

データサイエンスアワード2017  
プレゼンテーション資料

# 不動産データセット提供がもたらす 産学連携型オープンイノベーション

LIFULL Lab 主席研究員

清田 陽司

L LIFULL J

# LIFULLグループについて

## 沿革

- 1997年 株式会社ネクスト設立
- 2006年 東証マザーズ上場
- 2010年 東証第一部に市場変更
- 2017年 株式会社LIFULLに社名変更

## 概要

- 連結売上高 299億円 (2017年3月期)
- 主力事業
  - 総掲載物件数No.1の不動産・住宅情報サイト「LIFULL HOME'S」の運営
- その他事業
  - LIFULLブランドによる生活密着型情報サービスの提供
  - Trovit: 世界50ヶ国向けに展開する不動産・中古車・求職情報のアグリゲーションサービス (本社: スペイン・バルセロナ)



▼ LIFULL HOME'Sとは？



おうち見つかる、  
ホームズくん。



..... CM公開中 .....

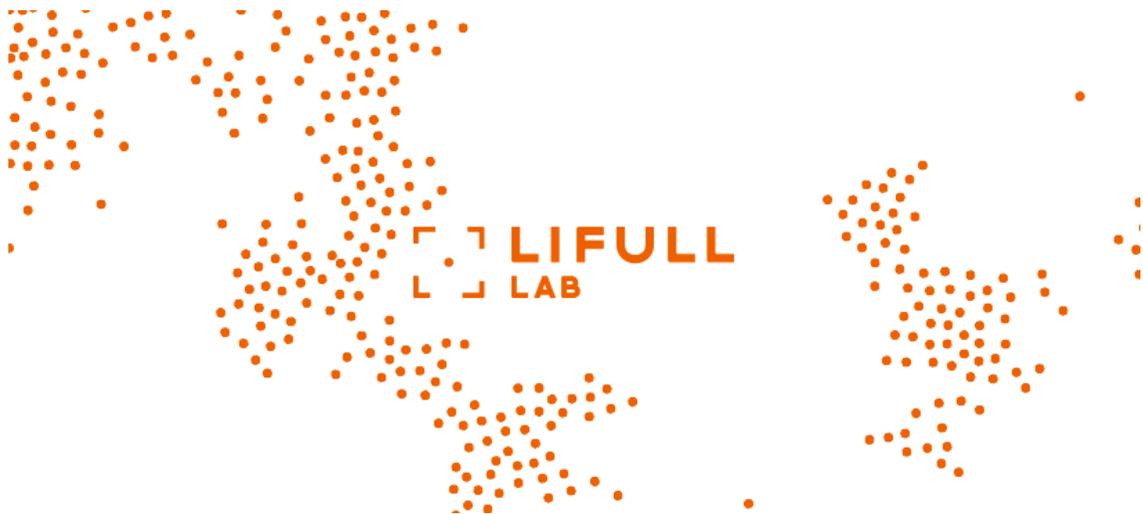
「ロッククライミング」 Ver.

30秒ver.



ロッククライミング版メイキング映像

# LIFULL Lab (リッテルラボラトリー)



- ルーツは2007年設立の  
東京大学発スタート  
アップ(株)リッテル

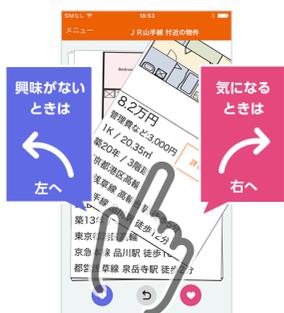
- **不動産テック**などに関  
わる最先端技術のR&D

- レコメンデーション
- AI関連技術 (ディープ  
ラーニングなど)
- ユーザーインタ  
フェース (UX)
- VR / AR / MR

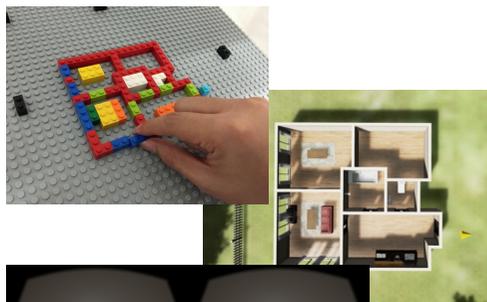
新感覚お部屋探しアプリ  
HOME'Sへヤサク!

部屋作りシミュレーション  
GRID VRICK

カードをサクサクめくって  
素敵なお部屋を見つけよう!



アプリがあなた好みの  
お部屋を学習!



**少人数 (2017年10月現  
在6名) で多岐にわた  
る分野をカバー**

# 不動産テックの位置づけ

(McKinseyのレポート)

