周研究室
知能メディア処理研究室
卒業研究の研究例

周　金佳
https://www.zhou-lab.info/

法政大学 HOSEI University
iMedia Lab.

周研の研究例

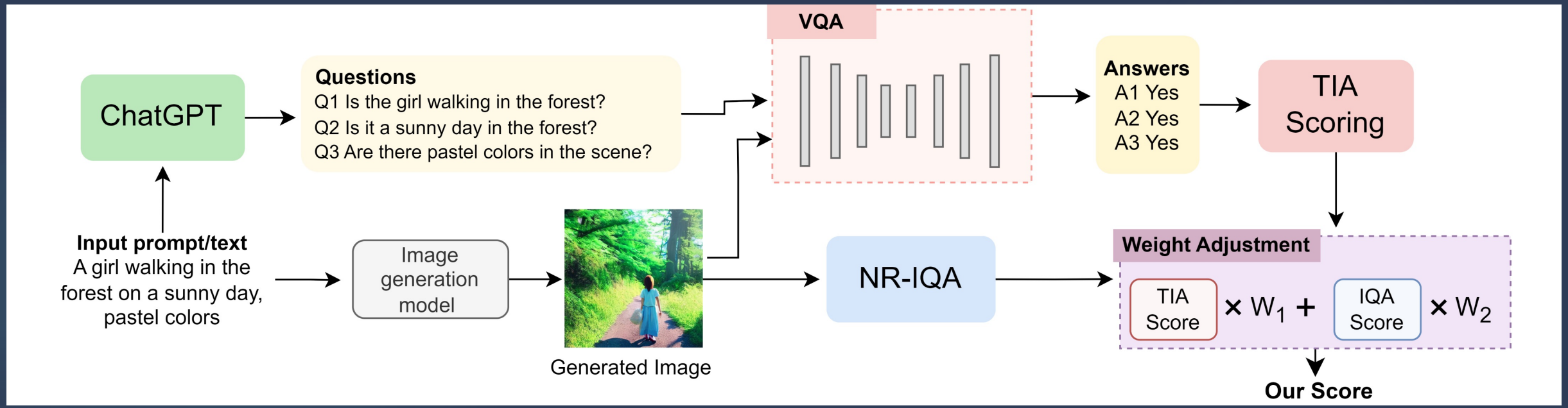
- テキストを用いた画像編集
- 画像説明文の自動生成
- 画像修復
- 動画圧縮

テキストから画像生成

Example images and corresponding captions of common Text-to-image(T2I) datasets

S. Frolov, "Adversarial text-to-image synthesis: A review", Neural Networks 2021

周研の研究例（2023年度卒業研究）：文章からの質問生成に対する画像質問応答ベースの評価指標
Visual question answering based evaluation metrics for text-to-image generation



- Visual Question Answering(VQA)モデル[8]を使用して、生成画像と Input text との細かな一致を評価
- Non-Reference Image Quality Assessment(NR-IQA) を使用することで画質の評価も同時に行う
- 最終スコアは画像テキスト間の関連性評価と画質の評価の比重を任意に調整可能

法政大学 HOSEI University
iMedia Lab.

従来法と比較した結果

Input Text	A room with a orange couch, white table set with dinnerware and a television.		A room with a orange couch, **brown** table set with dinnerware and a television.		A boy swinging a blue baseball bat at a game.		A boy swinging a **green** baseball bat at a game.	
								
GT(TIA)	1		2		1		2	
	Rank	Score	Rank	Score	Rank	Score	Rank	Score
CLIPScore[7]	2	0.598	1	0.599	2	0.759	1	0.812
Ours(TIA only)	**1**	**1.00**	**2**	**0.670**	**1**	**1.00**	**2**	**0.670**

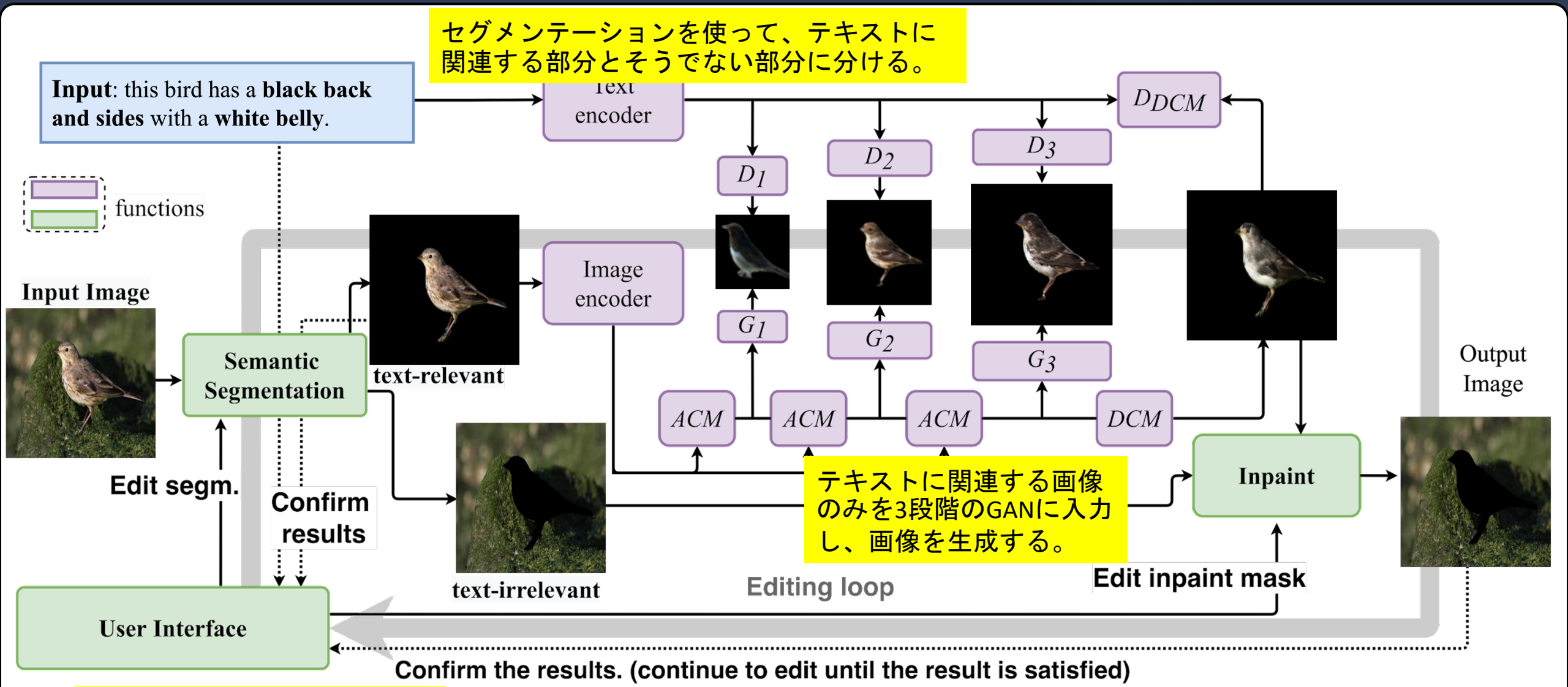
Input Text	A large pizza on a white plate sitting on a white table.		A large pizza on a white plate sitting on a **brown** table.		A large white bowl of many green apples.		A large **black** bowl of many green apples.	
								
