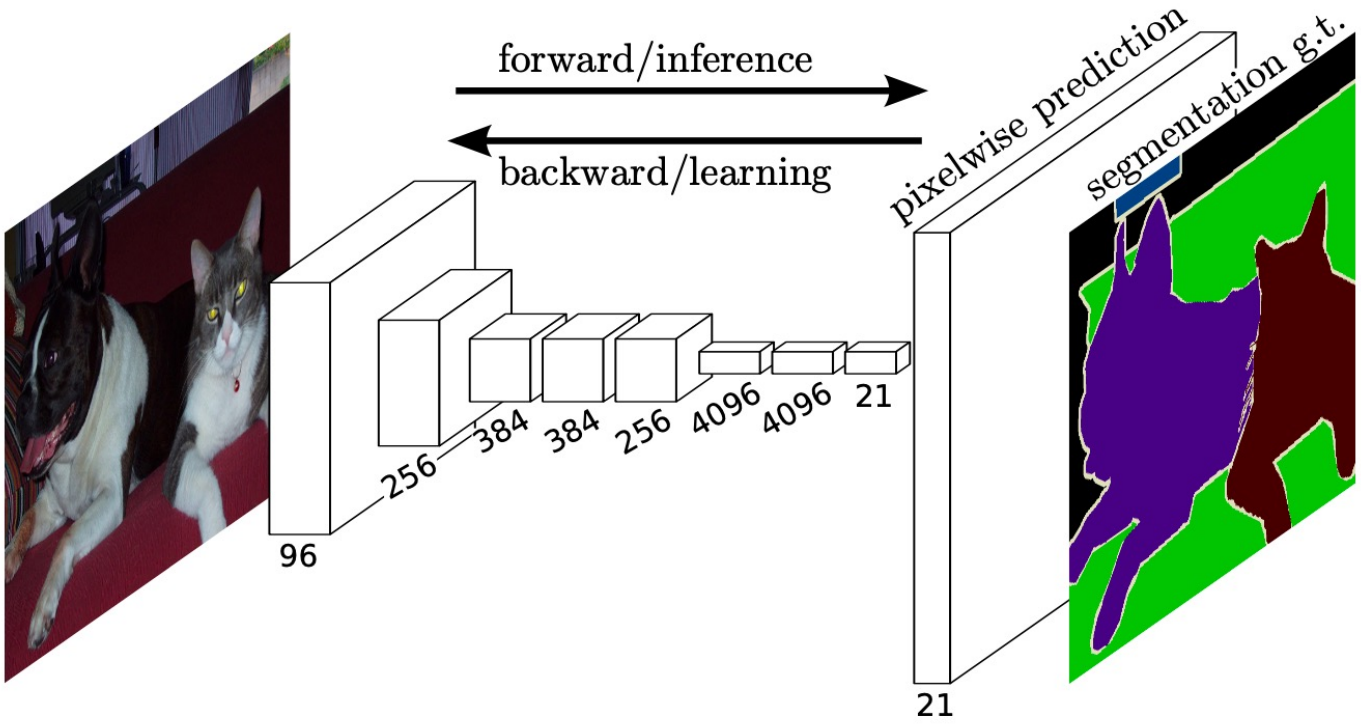


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

Generative Adversarial Networks