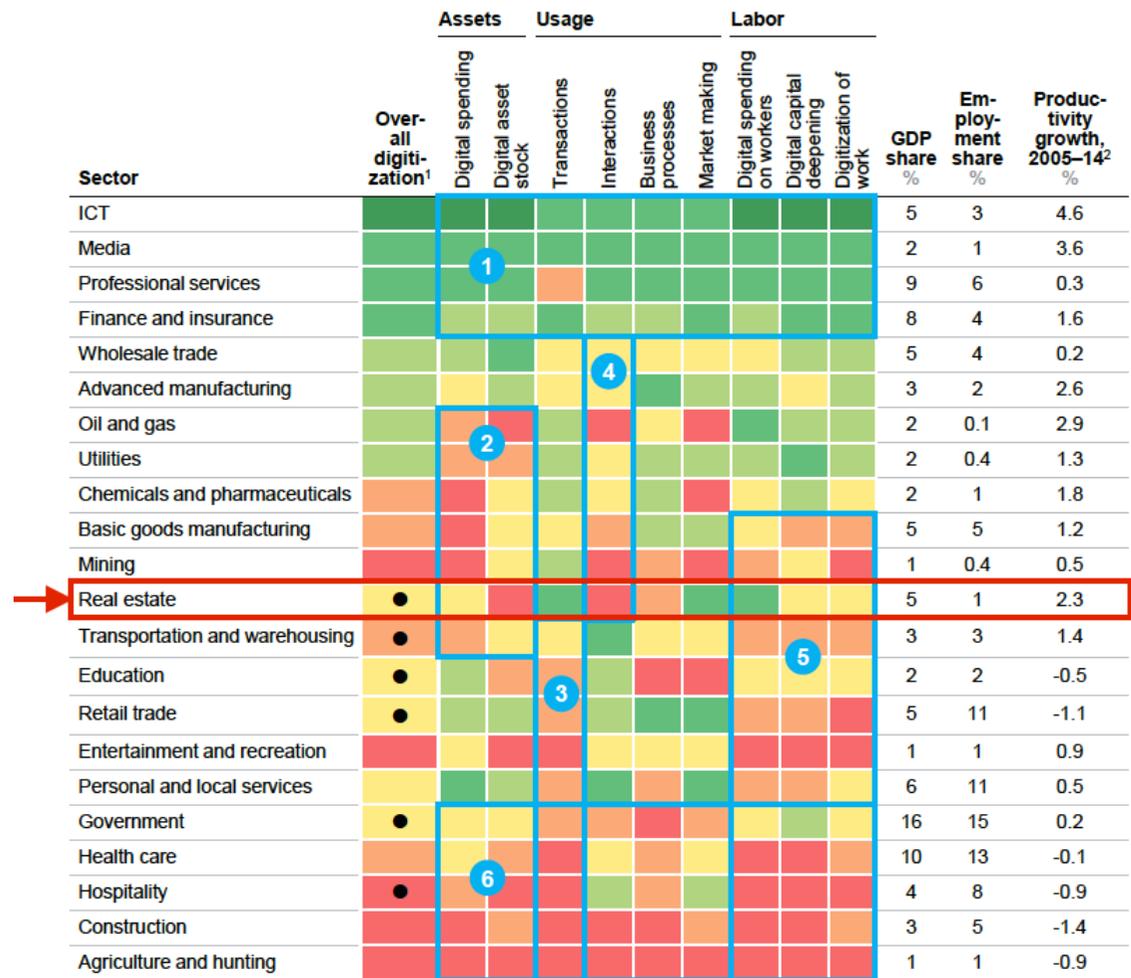
- 他の「Xテック」と比較するとデジタル化の度合いは中間的な位置づけ
  - ターゲティング広告など、すでに高度にデータ活用
  - 一方で、接客や内見などは人に頼る部分が多い
- 今後のデジタル化浸透にともなう変革が期待される分野

The MGI Industry Digitization Index

2015 or latest available data

Relatively low digitization  Relatively high digitization

● Digital leaders within relatively undigitized sectors



- 1 Knowledge-intensive sectors that are highly digitized across most dimensions
- 2 Capital-intensive sectors with the potential to further digitize their physical assets
- 3 Service sectors with long tail of small firms having room to digitize customer transactions
- 4 B2B sectors with the potential to digitally engage and interact with their customers
- 5 Labor-intensive sectors with the potential to provide digital tools to their workforce
- 6 Quasi-public and/or highly localized sectors that lag across most dimensions

# 不動産テック特有の課題

- 適切な物件価値の推定の難しさ
- 網羅的な物件データベースの未整備
- 消費者ニーズ把握の難しさ
- **Virtual Reality (VR) 活用の壁**

# VR活用の壁

- VRデバイスの普及
  - 「現地に行かなくても物件の状況を確認できる」というニーズを満たすことへの期待
- 3Dデータ作成が高コスト
  - 3Dスキャンを行うのにかかる時間は1物件あたり数時間
  - 適用分野は新築分譲マンションや高額物件など、コストに見合うものに限られる

MatterPort による物件3Dデータの作成



<https://matterport.com/how-it-works/>



<https://neighborhoodloans.com/what-is-matterport-and-why-you-need-it/>

# 不動産テックの課題解決に向けて

- 適切な物件価値の推定の難しさ
- 網羅的な物件データベースの未整備
  - **研究に使えるデータ資源が必要**
- 消費者ニーズ把握の難しさ
  - **協調フィルタリングなどの汎用的な推薦アルゴリズムとは異なるモデルの研究が必要**
- Virtual Reality (VR) 活用の壁
  - **すでに存在するデータ (写真や間取り図) から3Dデータを生成できないか？**

# LIFULL HOME'Sデータセット 提供の取り組み

# LIFULL HOME'Sデータセット

## 国内の賃貸不動産物件データ

**約530万件**

- 所在地 (郵便番号、最寄り駅など)
- 賃料、面積、築年数、部屋タイプ etc.
- 建物構造 (木造、鉄骨、鉄筋コンクリート etc.)
- 各種こだわり条件 (ペット可、楽器、カウンターキッチン、バス・トイレ別 etc.)

### 物件画像

**約8300万点**



重厚な感じの  
エントランス



日当たりの  
良いリビング

### 間取り図

**約510万点**



NII 国立情報学研究所  
National Institute of Informatics

情報学研究データリポジトリ

HOME データ一覧 組織 関連リンク お問い合わせ

> HOME > データ一覧 > HOME'Sデータセット

HOME'Sデータセット

国立情報学研究所が株式会社ネクストから提供を受けて研究者に提供しているデータセットです。  
2016/02/01 更新

データ概要

不動産・住宅情報サイトHOME'Sに掲載されたデータです。2015年9月時点のスナップショットとなっています。

- 賃貸物件データ (全国約533万件)  
賃料、面積、立地 (市区町村、郵便番号、最寄り駅、徒歩分)、築年数、間取り、建物構造、諸設備などのデータです。IDはユニーク番号に変換済みで、特定の物件に紐づく属性は含んでおりません。TSV形式のファイルで約1.6GBです。
- 画像データ (約8,300万ファイル)  
上記全物件に対する間取り図や室内写真などのデータです。最大横120ピクセル×縦120ピクセルのJPEG形式で、圧縮ファイルで約210GBとなります。メタデータはTSV形式のファイル (約7.3GB) で、「玄関」「キッチン」といった画像の種類や、一部にはフリーテキストによる説明が付与されています。
- 高精細度画像データ (約515万ファイル) 2016-02-01 新規  
上記画像データのうち、間取り図に関しての高精細度版の画像データです。JPEG形式で、圧縮ファイルで約140GBとなります。本データに関しては別途お申し込みが必要ですが (詳細は下記「お申し込み」欄をご覧ください)。

