AI社会実装の最前線
– 宇宙物理との関連も交えて

竹内 駿
富士通オーストラリア　リサーチディレクター
マッコーリー大学　客員准教授
2024年11月14日（木）

背景：ブラックホール天文学

- 輻射と磁場に支配されたBH周りのガス動力学

ST+ (2013, PASJ)
Jiao, ST+ (2015, APJ)
Kobayashi, ST+ (2018, PASJ)

Visual: NASA/STScI; X-ray: NASA/CXC/SAO; radio: NSF/NRAO/VLA

京都大学 理学研究科・理学部
Graduate School of Science / Faculty of Science, Kyoto University

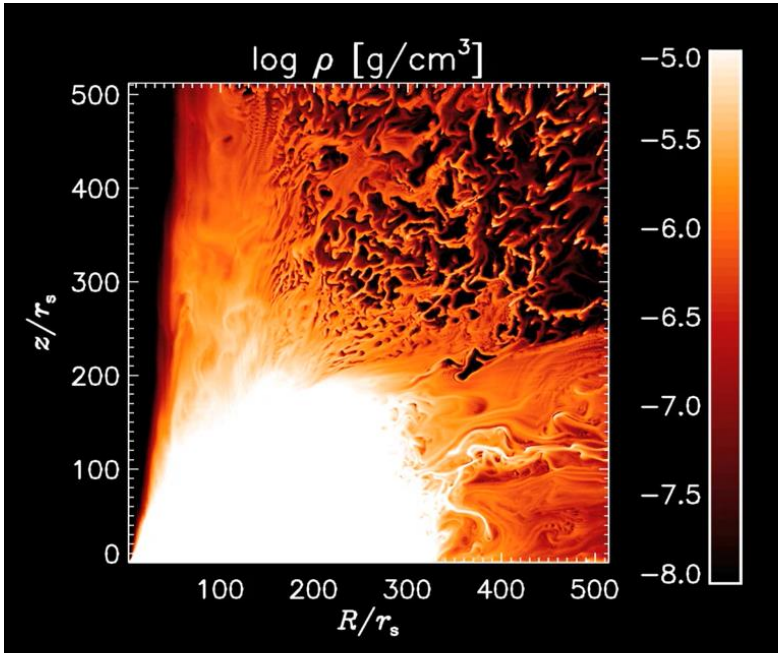
- 受験生の方
- 在学生の方
- 教職員の方
- 理学への支援のお願い
- 研究科・学部について
- 教育・研究紹介と研究者一覧
- 入試情報

研究紹介：ハイブリッド・ブラックホールジェット　-- 嶺重慎教授らのグループ

投稿日時： **2010.10.27**

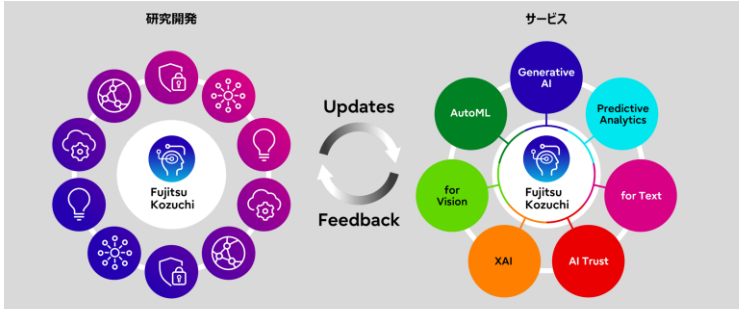
ハイブリッド・ブラックホールジェット：
スーパーコンピュータが解き明かした新タイプのジェット

国立天文台天文シミュレーションプロジェクト(CfCA)の大須賀健助教、京都大学大学院理学研究科宇宙物理学教室の竹内駿氏（元大学院生、現富士通）および嶺重慎教授の研究チームは、国立天文台のスーパーコンピュータ（Cray XT4）を用いた大規模シミュレーションにより、新しいタイプのブラックホールジェットを発見しました。



最近の活動：AI応用

- 技術開発



- 実証実験

- 論文・特許

- 産学連携

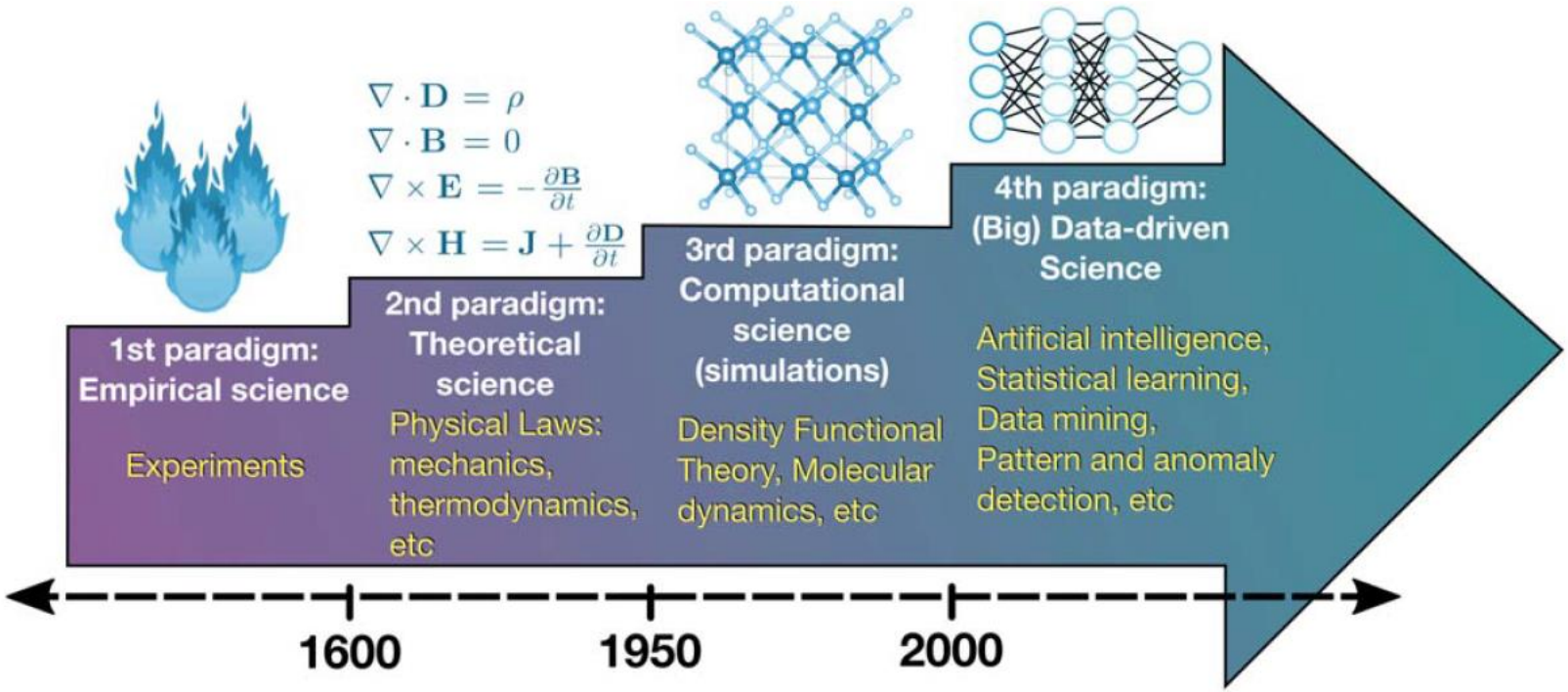
FUJITSU

AI社会実装を支える先端技術

科学技術の新しいパラダイム

FUJITSU

Schleder+ (2019, JPhys Materials)