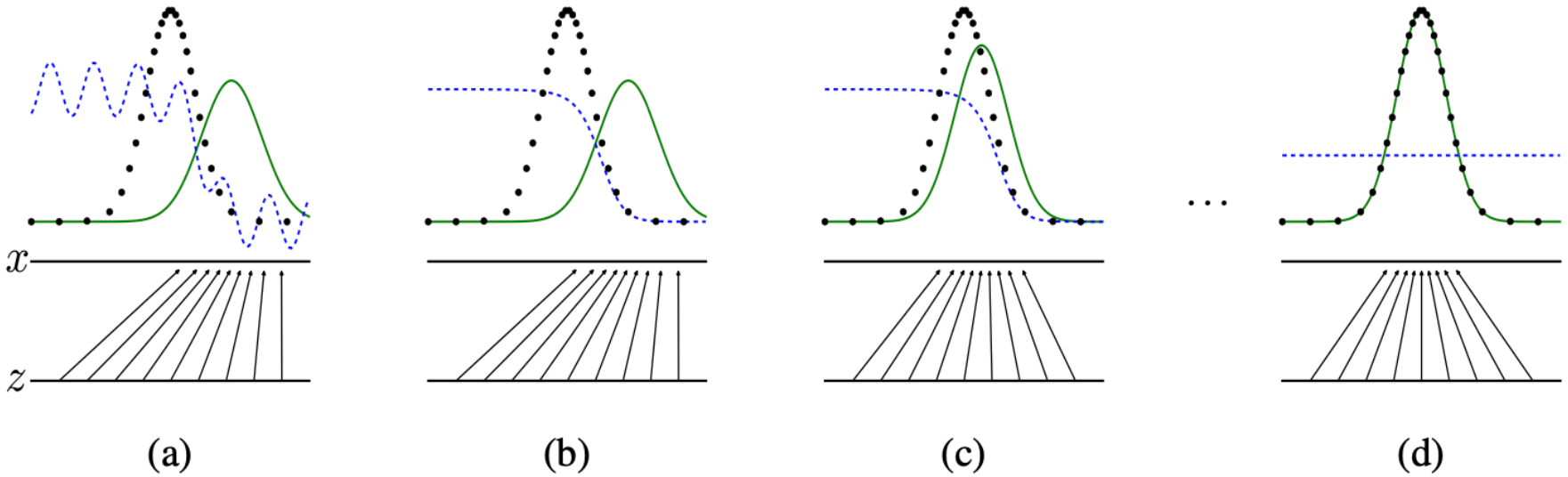
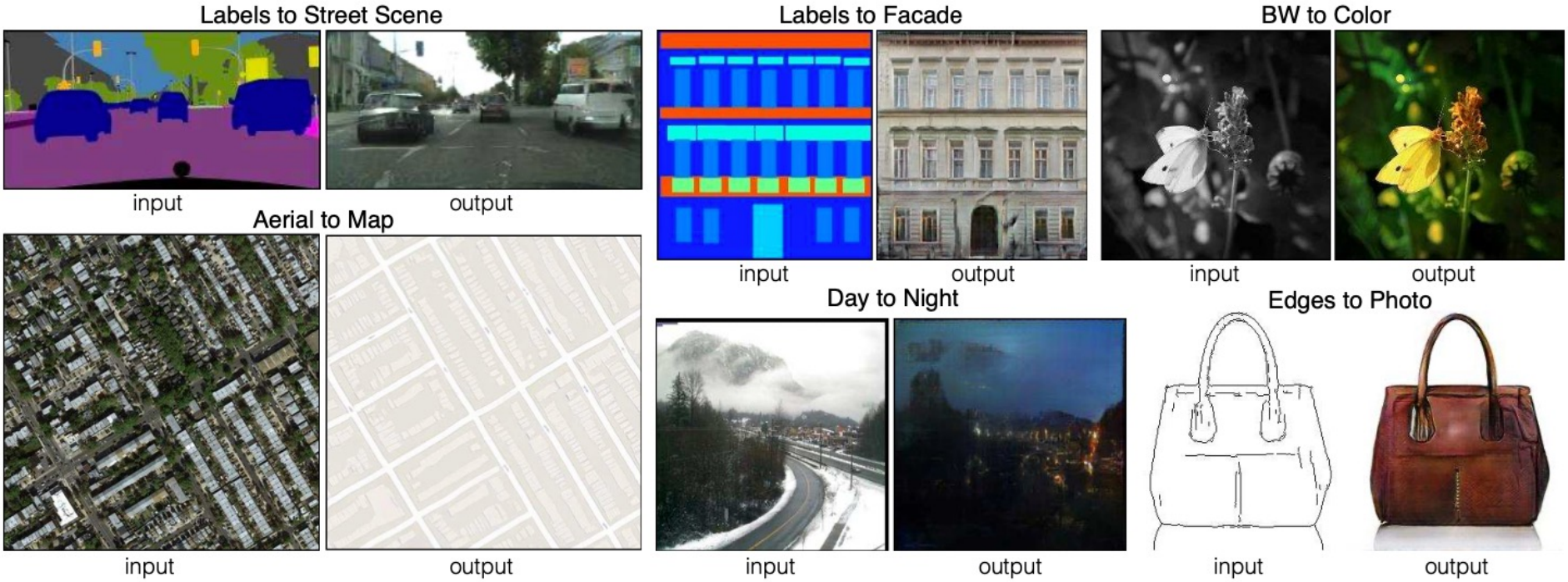
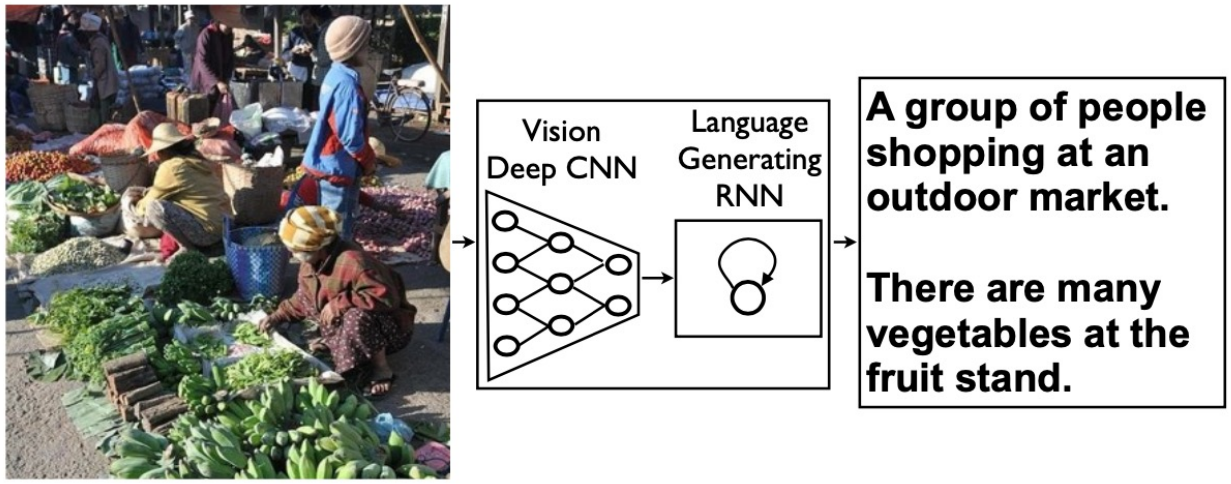