Yahoo!データセット  
楽天データセット  
ニコニコデータセット  
リクルートデータセット  
クックパッドデータセット  
HOME'Sデータセット  
不満調査データセット  
国文研データセット  
NTCIRテストコレクション  
音声コーパス

- **国立情報学研究所(NII)の協力を得て2015年11月より提供開始**
- **学術機関の研究者であれば無料で利用可能**

# データセット提供のねらい

- 限られた研究リソースをもとに、**オープンイノベーション**により成果を最大化すること
- 「不動産テック」に関連する**学際的研究コミュニティ**の活性化および人材育成への貢献
- 成熟しつつある**画像処理技術**（とくにディープラーニング）の応用研究への貢献

# データセット提供のリスクヘッジ

## 主なリスク

- 現居住者やオーナーのプライバシー侵害
- 地域情報のセンシティブ性
- データセット利用者の法的リスク (宅建業法など)
- 自社ビジネスへのマイナスの影響
- ステークホルダー（不動産会社など）からの反発

## どうやってヘッジするか？

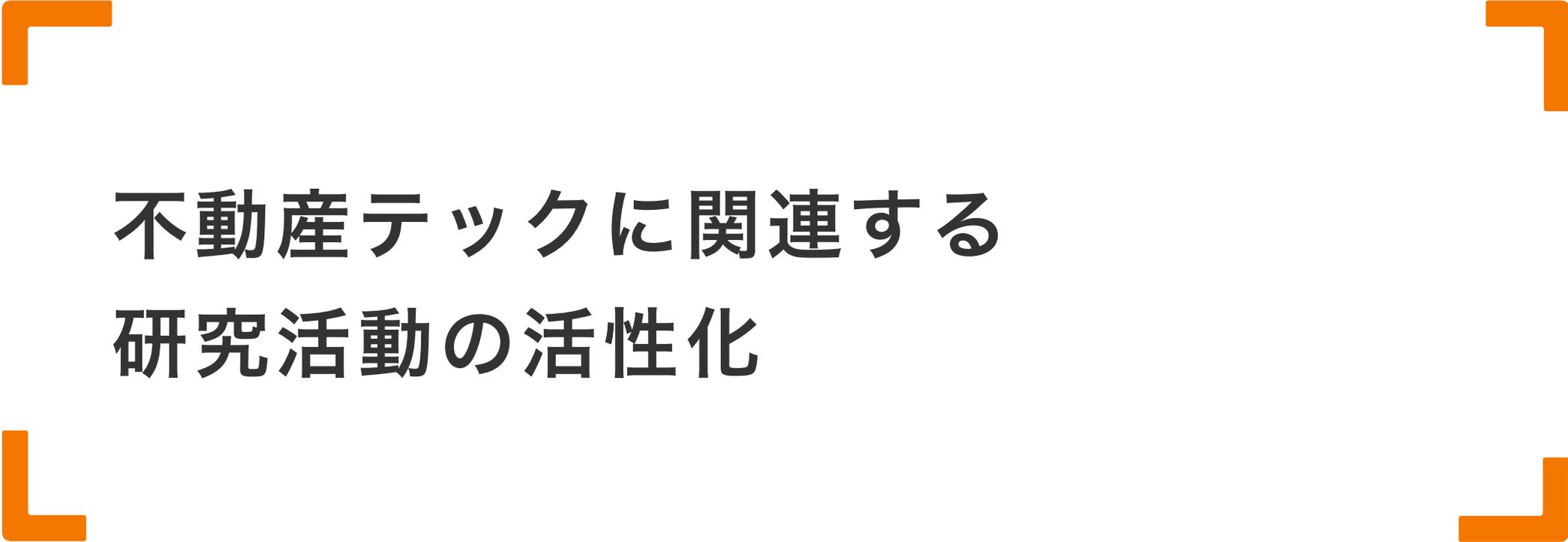
- 覚書・注意事項
- 物件を直接特定する情報は含めない
  - 最寄り駅、町域レベルまで
- 利用用途を学術研究目的に限定
- 特許の取得

# データセット利用の状況

**日本国内の50以上の大学研究室・  
研究機関ほか、米国・カナダでも  
利用**



データセット提供によって  
創出されたイノベーション



# 不動産テックに関連する 研究活動の活性化

# 主な研究発表

## データセット提供開始時に想定していなかった利用方法が次々に生み出されている

- 井上, 高橋, 裴, 高木: 受容度に基づく多数目的探索のお部屋探し, 第11回進化計算学会研究会資料集, pp. 183–189 (2016)
- 近藤, 和泉, 山田, 吉田: Twitter から抽出した地域特性による不動産価格の分析, NII IDRユーザフォーラム2016 <http://www.nii.ac.jp/dsc/idr/userforum/poster/IDR-UF2016 P10.pdf>
- 高田, 井上, 山崎, 相澤: 深層学習量を用いた類似間取り図検索, 人工知能学会全国大会 OS-15 「不動産とAI」 (2017)

# 人工知能学会誌上での 特集企画「不動産とAI」



不動産とAI

529

【特集】「不動産とAI」【エディトリアル】

## 不動産とAI

Real Estate and AI

清田 陽司 株式会社 LIFULL リッセルラボラトリー  
Yoji Kiyota Lital Laboratory, LIFULL Co., Ltd.  
kiyota@lital.t.u-tokyo.ac.jp, https://lifull.com/

山崎 俊彦 東京大学大学院情報理工学系研究科電子情報学専攻  
Toshihiko Yamazaki Department of Information and Communication Engineering, The University of Tokyo.  
yamazaki@hal.t.u-tokyo.ac.jp

諏訪 祐彦 奈良先端科学技術大学院大学  
Hirotaka Suwa Nara Institute of Science and Technology  
h-suwa@is.naist.jp, https://ubi-lab.naist.jp/

清水 千弘 日本大学スポーツ科学部、マサチューセッツ工科大学不動産研究センター  
Chikuo Shimizu College of Sports Science, Nihon University / Center for Real Estate Studies, Massachusetts Institute of Technology  
shimizu.chikuro@hon-u.ac.jp

**Keywords:** real estate, artificial intelligence, price estimation, computer vision, deep learning, data resource, heterogeneity, deregulation, human resource development.

