Creating a Better Future: Harnessing AI for Social and Environmental Responsibility

Prof. Sam Kwong

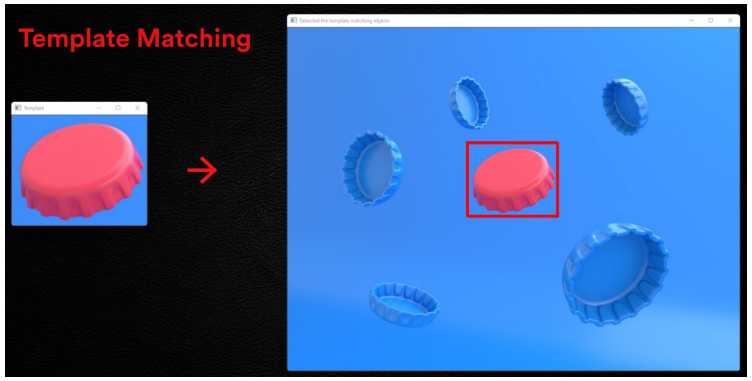
Lingnan University, Hong Kong SAR

December 2023

History and Evolution of AI

Early Days of Computer Vision (1950s)

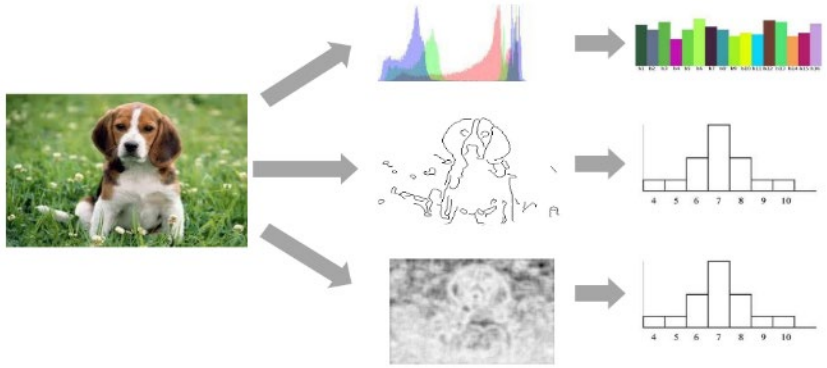
- Basic pattern recognition: **Template matching**



Source: analyticsvidhya.com

Pioneering Days (1960s and 1970s)

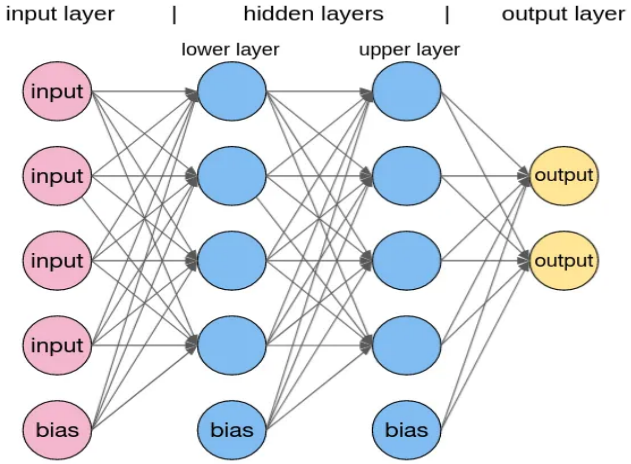
- Image recognition relied on **handcrafted features**



Source: paddlepaddle.org.cn

Rise of BP Neural Networks (1980s)

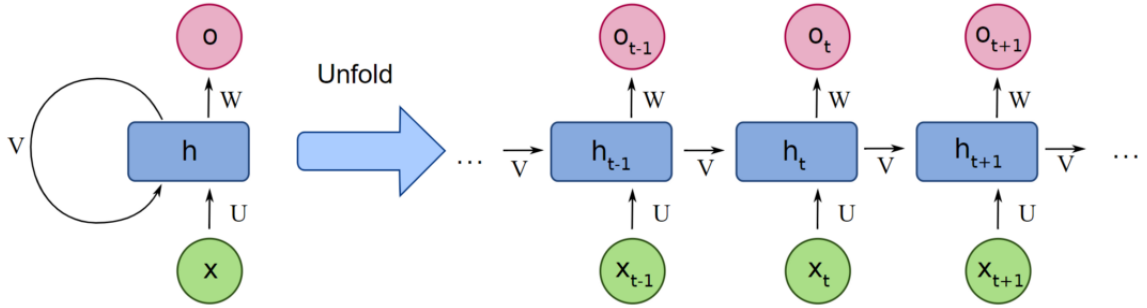
- Multi-layer Perceptron



Source: Medium.com

Rise of Recurrent Neural Networks(1990s)

- Image recognition relied on **handcrafted features**

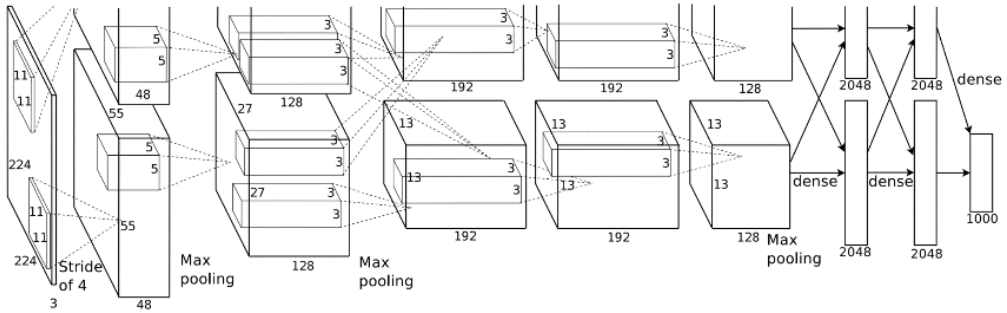


Source: Medium.com

History and Evolution of AI

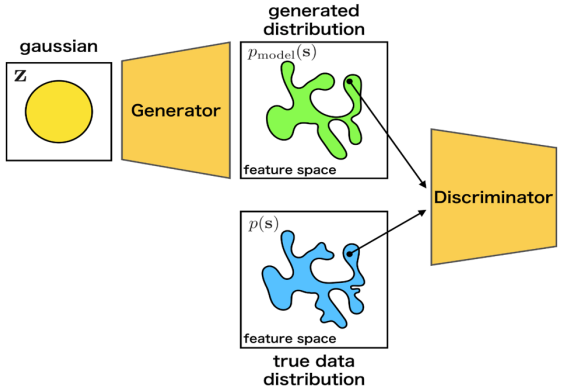
Deep Learning Revolution (2010s)

- Significant shift with the advent of **deep learning**



Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

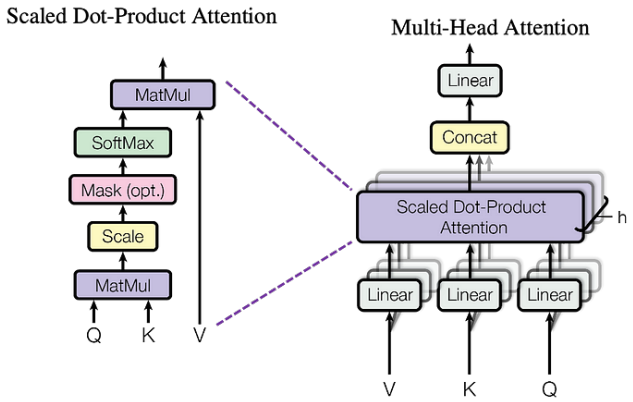
- Emergence of **Generative Adversarial Networks** (GANs)



Source: docs.chainer.org

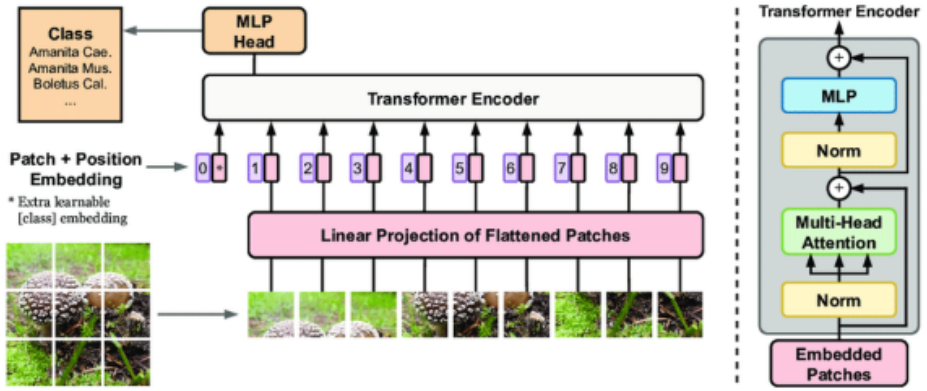
Recent Advancements

- Rise of **Attention Mechanisms**



Source: Medium.com

- **Transformers** Dominate Vision Tasks



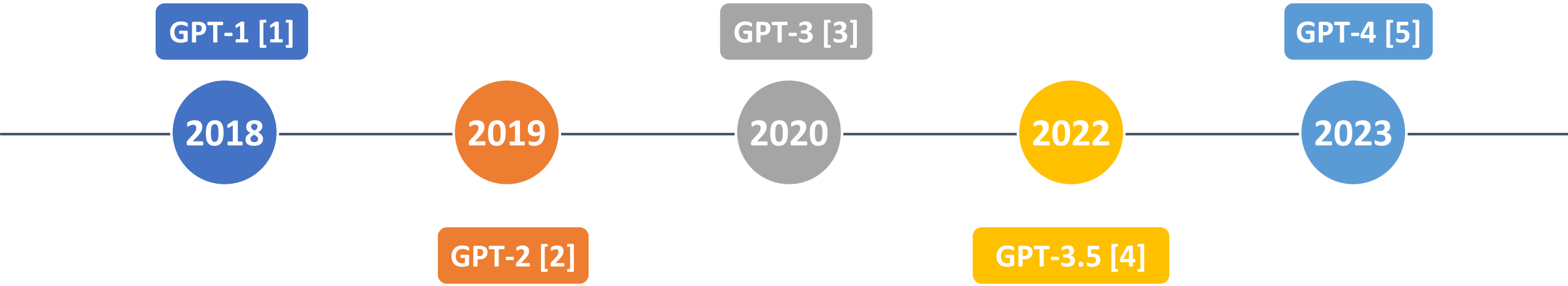
Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

History and Evolution of AI

Large Language Models and GPTs

- Recently, the research on **Large Language Models (LLMs)** has been largely advanced by both academia and industry, and a remarkable progress is the launch of Generative Pre-trained Transformers (**GPT)**

GPT Series Models



[1]. Radford, Alec, et al. "Improving language understanding by generative pre-training," Technical Report, OpenAI, 2018.
[2]. Radford, Alec, et al. "Language models are unsupervised multitask learners." Technical Report, OpenAI, 2019.
[3]. Brown, Tom, et al. "Language models are few-shot learners." ," In Proceedings of Advances in Neural Information Processing Systems (NIPS), no. 33, pp.1877-1901, 2020.
[4]. Ouyang, Long, et al. " Training language models to follow instructions with human feedback," In Proceedings of Advances in Neural Information Processing Systems (NIPS), no. 35, pp. 27730-27744, 2022.
[5]. OpenAI, "GPT-4 Technical Report", ArXiv, abs/2303.08774., 2023.