GT(TIA)	1		2		1		2	
	Rank	Score	Rank	Score	Rank	Score	Rank	Score
CLIPScore[7]	2	0.831	1	0.834	2	0.827	1	0.829
Ours(TIA only)	**1**	**1.00**	**2**	**0.670**	**1**	**1.00**	**2**	**0.50**

国際会議ISCASに採択た

Mizuki Miyamoto, Ryugo Morita, Jinjia Zhou, "Visual question answering based evaluation metrics for text-to-image generation". IEEE International Symposium on Circuits and Systems (ISCAS), 2024, Singapore.

法政大学 HOSEI University iMedia Lab.

周研の研究例（2022年度卒業研究）：テキストを用いた画像編集



テキスト関連部とそうでない部分を分けるUIを加える。

Ryugo Morita, Zhiqiang Zhang, Man M. Ho, Jinjia Zhou, "Interactive Image Manipulation with Complex Text Instructions"IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2023)

Interactive Image Manipulation with Complex Text Instructions

Ryugo Morita, Zhiqiang Zhang, Man M. Ho, Jinjia Zhou
Hosei University
Tokyo, Japan
ryugo.morita.7f@stu.hosei.ac.jp

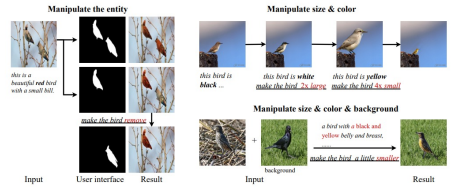
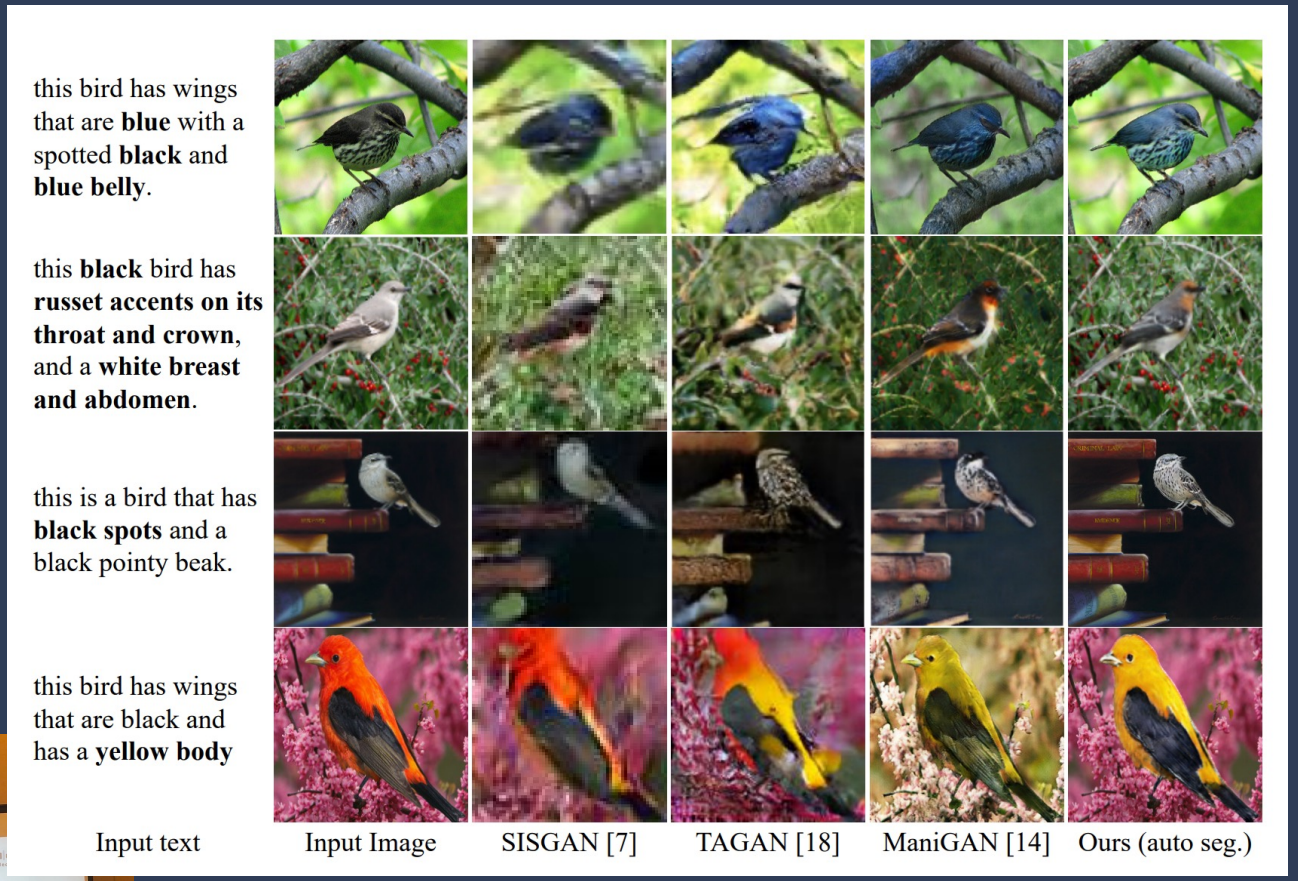