### 1. はじめに

ヒトが日常生活を営むうえで絶対に欠かせない基礎的な要件を、私達は「衣食住」という言葉で表現する。「衣」は寒さ・暑さをしのぐとともに、社会的動物としてのヒトが生きるために欠かせないものである。「食」は「生きる」ことと同義といってもよいであろう。「住」は雨風をさえぎるとともに、安心して眠りにつくことができる場を与えてくれるものであり、「衣」「食」と同様に、私達の生活に与える影響はきわめて大きい。当然の帰結として、「衣食住」に関連する産業は、それぞれ巨大な市場を形成している。

現在、あらゆる産業領域で急速に進行中の人工知能(AI)の活用の流れは、「衣食住」に関連する産業でも例外ではない。「衣」や「食」に関していえば、生産、在庫管理、物流の最適化はもちろんのこと、アルゴリズムによる商品の推薦、さらにはビッグデータ分析を通じて把握した消費者ニーズのマーケティングや商品開発への活用など、サプライチェーンのあらゆる段階でAIの活用は急速に進んでいる。

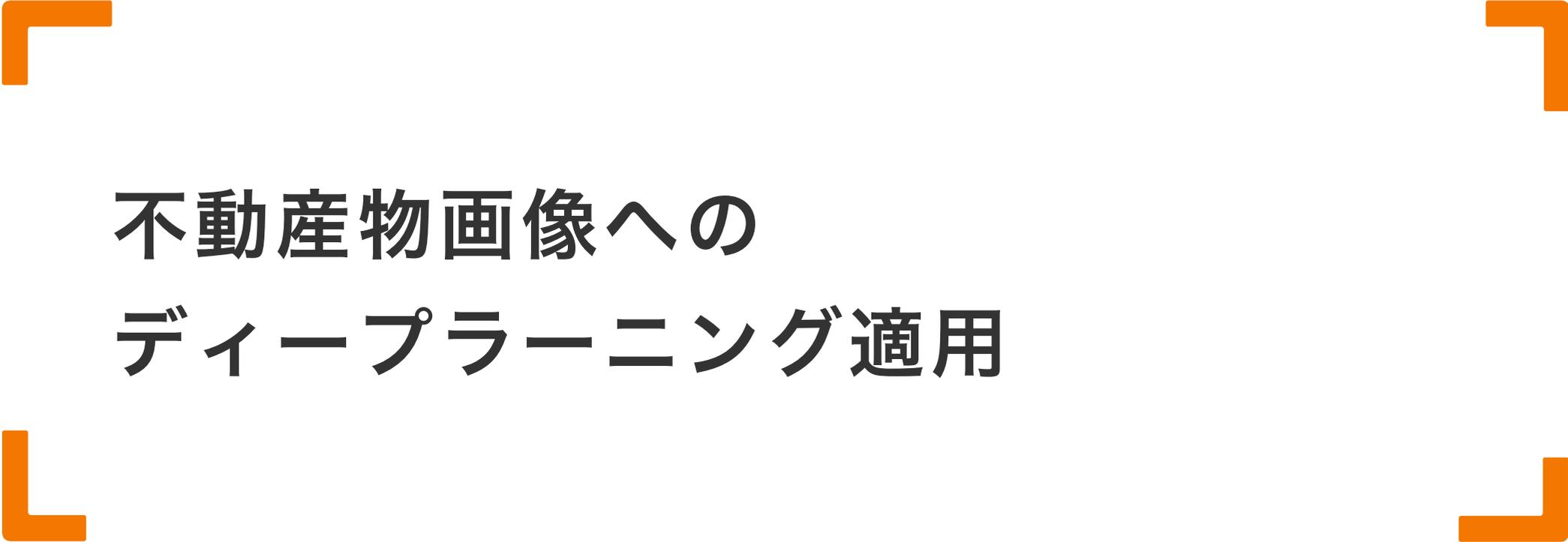
それでは、「住」に関連する産業である不動産でのAIの活用はどのような現状にあるのだろうか？ 現在、不動産業をビジネス領域とするAIを活用したユニコーン企業が世界中に数多く生まれており、日本でもAIを活用した不動産スタートアップ企業が立ち上がりつつある。インターネットによるマーケティングや集客を中

心に、テクノロジーの普及は著しい。一方で、不動産の購買プロセスの複雑さ、実世界の不動産の情報をどのように収集・取得するか、さらには独特の価値観や法規制など、AIの活用にあたって懸念となるさまざまな問題も存在している。

AIの活用にあたっては、あらゆる情報がデジタル化されていることが前提となる。山崎の記事に引用されているマッキンゼーによる各産業領域でのデジタル化の現状を調査したレポートの図を参照すると、PinTechというキーワードが人口に膾炙した金融などではあらゆる段階でデジタル化が進んでいる一方、ヘルスケアや農業などでどの段階でもデジタル化は低いレベルにとどまっていることがわかる。不動産業は、ちょうど両者の中間的な位置付けとなっている。取引のトランザクション管理やマーケティングなどではデジタル化が進んでいるのに対し、在庫管理や接客などまだアナログな部分も多く残っている。