History and Evolution of AI

Large Language Models and GPTs

	No. of Parameters	Training data
GPT 2 [1]	1.5 billion ($x10^9$)	8 million web pages (40GB)
GPT 3 [2]	175 billion ($x10^9$)	Web crawl + books + Wikipedia (45 TB) (300 billion tokens)
GPT 3.5 [3]	355 billion ($x10^9$)	Web crawl + books + Wikipedia (400 billion tokens)
GPT 4 [4]	1.8 trillion ($x10^{12}$)	Unknown (13 trillion tokens)
Infant Brain [5]	50 trillion synapses (86 billion neurons)	Streaming audio of mommy
3yo Child Brain [5]	1000 trillion (86 billion neurons)	0.75 million hours of streaming multimodal sensory data.
Adult Brain [5]	500 trillion (86 billion neurons)	Life experience

[1]. Radford, Alec, et al. "Language models are unsupervised multitask learners." Technical Report, OpenAI, 2019.
[2]. Brown, Tom, et al. "Language models are few-shot learners." ," In Proceedings of Advances in Neural Information Processing Systems (NIPS), no. 33, pp.1877-1901, 2020.
[3]. Ouyang, Long, et al. " Training language models to follow instructions with human feedback," In Proceedings of Advances in Neural Information Processing Systems (NIPS), no. 35, pp. 27730-27744, 2022.
[4]. OpenAI, "GPT-4 Technical Report", ArXiv, abs/2303.08774., 2023.
[5]. https://www.parentingforbrain.com/brain-development/

Green AI and AI Responsibility

Green AI

- AI relies on the extensive utilization of deep learning (DL) techniques
- Growing carbon emissions resulting from the increasing size of DL networks

Image source: Google Images

- Green AI [1] aims to minimize the **carbon footprint** of DL systems

AI Responsibility

Environmental Responsibility

Social Responsibility

- Apart from delivering convenience, AI should also shoulder the human environmental and social responsibilities

[1]. Kuo, C-C. Jay, and Azad M. Madni. "Green learning: Introduction, examples and outlook." Journal of Visual Communication and Image Representation Vol. 90:103685, 2023.

AI for Human

Everyday Convenience

- Unlocking our phones with facial recognition
- Getting fashion advice from AI-driven Apps

Enhanced Creativity

- Assist artists and designers in creating visuals, or even generating art

Social and environmental Responsibility

- Enhancing **security protocols**
- Pioneering breakthroughs in **medical diagnostics**
- Optimizing **traffic safety** algorithms
- Conducting in-depth analyses of **underwater ecosystems**

Facial recognition phone unlock (Source: livemint.com)

AI-assisted art creation (Source: wikipedia.org)

Economic Value and Social Impact

Boosting Productivity

- Faster decision-making and product development

Creating New Markets

- Augmented reality shopping
- AI-driven content creation

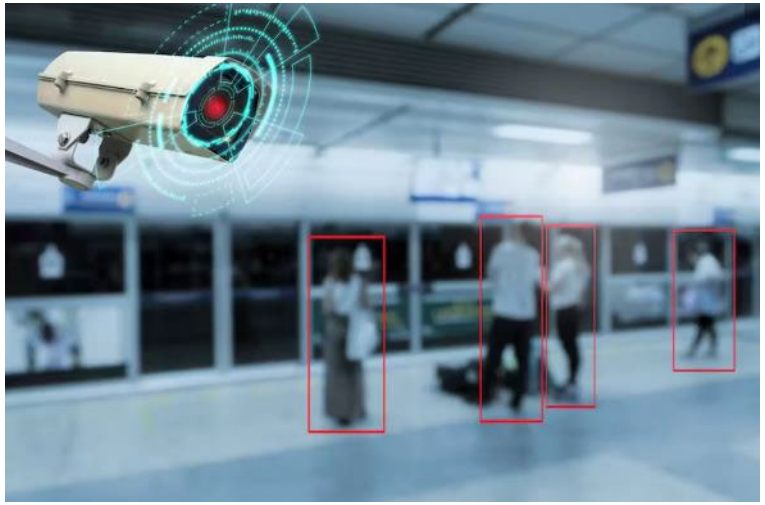
Improving Public Safety

- Analyze surveillance footage in real-time (crime prevention and public safety measures)

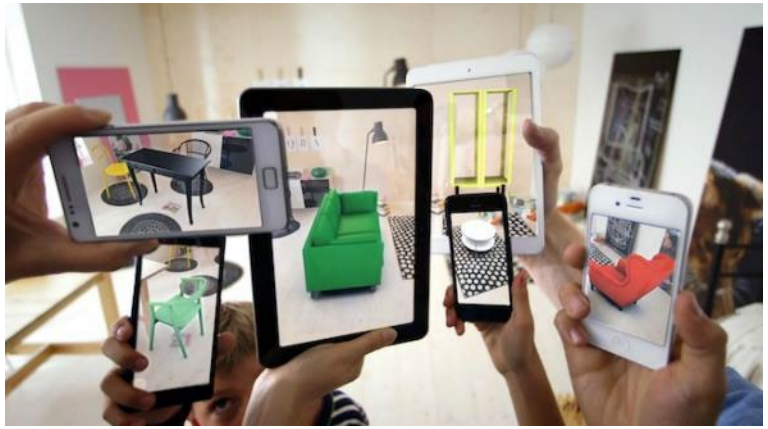
Enhancing Healthcare

- Detect diseases earlier and more accurately

AI surveillance camera (Source: istockphoto.com)



Augmented reality shopping AI (Source: amazon.com)



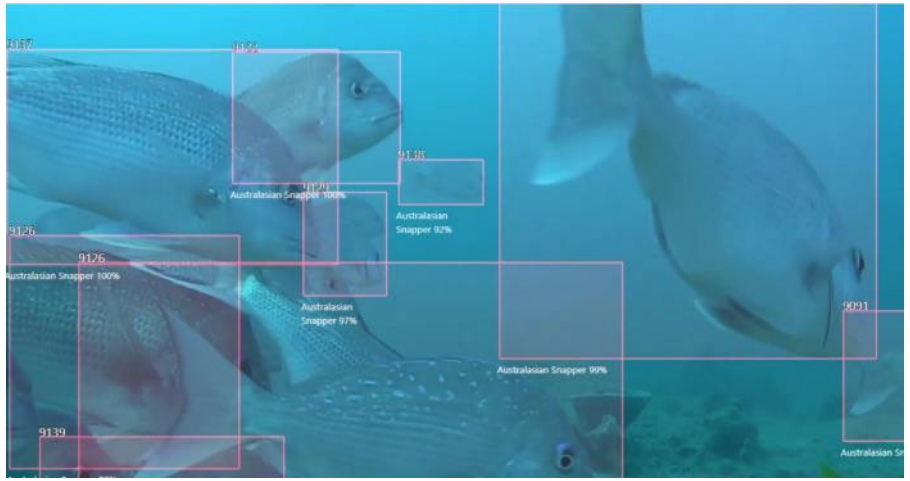
AI for Environmental Responsibility

Marine Life Monitoring

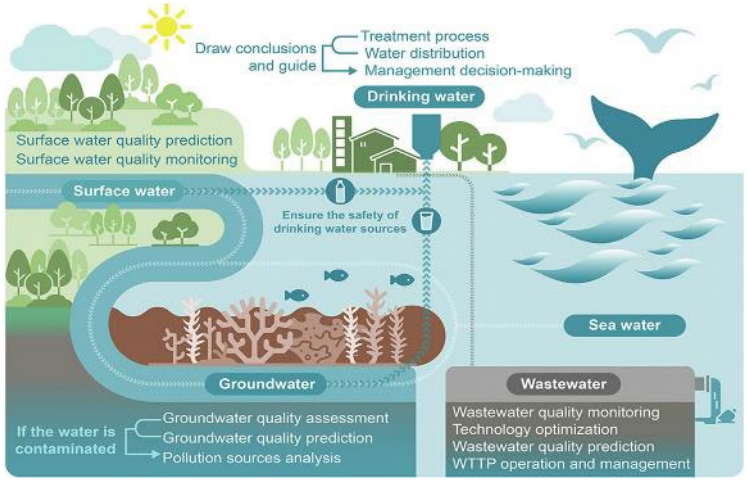
- **Biodiversity Assessment:** Analyze underwater images and videos to monitor marine biodiversity
- **Endangered Species Detection:** Pinpoint rare and endangered species, aiding in conservation efforts

Water Quality Analysis

- **Pollutant Detection:** Identify specific pollutants and potential sources, enabling targeted cleanup efforts
- **Ecosystem Health Assessment:** Provide a holistic view of the overall health and stability of underwater ecosystems



Underwater marine life monitoring [1]



Water quality analysis tools [2]

[1]. https://thefishsite.com/articles/umitron-launches-automatic-fish-measuring-system
[2]. Zhu, Mengyuan, et al. "A review of the application of machine learning in water quality evaluation." Eco-Environment & Health (2022).

Instance Segmentation for Underwater Imagery

Underwater Instance Segmentation

- **Image Segmentation :** dividing the interested part from the image background
- **Instance segmentation:** segment each instance (i.e., individual livings or objects)

Our Underwater Instance Segmentation dataset

Applications

- **Marine biological conservation:** Underwater instance segmentation helps identify the growth status of Marine life (e.g., coral reefs), and understand the interrelationships of different instances or species.

Challenge

- Lack of underwater instance datasets
- Occlusion due to gregarious habits of Marine life (like fishes, coral reefs)
- Quality degradation for general underwater images

We proposed the WaterMask: Instance Segmentation for Underwater Imagery (Accepted in ICCV2023)

Instance Segmentation for Underwater Imagery

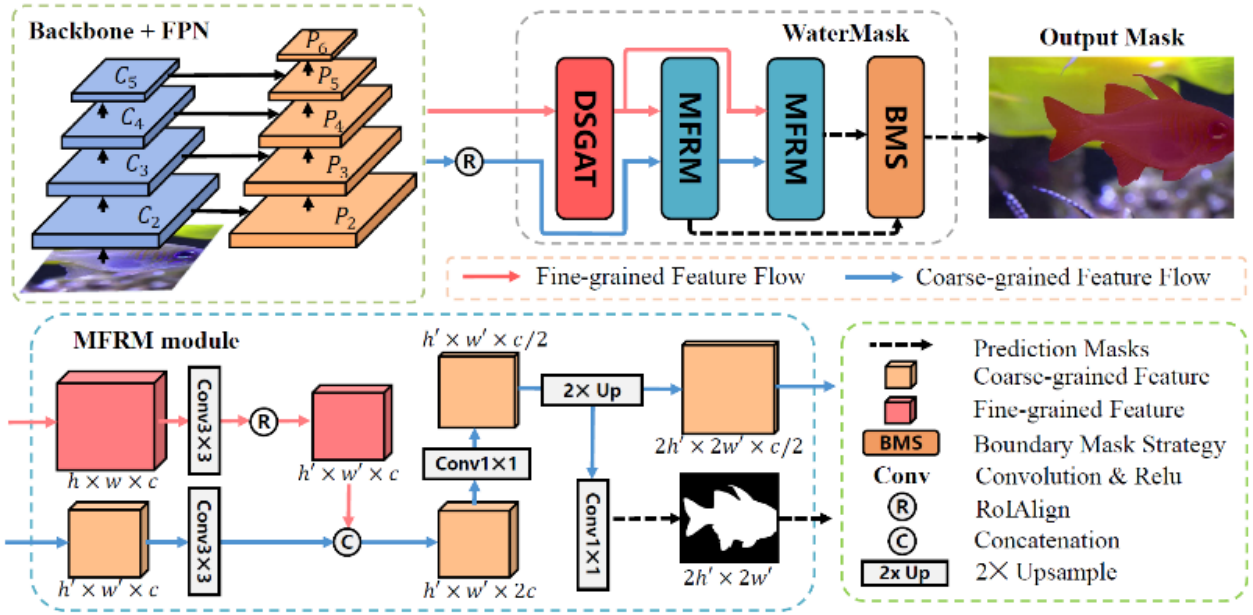
Motivations

- The quality of images often suffers **severe degradation**
- Diminishing the **visual appeal**
- Poses significant challenges for **image analysis tasks**

Objective

- Develop a method that can effectively perform **underwater instance segmentation** in such challenging environments

WaterMask



- We proposed the **first general Underwater Image Instance Segmentation dataset (UIIS)**, containing 4,628 images for 7 categories with annotations, including fish, coral reefs, aquatic plants, wrecks/ruins, etc.
- **Difference Similarity Graph Attention Module (DSGAT)**: Construct the graph and learn the degraded residual detailed features between patches according to the gregarious habits of Marine livings.
- **Multi-level Feature Refinement Module (MFRM)**: Refine multi-level segmentation by adding degradation details.
- **Boundary Mask Strategy (BMS):** Guides the network in predicting finer instance boundaries.