Figure 1: Generative adversarial nets are trained by simultaneously updating the **d**iscriminative distribution (D, blue, dashed line) so that it discriminates between samples from the data generating distribution (black, dotted line) $p_{\boldsymbol{x}}$ from those of the **g**enerative distribution p_g (G) (green, solid line). The lower horizontal line is the domain from which $\boldsymbol{z}$ is sampled, in this case uniformly. The horizontal line above is part of the domain of $\boldsymbol{x}$. The upward arrows show how the mapping $\boldsymbol{x} = G(\boldsymbol{z})$ imposes the non-uniform distribution p_g on transformed samples. G contracts in regions of high density and expands in regions of low density of p_g. (a) Consider an adversarial pair near convergence: p_g is similar to p_{data} and D is a partially accurate classifier. (b) In the inner loop of the algorithm D is trained to discriminate samples from data, converging to $D^*(\boldsymbol{x}) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}$. (c) After an update to G, gradient of D has guided $G(\boldsymbol{z})$ to flow to regions that are more likely to be classified as data. (d) After several steps of training, if G and D have enough capacity, they will reach a point at which both cannot improve because $p_g = p_{\text{data}}$. The discriminator is unable to differentiate between the two distributions, i.e. $D(\boldsymbol{x}) = \frac{1}{2}$.

