

間取り図を用いた賃料予測モデルに関する一検討

服部 凌典¹ 岡本 一志¹ 柴田 淳司²

¹ 電気通信大学 大学院情報理工学研究科 情報学専攻

² 産業技術大学院大学 産業技術研究科 情報アーキテクチャ専攻

はじめに

大野喜久之輔：継続賃料鑑定評価を再考する，
都市住宅学， 住宅新報社， 2016.

賃貸物件の特徴

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

賃貸物件の価格決定方法 [大野喜， 2016]

- ・ 積算法
- ・ 収益分析法
- ・ 賃貸事例比較法
 - ・ 対象物件の周囲の物件や類似する物件を参照して賃料を決定
 - ・ 決定支援にヘドニック・アプローチを採用

ヘドニック・アプローチ

ある商品の価格をその商品の変数の価値に関する集合とし,
その商品価格の予測モデルを線形回帰で構築する技術[唐渡, 2016]

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

賃料 占有面積 立地 階数 築年数

既存のヘドニック・アプローチは間取り図が考慮されていない

ヘドニック・アプローチを用いた関連研究

説明変数として使用されている変数のカテゴリ

	建物構造	立地 /アクセス	周辺環境	売買契約	画像
Ram+ 2019	○	○	○	×	×
Philipp+ 2016	○	○	○	×	×
Jun+ 2017	○	○	○	○	×
本研究	○	○	×	○	○

- Ram P. Dahal and Robert K. Grala and Jason S. Gordon and Ian A. Munn and Daniel R. Petrolia and J. Reid Cummings :
A hedonic pricing method to estimate the value of waterfronts in the Gulf of Mexico, Urban Forestry & Urban Greening, vol.41, pp. 184-194, 2019.
- Dr. Philipp Deschermeier and Björn Seipelt :
A Hedonic Rent Index for Student Housing in Germany, Cologne Institute for economic research, pp.1-12, 2016.
- yung-Jin Jun, Hee-Jae Kim : Measuring the effect of greenbelt proximity on apartment rents in Seoul, Cities, vol. 62, 2017.

間取り図の特性

清田陽司, 山崎俊彦, 諏訪博彦, 清水弘 :
エディトリアル : 「不動産とAI」, 人工知能,
vol.32, 2017.

間取り規格が同じであっても, 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]



同じ建物の間取り規格Kで間取り図が違う例

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

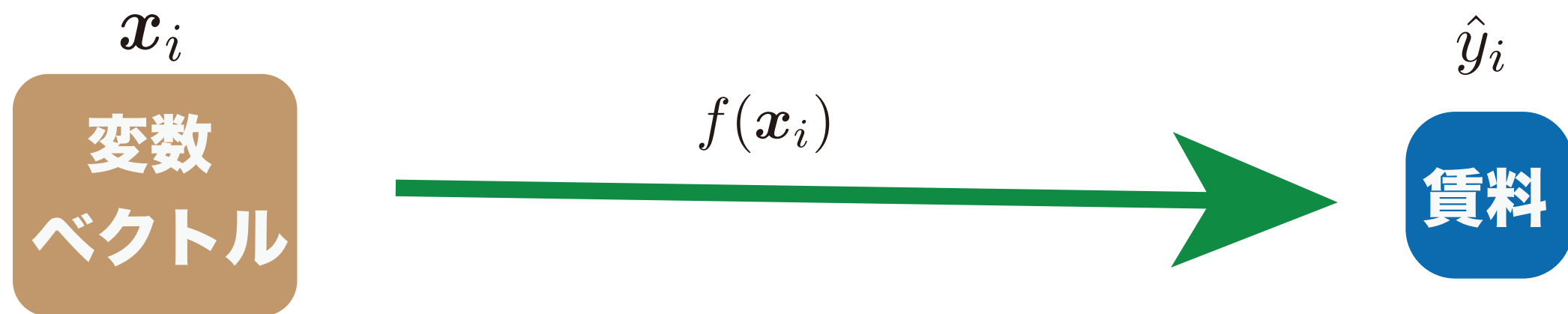
手段: 賃料の予測誤差を比較

- 間取り図を考慮しない予測モデル (LR)
- 間取り図を考慮した予測モデル
 - 線形変換による特徴量抽出法 (PCA-LR)
 - 非線形変換による特徴量抽出法 (VGG-LR)

