

間取り図を考慮した賃料予測モデル

服部凌典

学籍番号：1525059

先端工学基礎課程 IMCコース

岡本研究室

はじめに

唐渡 広志：ヘドニック・アプローチを利用した不動産価格指数の推定方法とその問題点,
都市住宅学, vol.2016, no.92, pp.17-20, 2016.

賃貸物件の特徴

- ・ 同じ物件が存在しない
- ・ 間取り, 階数, 立地などの物件属性が賃料に影響を与える

ヘドニック・アプローチ [唐渡, 2016]

商品の価格をその属性の価値に関する集合とし,
価格予測モデルを線形回帰で構築する手法

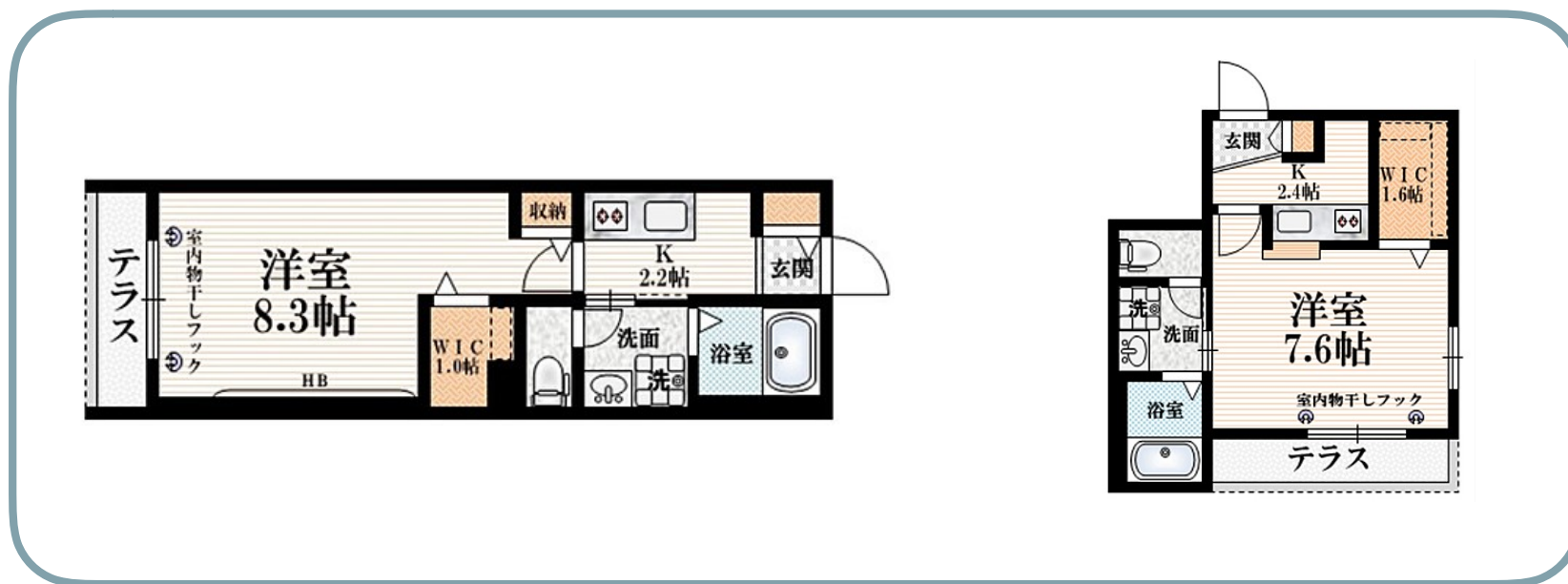
$$\hat{y}_i = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots + \alpha_n x_n + \beta$$

賃料	占有面積	立地	階数	築年数
----	------	----	----	-----

間取り図の特性

清田陽司, 山崎俊彦, 諏訪博彦, 清水弘: 不動産とAI, 人工知能, vol.32, pp.529-535, 2017.