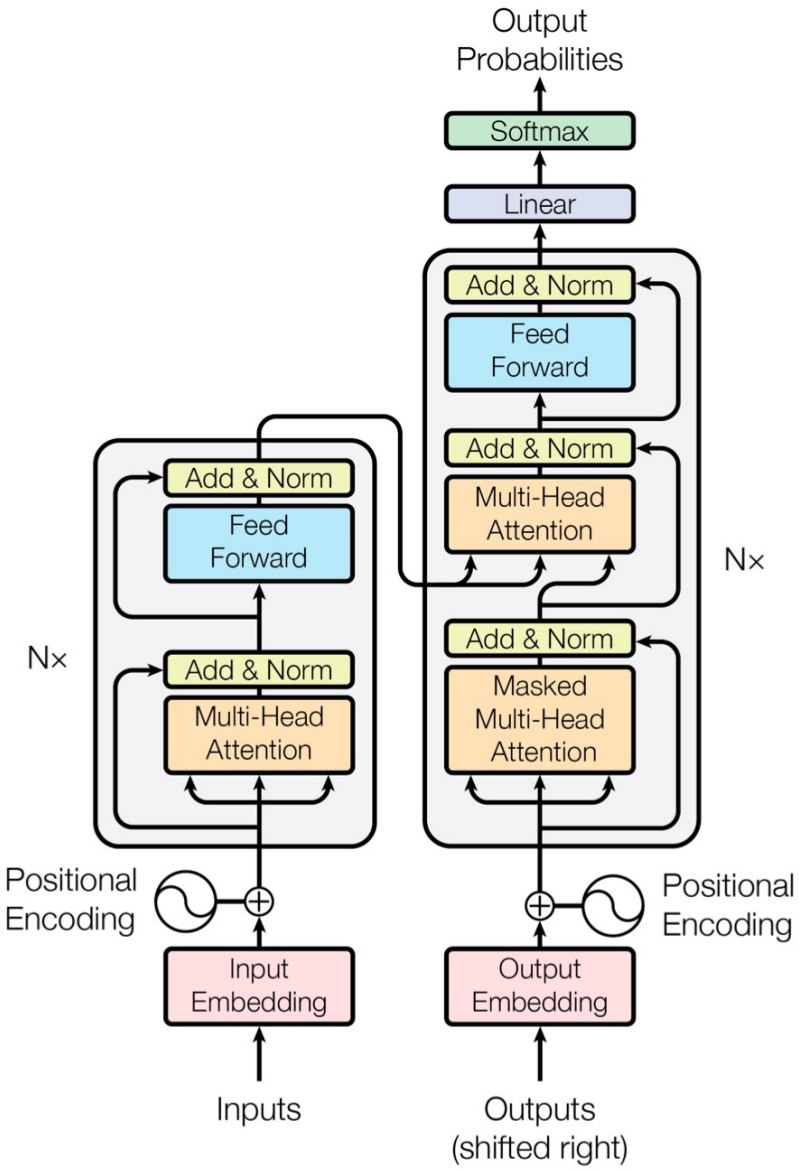
Image to Image Translation



Vision and Text



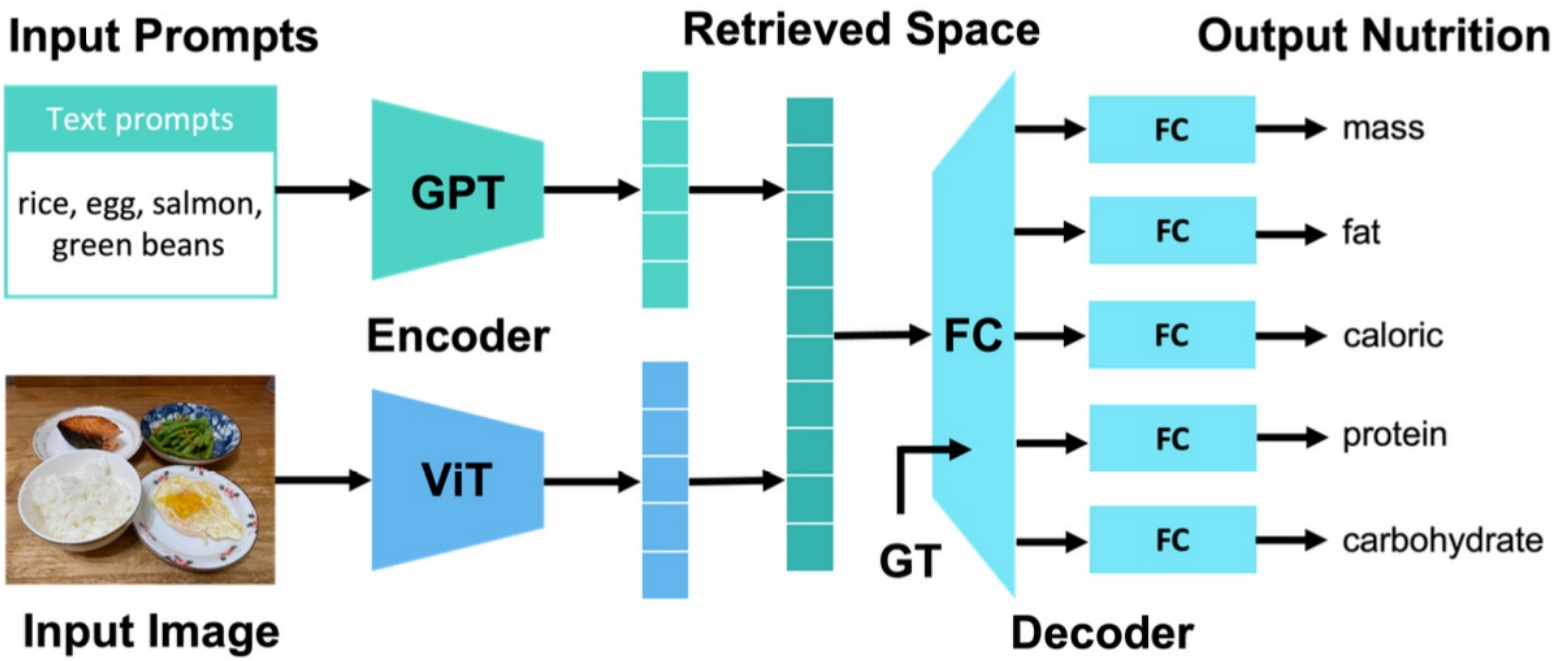
Transformer



Previous Projects

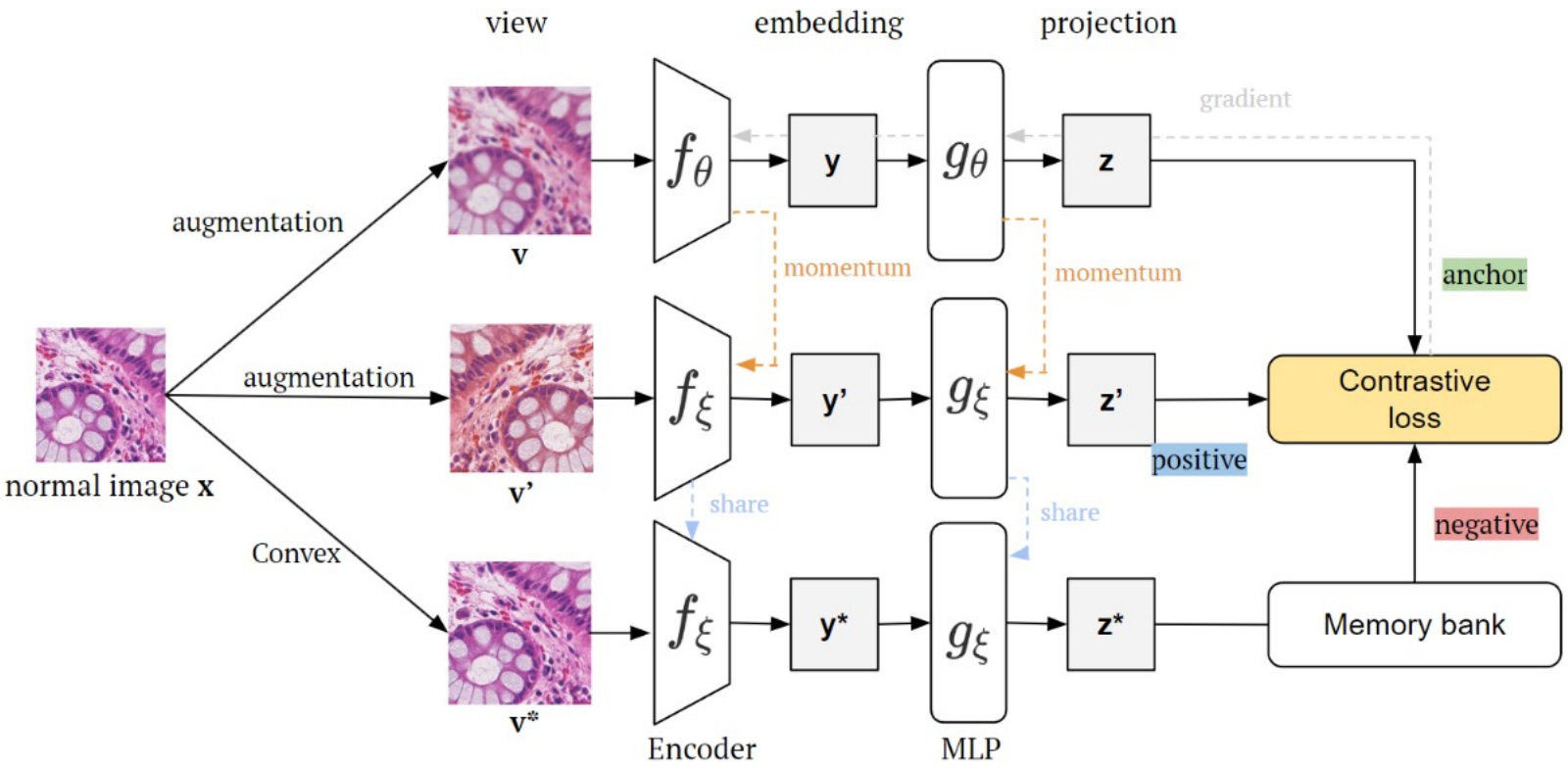
Projects in 2023