LR (Linear-Regression)

一般的なヘドニック・アプローチ

- 予測式： $\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\alpha}^T \mathbf{x}_i + \beta$
- 説明変数に間取り図を含まない
- scikit-learn (version : 0.20.2) のLinearRegressionクラス使用



このLRと間取り図を考慮したPCA-LRとVGG-LRを比較する

PCA (Principal Component Analysis) -LR

LRの説明変数に間取り図を追加したモデル

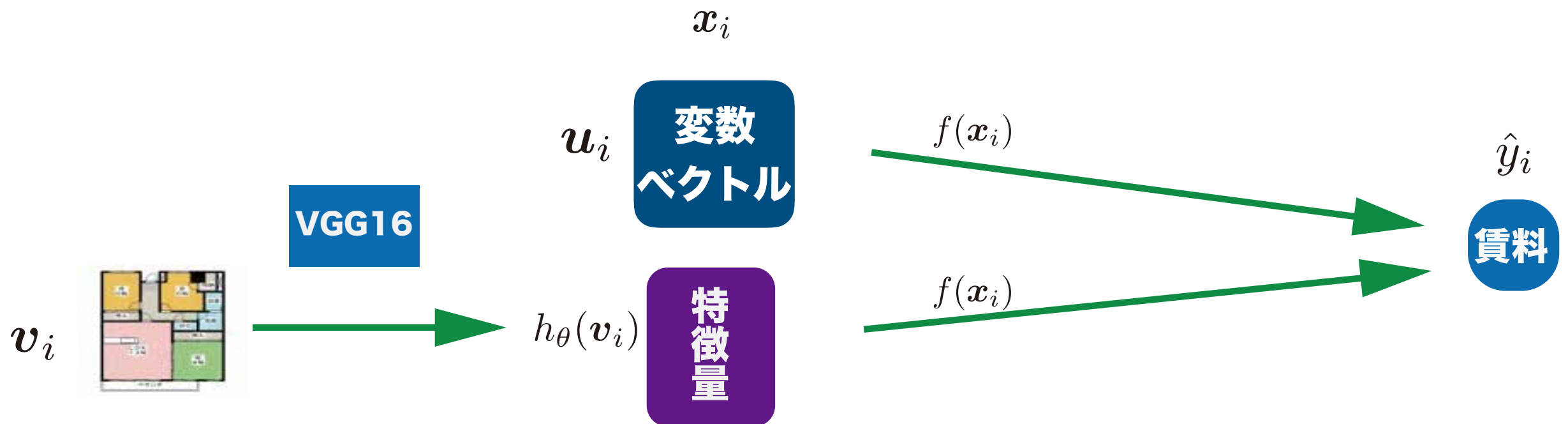
- 主成分分析による特徴量抽出
- 抽出した間取り図の特徴量と変数ベクトルを組み合わせ、線形回帰で賃料予測
- 間取り図の特徴量: 64, 128, 256, 512, 1024, 2048次元の6種類
- scikit-learn (version : 0.20.2) のPCAクラス使用
- $\mathbf{x} = [\mathbf{u}, \phi_{\theta}(\mathbf{v})]$