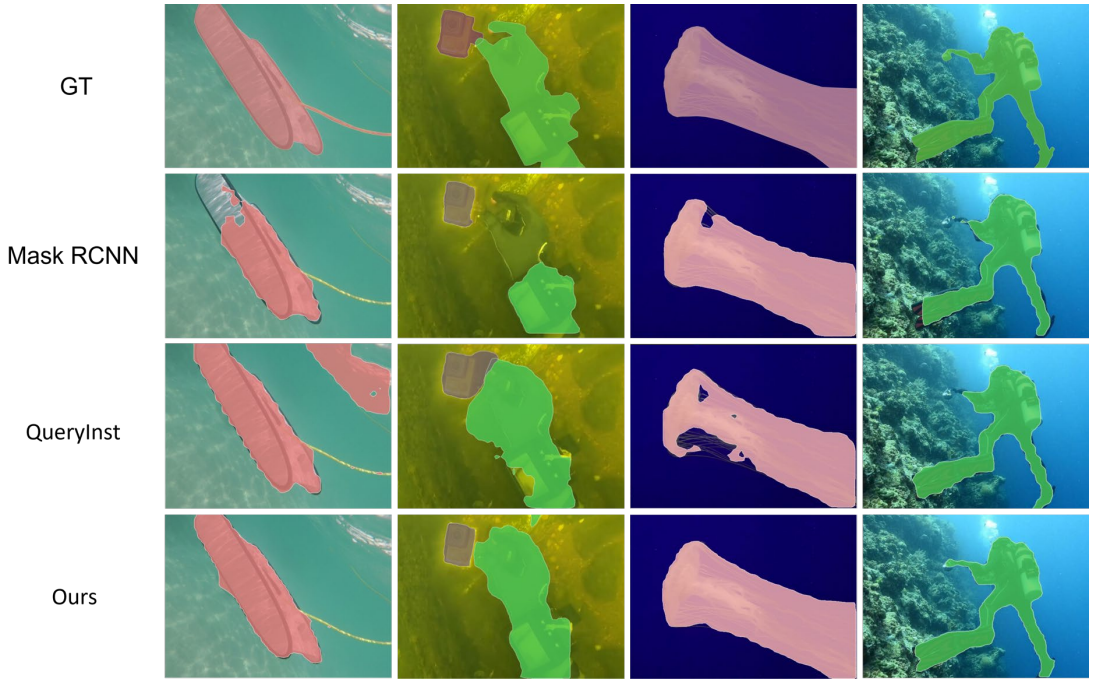
Shijie Lian, Hua Li, Runmin Cong, Suqi Li, Wei Zhang, Sam Kwong, "WaterMask: Instance Segmentation for Underwater Imagery." In Proceedings of International Conference on Computer Vision (ICCV), 2023.

Instance Segmentation for Underwater Imagery

Contributions

- WaterMask surpasses state-of-the-art methods across various metrics
- In scenarios with **high saturation and extensive degradation**, WaterMask stands out in its performance
- Identification and counting of individual marine species for **monitoring the health and diversity of marine ecosystems**

Results



Method	Backbone	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AP_f	AP_h	AP_r	Params
Mask R-CNN‡ [12]	ResNet-101	23.4	40.9	25.3	9.3	19.8	32.5	43.6	49.0	18.0	63M
Mask Scoring R-CNN‡ [13]	ResNet-101	24.6	41.9	26.5	8.4	20.0	34.3	44.2	52.8	16.0	79M
Cascade Mask R-CNN‡ [3]	ResNet-101	25.5	42.8	27.8	7.5	20.1	35.0	43.9	52.9	22.3	88M
BMask R-CNN‡ [7]	ResNet-101	22.1	36.2	24.4	5.8	17.5	35.0	40.7	50.0	17.7	66M
Point Rend [20]	ResNet-101	24.8	41.7	25.4	7.8	21.6	34.2	44.8	50.4	18.6	75M
Point Rend‡ [20]	ResNet-101	25.9	43.4	27.6	8.2	20.2	38.6	43.3	54.1	20.6	75M
R^3-CNN‡ [28]	ResNet-101	24.9	40.5	27.8	9.7	21.4	33.6	45.4	52.2	20.2	77M
SOLOv2 [29]	ResNet-101	24.5	40.9	25.1	5.6	19.4	37.6	36.4	48.3	20.6	65M
QueryInst‡ [9]	ResNet-101	26.0	42.8	27.3	8.2	21.7	35.1	43.3	54.1	20.6	191M
Mask Transfiner‡[19]	ResNet-101	24.6	42.1	26.0	7.2	19.4	36.1	43.8	26.3	19.8	63M
Mask2Former‡ [6]	ResNet-101	25.7	38.0	27.7	6.3	18.9	38.1	41.1	51.9	23.1	63M
WaterMask R-CNN	ResNet-101	25.6	41.7	27.9	8.8	21.3	36.0	45.3	53.9	19.0	67M
WaterMask R-CNN‡	ResNet-101	27.2	43.7	29.3	9.0	21.8	38.9	46.3	54.8	20.9	67M
Cascade WaterMask R-CNN‡	ResNet-101	27.1	42.9	30.4	8.3	21.0	38.9	47.0	55.8	22.5	107M

☐ Qualitative & Quantitative Comparison with the State-of-the-art Methods on UIIS dataset

Shijie Lian, Hua Li, Runmin Cong, Suqi Li, Wei Zhang, Sam Kwong, "WaterMask: Instance Segmentation for Underwater Imagery." In Proceedings of International Conference on Computer Vision (ICCV), 2023.

Image Reconstruction based on Compressive Sensing

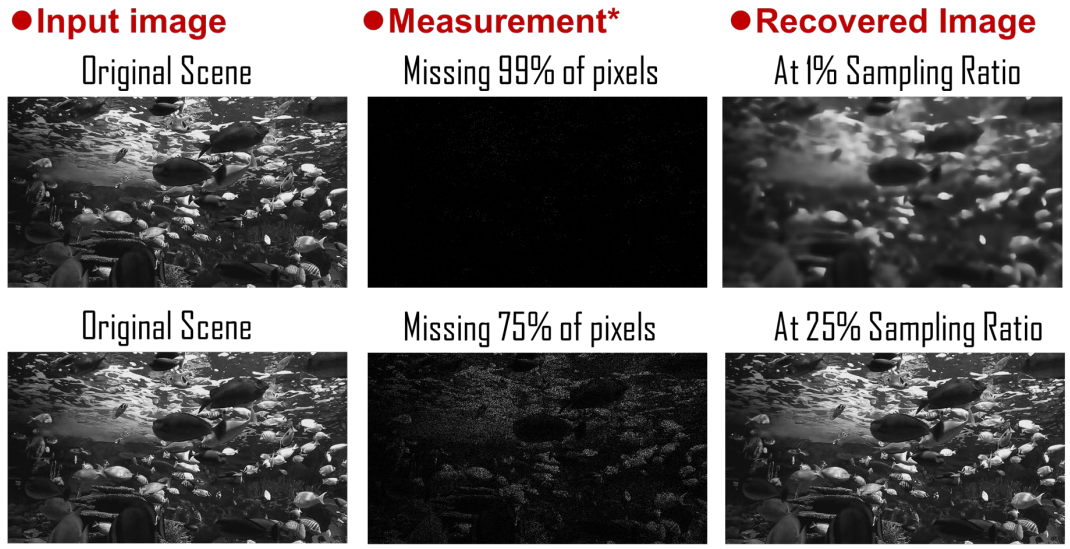
Introduction

Compressed sensing theorems [1] state that, for signals that are sparse in some domains, they can be fully reconstruct using **only few measurement**.

- Reducing sampling rate
- low-cost and efficient data compression
- Relieving data storage and transmission bandwidth burden

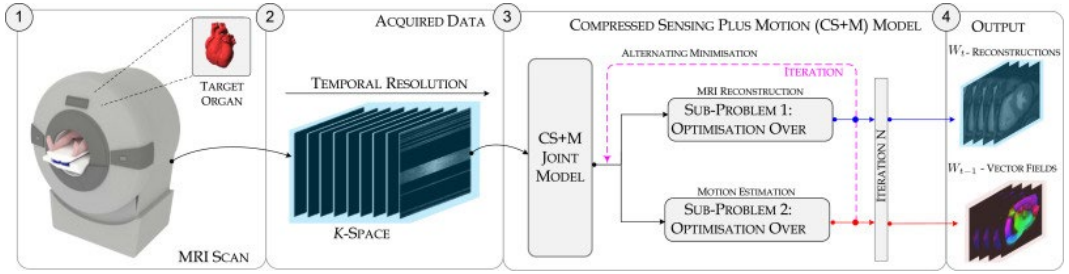
Challenge in Reconstruction

- The limitation of existing methods in capturing long-rang dependencies



* Presented in the form of images for ease of understanding, but actually numerical values.

Applications



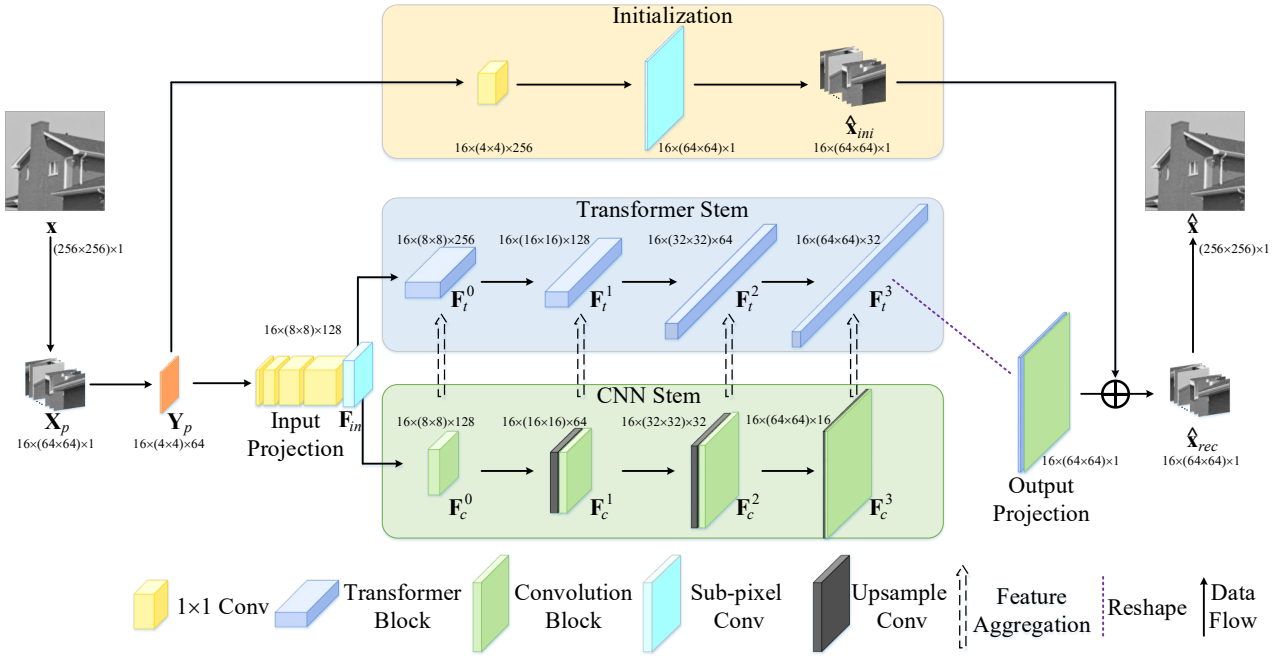
☐ Medical Imaging [2]

We proposed a Compressive Sensing Framework based on CNN and Transformer (Accepted in TIP 2023)

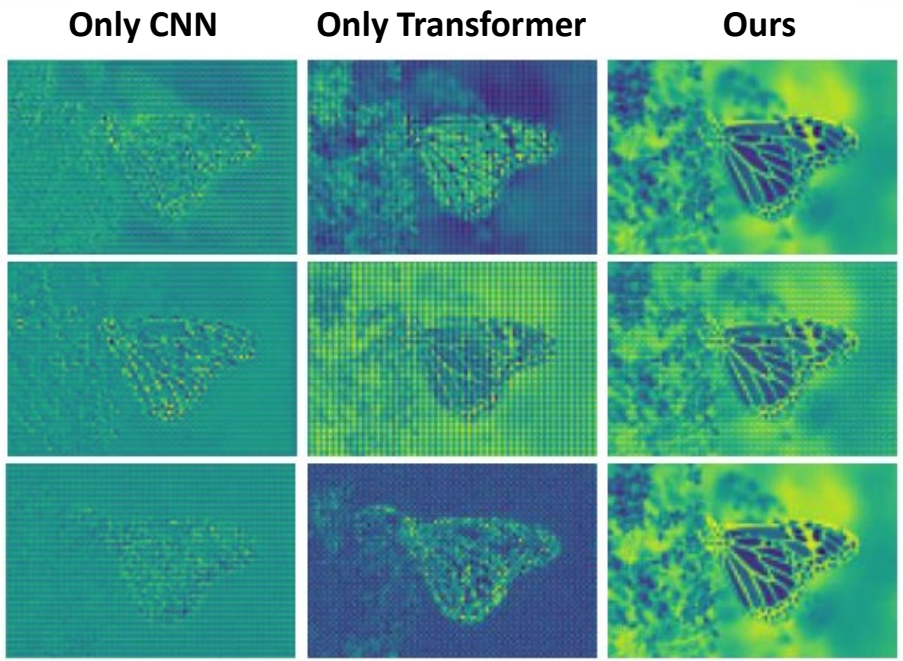
[1] D. L. Donoho, "Compressed sensing," IEEE Transactions on Information Theory , vol. 52, no. 4, pp. 1289–1306, 2006.
[2] Aviles-Rivero, et al.. "Compressed sensing plus motion (CS+ M): a new perspective for improving undersampled MR image reconstruction." *Medical Image Analysis, vol.* 68 , pp. 101933, 2023.