間取りタイプが同じであっても, 部屋の形状が違うものが存在
希望の賃貸物件を探す際には間取り図を見る習慣がある[清田+, 2017]



賃料予測において間取り図の影響を明らかにすることは重要

研究内容

目的: 賃料予測における間取り図の影響を明らかにする

賃料の予測精度を比較

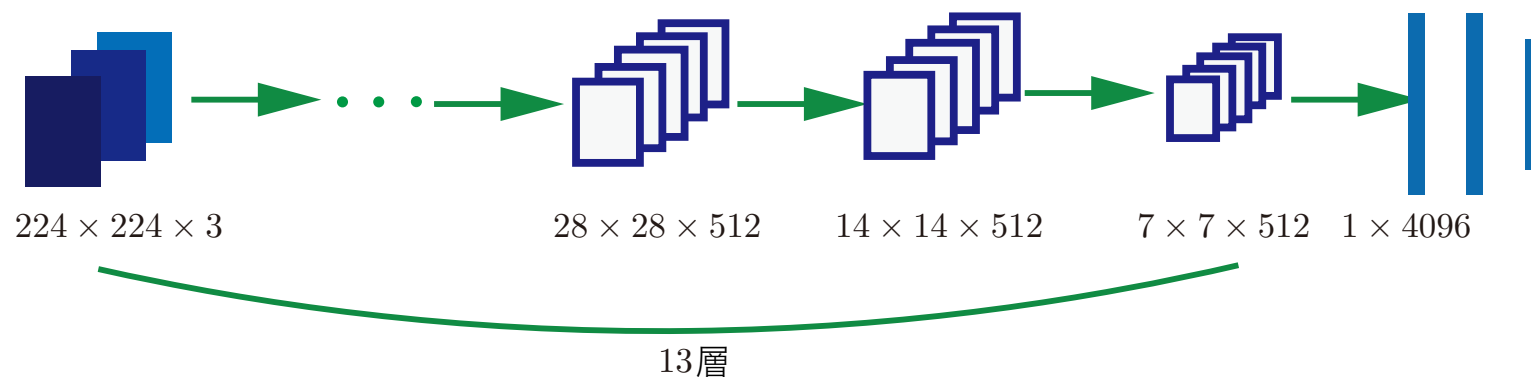
- ・ 間取り図を考慮しない予測モデル
- ・ 間取り図を考慮した予測モデル

モデル名	予測式	パラメータ
間取り図を考慮しない 予測モデル (LR)	$\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\alpha}^T \mathbf{x}_i + \beta$	$\boldsymbol{\alpha}, \beta$ を更新
主成分分析を用いた 予測モデル (PCA-LR)	$\hat{y}_i = f(\mathbf{u}_i, \phi_{\theta}(\mathbf{v}_i))$ 間取り図	θ を更新した後、 $\boldsymbol{\alpha}, \beta$ を更新
VGGを用いた 予測モデル (VGG-LR)	$\hat{y}_i = h_{\theta}(\mathbf{v}_i)$ $\hat{y}_i = f(\mathbf{u}_i)$ 物件属性	$\boldsymbol{\alpha}, \beta, \theta$ を更新

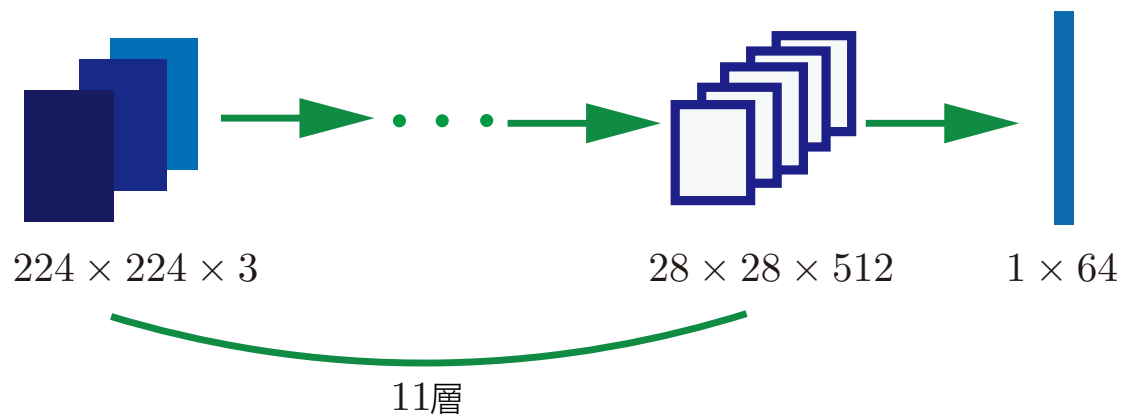
VGG16の構造

Karen Simonyan, and Andrew Zisserman : Very Deep Convolutional Networks for Large-Scale Image Recognition, Computing Research Repository, vol.abs/1409.1556, 2014.