VGG-LR

間取り図を非線変換で特徴抽出するモデル

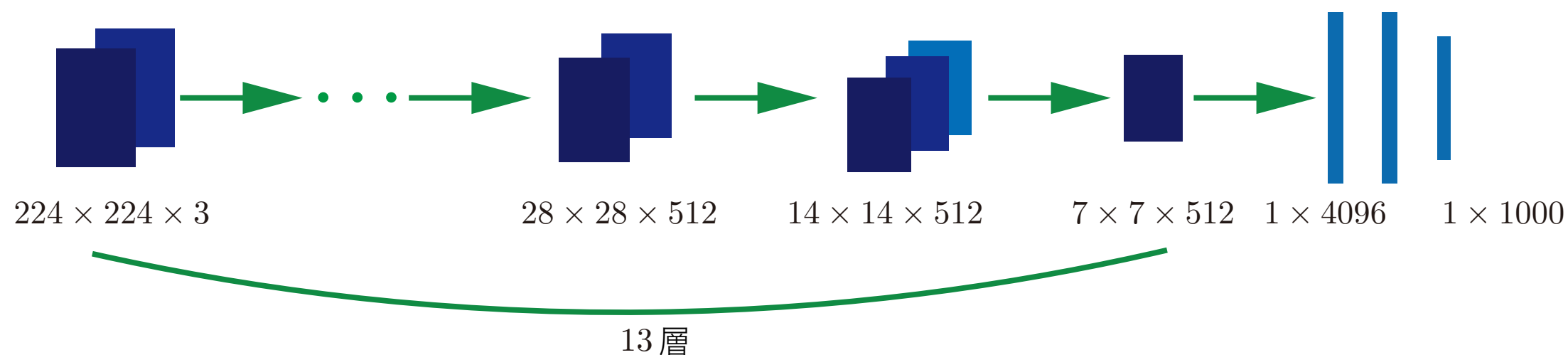
- ・ ミニバッチ＋確率的勾配降下法を適用
- ・ batch size: 100, epoch: 100
- ・ 間取り図の特徴量抽出: 64次元
- ・ 活性化関数: ReLU
- ・ Keras (version : 2.2.4) のFunction API
- ・ $\mathbf{x} = [\mathbf{u}, h_{\theta}(\mathbf{v})]$



VGG16の構造

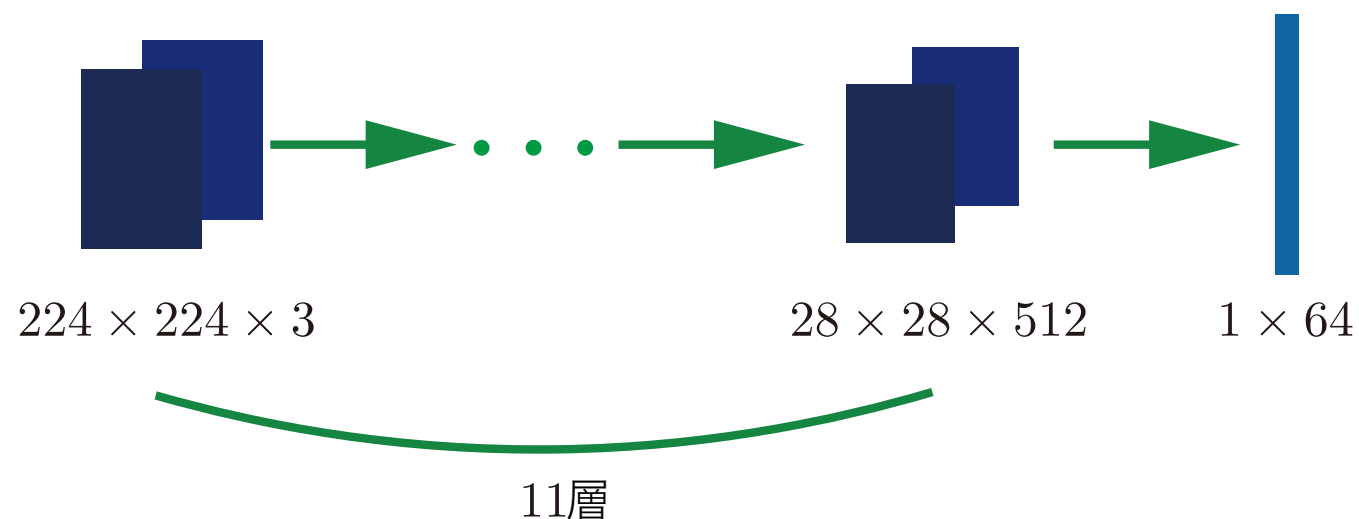
Karen Simonyan, and Andrew Zisserman : Very Deep Convolutional Networks for Large-Scale Image Recognition, Computing Research Repository, vol.abs/1409.1556, 2014.