Image Reconstruction based on Compressive Sensing

Framework



Feature Visualization



- Coupling transformer with CNN for adaptive sampling and reconstruction
- Inheriting local features from CNN and global representations from the transformer

Dongjie Ye, Zhangkai Ni, Hanli Wang, Jian Zhang, Shiqi Wang, Sam Kwong, "CSformer : Bridging Convolution and Transformer for Compressive Sensing," IEEE Transactions on Image Processing (TIP), vol. 32, pp.2827-2842, 2023.

Image Reconstruction based on Compressive Sensing

Lingnan University 嶺南大學 香港 Hong Kong

Experimental Results

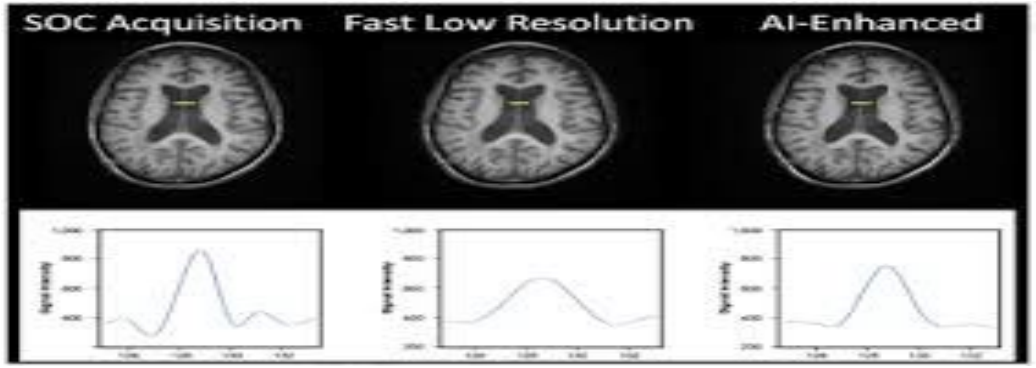
Dataset	CS ratio	CSNet+	DPA-Net	OPINE-Net	AMP-Net	Our Method
Set11	1%	21.03/0.5566	18.05/0.5011	20.15/0.5340	20.57/0.5639	21.95/0.6241
	4%	-	23.50/0.7205	25.69/0.7920	25.26/0.7722	26.93/0.8251
	10%	28.37/0.8580	26.99/0.8354	29.81/0.8884	29.45/0.8787	30.66/0.9027
	25%	-	31.74/0.9238	34.86/0.9509	34.63/0.9481	35.46/0.9570
	50%	38.52/0.9749	36.73/0.9670	40.17/0.9797	40.34/0.9807	41.04/0.9831
BSD68	1%	22.36/0.5273	18.98/0.4643	22.11/0.5140	22.28/0.5387	23.07/0.5591
	4%	-	23.27/0.6096	25.20/0.6825	25.26/0.6760	25.91/0.7045
	10%	27.18/0.7766	25.57/0.7267	27.82/0.8045	27.86/0.7926	28.28/0.8078
	25%	-	29.68/0.8763	31.51/0.9061	31.74/0.9048	31.91/0.9102
	50%	35.42/0.9614	32.89/0.9373	36.35/0.9660	36.82/0.9680	37.16/0.9714
Urban100	1%	20.75/0.5273	16.36/0.4162	19.82/0.5006	20.90/0.5328	21.94/0.5885
	4%	-	21.64/0.6498	23.36/0.7114	24.15/0.7029	26.13/0.7803
	10%	26.52/0.8053	24.54/0.7851	26.93/0.8397	27.38/0.8270	29.61/0.8762
	25%	-	28.81/0.8951	31.86/0.9308	32.19/0.9258	34.16/0.9470
	50%	35.25/0.9621	32.09/0.9454	37.23/0.9747	37.51/0.9734	39.46/0.9811
Set5	1%	24.18/0.6478	19.02/0.5133	21.89/0.6101	23.48/0.6518	25.22/0.7197
	4%	-	26.63/0.7767	27.95/0.8209	29.01/0.8359	30.31/0.8686
	10%	32.59/0.9062	30.32/0.8713	32.51/0.9058	33.42/0.9140	34.20/0.9262
	25%	-	33.96/0.9360	36.78/0.9510	38.03/0.9586	38.30/0.9619
	50%	41.79/0.9803	39.57/0.9716	41.62/0.9779	42.72/0.9818	43.55/0.9845
Set14	1%	22.92/0.5630	18.30/0.4616	21.36/0.5340	22.79/0.5751	23.88/0.6146
	4%	-	23.69/0.6534	25.50/0.6974	26.67/0.7219	27.78/0.7581
	10%	29.13/0.8169	26.28/0.7693	28.77/0.8129	29.92/0.8312	30.85/0.8515
	25%	-	30.15/0.8813	33.12/0.9102	34.31/0.9213	35.04/0.9316
	50%	37.89/0.9631	33.78/0.9440	38.09/0.9621	39.28/0.9684	40.41/0.9730
Direct Average	1%	22.25/0.5630	18.14/0.4713	21.07/0.5386	22.00/0.5725	23.21/0.6212
	4%	-	23.75/0.6820	25.54/0.7408	26.07/0.7418	27.41/0.7873
	10%	28.76/0.8326	26.94/0.7976	29.17/0.8503	29.61/0.8487	30.72/0.8729
	25%	-	30.87/0.9025	33.63/0.9298	34.18/0.9317	34.97/0.9415
	50%	37.77/0.9684	35.01/0.9531	38.69/0.9721	39.33/0.9745	40.32/0.9786
Weighted Average	1%	21.56/0.5310	17.56/0.4431	20.79/0.5122	21.55/0.5426	22.55/0.5855
	4%	-	22.57/0.6434	24.39/0.7077	24.89/0.7022	26.32/0.7574
	10%	27.19/0.8017	25.80/0.7689	27.67/0.8301	27.99/0.8206	29.42/0.8537
	25%	-	29.50/0.8903	32.12/0.9225	32.47/0.9203	33.63/0.9342
	50%	35.84/0.9631	32.93/0.9444	37.26/0.9712	37.69/0.9718	38.93/0.9774

Dongjie Ye, Zhangkai Ni, Hanli Wang, Jian Zhang, Shiqi Wang, Sam Kwong, "CSformer : Bridging Convolution and Transformer for Compressive Sensing," IEEE Transactions on Image Processing (TIP), vol. 32, pp.2827-2842, 2023.

AI for Social Responsibility: Healthcare

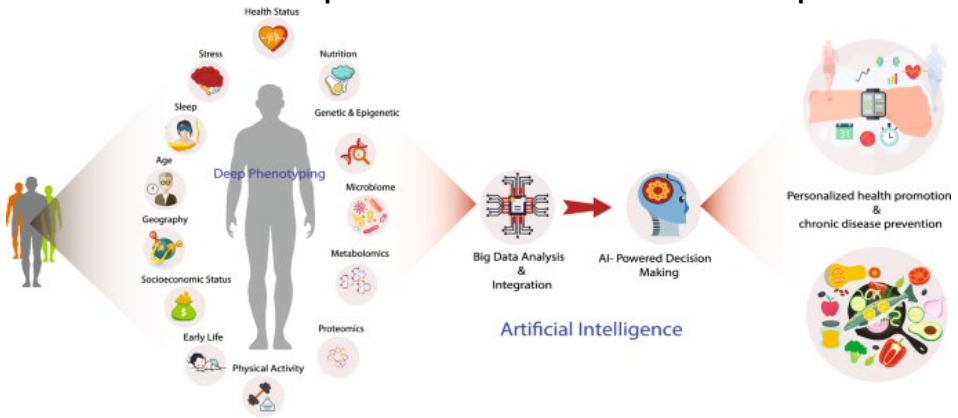
Medical Imaging [1]

- AI enhances interpretation of scans for improved accuracy



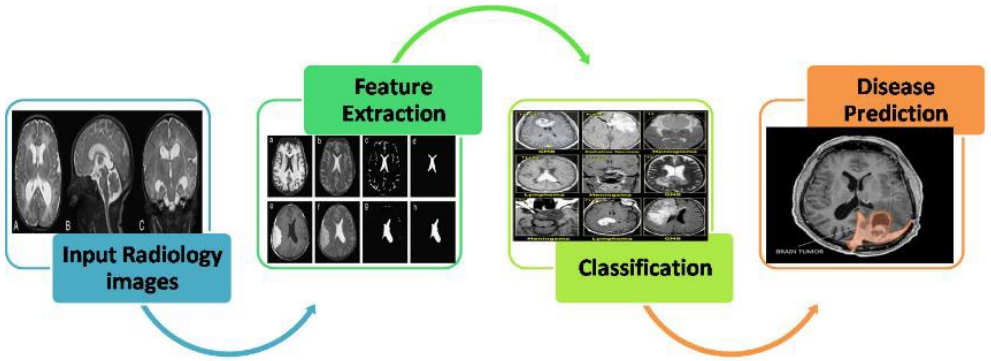
Personalized Treatment [2]

- AI tailors treatment plans based on individual patient data



Disease Diagnosis [3]

- AI analyzes medical images for accurate and swift diagnosis



Remote Monitoring [4]

- AI-powered wearables provide continuous health tracking

[1]. J.D. Rudie, et al. "Clinical assessment of deep learning–based super-resolution for 3D volumetric brain MRI". Radiology: Artificial Intelligence, vol. 4, no.2, pp. e210059, 2022.
[2] M. Subramanian, et al., "Precision medicine in the era of artificial intelligence: implications in chronic disease management". Journal of Translational Medicine, vol. 18, no.1, pp. 472, 2020.
[3] https://www.aitimejournal.com/dominance-of-ai-in-healthcare/17749
[4] https://mobisoftinfotech.com/resources/blog/wearable-technology-in-healthcare