Figure 1. Given an image and text instruction that reveals desired modifications to the image, our method first tries to understand the image and localize where should be modified, then makes appropriate manipulations. In addition, our network design also allows users to adjust the affected area and add/redo image manipulations for undesired results. As an advantage, this work provides effective image manipulations with high controllability, such as changing an object's attributes (e.g., colors and texture), enlarging, dwindling, removing objects, and replacing the background.

Abstract

Recently, text-guided image manipulation has received increasing attention in the research field of multimedia processing and computer vision due to its high flexibility and tasks possible, we apply three strategies. First, the given image is divided into text-relevant content and text-irrelevant content. Only the text-relevant content is manipulated and the text-irrelevant content can be maintained. Second, a super-resolution method is used to enlarge the manipulation region to further improve the operability and to help manipulate the object itself. Third, a user interface is introduced for editing the segmentation map interactively to re-modify the generated image according to the user's desires. Extensive experiments on the Caltech-UCSD Birds-200-2011 (CUB) dataset and Microsoft Common Objects in Context (MS COCO) datasets demonstrate our proposed method can enable interactive, flexible, and accurate image manipulation in real-time. Through qualitative and quantitative evaluations, we show that the proposed model outperforms other state-of-the-art methods.

従来法と比較した結果



国際会議WACVに採択された

法政大学 HOSEI University
iMedia Lab.

Deep Learningを用いた画像から説明文の自動生成について

- 3D Dense Captioning とは 3D画像を自然言語により密に説明する技術

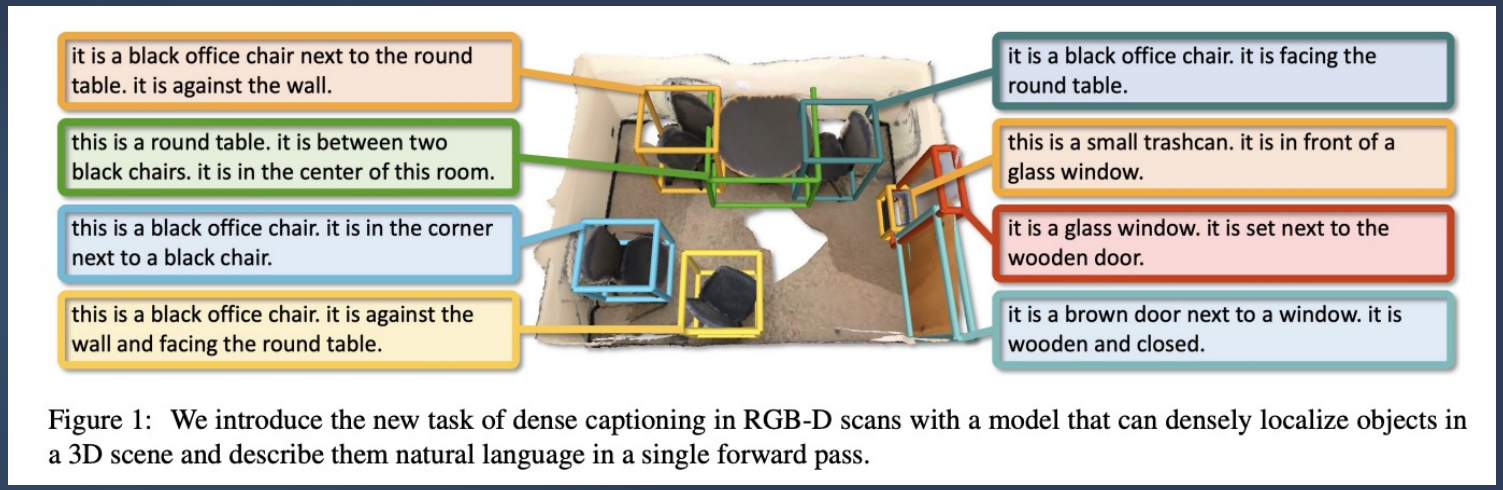
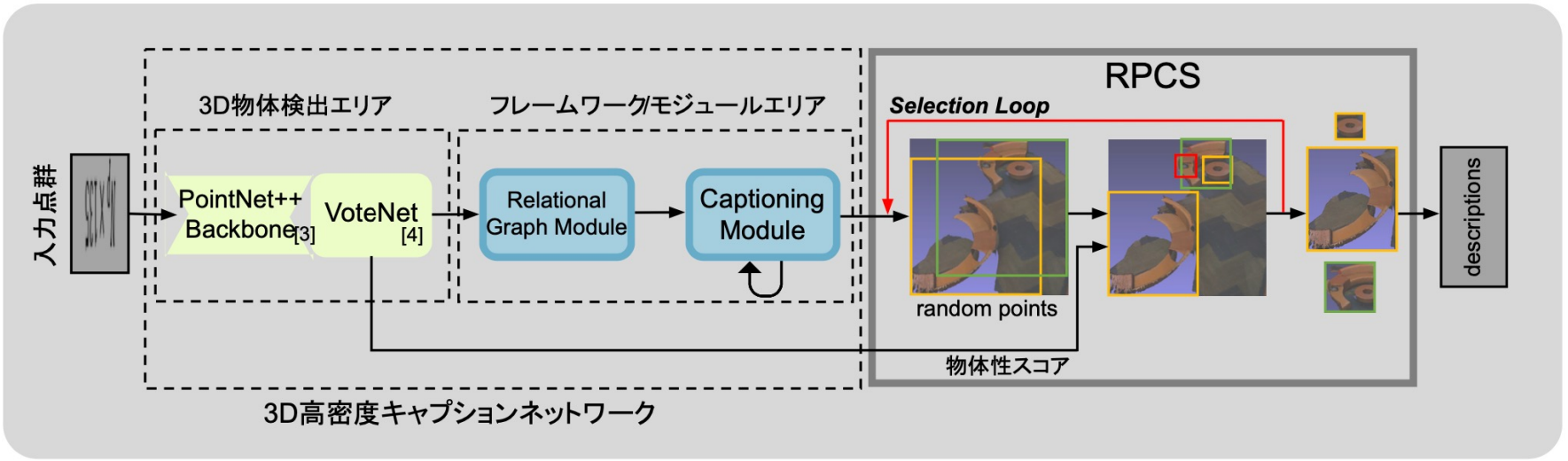