VGG16 [Karen+, 2014]: オックスフォード大学のVGGチームによって開発された学習済みのニューラルネットワーク



使用するVGG16

- ・ 後ろの5層を除去
- ・ 64次元の全結合層追加
- ・ 後ろ3層のみ学習



実験環境

LIFULL HOME'Sデータセット (2015年9月時点)

- 賃貸物件データ (70変数, 533万件)
- 120×120ピクセルの画像データ (8,300万ファイル)

使用する計算機 (OS: Ubuntu 18.04.1)

- CPU : Xeon(R) CPU E5-2650 v3 @ 2.30GHz
- GPU : GeForce GTX 1080 Ti 11GB
- メモリ : 64GB

データセット

使用賃貸物件データ（14変数, 9万件）

- ・ 東京都のデータを使用（∵ 47都道府県で最もデータ量が多いため）
- ・ 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
- ・ 間取り図：グレースケール変換

データセット分割

- ・ 使用賃貸物件データを間取り規格を基準にK, R, DK, LD, LDKの5種類に分割
- ・ LDはデータ数が35件のため使用しない
- ・ 20%をテストデータに, 80%を開発用データに分割

実験概要

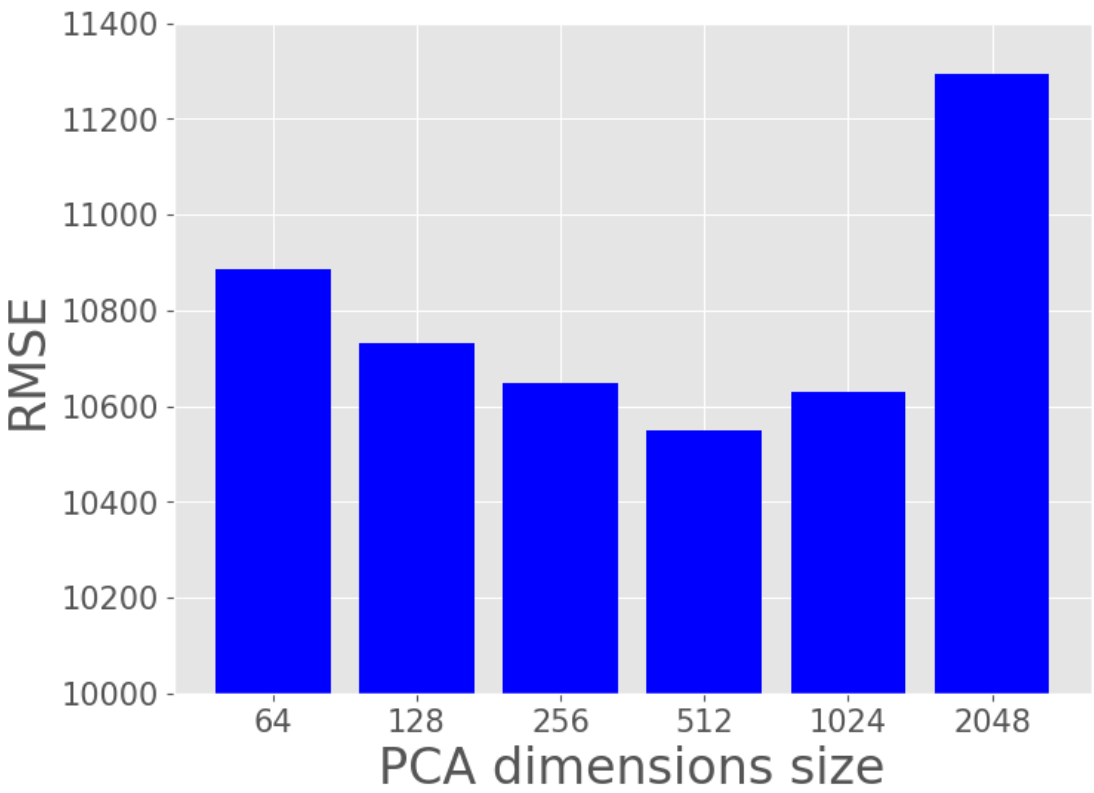
1. ハイパーパラメータチューニング

- PCA-LRにおける適切な間取り図の次元数を探索
- 64, 128, 256, 512, 1024, 2048次元の6種類
- 開発用データに10-fold cross-validationを適用
- 評価指標：平方平均二乗誤差（RMSE: Root Mean Square Error）