Low-Light Image Enhancement with External Memory

Target

Before enhancement

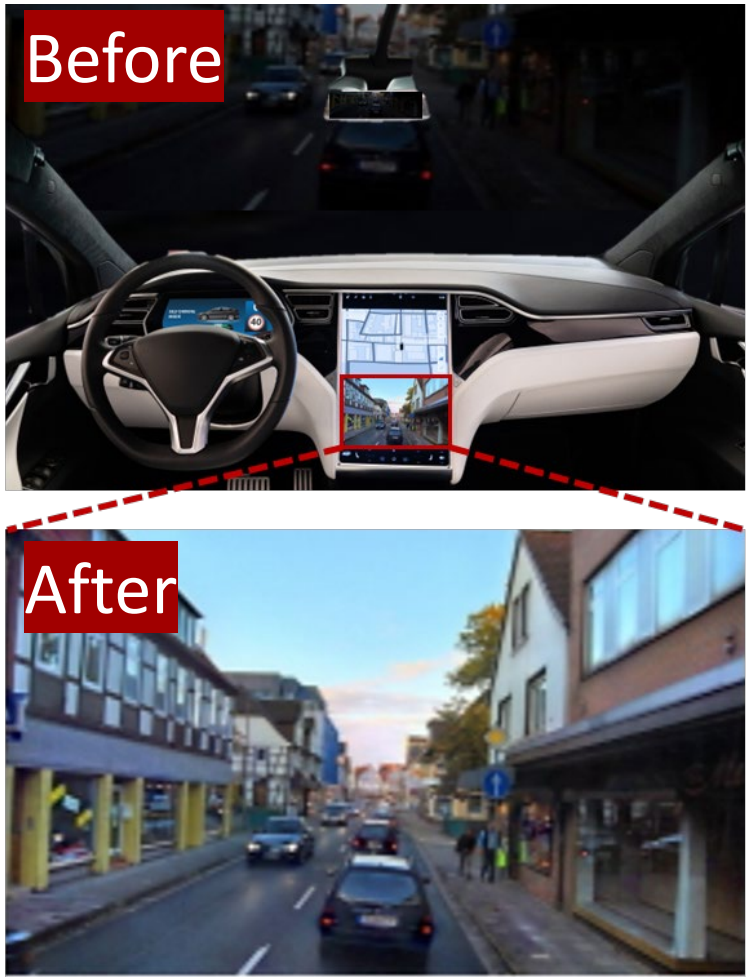
Enhancement result

Challenge

- Ineffective enhancement in the testing phase for the samples that differ significantly from the training distribution

We proposed the Low-light Image Enhancement framework based on the External Memory (Accepted in TMM 2023)

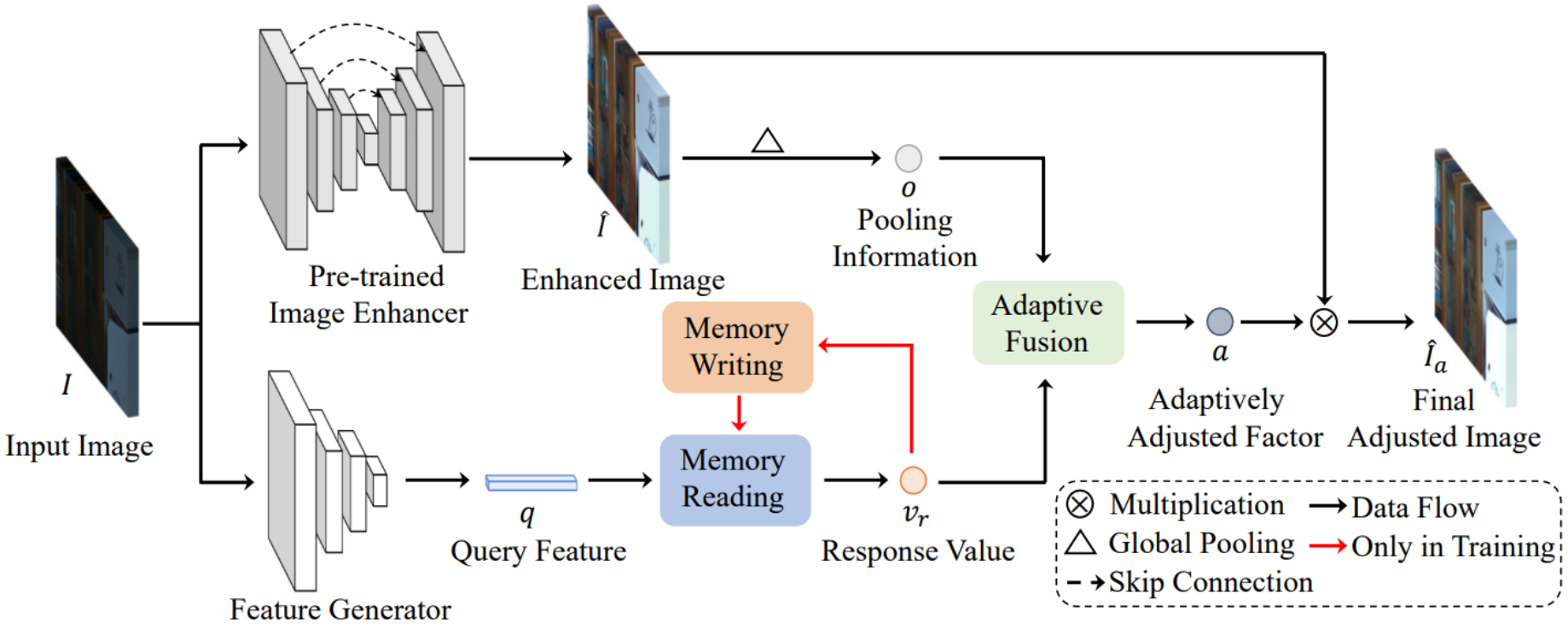
Application



Driving Condition Enhancement

Low-Light Image Enhancement with External Memory

Framework



- During the **training cycle**, the response value v_r is fed into the memory writing module to **learn the correct memory value**
- During the **testing cycle**, v_r and the pooling information o are taken as input to generate the adaptively adjusted factor a in the adaptive fusion module. Finally, we can get the **final adjusted image** I_a by tuning $\hat{I}$ with **adaptively adjusted factor** a

Dongjie Ye, Zhangkai Ni, Wenhan Yang, Hanli Wang, Shiqi Wang, Sam Kwong, "Glow in the Dark : Low-Light Image Enhancement with External Memory," IEEE Transactions on Multimedia (TMM), Early Access, 2023.

Low-Light Image Enhancement with External Memory

Low-Light Image Enhancement with External Memory

Experimental Results

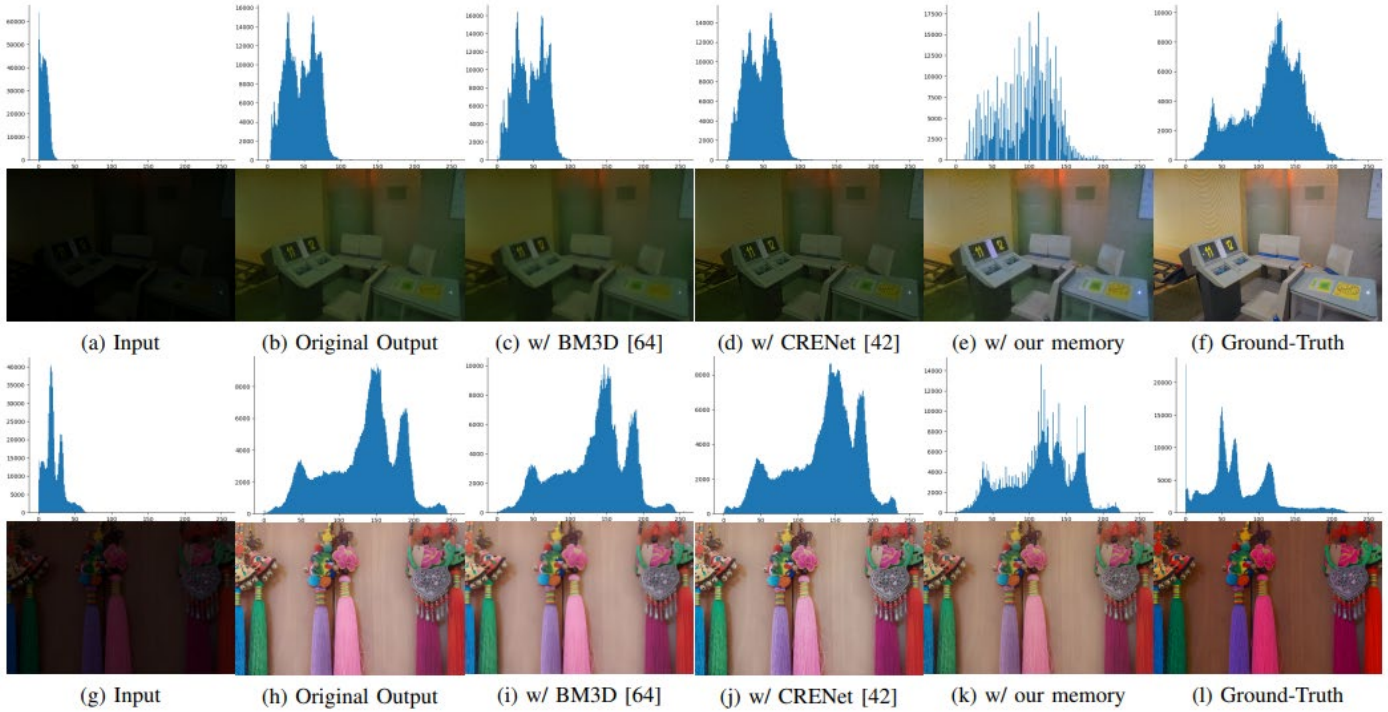


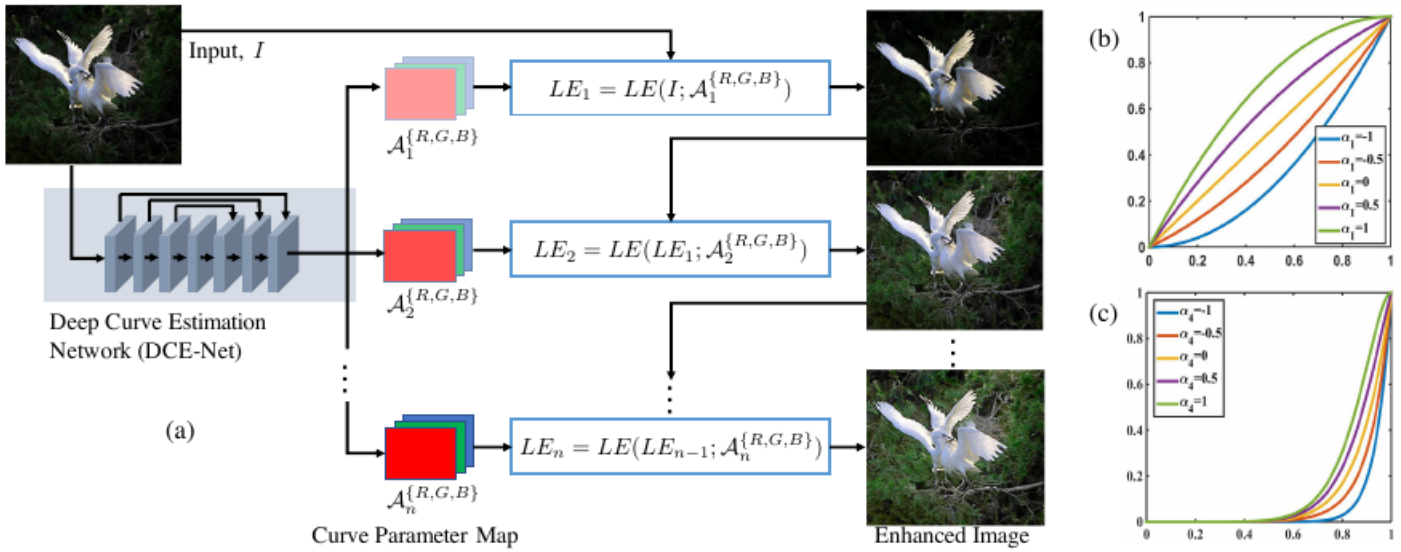
TABLE I
QUANTITATIVE COMPARISONS BETWEEN THE PROPOSED EMNET AND OTHER METHODS ON REAL TESTING IMAGES OF LOL-V1 DATASET [5]. THE BEST ONE IS SHOWN IN RED AND THE SECOND BEST IN BLUE

Method	PSNR↑	SSIM↑	LPIPS↓	FSIM↑	VIF↑
Dong [11]	16.72	0.4784	0.3770	0.8798	0.4304
DHECI [8]	16.39	0.4056	0.4441	0.8371	0.4603
CRM [63]	17.20	0.6230	0.3157	0.9362	0.4923
EFF [2]	13.88	0.5950	0.3191	0.9172	0.4650
LIME [3]	16.05	0.4860	0.3775	0.8532	0.4702
JED [12]	13.69	0.6510	0.2797	0.8729	0.3815
MemNet [57]	19.39	0.7679	0.2171	0.9283	0.4483
DRN [5]	16.77	0.4250	0.4649	0.8491	0.3722
KinD++ [31]	18.45	0.7785	0.1765	0.8973	0.5019
LLFlow [32]	21.15	0.8523	0.1194	0.9608	0.5382
Uformer [46]	19.61	0.7554	0.2400	0.9268	0.4947
IMLUT [58]	21.70	0.6914	0.3108	0.9447	0.5130
URN [4]	20.45	0.8218	0.1244	0.9538	0.5168
DCC [23]	22.98	0.8475	0.1377	0.9588	0.5279
SNRA [24]	24.61	0.8401	0.1502	0.9581	0.4884
EMNet (Ours)	25.37	0.8686	0.0860	0.9676	0.5599

Dongjie Ye, Zhangkai Ni, Wenhan Yang, Hanli Wang, Shiqi Wang, Sam Kwong, "Glow in the Dark : Low-Light Image Enhancement with External Memory," IEEE Transactions on Multimedia (TMM), Early Access, 2023.