ヒューマンセンシング技術

FUJITSU

課題 1 学習データ準備や検証が手間

解決策 1 **基本動作認識**

約100種類の基本動作学習済モデル

課題 2 導入後のカスタマイズが難しい

解決策 2 **上位行動推定**

基本動作の組み合わせで複雑な行動を簡単に定義

例：リアル店舗の購買行動分析

FUJITSU

Group Customers

Female, Age:30s, Customer
- pickup ITEM2
- talk
- return ITEM2
- pickup ITEM6
- talk

Male, Age:30s, Customer
- talk
- cross arms
- talk

Vision AI meets LLM

FUJITSU

FUJITSU Vision AI Assistant Demo App

RUNNING... Stop

Surveillance cameras

Camera A Camera B

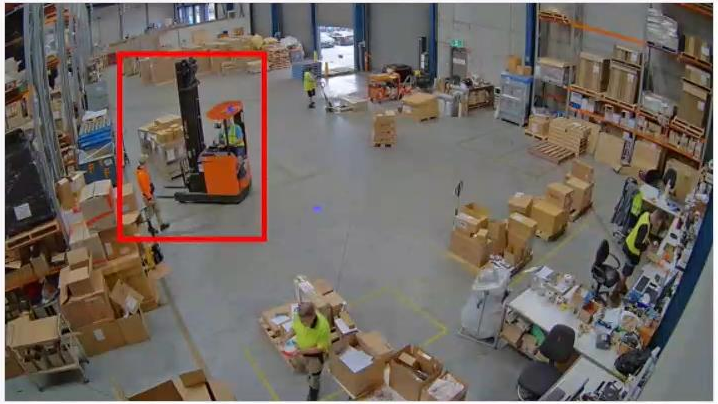
Camera C Camera D

Options

use visual prompt

Message examples

For the last three months, show



Please tell me the cause of the approaching event enclosed in the box based on the scene situation. And what kind of education and measures can we take to reduce this kind of approach?

Context-aware_video_analysis: Please analyze the context of the video and provide the caus...

Send a message

FUJITSU

教師なしドメイン適応学習

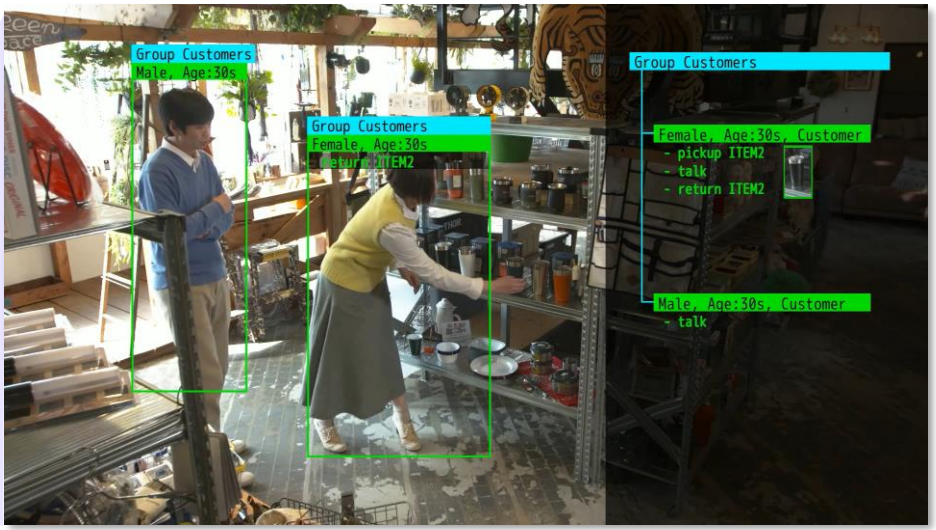
Takeuchi et al., “Unsupervised domain-adaptive person re-identification with multi-camera constraints,” in IEEE ICIP, 2022

Video-based Behavior Analysis

Widely applied to many business fields such as retail, manufacturing, mobility, logistics

Retail scene

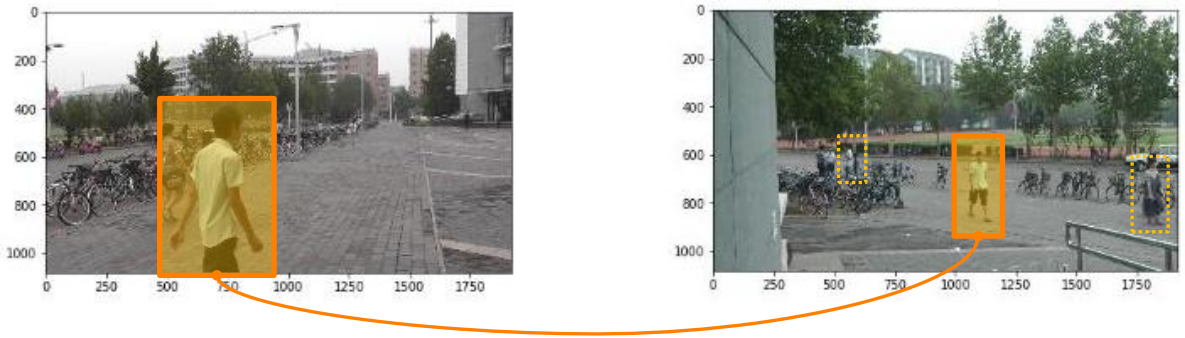
- consumer movement pattern
- product shelves of interest
- wants within the group

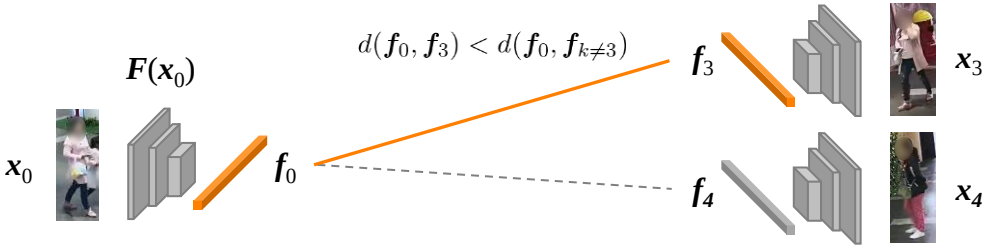


"Enhancing the Retail Experience with Actlyzer" (2022, Fujitsu)

Person Re-identification

Retrieve the same person (i, j) based on a distance $d(\boldsymbol{f}_i, \boldsymbol{f}_j)$ between the features $\boldsymbol{f}_k = F(\boldsymbol{x}_k)$



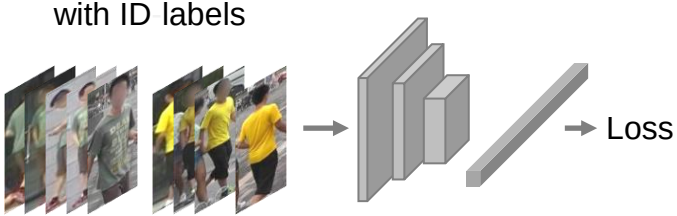


Mang et al. “Deep learning for person re-identification: A survey and outlook” (2021, PAMI)

Domain Gap

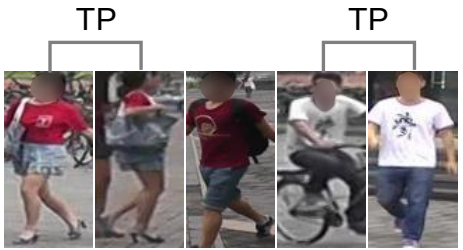
Remains a major challenge for practical application, $\mathcal{L}_{\text{source}}^{\text{train}} \sim \mathcal{L}_{\text{source}}^{\text{test}} < \mathcal{L}_{\text{target}}^{\text{test}}$

Market-1501
with ID labels



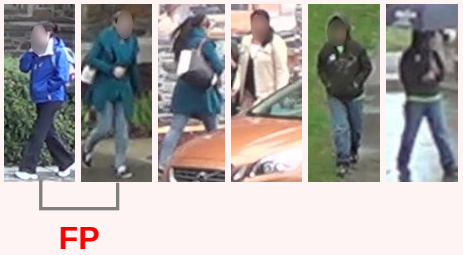
The annotation cost of ID label is
relatively high compared to other tasks.