- Nutrition estimation from food images



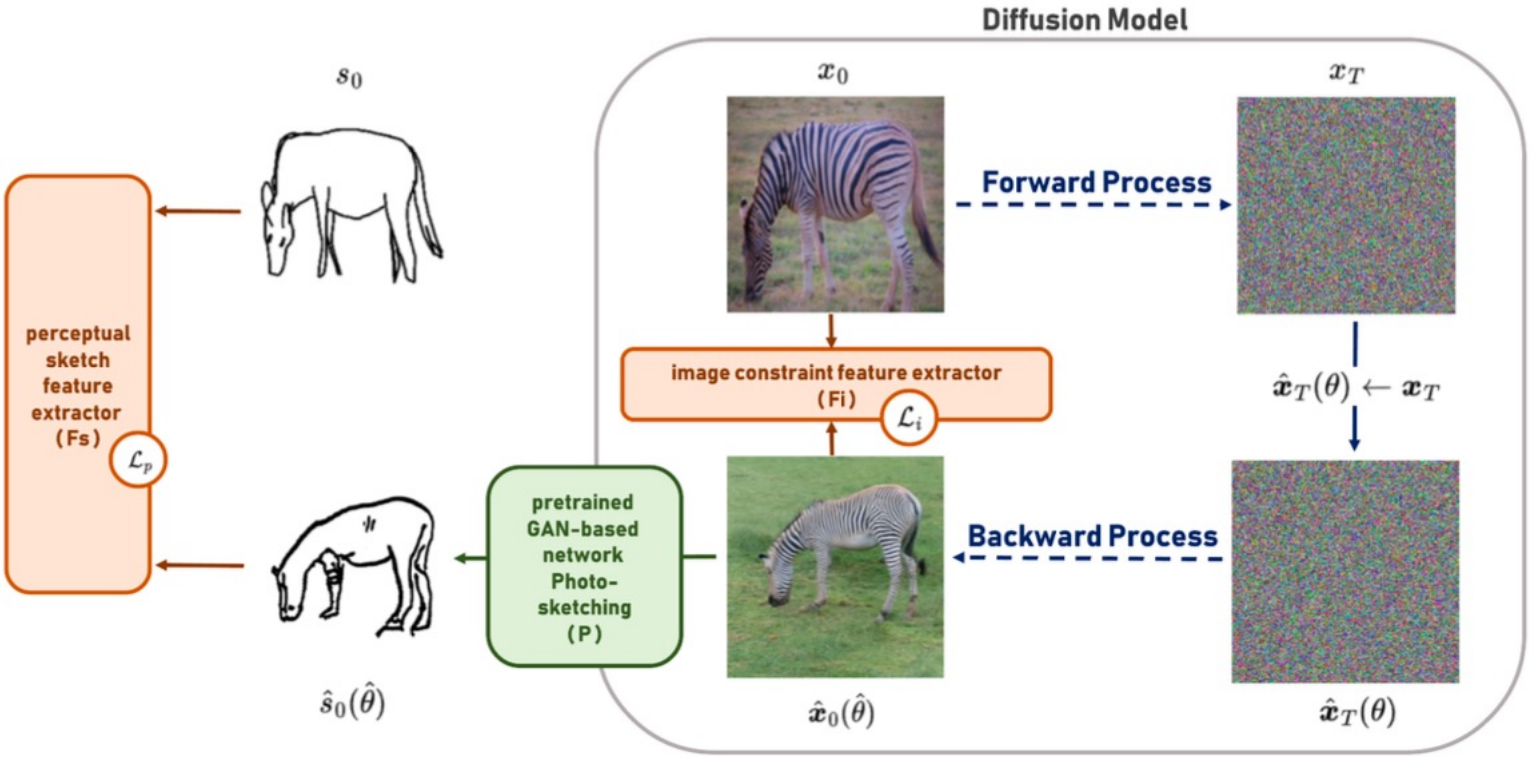
Projects in 2023

- Anomaly Detection in Pathology images



Projects in 2023

- Sketch-Guided Image Synthesis with Diffusion Model



Projects in 2023