Zero-reference Low-Light Image Enhancement

Pipline



Highlights

- We propose the first low-light enhancement network that is **independent of paired and unpaired training data.**
- We design an image-specific curve that is able to **approximate pixel-wise and higher-order curves by iteratively applying itself**.

Results

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. "Zero-reference deep curve estimation for low-light image enhancement." In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1780-1789. 2020.

AI for Social Responsibility: Smart Transportation

Traffic Optimization

- AI analyzes real-time traffic data to predict congestion, optimizing signal timing for smoother flow

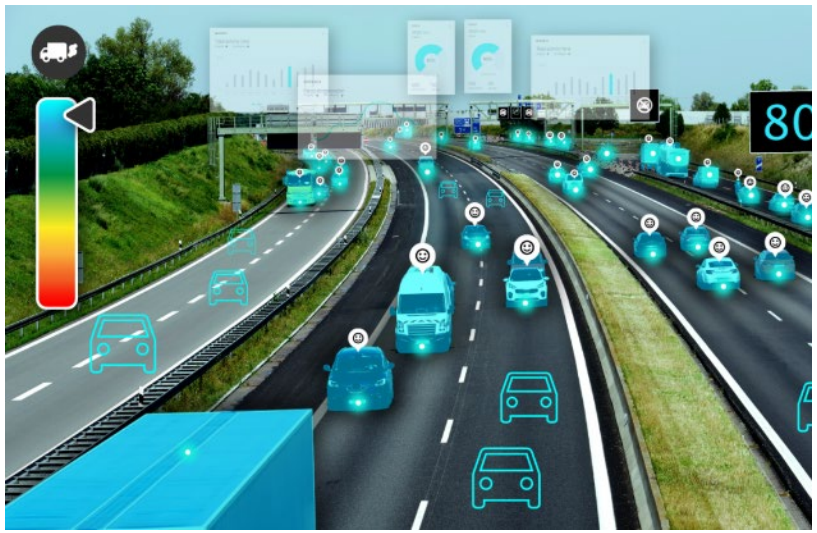
Autonomous Driving

- AI-driven sensors enable safer and self-driving vehicles, reducing accidents and enhancing road safety

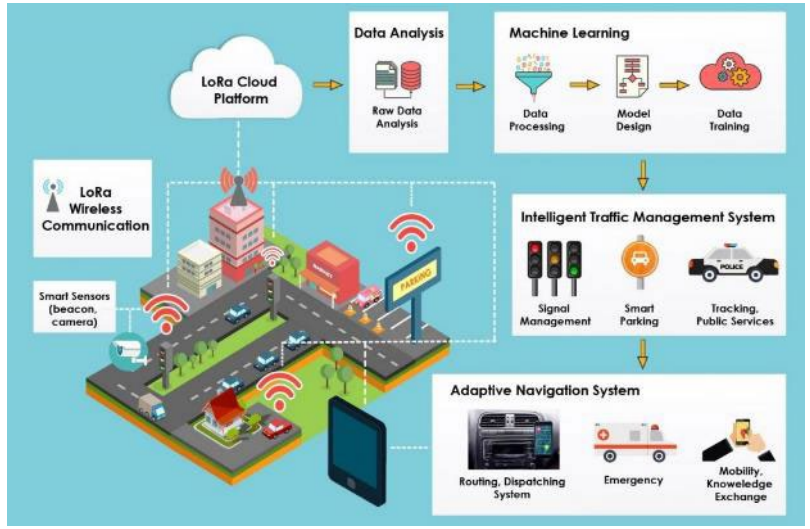
Accident Prevention

- AI detects risky behavior, issuing warnings and predictions to prevent accidents

Autonomous driving



Road monitoring [1]



Smart traffic management [2]

[1] https://www.edi.gmbh/en-solutions/ai-based-control-of-traffic-monitoring-systems
[2] https://cmte.ieee.org/futuredirections/tech-policy-ethics/2018articles/smart-town-traffic-management-system-using-lora-and-machine-learning-mechanism

IP Video – Cisco 2020 Report

- IP Video Globally
- IP video traffic will grow 3-fold from 2015 to 2020, a compound annual growth rate of 26%.
- IP video traffic will reach 158.9 Exabytes per month in 2020, up from 51.0 Exabytes per mo nth in 2015.(1 Exabytes = 1073741824 Gigabytes = 1 e+9 Gigabytes))
- IP video will be 82% of all IP traffic in 2020, up from 70% in 2015.
- Ultra High Definition HD will be 16.4% of IP Video traffic in 2020, up from 2.0% in 2015
- HD will be 63.1% of IP Video traffic in 2020, up from 51.2% in 2015 (30.9% CAGR).
- Standard Video SD will be 20.6% of IP Video traffic in 2020, compared to 46.7% in 2015
- consumer IP video traffic will be 85% of consumer IP traffic in 2020, up from 77% in 2015.
- Business IP video traffic will be 64% of business IP traffic in 2020, up from 42% in 2015

Cisco report: https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2020_Forecast_Highlights.pdf

Why do we need video compression?

- Video: A group of pictures
  - 1920x1080, 24 bit (RGB), size: 5.9MB

1920x1080x3=5.9MB

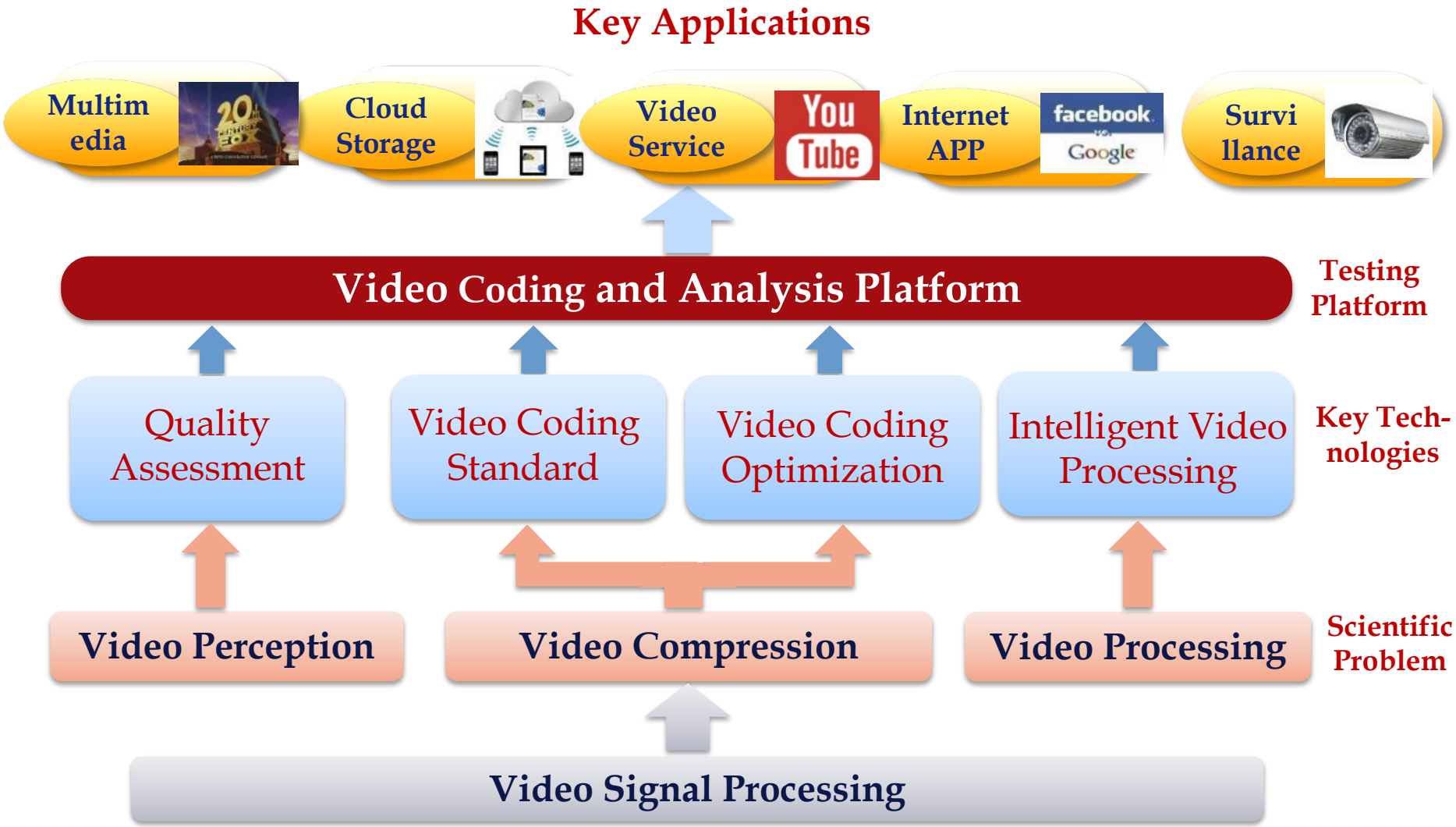
30fps

5.9 MB x 30 fps = 177 MB/s → **Transmission issue!**

90 min

177 MB/s x (90 x 60) s ≈ 933 GB → **Storage issue!**

Video compression and applications

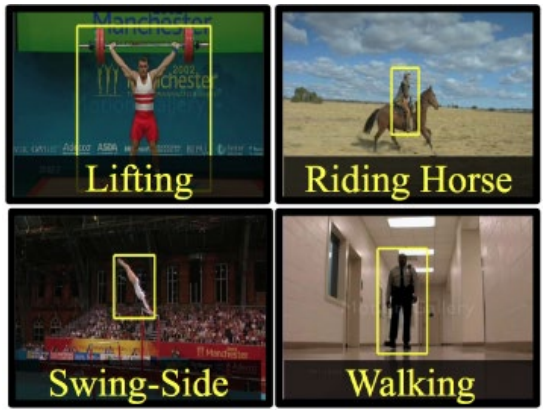


Visual Saliency Detection for Stereoscopic Video

Prof. Sam Kwong

Qiudan Zhang, Xu Wang, Shiqi Wang, Shikai Li, Sam Kwong, Jianmin Jiang, Learning to Explore Intrinsic Saliency for Stereoscopic Video, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR19)*, Long Beach, CA, Jun. 16-20, 2019.