✓ Market-1501



✘ **DukeMTMC-reID**

different
- date and location
- camera specification
- personal attribute



Yang et al. “Cross dataset person re-identification” (2014, ACCV)

Related study: UDA

Self-training scheme

1. trains a base model using labeled data such as public datasets $\mathcal{D}_{\mathrm{s}} = \{\boldsymbol{x}_i^{\mathrm{s}}, \boldsymbol{y}_i^{\mathrm{s}}\}$
2. infers pseudo-labels of target domain data by using the pre-trained model $\tilde{\boldsymbol{y}}^t = F_{\mathrm{clust}}(d_{\mathrm{J}})$
3. trains a model for the target domain by using the pseudo-labels $\mathcal{D}_{\mathrm{t}} = \{\boldsymbol{x}_i^{\mathrm{t}}, \tilde{\boldsymbol{y}}_i^{\mathrm{t}}\}$

➡ **However, the pseudo-label noise is an obstacle to the domain adaptation**

Song et al. “Unsupervised domain adaptive re-identification: Theory and practice” (2020, Pattern Recognition)

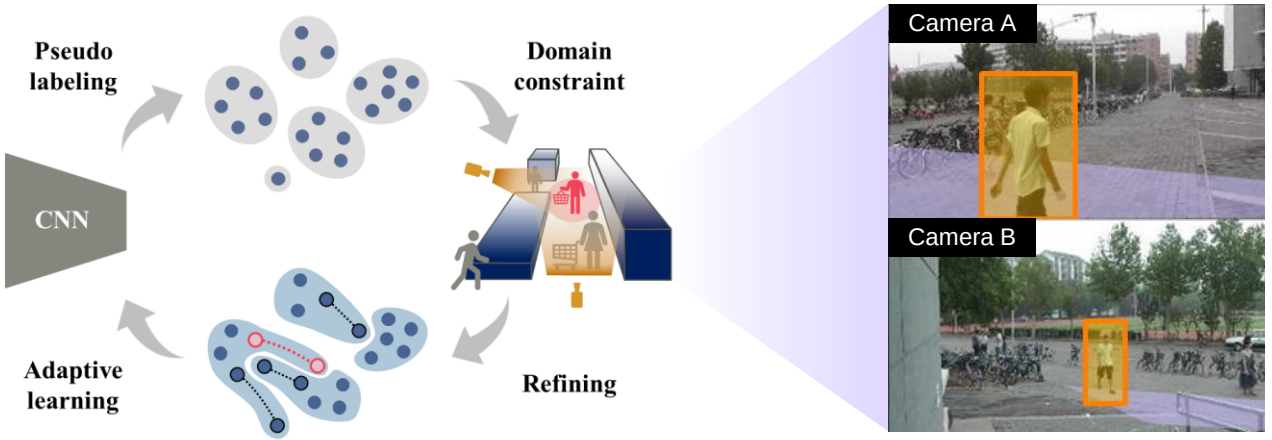
Our Idea

Focus on acquiring the same person images from the overlapping camera views

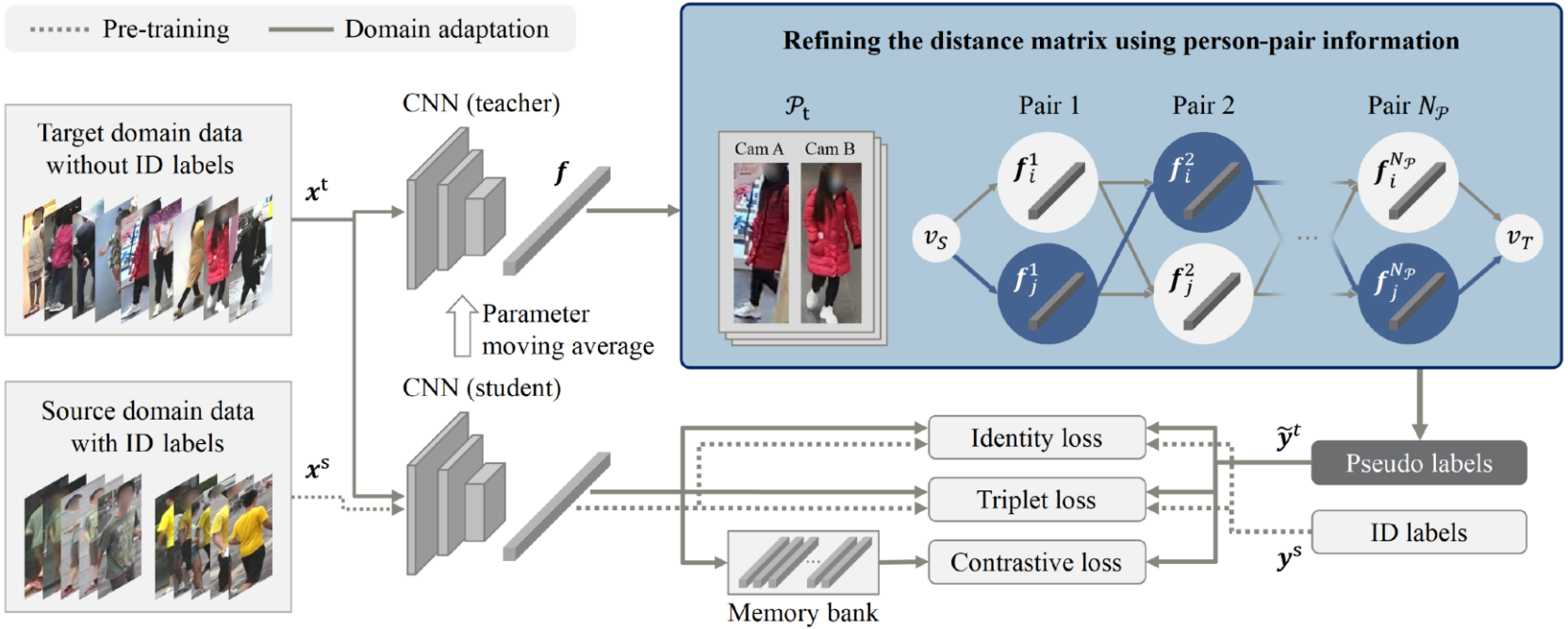
Retail scene

- many surveillance cameras are installed in physical stores to analyze customer behaviors
- adjacent camera views frequently overlap

➡ **incorporate the person-pair info into an UDA ReID model to refine the pseudo-labels**