VGG16 [Karen+, 2014]: VGG (Visual Geometry Group) によって開発された
畳み込みニューラルネットワーク



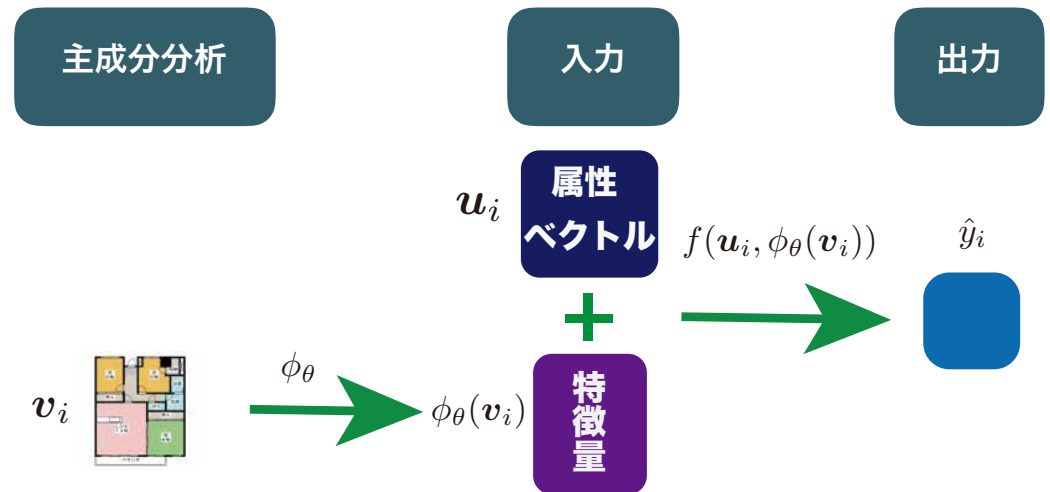
使用するVGG16



間取り図を考慮した予測モデル

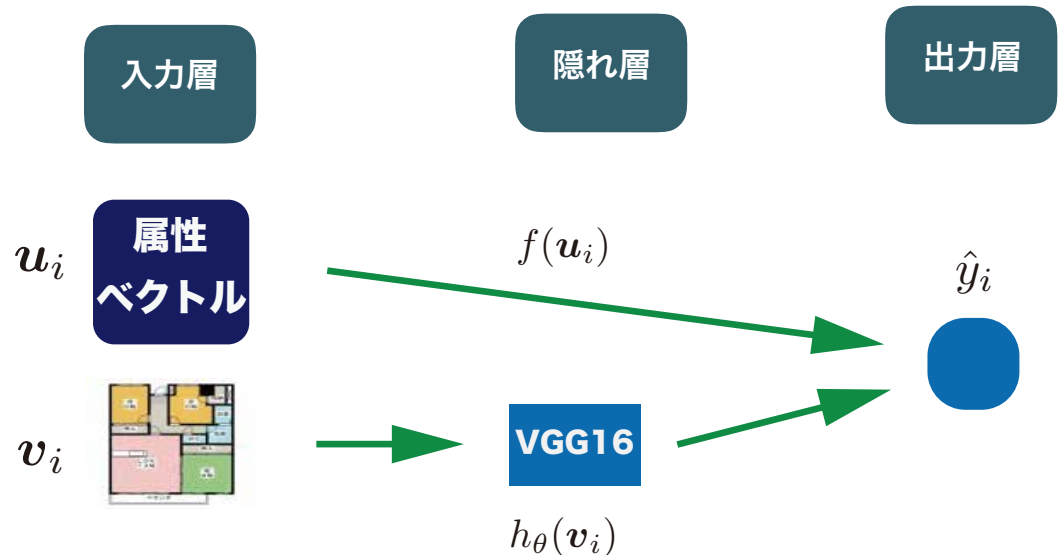
PCA-LR

- 主成分分析による特徴抽出
- 抽出した間取り図の特徴量と属性ベクトルを組み合わせ，線形回帰で賃料予測
- 間取り図の特徴量: 64, 128, 256, 512, 1024, 2048次元の6種類



VGG-LR

- ミニバッチ毎に確率的勾配降下法を適用
 - batch size: 100
 - epoch: 100
- 活性化関数: ReLU
- 間取り図の特徴抽出: 64次元