Figure 1: We introduce the new task of dense captioning in RGB-D scans with a model that can densely localize objects in a 3D scene and describe them natural language in a single forward pass.

[1] Z. Chen, A. Gholami, M. Niessner, and A. X. Chang. Scan2Cap: Context-Aware Dense Captioning in RGB-D Scans. In CVPR, 2021.

周研の研究例（2022年度卒業研究）： 3D Dense Captioning

提案：循環点群選択法(Recurrent Point Clouds Selection: RPCS)



ステップ1：点群からランダムな点を選択し、物体性スコアを用いて生成された説明文の品質を考慮、保存

ステップ2：いくつかの物体性スコアが閾値より低い場合、ステップ1を再び処理

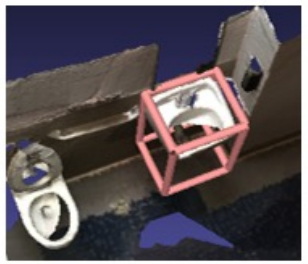
ステップ3：ステップ2でより豊かな記述が生成された場合、保存されている品質が低い記述を削除

[3] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In NeurIPS, 2017
[4] C. R. Qi, O. Litany, K. He, and L. J. Guibas. Deep hough voting for 3D object detection in point clouds. In ICCV, 2019

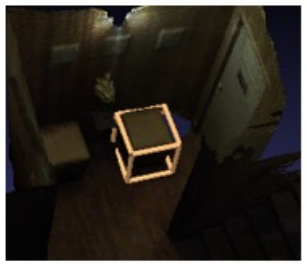
従来法と比較した結果

GT: this is a black refrigerator. it sets in the corner of the room, by the counter.
Scan2Cap: this is a white refrigerator. it is to the right of the stove.
Scan2Cap+Ours: this is a stainless steel refrigerator. it is to the right of the kitchen cabinet.

GT: there is a window with blinds covering it. it has a baseboard heater below it.
Scan2Cap: the object is a drawer. it is located under the bed on the right side of the bed.
Scan2Cap+Ours: this is a window with blinds. it is to the left of the desk.



GT: the sink is below the mirror and right of the toilet. it is below and to the left of the black soap dispenser.
Scan2Cap: this is a white sink. it is to the left of the toilet.
Scan2Cap+Ours: this is a white sink. it is to the left of a soap dispenser.

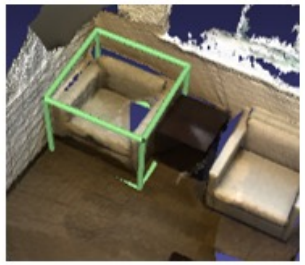


GT: this is a ottoman. the ottoman is near the wall.
Scan2Cap: this is a brown hamper. it is to the left of a chair.
Scan2Cap+Ours: this is a ottoman. the ottoman is near the wall.

GT: the coffee table is in the center of the room surrounded by a sofa. the coffee table is a brown rectangle.
Scan2Cap: the couch is surrounding the coffee table. the couch is a long, gray, and rectangular.
Scan2Cap+Ours: the coffee table is in the center of the room. the coffee table is a brown rectangle.

GT: the object is the bed with a green spread, with a laptop computer resting on it. the bed is to the left of the nightstand.
Scan2Cap: there is a bed with a brown headboard. it is to the right of the desk.
Scan2Cap+Ours: the bed is in the middle of the room. the bed is to the left of the nightstand.

GT: the trash can is right of the blue covered trash can. the trash can is a gray rectangular prism.
Scan2Cap: the trash can is the second rightmost one out of a row of trash cans. the trash can is a cylinder.
Scan2Cap+Ours: the trash can is the farthest right one out of a row of trash cans. the trash can is a gray cylinder.



GT: to the right of this armchair is a black table. this gray armchair is located near the left wall.
Scan2Cap: this is a brown ottoman. it is to the left of the chair.
Scan2Cap+Ours: this is a brown armchair. it is to the left of the table.