Network architecture of ECA-Net

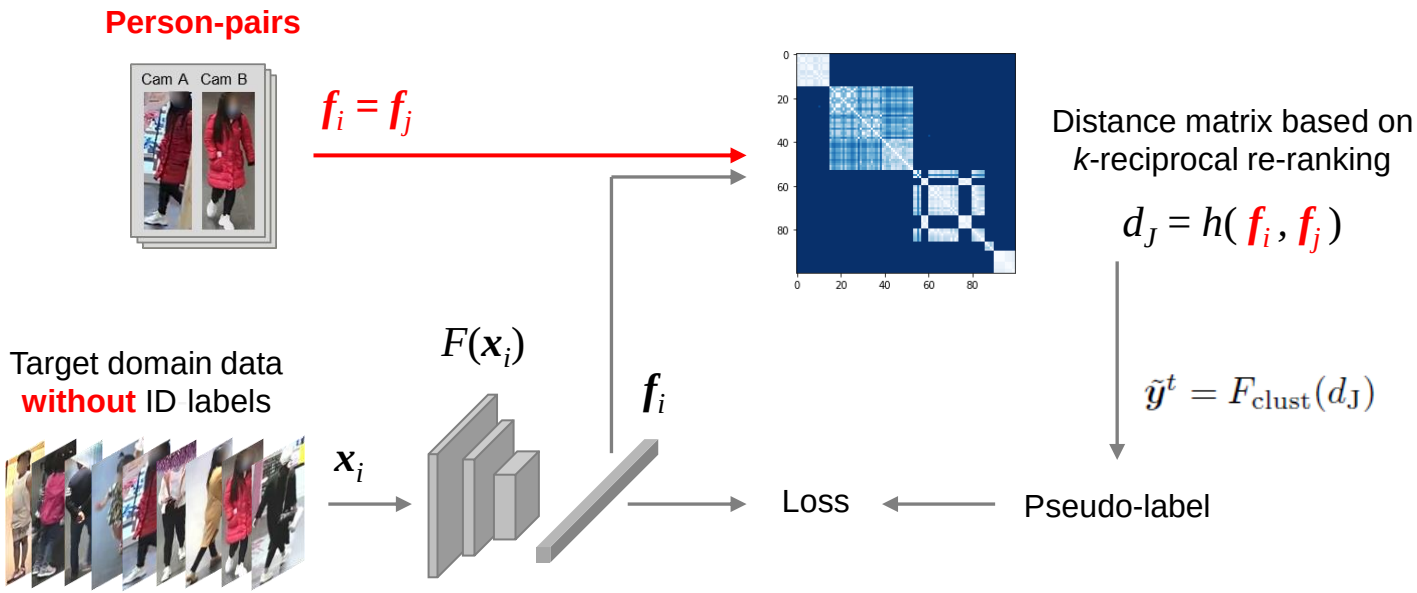


ECA-Net: Constrained UDA ReID

How to incorporate the person-pair info without ID labels into the model?

Distance matrix is calculated based on the *k*-reciprocal re-ranking as functions of (f_i , f_j)

➡ propose the constraint to obtain the same pseudo-label, $f_i = f_j$

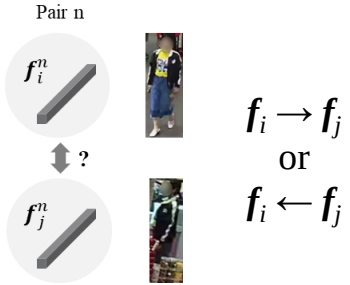


ECA-Net: Feature selection

How to select the optimal feature from a parson-pair?

The *k*-reciprocal re-ranking depends on the features of all samples

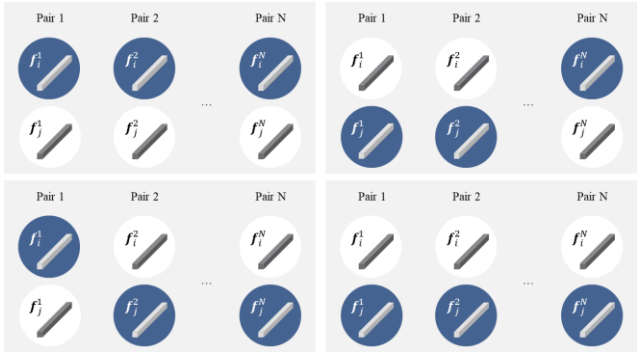
➡ propose an approximate solution that depends on that of some samples



Exact solution

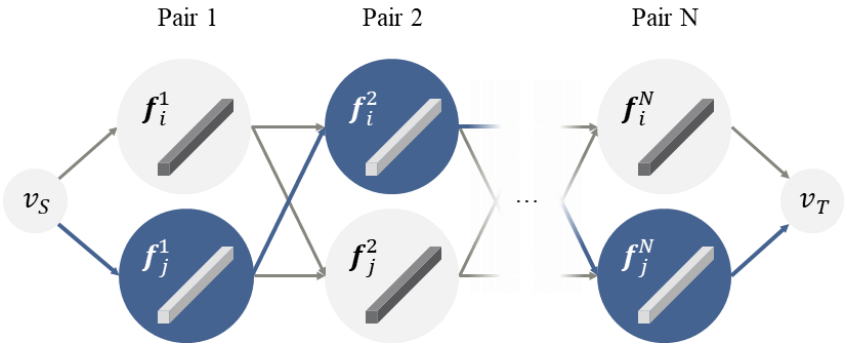
Brute-force search of d_J for the number of pairs *N*, $O(2^N)$

e.g.



Approximate solution

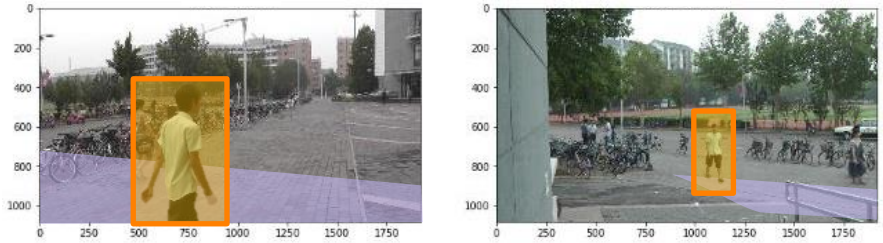
Dijkstra's algorithm that depends on the features of some samples, $O(N^2)$



Experiments

Datasets		#cam	Overlap	#imgs	#Train IDs	#Test IDs
Market-1501	Public	6	**0.86**	32,217	751	750
DukeMTMC-reID	Public	8	**0.007**	36,411	702	702
MSMT17	Public	15	**0.24**	126,441	1,041	3,060
Shopping mall	Private	3	**0.13**	37,971	1,466	1,370

Market-1501 (PRW)



DukeMTMC-reID

Zheng et al. “Scalable person re-identification: A benchmark” (2015, ICCV)
Zheng et al. “Unlabeled samples generated by GAN improve the person re-identification baseline in vitro” (2017, ICCV)
Wei et al. “Person transfer GAN to bridge domain gap for person re-identification” (2018, CVPR)

Comparison with SOTA methods

FUJITSU

ECA-Net outperforms SOTA methods in domains with overlapping camera views

Datasets	Overlap
Market-1501	**0.86**
DukeMTMC-reID	0.007
MSMT17	**0.24**
Shopping mall	**0.13**

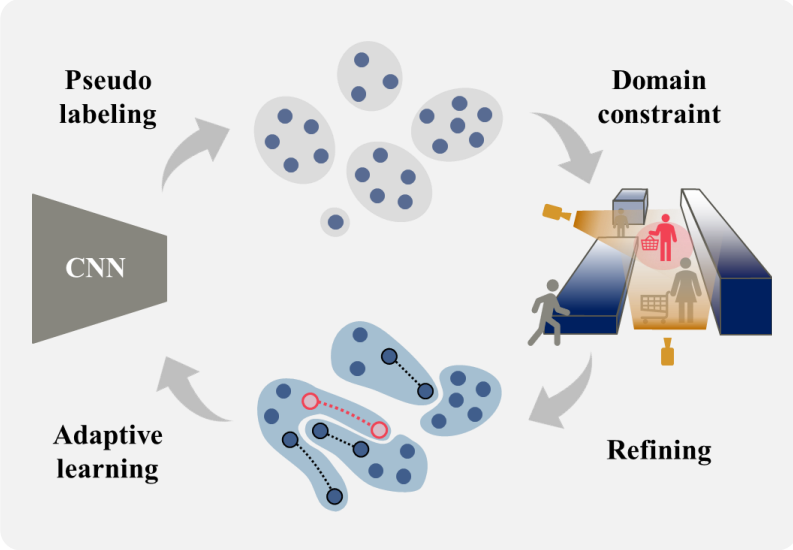
Methods	Reference	Duke → Market				Market → Duke			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
AD-Cluster	CVPR20	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
MMT	ICLR20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
NRMT	ECCV20	71.7	87.8	94.6	96.5	62.2	77.8	86.9	89.5
MEB-Net	ECCV20	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
GLT	CVPR21	79.5	92.2	96.5	97.8	**69.2**	82.0	90.2	92.8
UNRN	AAAI21	78.1	91.9	96.1	97.8	69.1	82.0	**90.7**	**93.5**
ECA-Net (Ours)	-	**83.4**	**94.0**	**97.5**	**98.3**	68.3	**82.3**	90.4	92.5