データセット

LIFULL HOME'Sデータセット（2015 年 9 月時点）

- ・ 賃貸物件データ（70属性, 533万件）
- ・ 画像データ（8,300万ファイル）

使用賃貸物件データ（14属性, 9万件）

- ・ 目的変数: 賃料 + 共益費
- ・ 説明変数: 徒歩距離, 立地, 階数（地上）, 階数（地下）, 市町村区, 部屋階数, 契約期間, 建物構造, 駐車場料金, 新築・未入居フラグ, 占有面積, 築年数, 間取り図

使用賃貸物件データを間取りタイプを基準にK, R, DK, LDKの4種類に分割
(間取りタイプがLDの物件も存在するが, 数十件のみのため本実験では対象外とする)

予測精度の評価方法

使用モデル

- LR, VGG-LR, PCA-LR (予測精度の高いハイパーパラメータを使用)

予測精度の評価

- データセットの20%をテストデータに, 80%を開発用データに分割
- 間取りタイプごとに, 開発用データを学習に使用
- テストデータを予測

PCA-LRにおけるハイパーパラメータのチューニング

- 間取り図の特徴量: 64, 128, 256, 512, 1024, 2048次元の6種類
- 開発用データに10分割交差検証法を適用

評価指標

- 平方平均二乗誤差 (RMSE: Root Mean Square Error)

PCA-LRのチューニング結果

RMSEの平均	64	128	256	512	1024	2048
K	10,886	10,733	10,649	10,548	10,630	11,293
R	9,134	9,104	9,054	8,967	8,916	8,790
DK	11,207	11,189	11,144	11,096	11,069	11,148
LDK	20,027	19,958	19,919	19,807	19,733	19,668

テストデータに対するRMSEと学習時間

テストデータに対するRMSE[円]

	データ数	VGG-LR	PCA-LR	LR
K	[1,841]	16,862	11,689	11,602
R	[7,997]	12,541	8,727	9,217
DK	[3,795]	16,467	10,957	11,156
LDK	[7,319]	22,878	19,646	20,121

開発用データの学習時間[秒]

	データ数	VGG-LR	PCA-LR	LR
K	[9,205]	2,897	11	0.06
R	[39,985]	14,566	48	0.08
DK	[18,975]	6,471	23	0.05
LDK	[36,595]	13,026	45	0.07

テストデータに対する価格帯別RMSE (1/2)

間取りタイプ: K

データ数	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	254		947		420		54		3		0	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	6,308	1,927	6,466	2,811	9,288	2,487	22,740	7,315	44,833	11,650	-	-
PCA-LR	5,140	3,935	5,863	4,577	8,508	6,256	24,556	22,291	54,623	56,182	-	-
LR-LR	4,959	4,118	6,059	4,939	10,060	8,997	31,256	33,133	76,081	77,904	-	-

間取りタイプ: R

データ数	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	964		5,361		944		65		7		0	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	6,308	1,927	6,466	2,811	9,288	2,487	22,740	7,315	44,833	11,650	-	-
PCA-LR	5,140	3,935	5,863	4,577	8,508	6,256	24,556	22,291	54,623	56,182	-	-
LR-LR	4,959	4,118	6,059	4,939	10,060	8,997	31,256	33,133	76,081	77,904	-	-

テストデータに対する価格帯別RMSE (2/2)

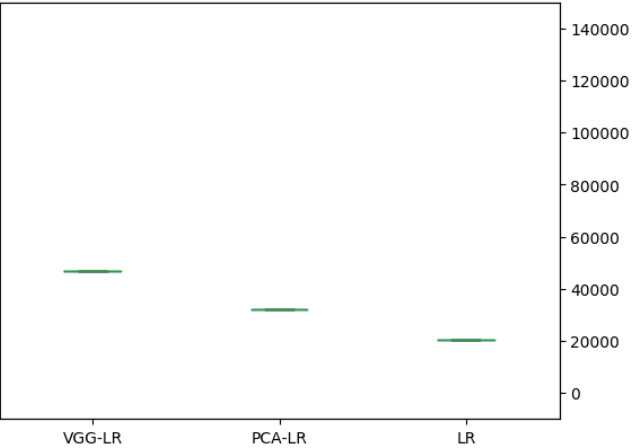
間取りタイプ: DK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
データ数	254		947		420		54		4		0	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	10,033	2,061	8,688	4,825	13,626	7,203	19,647	5,157	12,665	12,665	-	-
PCA-LR	6,377	4,690	7,398	5,719	9,708	7,987	15,883	14,587	69,914	69,914	-	-
LR-LR	6,456	4,893	7,254	5,699	10,331	8,647	19,445	20,853	75,228	75,228	-	-