※ 緑色：正しい記述
赤色：誤り記述

国際会議ICONIPに採択された

Shinko Hayashi, Zhiqiang Zhang, and Jinja Zhou, " HA Recurrent Point Clouds Selection Method for 3D Dense Captioning". The 29th International Conference on Neural Information Processing (ICONIP 2022)

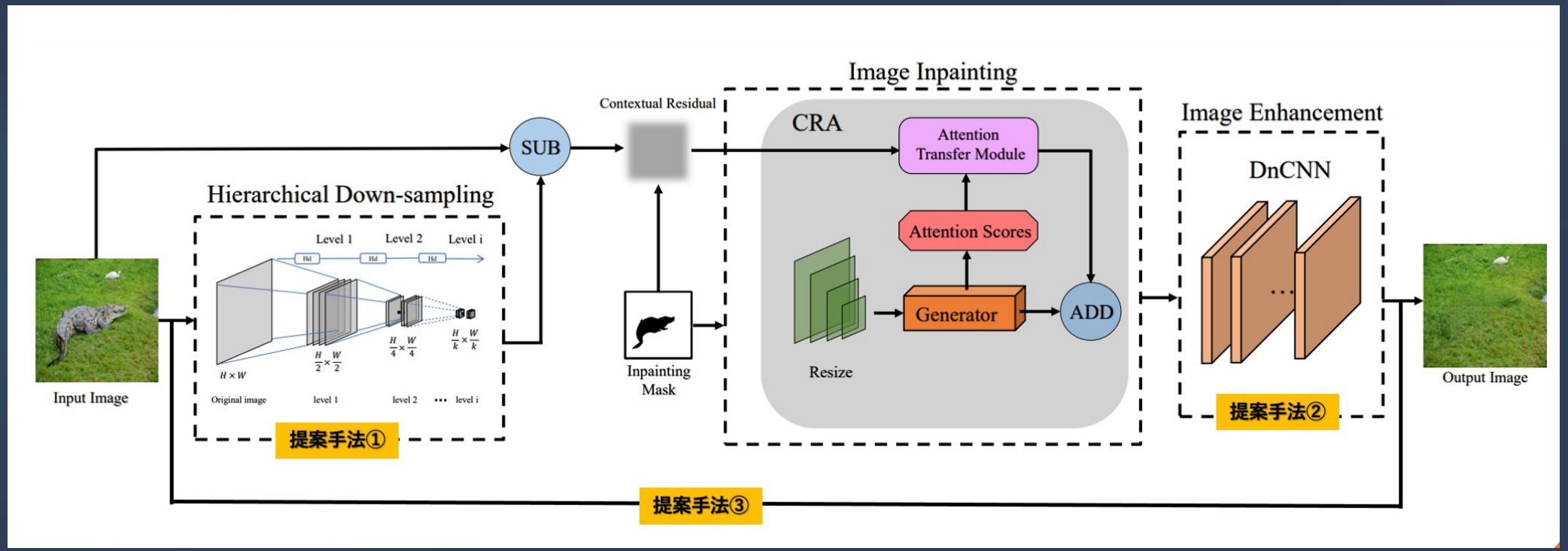
法政大学 HOSEI University
iMedia Lab.