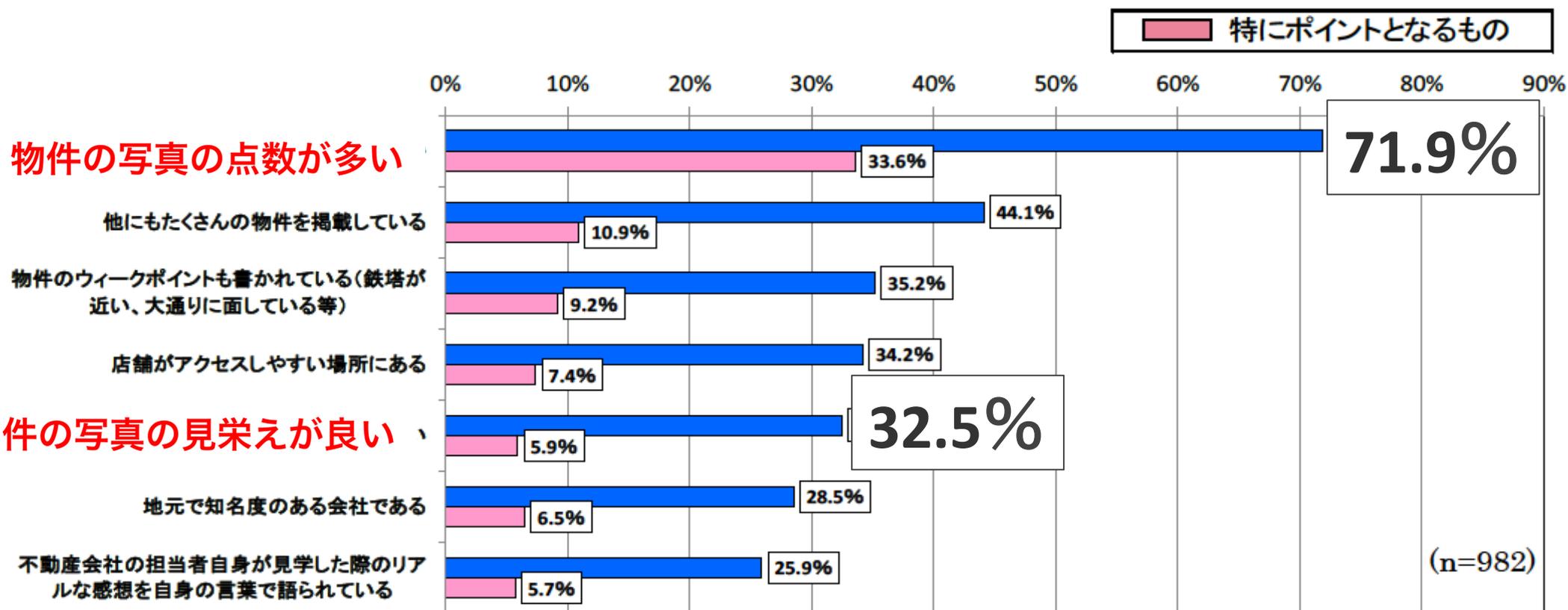
複雑な実社会の問題をデジタルの世界に載せ、AIなどのテクノロジーを活用していくうえで、研究コミュニティの果たす役割、特に学際的な研究の活性化は欠かせない。AI活用の先導である金融産業でアルゴリズムの利用が急激に浸透した背景には、金融工学という新たな分野に、コンピュータ科学、数学、物理学、心理学などの分野からの人材が大量に流入し、さまざまな分野の知見が生かされたことがあった。不動産業に関する学際分野としては、不動産学のほか、経済学、建築学、都市学などが存在し、コンピュータ科学やAIの分野でも、

2017年7月号に掲載、記事11本・61ページ



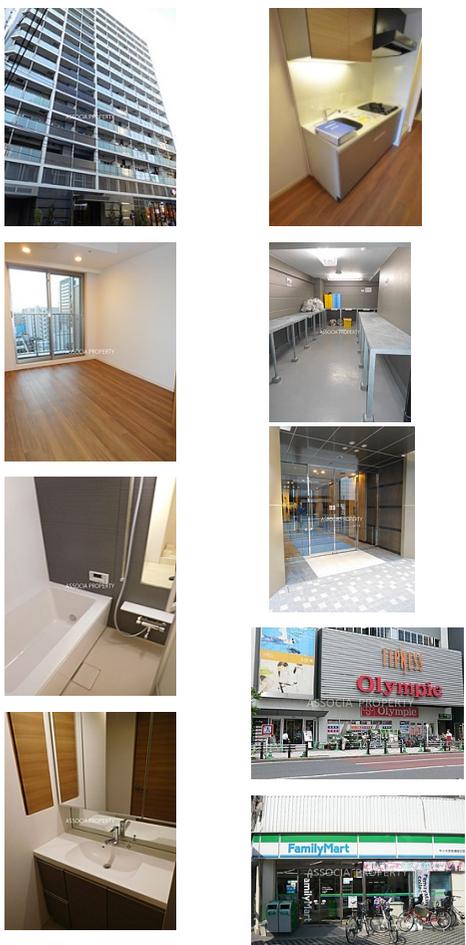
# 不動産物画像への ディープラーニング適用

# 不動産情報サイトで不動産会社を選ぶ際のポイントは？ (RSC 2015年調査)

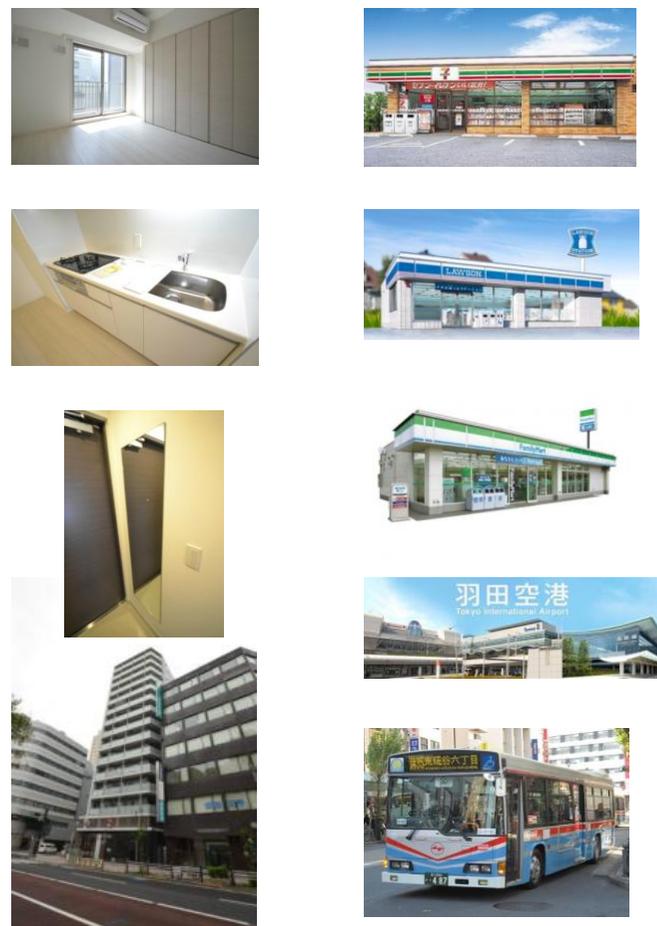


# どちらの物件情報がユーザーにとってより嬉しい？

## A社



## B社



# 画像種別の深層学習(CNN)による判別

- > どの種類の物件画像なのかを判定する
- > LIFULL HOME'Sデータセット画像に付けられているタグを正解データとして使用
  - 13タグ分の画像を使用
  - それぞれ訓練用1万枚、検証用1000枚