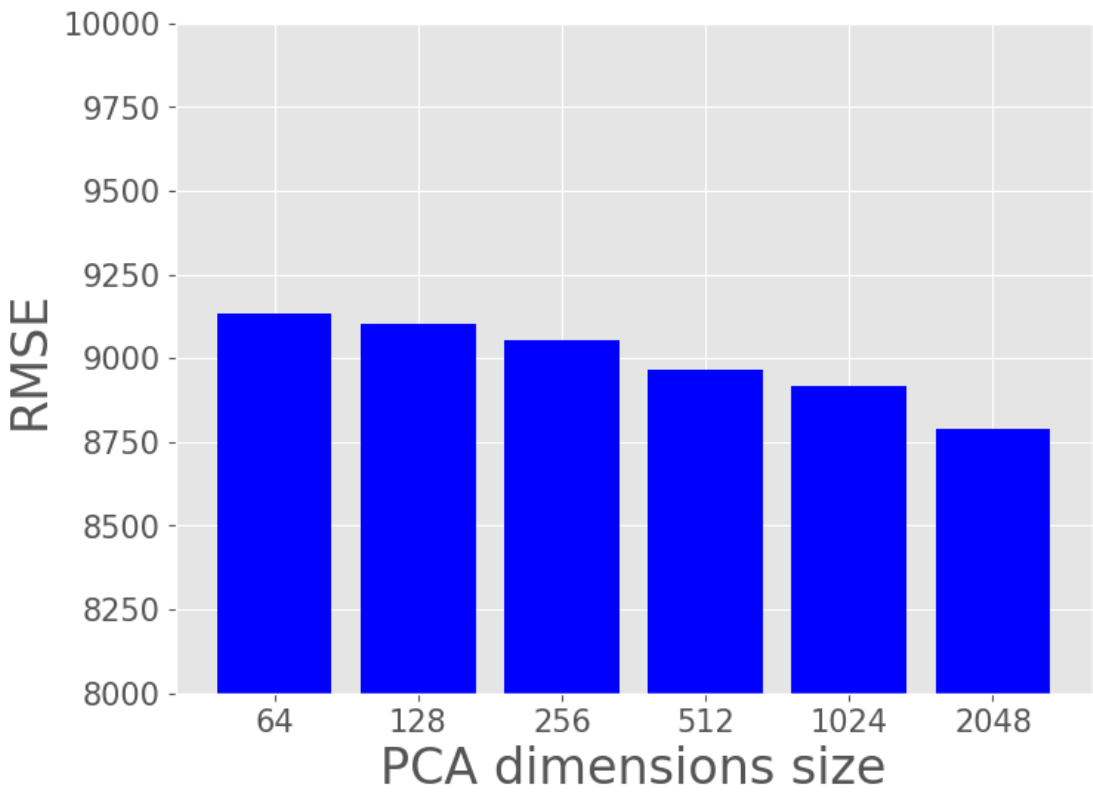
2. テストデータに対するLRとVGG-LR, PCA-LRの評価

1. RMSEによる評価と学習時間
2. 予測値の可視化

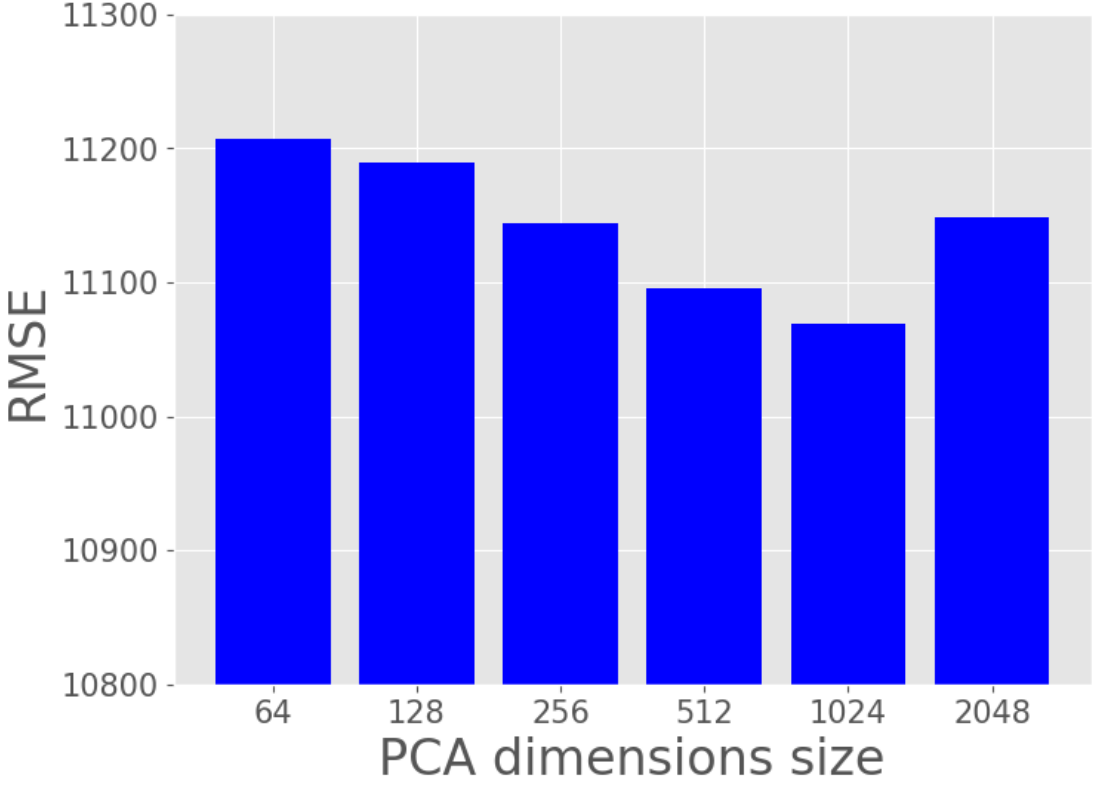