画像インペインティングとは？

写真の折り目やキズを目立たなくし たり，観光地で撮影した写真に写って
しまった不要オブジェクトの除去をしたりする手法

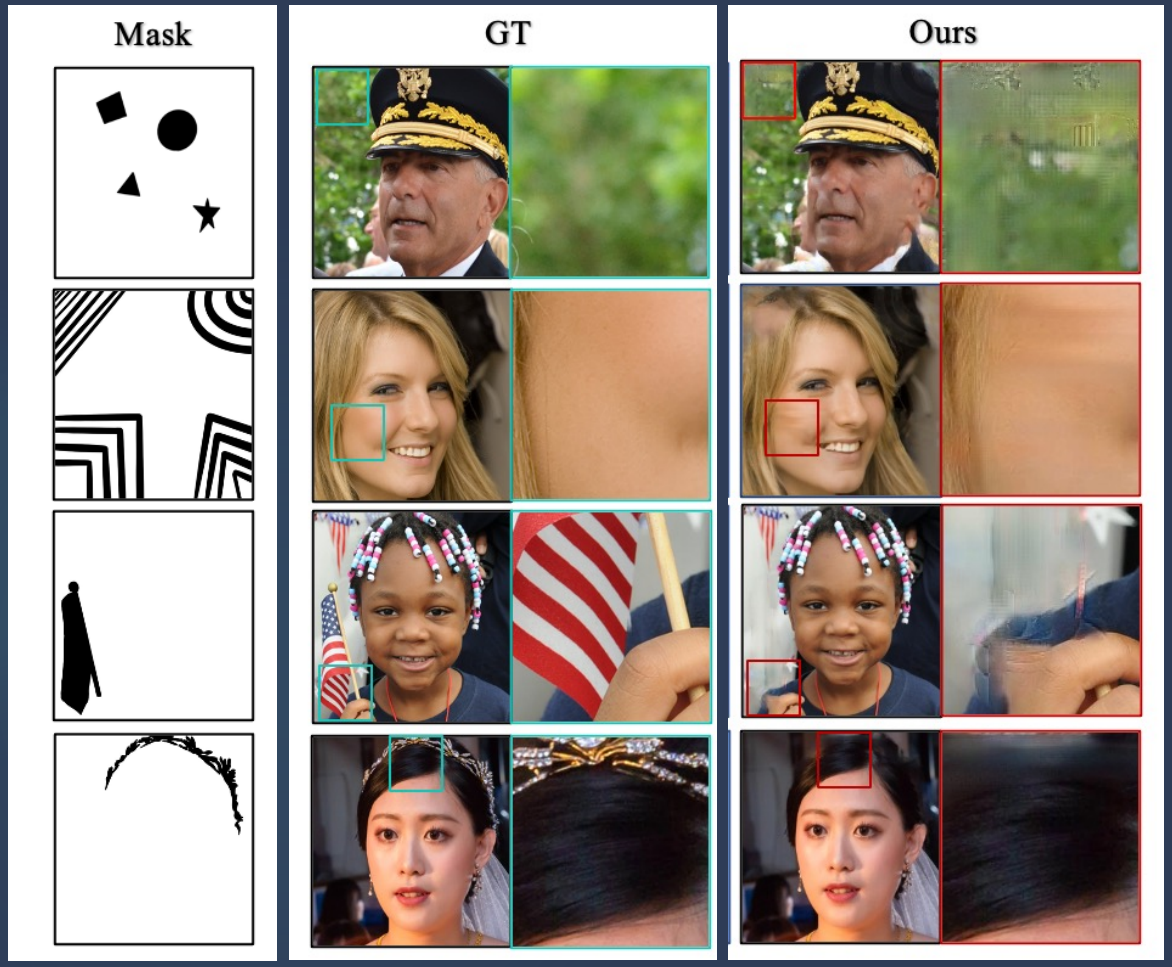
周研の研究例（2022年度卒業研究）：画像インペインティング

提案：コンテクスト残差集計法に基づく階層的ダウンサンプリング画像インペインティング



法政大学 HOSEI University
iMedia Lab.

実験結果



国際会議ICONIPに採択された

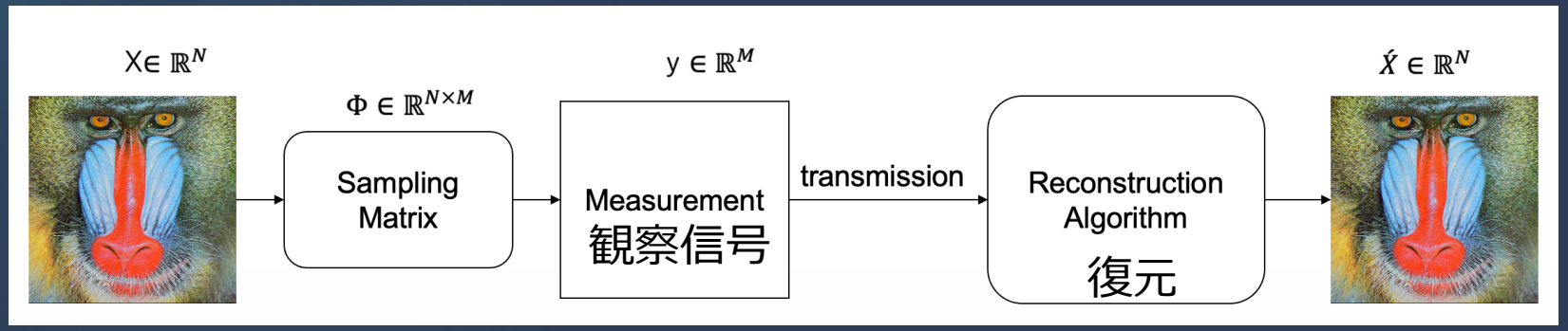
Haixin Wang, Jian Yang and Jinjia Zhou, " Hierarchical down-sampling based ultra high-resolution image inpainting". The 29th International Conference on Neural Information Processing (ICONIP 2022)

法政大学 HOSEI University

iMedia Lab.

動画圧縮技術：圧縮センシング

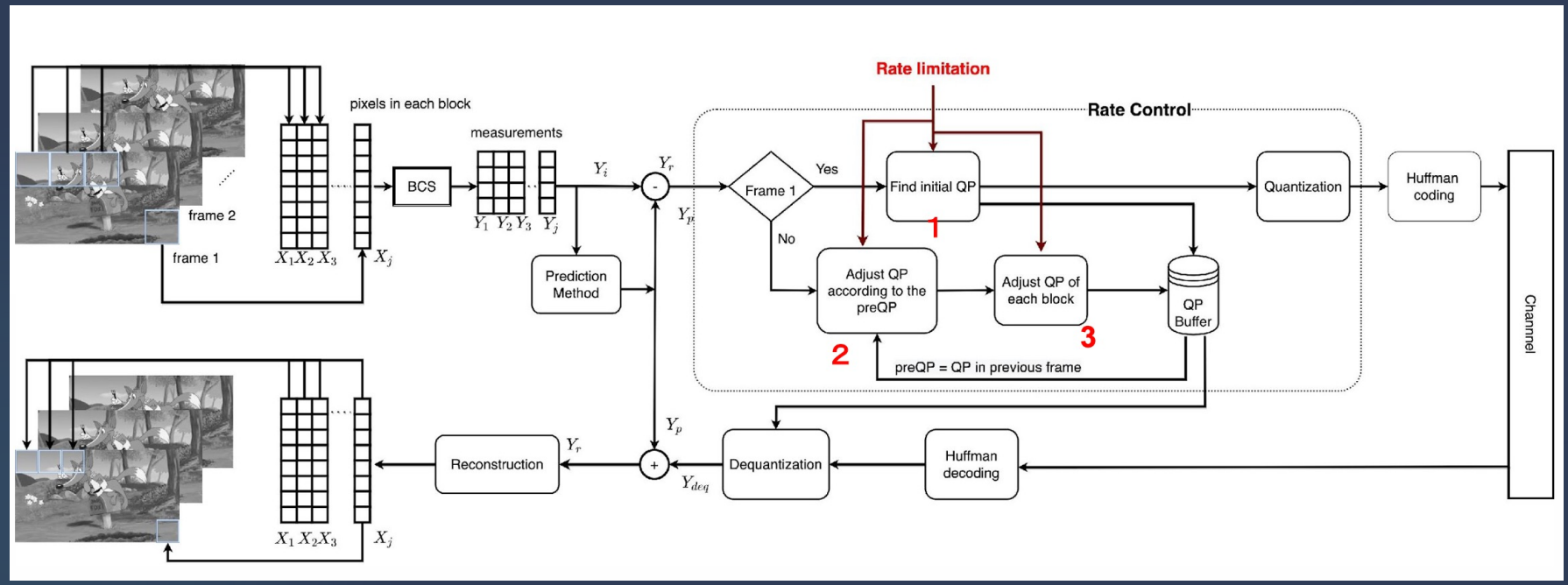
- 圧縮センシングは，数学，工学，コン ピュータ科学，天文学，医学など，大量 のデータを扱う分野で応用されることが期待され，2006 年頃から急速にその研究が展開された.
- 圧縮センシングは，対象となる信号をできるだけ少ない観察から，復元する技術である。圧縮技術の多くは，いったん観察信号を大容量のデータとして取得し た後に圧縮削減するが，圧縮センシング は観察と圧縮を同時に行い，効率的に データの取得を行うため，大容量で冗長 なデータ取得を制限できる利点がある。