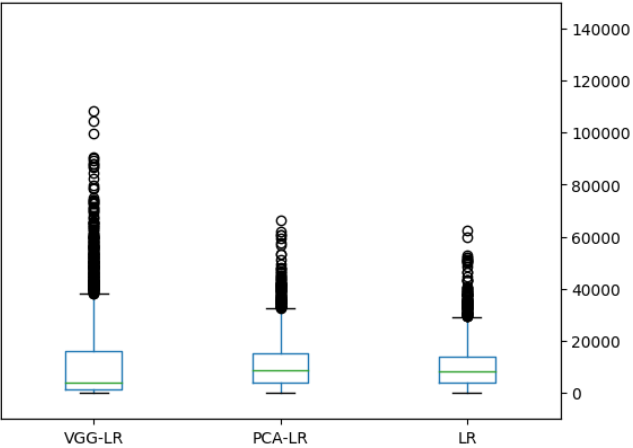
間取りタイプ: LDK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
データ数	1		2,650		2,013		1,527		429		30	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	47,045	47,045	11,404	3,853	14,858	7,953	12,641	3,870	23,011	7,323	47,122	47,218
PCA-LR	32,216	32,216	11,044	8,630	15,048	11,665	12,336	7,985	28,294	27,003	58,969	55,374
LR-LR	20,233	20,233	10,219	8,163	15,133	11,925	13,705	9,635	33,436	33,540	61,898	65,562

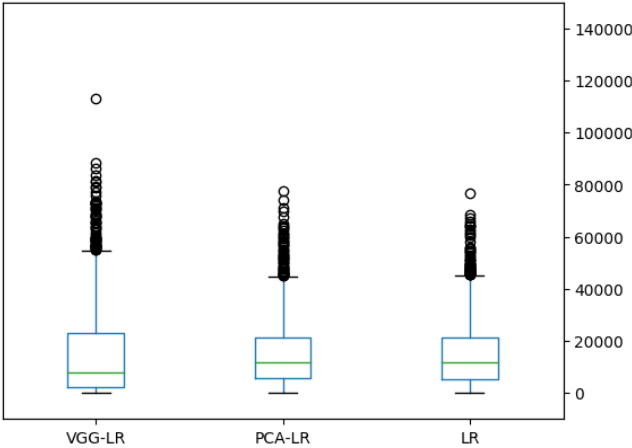
LDKにおける価格帯別RMSE



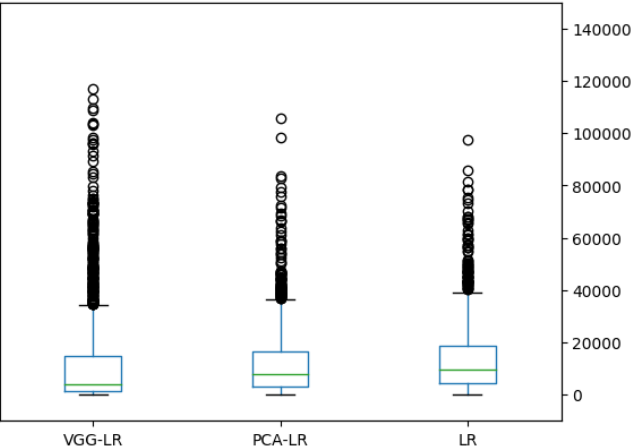
賃料: 5万円未満
(物件数: 1)



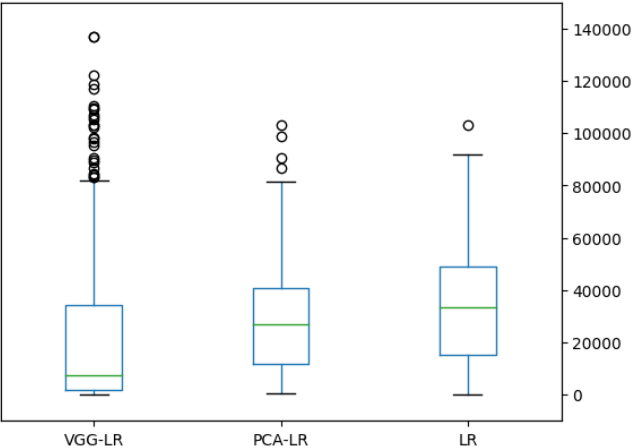
賃料: 5万円以上10万円未満
(物件数: 2,650)