- Learning to Construct a Multi-View Wire Art with Neural Radiance Field

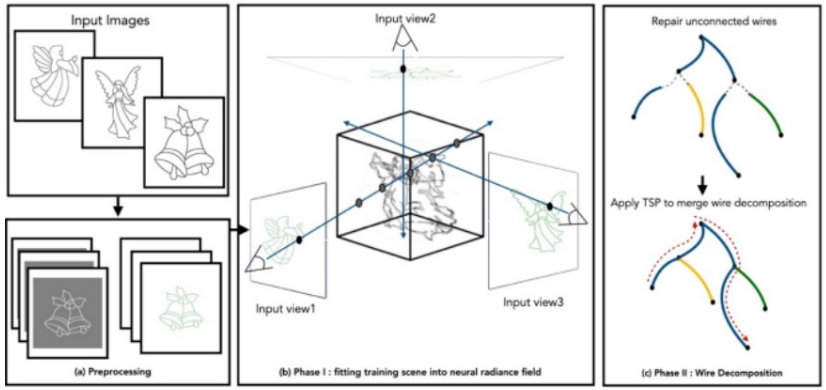


Figure 6. System framework

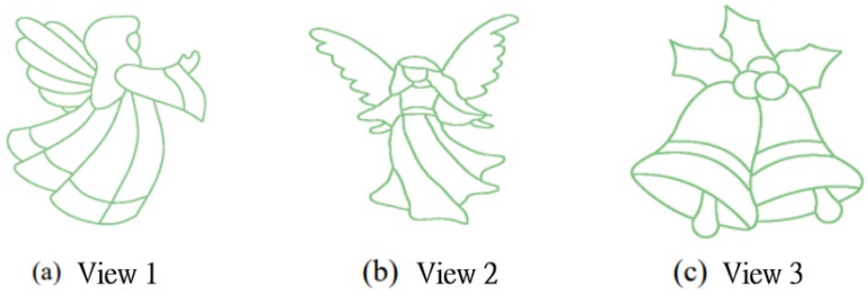
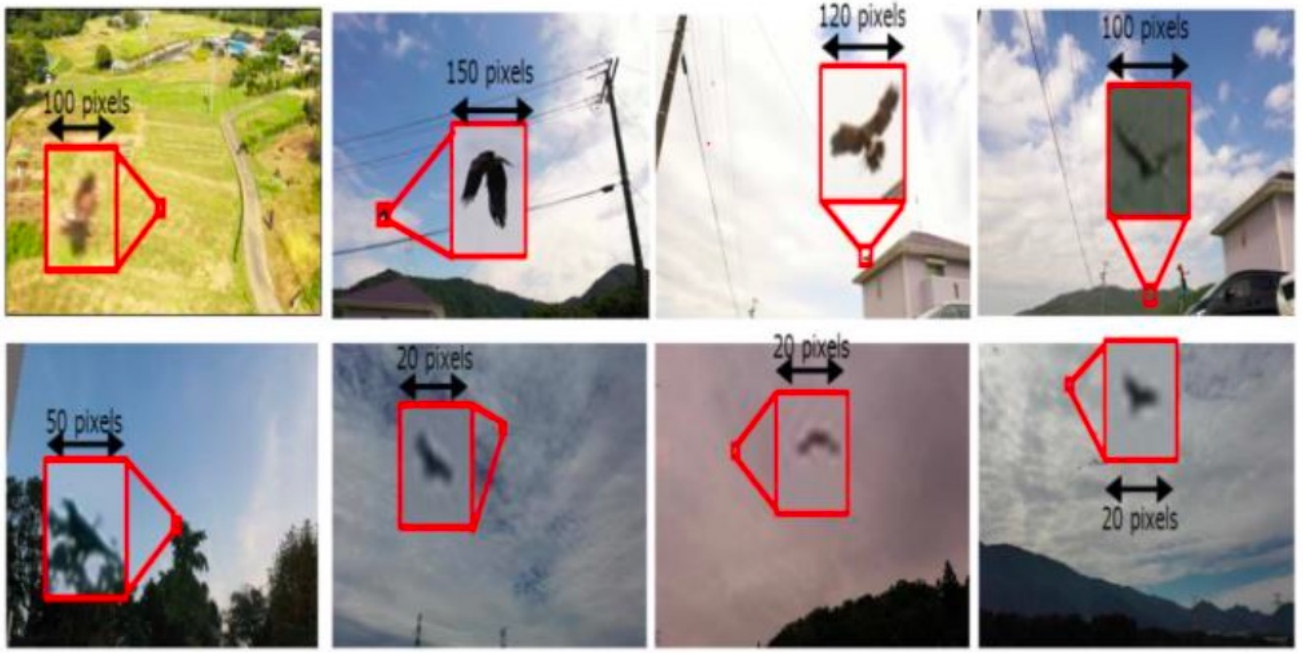