Methods	Reference	Market → MSMT				Duke → MSMT			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
NRMT	ECCV20	19.8	43.7	56.5	62.2	20.6	45.2	57.8	63.3
ABMT	WACV21	23.2	49.2	-	-	26.5	54.3	-	-
GLT	CVPR21	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
UNRN	AAAI21	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
ECA-Net (Ours)	-	**35.0**	**65.1**	**76.1**	**80.2**	**36.6**	**66.9**	**78.0**	**82.0**

Methods	Reference	Market → Shopping mall				Duke → Shopping mall			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
UNRN	-	45.6	55.1	72.4	78.0	42.4	53.3	69.7	74.9
ECA-Net (Ours)	-	**57.4**	**67.8**	**79.8**	**83.7**	**56.6**	**67.5**	**81.5**	**85.5**

Conclusions

- The first work on UDA ReID modeling with multi-camera constraints
- SOTA in domains with overlapping camera views
- Promising approach in practical situations, since the acquisition cost of the pair images is low



FUJITSU

自己教師あり学習

Kikuchi and **Takeuchi**, “Self-supervised Human-Object Interaction of Complex Scenes with Context-aware Mixing: Towards In-store Consumer Behavior Analysis,” in WACVW, 2024

Background: Consumer Behavior Analysis

- Consumer behavior analysis in physical retail store using CCTV camera provides important information for the retailers to promote sales.
  - Common for online-store and is shown its effectiveness.
  - Able to provide information that cannot be achieved from POS data.

July 6, 2022

AEON RETAIL's Mission To Create the Data-Driven Store of the Future

Retail AI

February 28, 2023

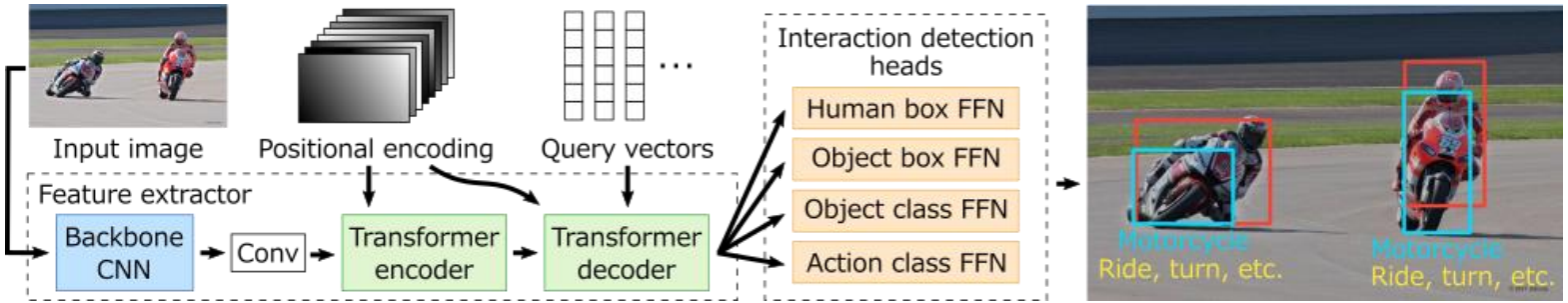
Analyzing Pre-purchase Behavior using AI and Consumer Behavior Theory to Create a Customer-friendly Store

Retail AI Converging technologies

Related Work: Human-Object Interaction Detection

FUJITSU

- Detect the person's action towards an object.
  - Task to detect <Human Bbox, Object Bbox, Interaction class> triplet
- Related Work
  - Transformer-based method (Based on DETR)
    Input: Image, Output: HOI triplet. End-to-End model.



GitHub - hitachi-rd-cv/qpic: Repo for CVPR2021 paper "QPIC: Query-Based Pairwise Human-Object Interaction Detection with Image-Wide Contextual Information"

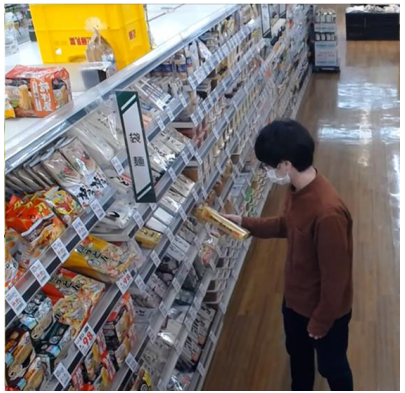
Issues when applying to physical retail stores

- Accuracy decreases when applied to complex scenes with background objects.
- Fine tuning is necessary but creation of dataset but time consuming.
  - Annotation of human, object and the interaction relationship is needed.


Most of the images in public dataset are having person and object focused in middle

Cases when accuracy decreases

Many similar objects

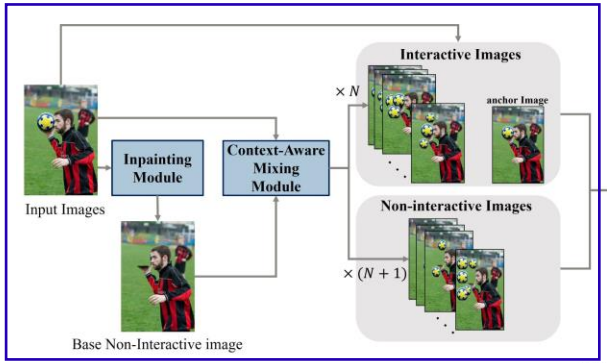


Have many objects in the image

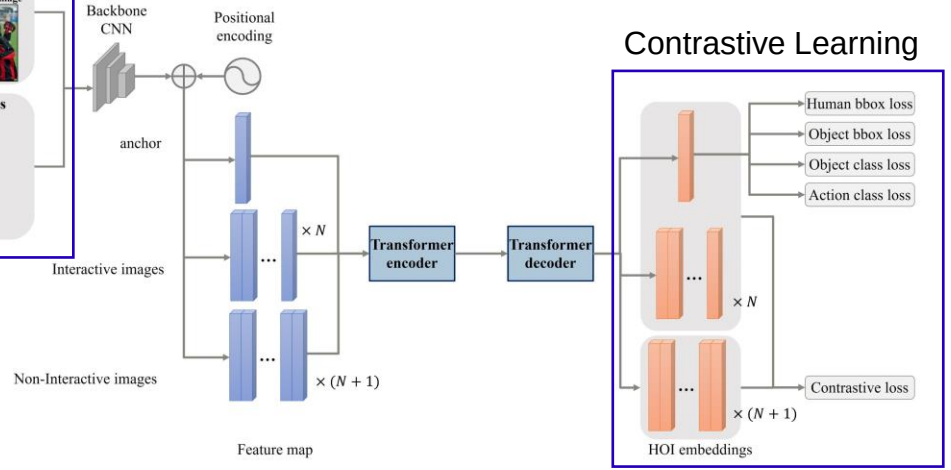
Motivation: Is there a way to improve generalization ability for these cases?