間取り



風呂



居間



外観



周辺



## 使用するタグ

- ◆ 間取り
- ◆ 洗面
- ◆ 地図
- ◆ 収納
- ◆ 玄関
- ◆ 設備
- ◆ 居間
- ◆ バルコニー
- ◆ キッチン
- ◆ エントランス
- ◆ 風呂
- ◆ 駐車場
- ◆ トイレ

## 使用しないタグ

- ◆ 外観
- ◆ 周辺
- ◆ 内装
- ◆ その他
- ◆ 寝室
- ◆ 子供部屋

# 画像種別の判別精度

> テストデータのError rate: 0.143

正答サンプル



キッチン97.3%



リビング52.0%



間取り91.0%

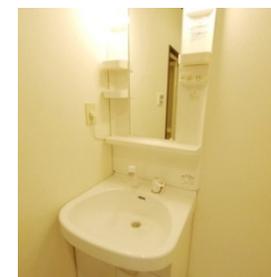


風呂100.0%

不正答サンプル



正解：リビング  
結果：収納  
収納 40.1%  
リビング 20.6%  
洗面 12.1%



正解：洗面  
結果：風呂  
風呂 64.7%  
洗面 26.2%  
トイレ 3.1%



正解：収納  
結果：風呂  
風呂 75.0%  
トイレ 9.2%  
バルコニー 3.7%

# NEWS

HOME > NEWS > サービス・商品 >  
AIによる物件の不整合画像検出を開始

## AIによる物件の不整合画像検出を開始

2016年12月19日

 Tweet  Like 0  B! 0

株式会社ネクストは、この度ディープラーニングを活用した物件データベースの不整合画像検出を開始いたします。

### ■ディープラーニング(深層学習)による不整合画像検出の概要

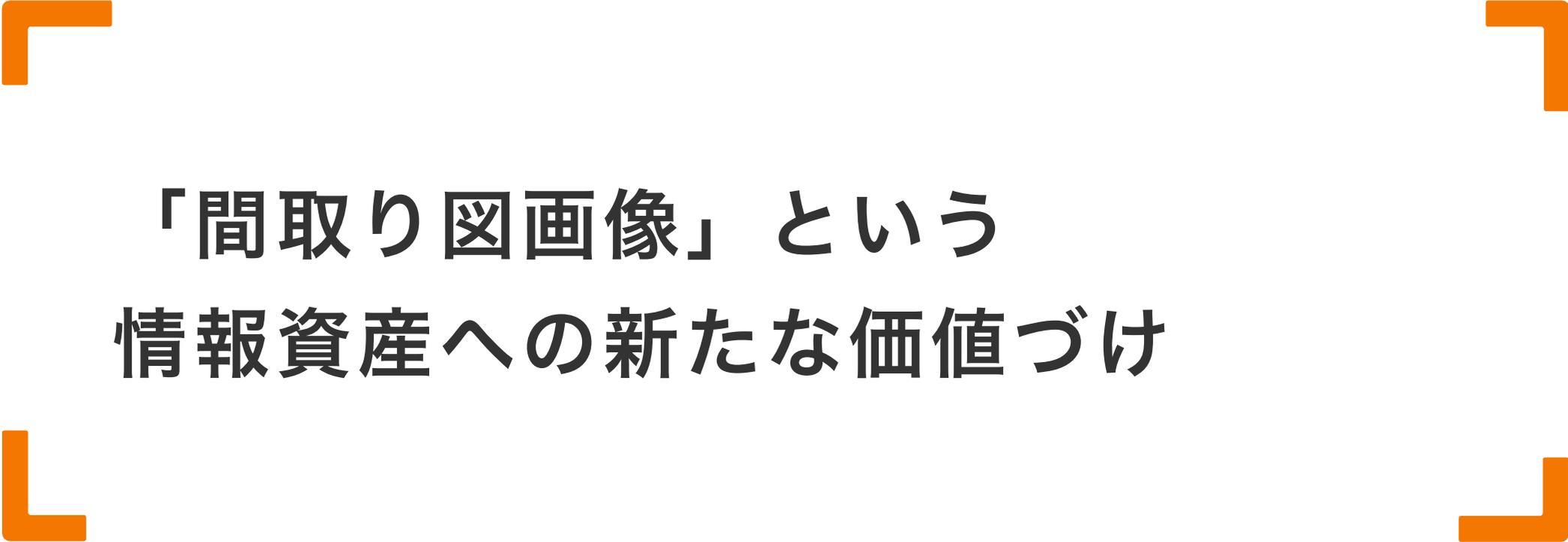
ディープラーニングとは、人間の脳の神経回路の仕組みを模して開発されたアルゴリズムです。予め定義したルールに基づく従来の機械学習とは異なり、自動で特徴を識別することで、認識精度の飛躍的な向上が可能となりました。

この度導入する物件データベースの不整合画像検出では、まず『HOME'S』が大量に保有している「間取り」、「キッチン」、「居間」などの物件に関する画像データをシステムに学習させます。次に学習した画像の特徴と不動産事業者が登録した画像を照合し、登録した物件と画像が同一か、登録項目と画像の相違が大きいかを判定し、相違が大きいものを「不整合画像」として検出します。

入稿画像	登録項目	整合率	判定結果
	<p>キッチン</p>	<p>キッチン 70.3% 玄関 23.7% 設備 2.804%</p>	
	<p>キッチン</p>	<p>キッチン 97.3%</p>	
	<p>キッチン</p>	<p>キッチン 0% 玄関 0%</p>	

# 不動産物件画像の 不整合検出による効果

- 室内の写真が充実した物件情報など、**エンドユーザーにとって価値の高い情報を優先して正しく届けることが可能に**
- 日々数百万点にのぼる入稿物件写真を**最小限の人手でチェック**できるように