AI Based Video Saliency Prediction



Action recognition

Robotic navigation

Autonomous driving

Crowd Analysis

AI Based Video Saliency Prediction

Proposed Deep Learning based Saliency Detection Model

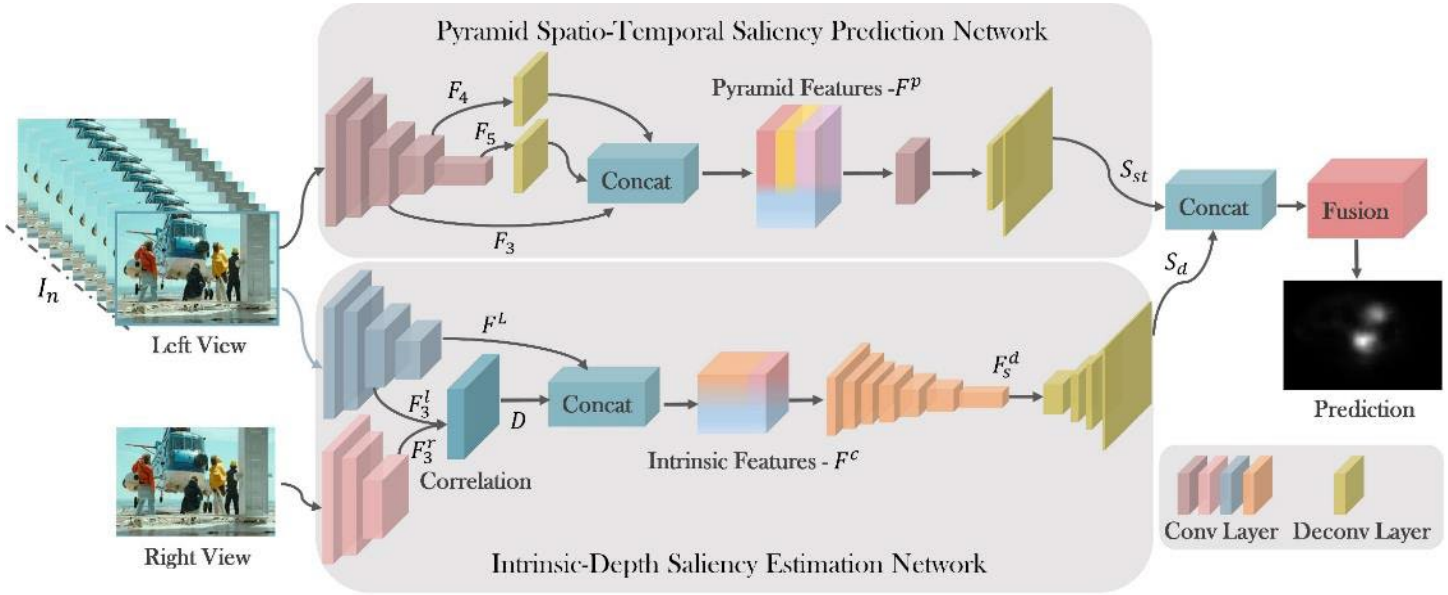


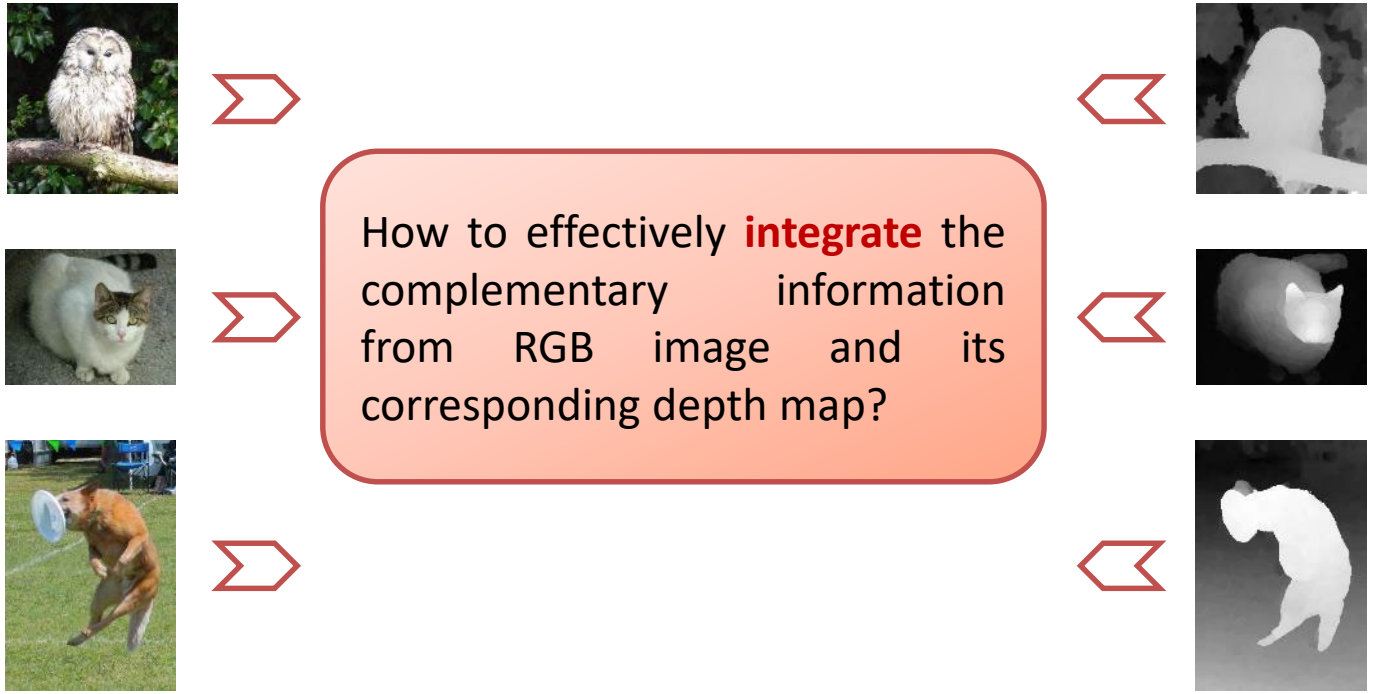
Figure 2. The overall architecture of the proposed stereoscopic video saliency prediction model.

ASIF-Net: Attention Steered Interweave Fusion Network for RGB-D Salient Object Detection

Chongyi Li, Runmin Cong, Sam Kwong, Junhui Hou,
Huazhu Fu, Guopu Zhu, Dingwen Zhang, and Qingming Huang

IEEE Transactions on Cybernetics, vol. 50, no. 1, pp. 88-100, 2021

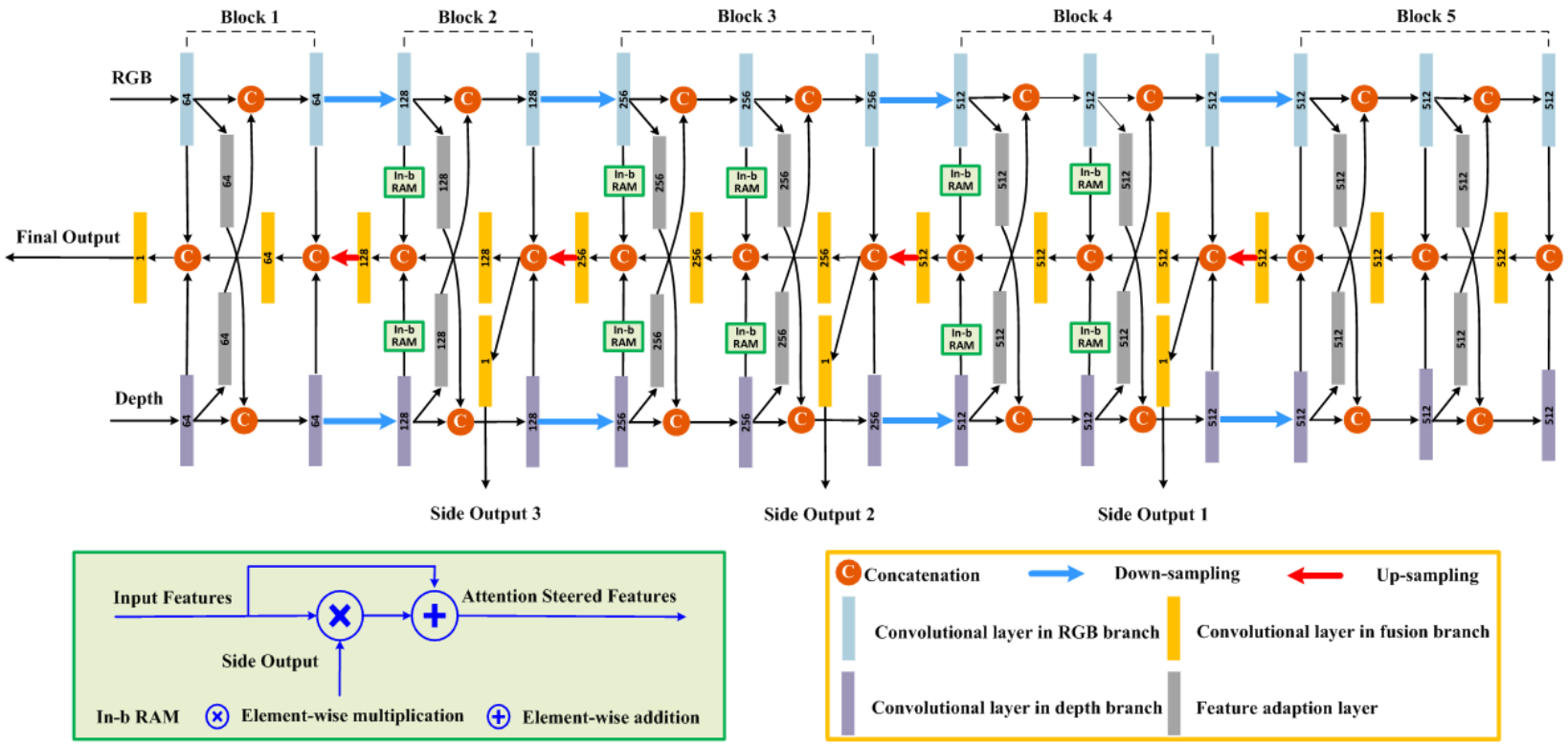
Challenges


How to effectively integrate the complementary information from RGB image and its corresponding depth map?

Contributions

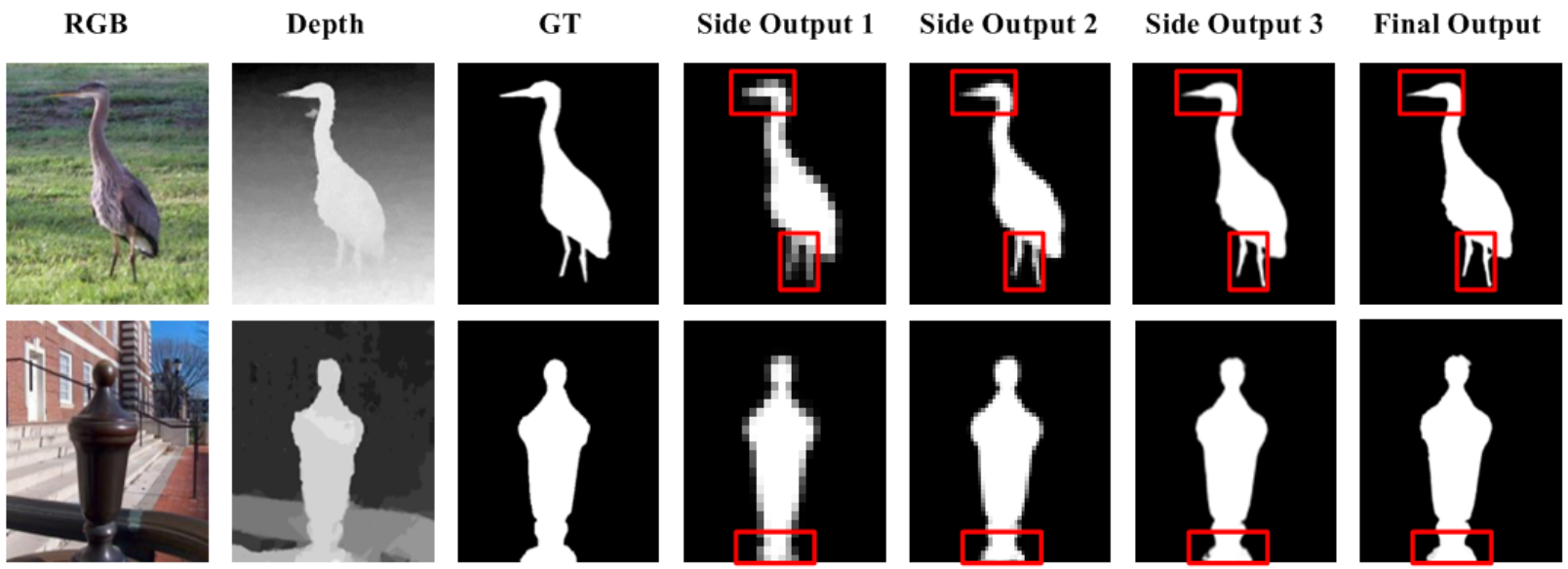
1. We propose an attention steered interweave fusion network architecture (ASIF-Net), which progressively and interactively captures complementarity from RGB-D images via the interweave fusion and weights the saliency regions by the steering of deeply supervised attention mechanism.
2. We design an in-block residual attention module to avoid the blur induced by upsampling interpolation, and introduce a global semantic constraint to our network by using adversarial learning.
3. Without the use of any pre-processing and post-processing strategies, the proposed network is trained from scratch and achieves the competitive performance against 15 state-of-the-art saliency detectors on four benchmark datasets.