Overview of Proposed Method

- We propose transformer-based HOI method with contrastive learning technique to differentiate the target and background objects.
- We introduce an image mixing method (Context-aware mixing) to generate challenging images with non-interactive objects nearby.

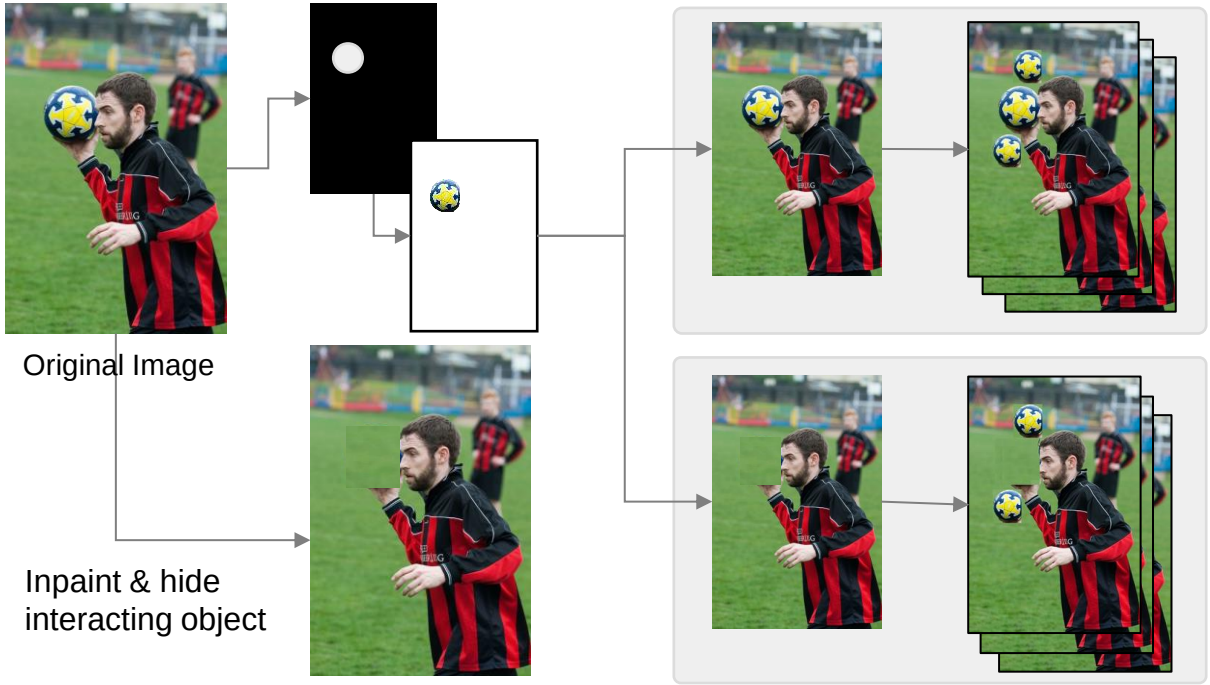


Context-aware Mixing



Context-aware Mixing

- The original objects are augmented near to the actual interacting objects.
- The images are generated using inpainting and segmentation technique.



Hyperparameter
Distance Δd
Number of objects N

Example and the effect of Context-Aware Mixing

Images from the Training Dataset: HICO-DET

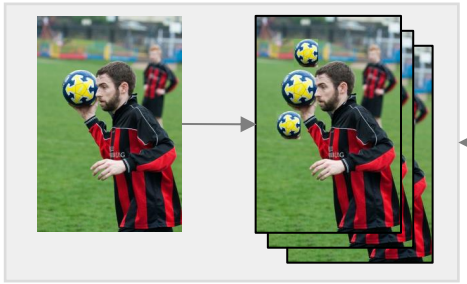
Augmented image

Original image

(c) shows that by adding objects nearby changes the output of the model.

Contrastive Learning

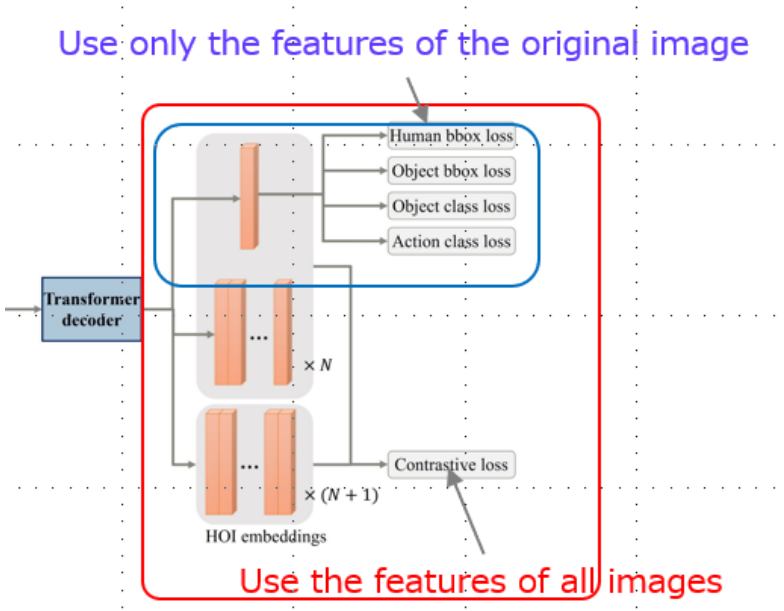
- We introduce supervised contrastive learning to our method.
  - Contrastive learning is carried out between the original images that are the same.
- We use contrastive learning as multi-task for training the HOI detection task.



With interaction

Differentiate the features between the two groups

Without interaction

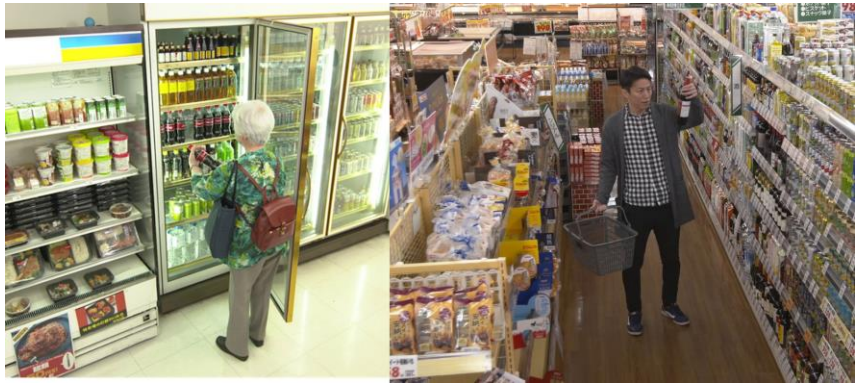


Evaluation

FUJITSU

Evaluation	Training Dataset	Test Dataset	Target Action
Public Dataset	**Public Dataset HICO-DET**	**Public Dataset HICO-DET**	“Hold”–“Something”
Private Dataset 1		**Multiple store dataset** Images from several stores and areas	
Private Dataset 2		**Complex retail shelf** Image from the Shelf that the base model struggled in pre-study	

- Focused for the case in retail store
  - Focused on action **‘Hold’**
  - All Object is grouped to **‘Something’**
    - In order to maintain enough data



Result

Method	Mean Average Precision		
	HICO-DET	Multiple Retail Store	Complex Retail shelf
QPIC (Base model)	58.6	34.8	28.1
CDN	61.2	36.6	30.7
GEN-VLTK	60.5	35.8	37.5
Our Method	**53.3**	**47.8**	**39.2**

- Our method showed **better result** when applied to the **retail store data**.
  - Able to improve the generalization ability for scenes with complex background.
- Our method showed **lower result** when applied to the **public dataset**
  - Public dataset has smaller amount of the scenes with complex background.