Figure 7. 3-views Wire Art model datasets

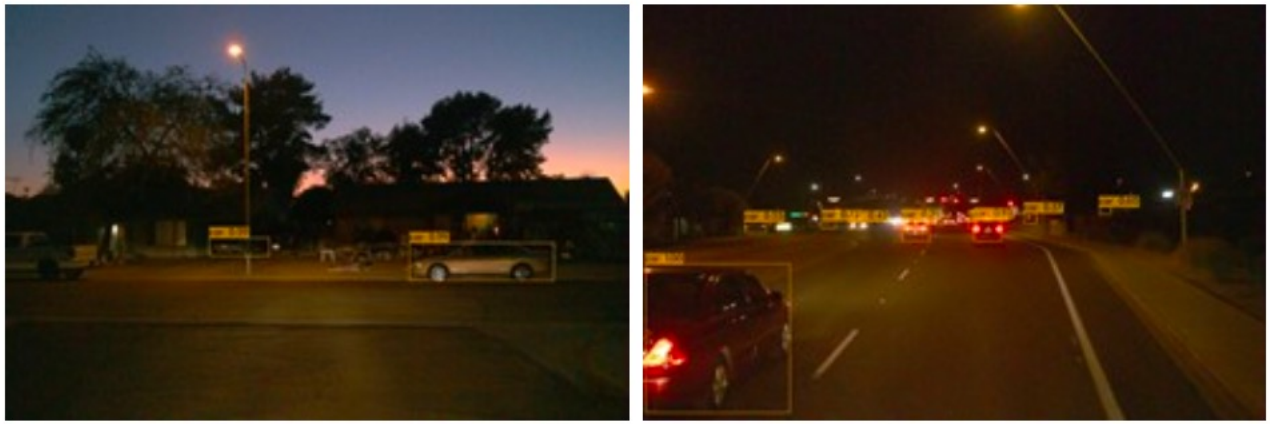
Projects in 2023

- Small Object Detection for Spotting Birds



Projects in 2022

- Lane line Detection and Vehicle Tracking

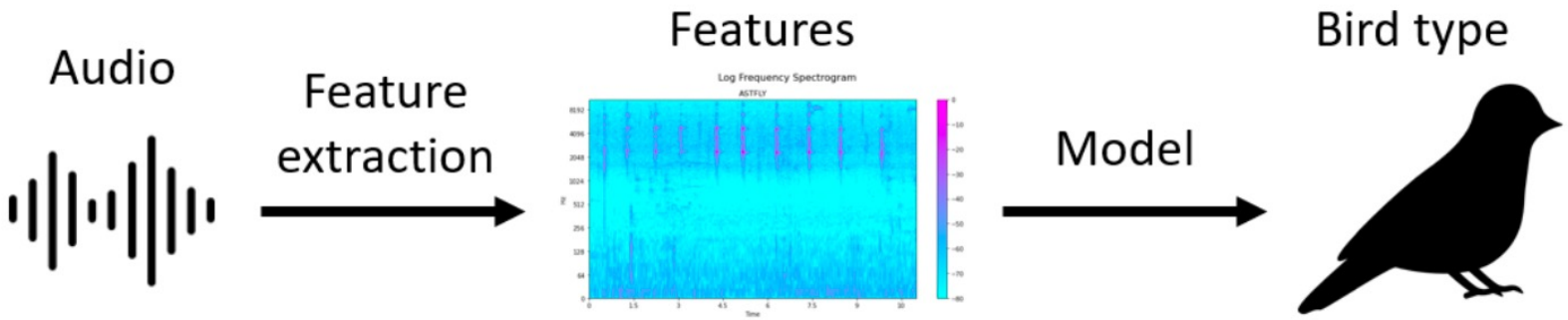


(a) result 1 (b) result 2

(c) result 3 (d) result 4

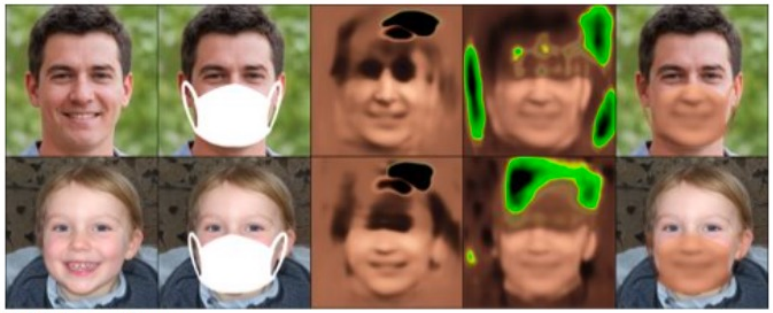
Projects in 2022

- Identify Birds with Sound