Network


Block 1
Block 2
Block 3
Block 4
Block 5
RGB
Depth
Final Output
In-b RAM
Side Output 3
Side Output 2
Side Output 1
Input Features
Attention Steered Features
Side Output
In-b RAM
Element-wise multiplication
Element-wise addition
Concatenation
Down-sampling
Up-sampling
Convolutional layer in RGB branch
Convolutional layer in fusion branch
Convolutional layer in depth branch
Feature adaption layer

Network

- **Side Output**

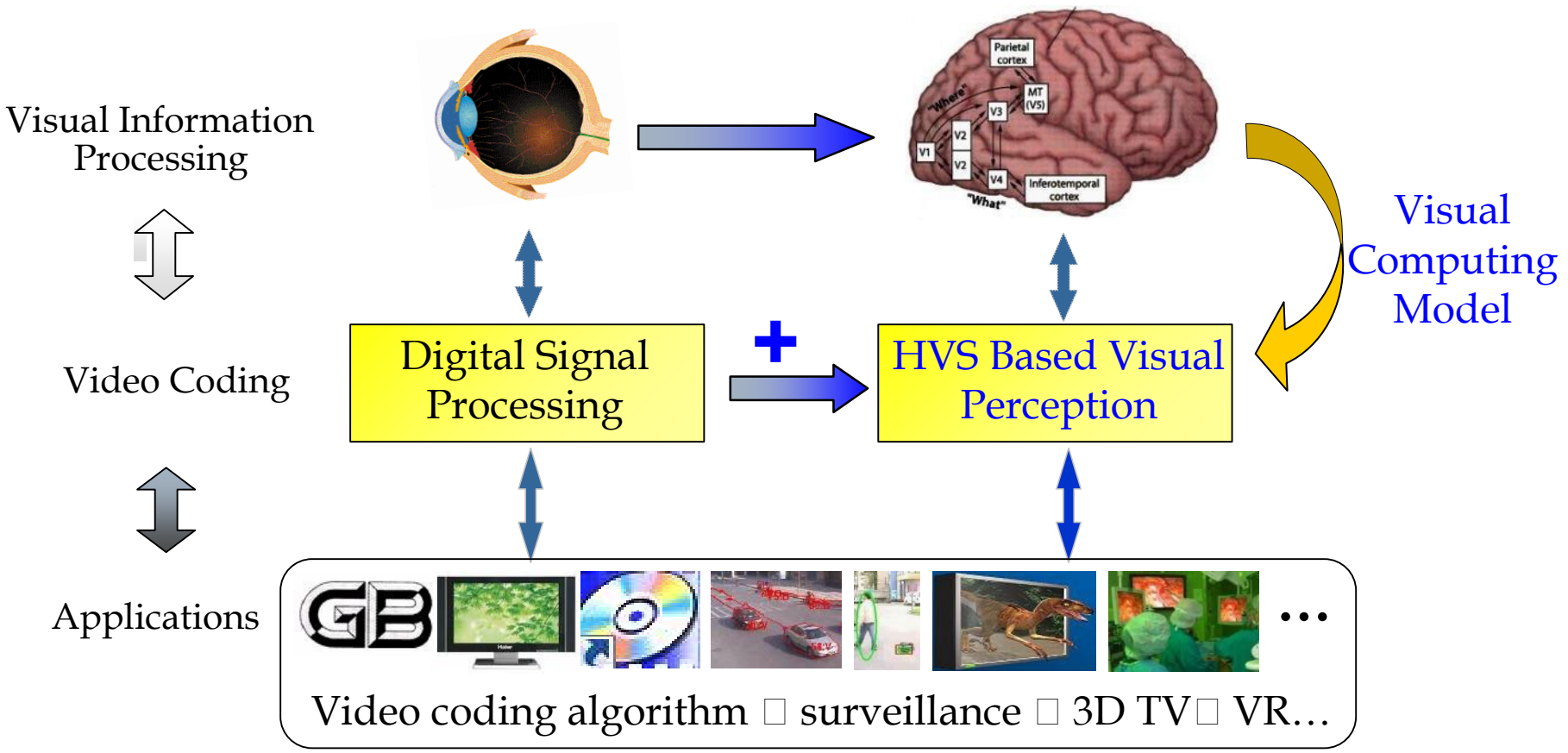


Experiments

Visual results:

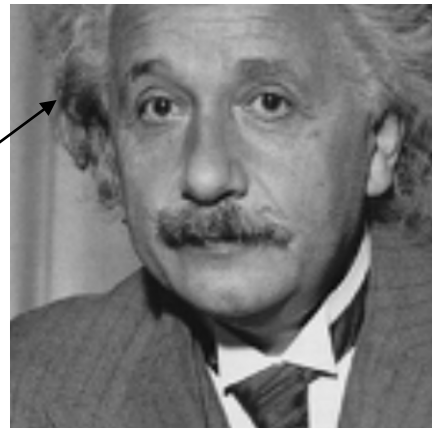
The proposed method achieves superior visual performance with highlighted foreground, complete structure, sharp boundaries, and clean background.

Perceptual Visual Processing

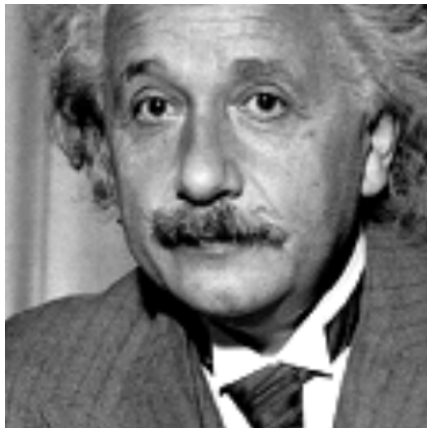


Perceptual Visual Processing

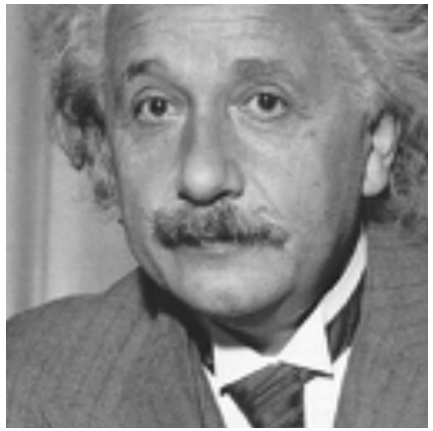
Original Image



MSE=0, MSSIM=1

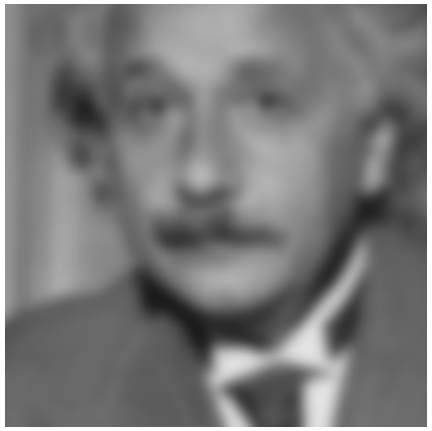


MSE=144, MSSIM=0.913

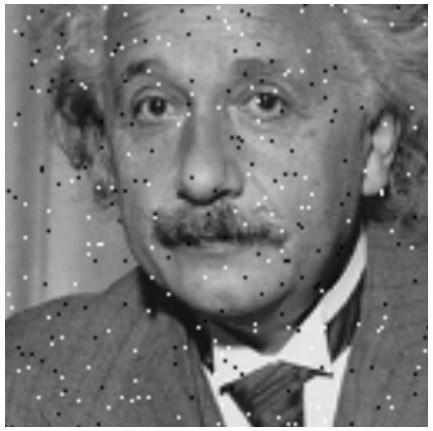


MSE=144, MSSIM=0.988

MSE=142, MSSIM=0.662



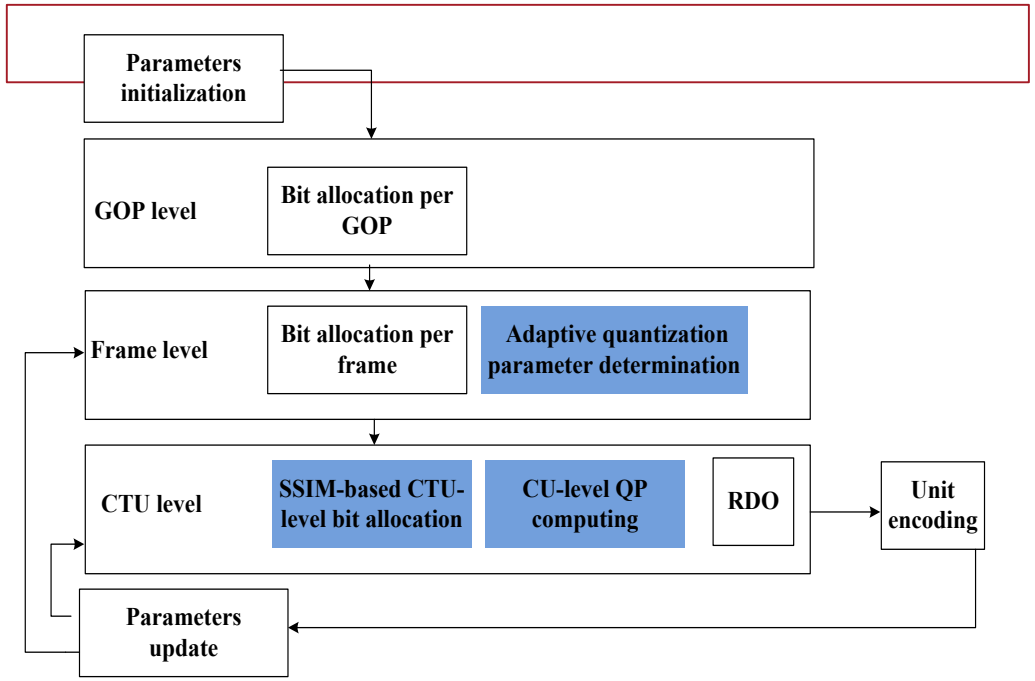
MSE=144, MSSIM=0.694



MSE=144, MSSIM=0.840

Z. Wang and A. C. Bovik, " Modern Image Quality Assessment, in Syntheses Lectures on Image, Video and Multimedia Processing", *Morgan & Claypool Publishers*, Mar. 2006.

Perceptual Rate Control for HEVC



Framework of the perceptual rate control

Frame level QP is clipped within the range :

$$[QP_{L0} - |\Delta Q_{Li}|,\ QP_{L0} + |\Delta Q_{Li}|]$$

Optimal bit allocation using SSIM as the quality metric:

$$D'(R) = \ln(c \times R^{-k})/f$$

GOP level bit allocation :

$$T_G = N_G * \left(\frac{\frac{R_{target}}{f}(N_c + S_\omega) - T_c}{S_\omega}\right)$$

Frame level bit allocation

$$T_{Pic_p} = \frac{T_G - \bar{T}_G}{\sum_{j>p} \omega_j} * \omega_p$$

ΔQP determination:

$$\Delta Q_{Li} = 3\log_2\left(\frac{\lambda_{Li}}{\lambda_{L0}} \cdot \frac{q_{factor,L0}}{q_{factor,Li}}\right)$$

Thanks for Your Attention!