法政大学 HOSEI University
iMedia Lab.

周研の研究例（2022年度卒業研究）：

提案：ビデオ圧縮センシング向けのフレーム適応型レートコントロール



まず最初のフレームに対して適切なQPを求める。2枚目以降のフレームは前のフレームのQPを活用してQPを決める。最後にブロックごとにQPを調整する方法を提案する。

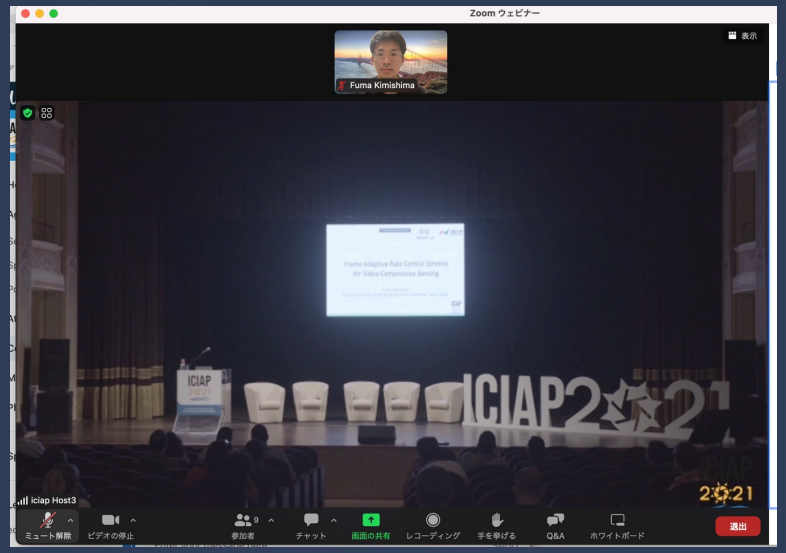
法政大学 HOSEI University iMedia Lab.

既存研究とのビデオ品質と処理時間の比較

SR=0.5	BL=0.5				BL=1.0			
Sequences	TMM'21[10]		This work		TMM'21[10]		This work	
	PSNR	Time(min)	PSNR	Time(min)	PSNR	Time(min)	PSNR	Time(min)
WILD_TRACK	27.61	17.0	27.5	14.4	31.32	17.8	31.54	14.5
videoSRC20_28	33.28	34.2	33.55	31.2	36.55	68.3	36.62	42.2
videoSRC02_16	41.06	45.7	41.17	34.4	43.02	66.8	43.02	39.6
videoSRC16_06	41.99	33.1	42.08	25.6	42.60	34.3	42.56	26.0
videoSRC30_20	38.97	89.9	40.32	48.6	40.18	120.0	41.59	56.4
videoSRC04_30	24.72	29.2	24.85	24.7	27.23	38.1	27.26	30.7
average	34.61	41.5	34.91	29.8	36.82	57.6	37.10	34.9
SR=0.75	BL=0.5				BL=1.0			
Sequences	TMM'21[10]		This work		TMM'21[10]		This work	
	PSNR	Time(min)	PSNR	Time(min)	PSNR	Time(min)	PSNR	Time(min)
WILD_TRACK	25.90	22.3	26.3	15.8	31.13	21.4	31.18	17.3
videoSRC20_28	31.92	44.5	32.08	41.7	37.09	69.3	37.20	48.3
videoSRC02_16	39.60	58.0	39.81	53.0	43.16	74.1	43.18	51.9
videoSRC16_06	41.90	50.5	42.01	30.7	43.18	44.2	43.15	36.1
videoSRC30_20	38.44	68.3	39.44	55.2	40.47	117.6	42.03	70.4
videoSRC04_30	23.58	37.8	24.20	34.0	27.49	47.4	27.46	39.3
average	33.56	46.9	33.97	38.4	37.09	62.3	37.37	43.9