Example of Output

Our method enabled to extract only the interacting objects showing that our method was effective to differentiate with background objects.

Conclusion

- We propose transformer-based HOI method with contrastive learning technique to differentiate the target and background objects.
- We will apply our sensing technology to contextual marketing in physical retail stores to understand consumer behaviors for better retail solutions.

FUJITSU

宇宙物理学との関連

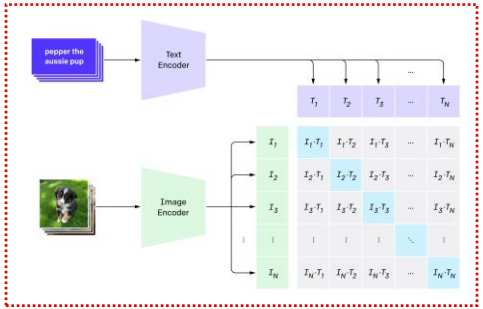
宇宙物理学と情報工学

	宇宙物理学	情報工学（AI系）
主な成果	ジャーナル論文	国際会議論文
国際会議	ネットワーキング	論文出版
特許	基本なし	基本あり
学会	分野固定：天文学会、物理学会	横断的：AI、電子情報、精密工学…
和文誌	（Stars and Galaxies）	多い
共同研究者	理論では単著も	ほとんど共著
目的	真理の追求	人類の幸福
手段	原理や現象の発見、解明	モノやコトの発明、実用化
価値観	世界で最初、追究	役に立つこと、実用性

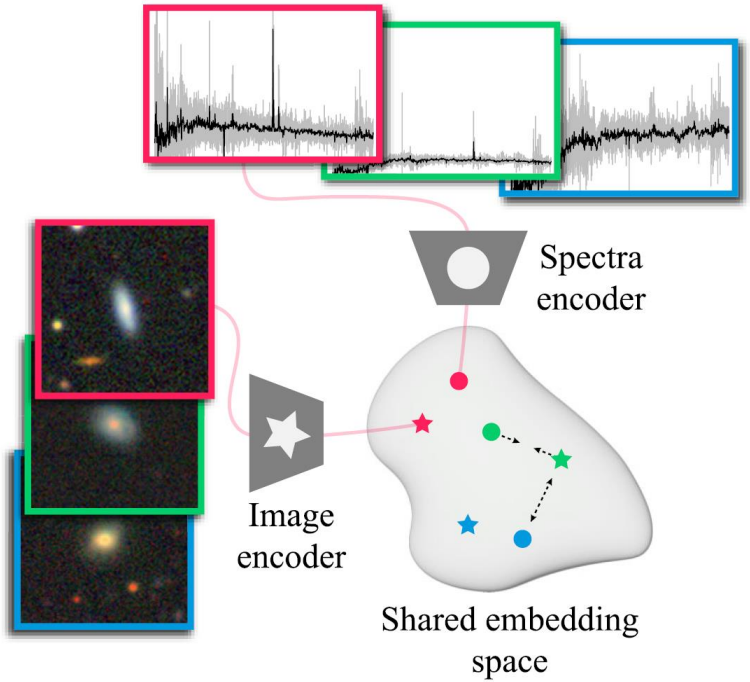
AstroCLIP

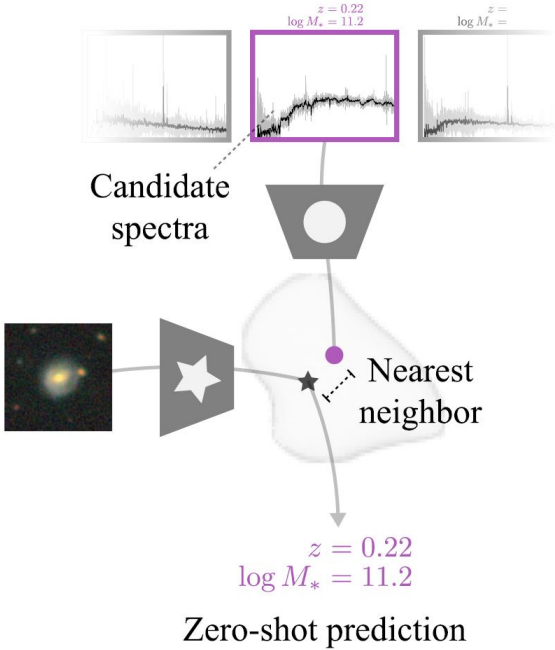
- Cross-modal foundation model for galaxies

Parker et al. (2024, MNRAS)



CLIP: Contrastive Language-Image Pre-training (Radford+ 2021, ICML)

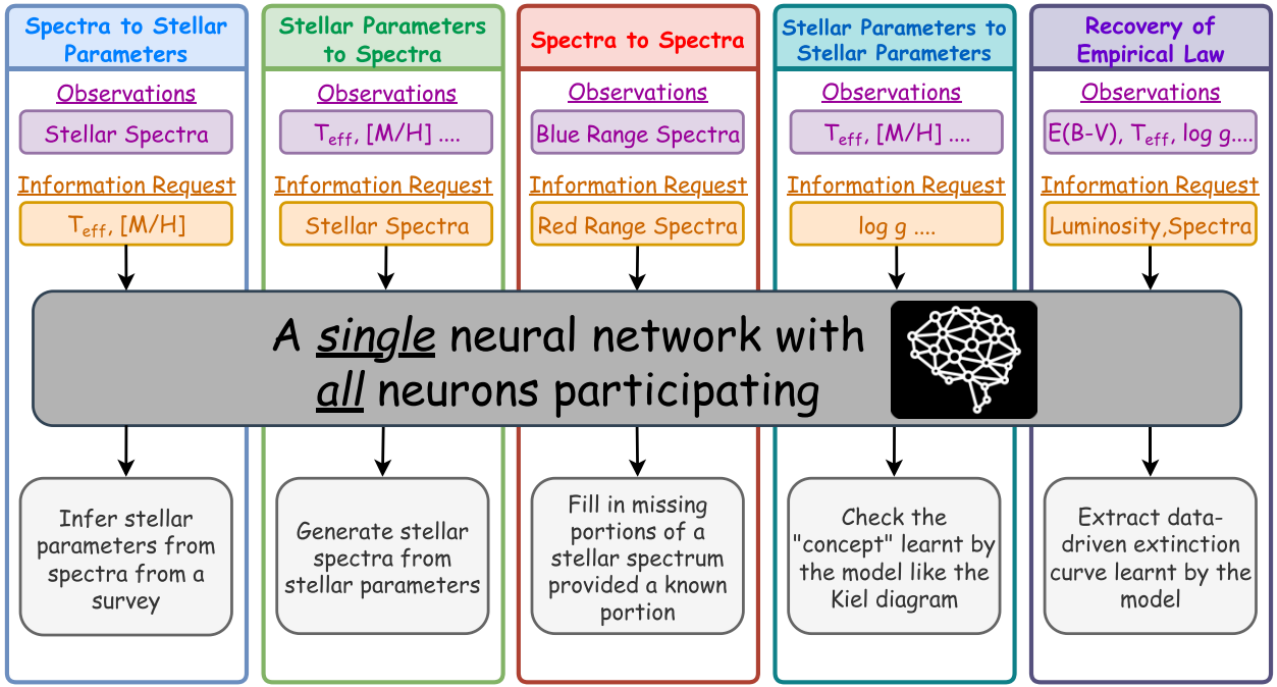


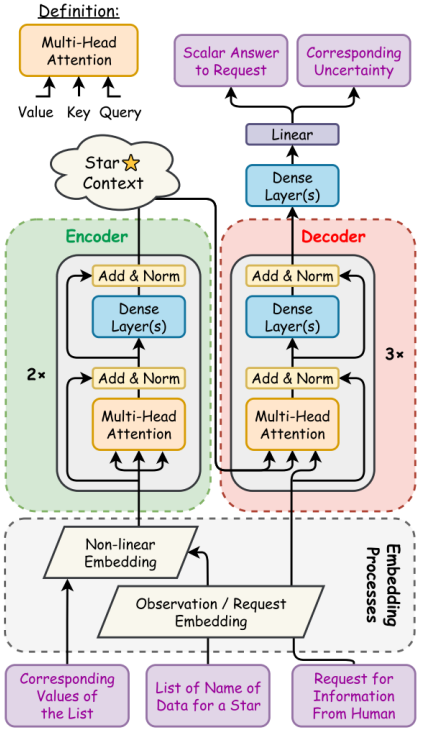


Foundation model for stars

- スペクトルデータに関する系列変換モデル

Leung & Bovy (2024, MNRAS)