# 「間取り図画像」という 情報資産への新たな価値づけ

# Washington Univ. in St. Louis 古川泰隆准教授のグループによる LIFULL HOME'Sデータセット利用研究

<https://arxiv.org/pdf/1612.01225.pdf>

## Deep Multi-Modal Image Correspondence Learning

Chen Liu<sup>1</sup>

Jiajun Wu<sup>2</sup>

Pushmeet Kohli<sup>3</sup>

Yasutaka Furukawa<sup>1</sup>

<sup>1</sup>Washington University in St. Louis

<sup>2</sup>Massachusetts Institute of Technology

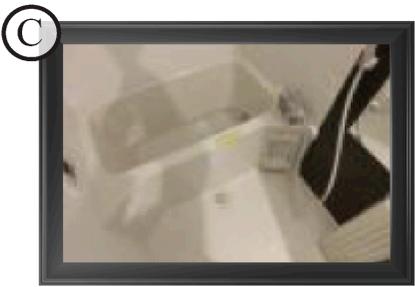
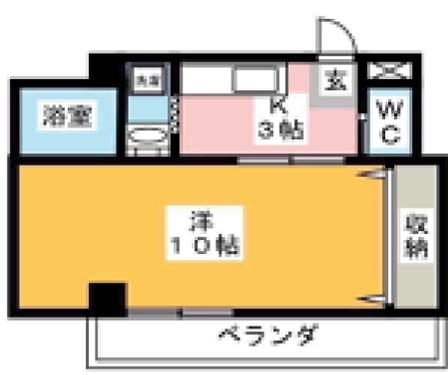
<sup>3</sup>Microsoft Research



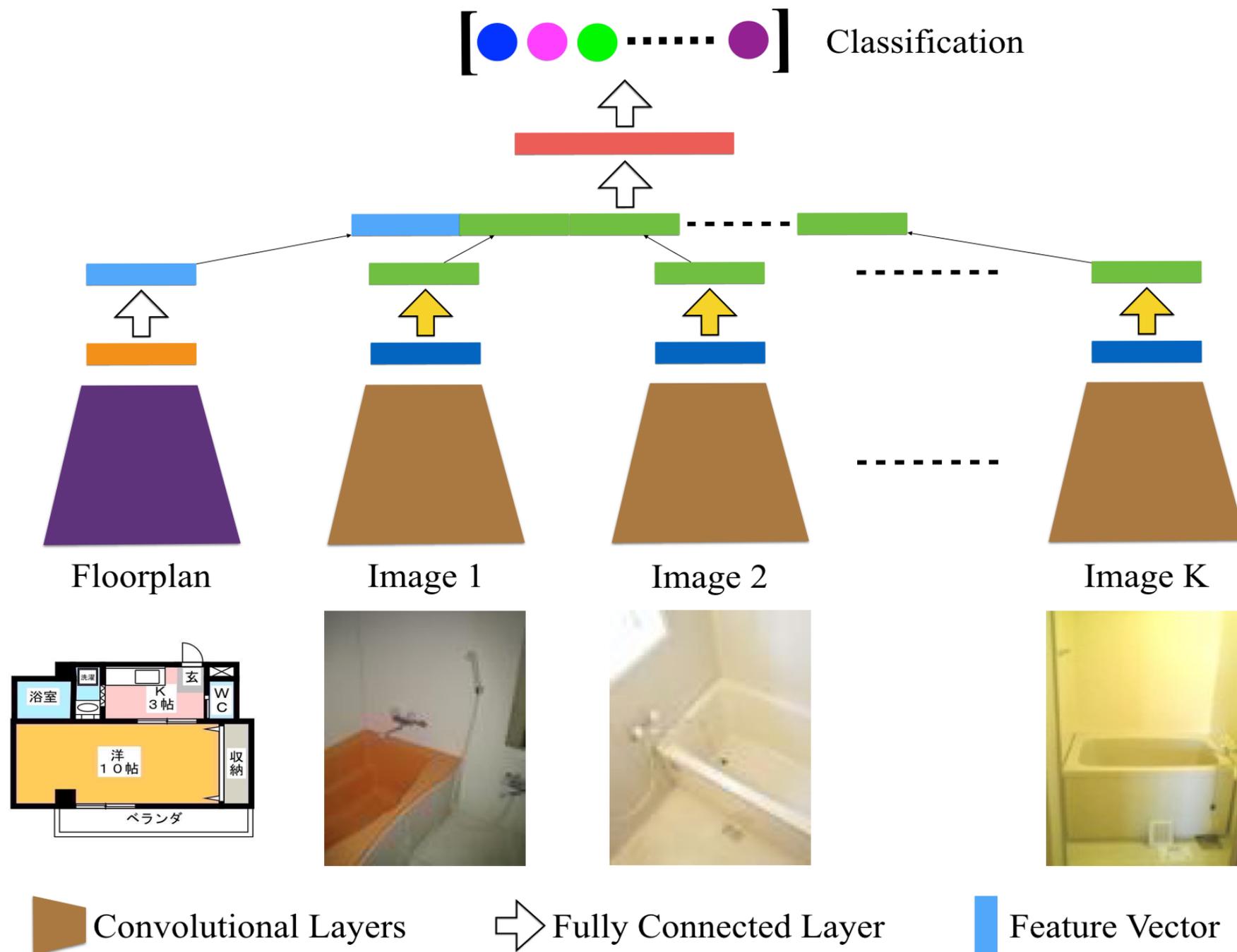
Figure 1: Which of the four photographs corresponds to the left floorplan? The task requires careful and sophisticated human reasoning, unlike many other computer vision problems that only require instant human attention. This paper explores the potential of deep-neural networks in solving such a problem. The answer is in the footnote<sup>2</sup>.