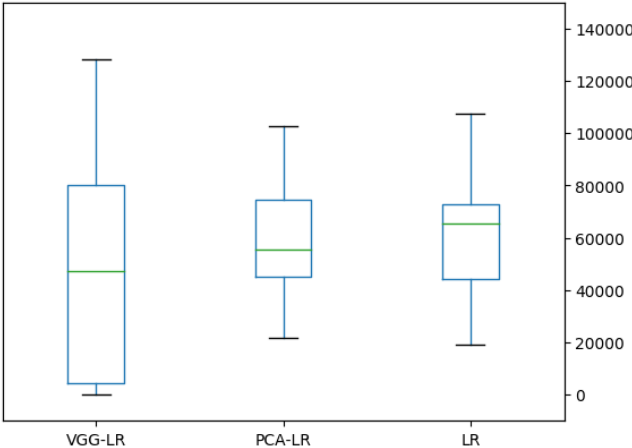
賃料: 10万円以上15万円未満
(物件数: 2,013)



賃料: 15万円以上20万円未満
(物件数: 1,527)



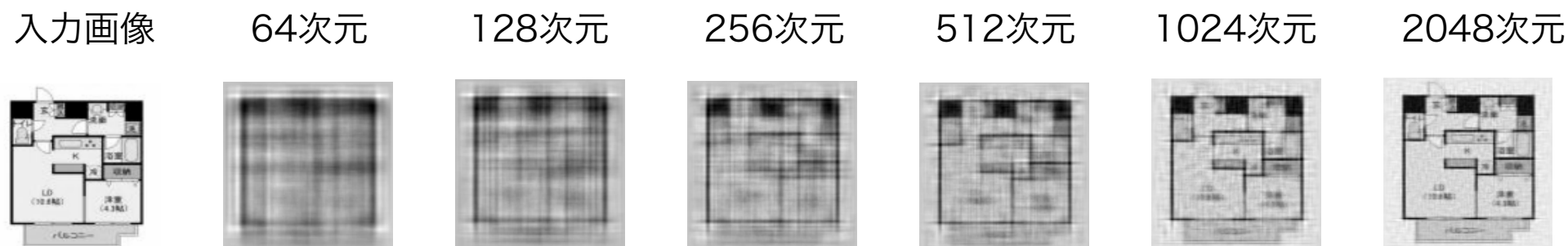
賃料: 20万円以上25万円未満
(物件数: 429)



賃料: 25万円以上
(物件数: 30)

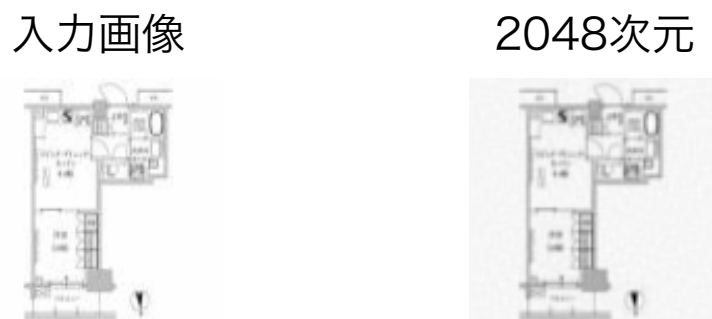
PCAによる特徴抽出量からの間取り図の復元

検証データ（LDK）に対する各次元数からの復元図



1024次元の特徴量で元の画像を認識できる程度に復元できている

テストデータに対する2048次元からの復元図



おわりに

目的：賃料予測における間取り図の影響を明らかにする

実験結果

- ・ 間取り図の考慮が予測精度の向上に寄与
- ・ 主成分分析を用いると予測精度が向上する
 - ・ 主成分分析を用いる際、その特徴量は1024, 2048次元が望ましい
- ・ VGG16を用いた場合、価格帯別の予測の中央値が向上

今後の展望

- ・ データセットの拡大
- ・ VGG16に追加する全結合層の次元数の影響の確認