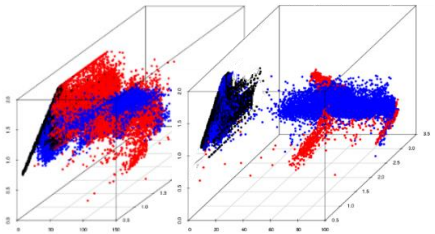
故障予兆検知

- 機械故障の予兆を検知し予防的保全を実現
  - データ不足により、汎用的な学習器を構築できず、個体差や環境変化に対応できない

    ⇒ 故障の物理特性に基づいて、限られたデータから適用可能な学習器を構築

ST+ (2017, AISI)
ST+ (2018, ICMLA)

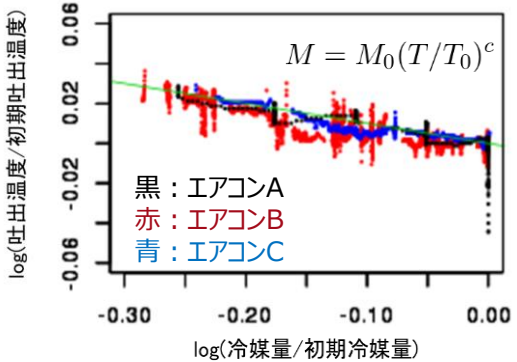
エアコンA（室内機2台）　エアコンB（室内機8台）

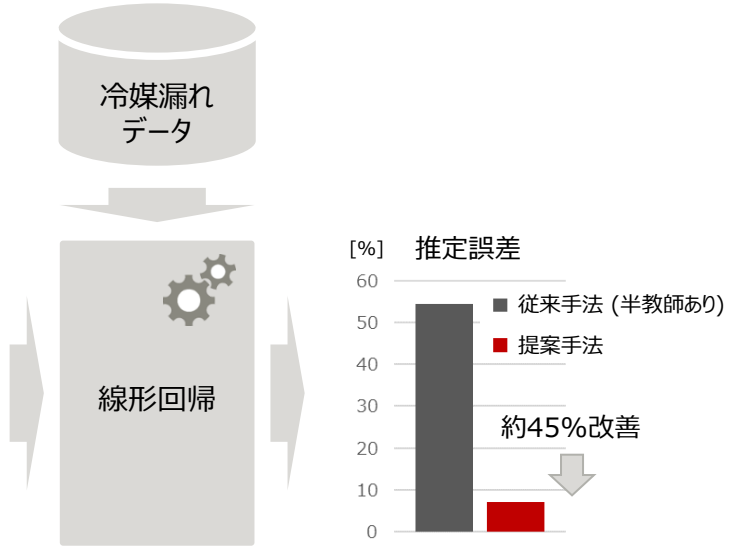


エアコンごとにセンサ値分布が異なり、学習器を他エアコンに適用できない

物理的知識（自然科学）

冷媒漏れが従う流体力学的なスケーリング則



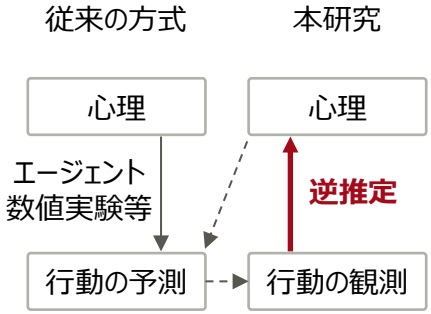


心理構造推定

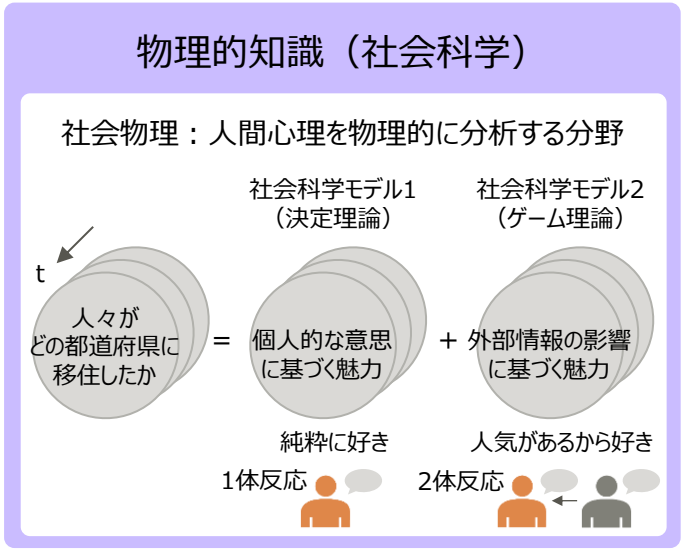
ST & Aguilar (2019, ICGDA)
ST+ (2020, CSSJ workshop)

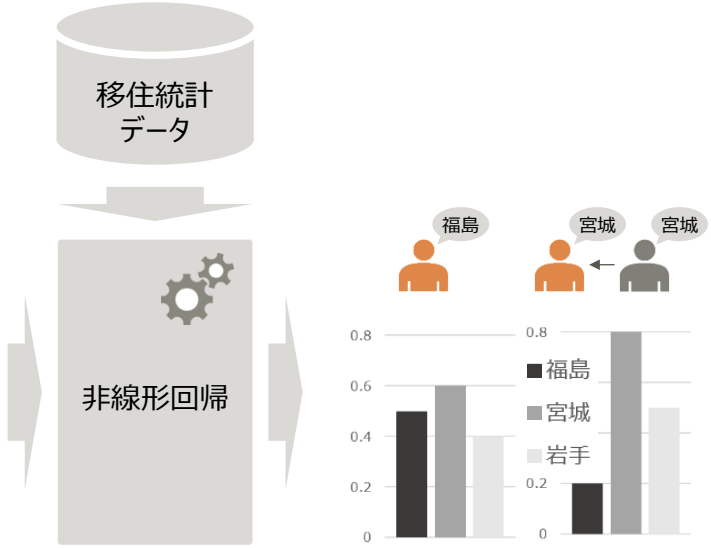
- 人間心理を如何に推定するか
  - 2つの社会科学モデルを統一的に定式化した物理モデルを採用

⇒ 心理の基礎理論に基づいて、汎用的に心理状態を推定する手法を開発



人間心理は直接測定不可
心理情報は人手で与えれてきた





まとめ

- **AI社会実装**　Vision AIによる行動分析、LLMとの統合
- **学習手法の高度化**　教師なしドメイン適応学習、自己教師あり学習
- **宇宙物理とAI**　基盤モデルへのシフト、AI研究者との連携

Thank you