53% minute ↓
1.41dB ↑

40% minute ↓
1.56dB ↑

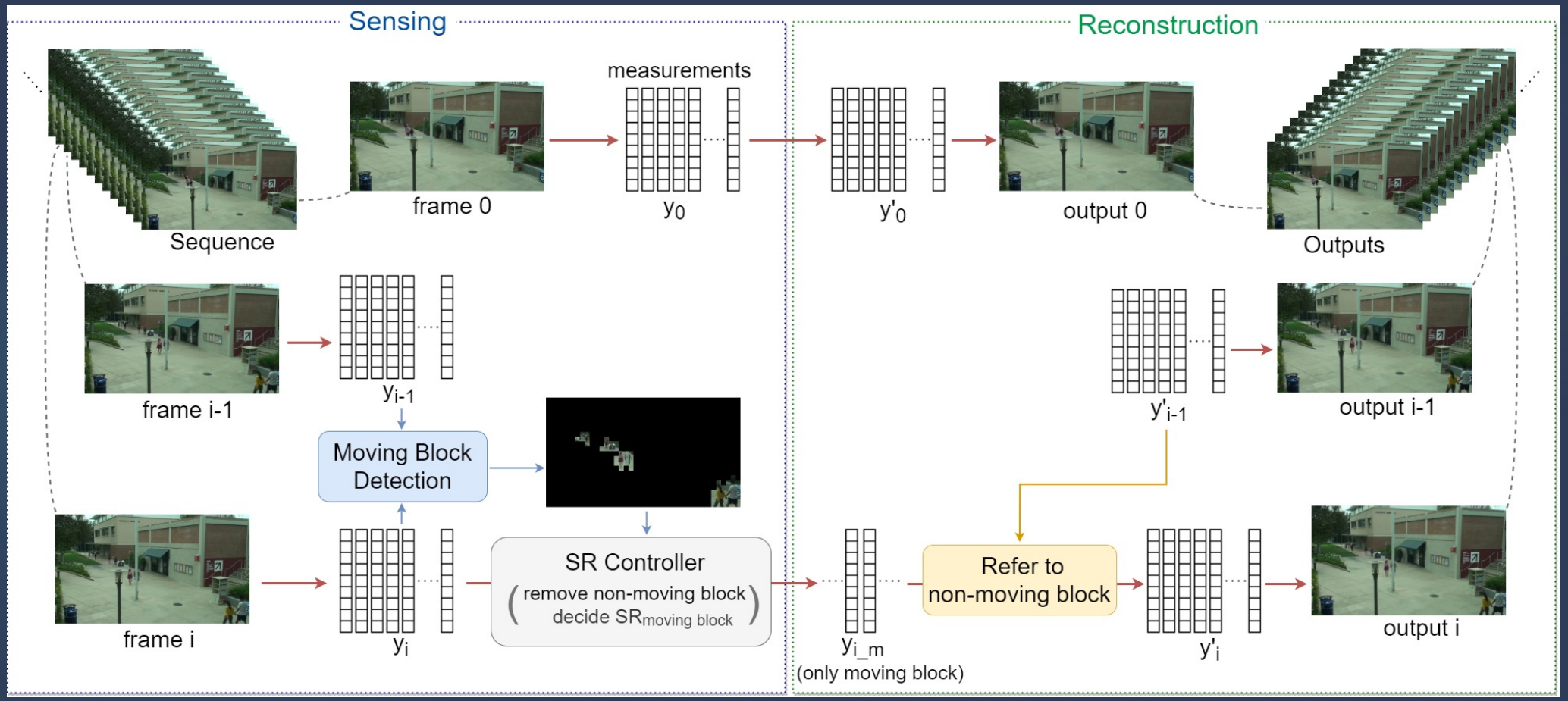


国際会議ICIAPに採択された

Fuma Kimishima, Jian Yang, Thuy Thi Thu Tran, Jinjia Zhou, "Frame Adaptive Rate Control Scheme for Video Compressive Sensing", The 21st International Conference on Image Analysis and Processing (ICIAP), May 2022. Pages 247–256.

法政大学 HOSEI University
iMedia Lab.

周研の研究例（2023年度卒業研究）：
サンプリングレート制御を用いた適応的なブロックベース圧縮センシング



法政大学 HOSEI University
iMedia Lab.

実験結果

VIRATデータセット[5]における最先端動画像CS手法(Multi-rate-VCS[1], VCSL[2])とのPSNR及びSSIMの比較

Video	Multi-rate-VCS [5]		VCSL [14]		Ours	
	average SR	PSNR/SSIM	average SR	PSNR/SSIM	average SR	PSNR/SSIM
S_0401	1.13%	33.35/0.9393	0.95%	34.48/0.9488	0.99%	34.62/0.9505
S_0502	1.05%	31.86/0.9045	0.73%	34.07/0.9420	0.96%	35.42/0.9495
S_0002	1.01%	35.46/0.9463	0.84%	36.56/0.9539	0.98%	37.06/0.9548
S_0102	1.02%	34.52/0.9242	1.19%	36.23/0.9537	0.99%	36.17/0.9465
S_0100	0.91%	30.61/0.8856	0.83%	34.25/0.9453	0.94%	34.98/0.9460
S_0101	1.22%	32.47/0.9404	0.85%	33.55/0.9533	0.98%	34.18/0.9533
average	1.06%	33.05/0.9234	0.90%	34.86/0.9495	0.97%	35.41/0.9501

国際会議ACM MM Asiaに採択された

Kosuke Iwama, Ryugo Morita, Jinjia Zhou, "Block based Adaptive Compressive Sensing with Sampling Rate Control", ACM Multimedia Asia 2023, Tainan, Taiwan

法政大学 HOSEI University

iMedia Lab.

ほかの卒業研究テーマ

- Zigzag Ordered Walsh Matrix for Compressed Sensing Image Sensor（国際会議DCCに採択された、卒研）
- An Iterative Image Inpainting Method Using Mask Shrinking（国際会議ISCASに採択さらた、卒研）
- High Frequency Feature Distillation Network for Compressive Sensing Reconstruction（国際会議IJCNNに採択された、卒研）
- Video frame prediction based measurement interpolation for compressed sensing（国際会議IJCNNに採択され、卒研）

周研究室

知能メディア処理 Intelligent Media Processing

https://www.zhou-lab.info/

法政大学 HOSEI University

iMedia Lab.