# ディープ多形態対応学習

問: 左の間取り図に対応する正しい浴室写真はどれ？



# どの画像が左の間取り図に対応するかを判別するNN



# 判別精度

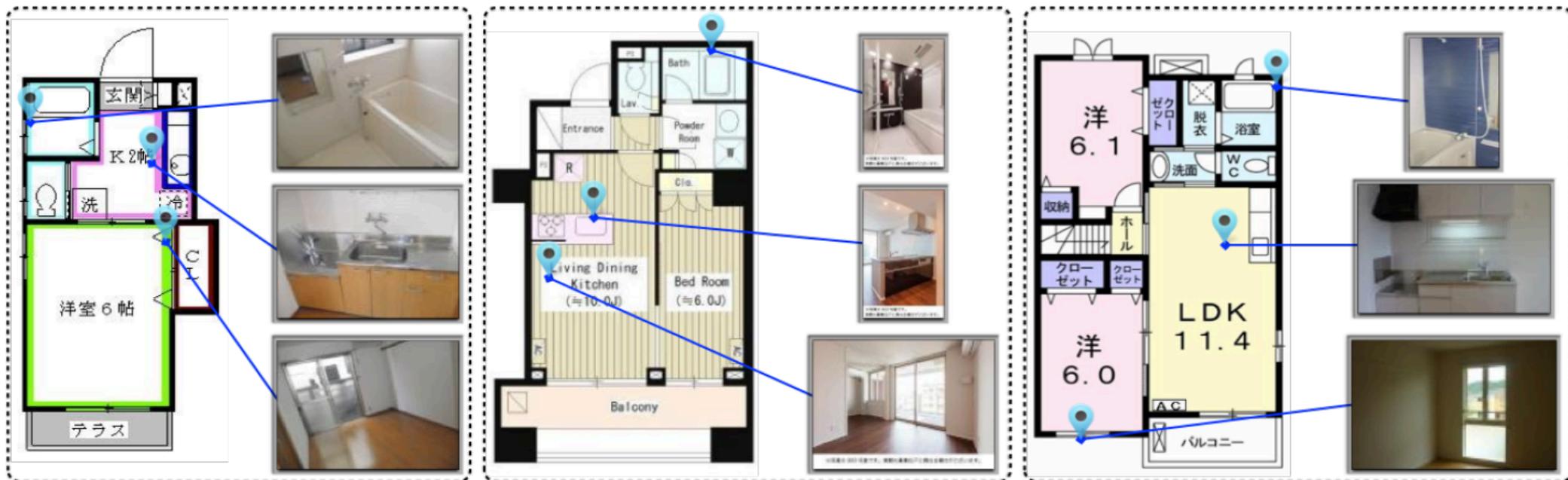
候補の数 ( $\alpha$ )	実画像の種類 ( $\beta$ )			
	浴室	台所	リビング	全三種類
pair (1-way)	82.3 $\pm$ 1.6 (51.7)	81.8 $\pm$ 2.1 (58.9)	77.8 $\pm$ 1.8 (59.5)	85.3 $\pm$ 3.4 (61.5)
2-way	86.2 $\pm$ 1.4 (64.1)	84.8 $\pm$ 3.5	81.2 $\pm$ 1.6	91.0 $\pm$ 1.5
4-way	72.4 $\pm$ 3.6 (43.0)	72.4 $\pm$ 1.8	66.5 $\pm$ 1.7	77.8 $\pm$ 2.5
8-way	56.9 $\pm$ 1.8 (42.0)	59.3 $\pm$ 1.9	54.0 $\pm$ 3.9	61.4 $\pm$ 2.5

※**緑色の数字**はAmazon Mturkのワーカーによる判別精度

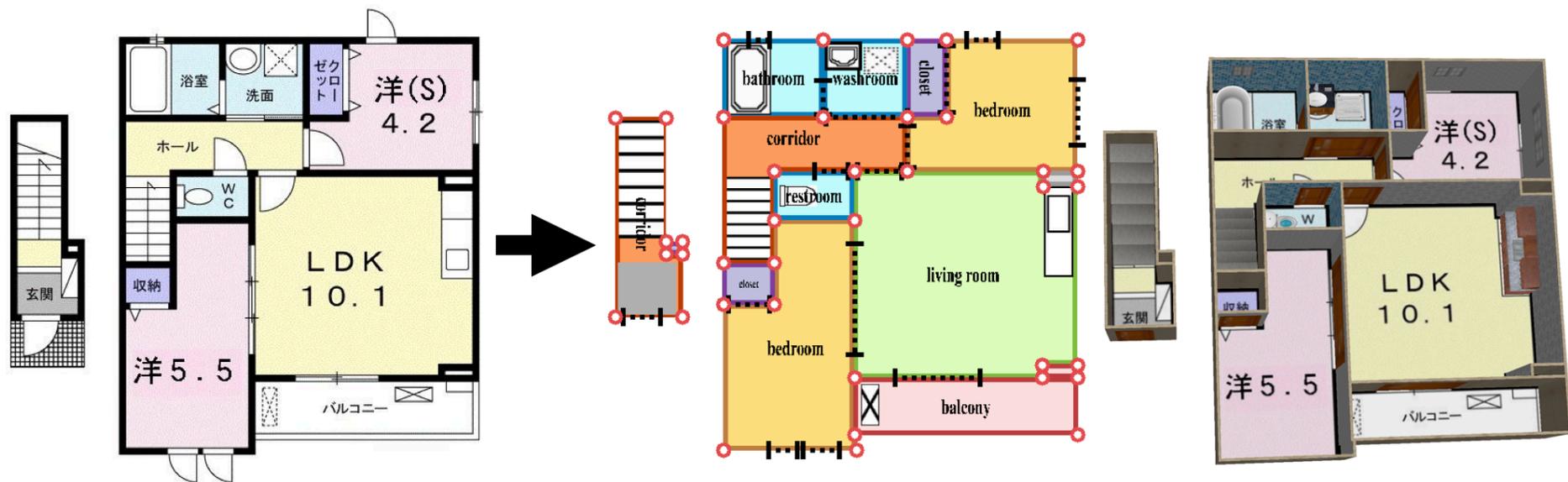
# 画像にノイズを加えて、写真が間取り図のどこに対応するかを識別



# 間取り図ナビゲーションへの応用



# 間取り図画像を解析して 3Dモデルを生成



古川准教授のグループによるLIFULL HOME'Sデータセット利用研究  
<https://www.cse.wustl.edu/~chenliu/floorplan-transformation.html>

→ 画像処理のトップレベル国際会議 ICCV 2017に採択

「」 LIFULL

あらゆる LIFE を、FULL に。