Projects in 2022

- Generating Face Behind the Mask



500 iterations

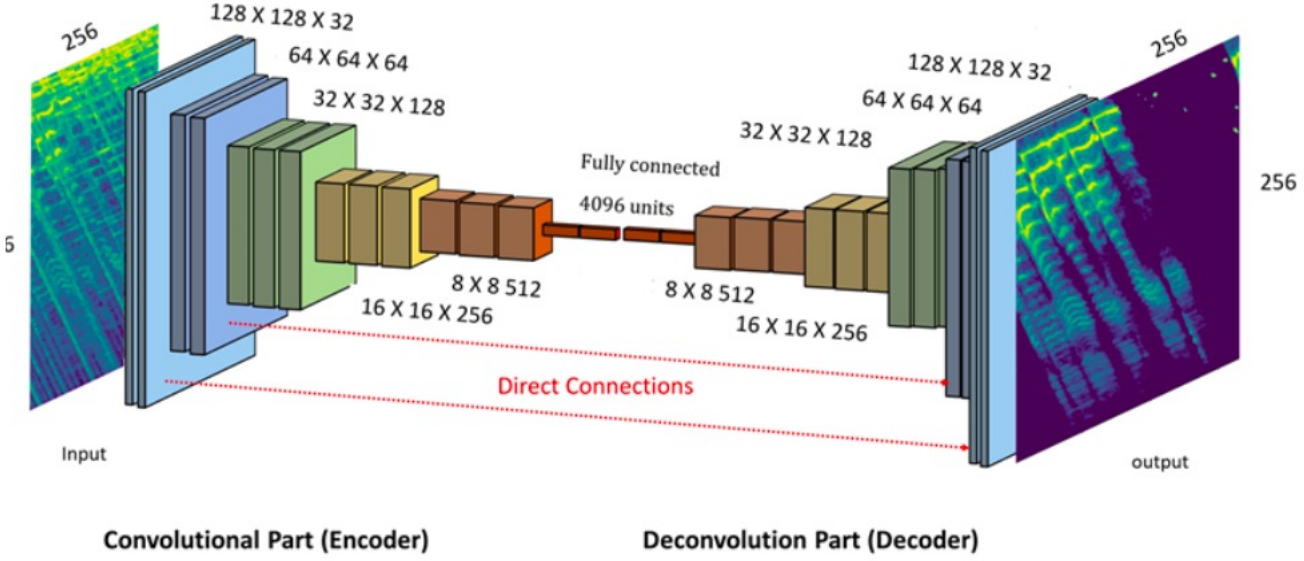
2000 iterations

4000 iterations

9500 iterations

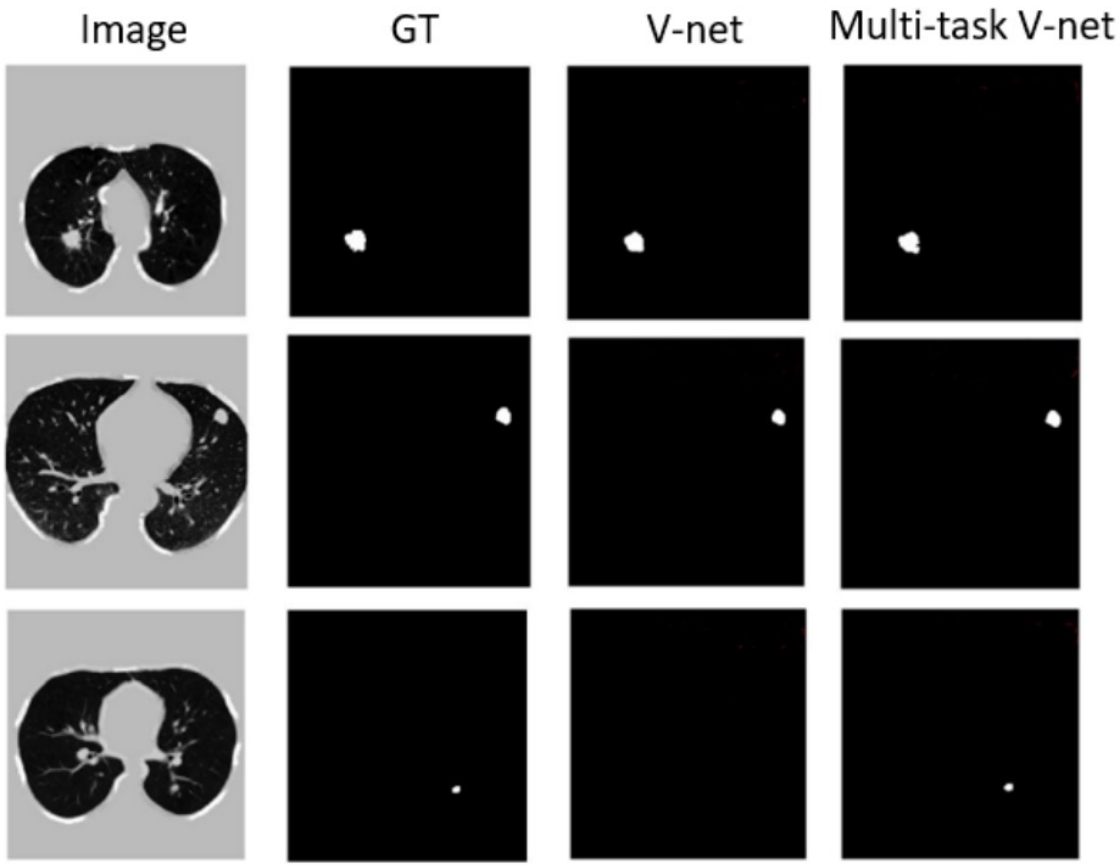
Projects in 2022

- Speech Denoising Neural Networks



Projects in 2022

- Lung Nodule Segmentation and Classification



Projects in 2022

- CNN-based Facial Age Prediction Model