1. ハイパーパラメータチューニング



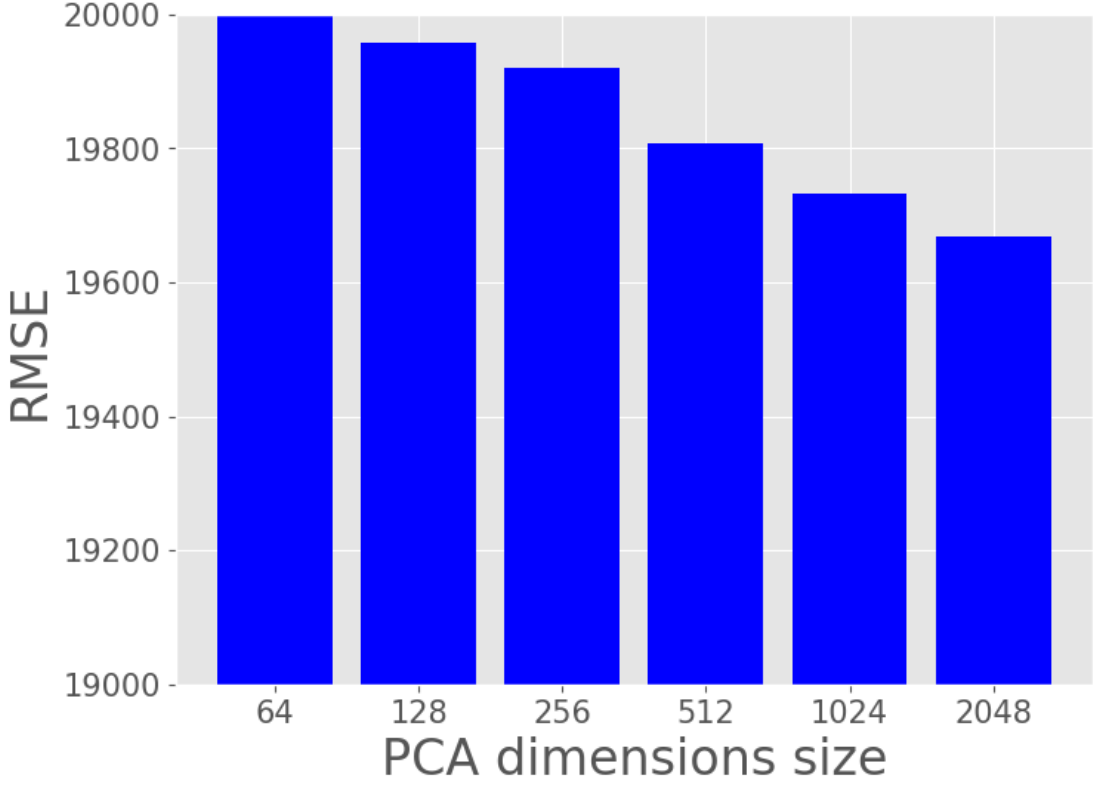
K



R



DK



LDK

2-1. RMSEによる評価と学習時間

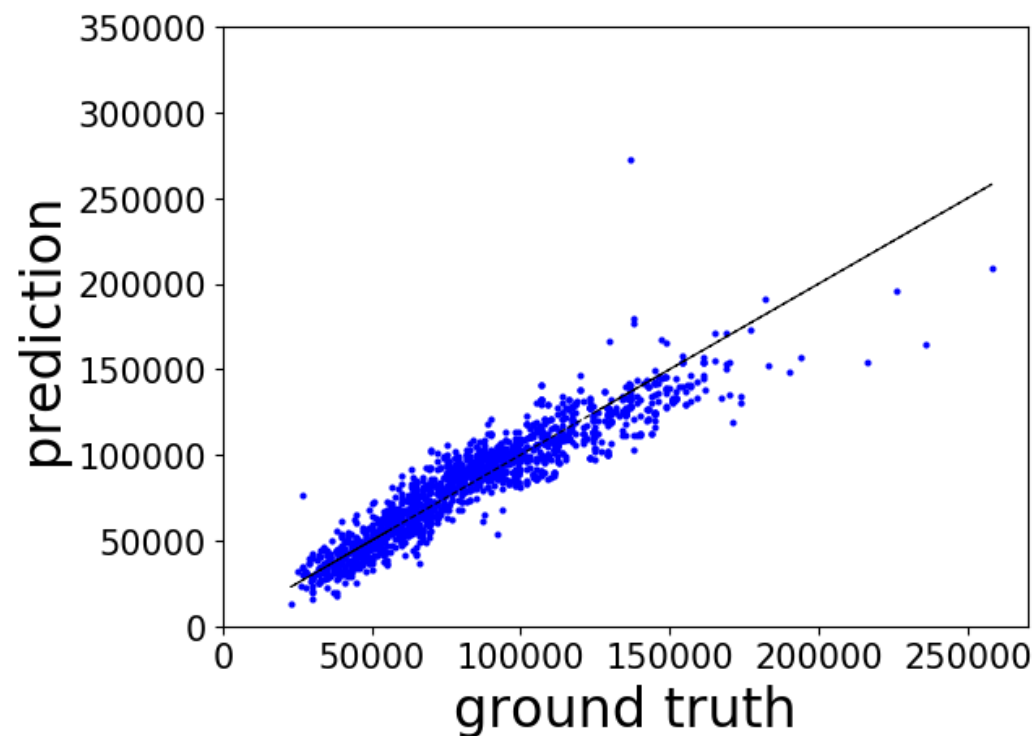
テストデータに対するRMSE[円]

	データ数	LR	PCA-LR	VGG-LR
K	1,712	11,602	11,689	16,862
R	7,525	9,217	8,727	12,541
DK	3,548	11,156	10,957	16,467
LDK	6,842	20,121	19,646	22,878

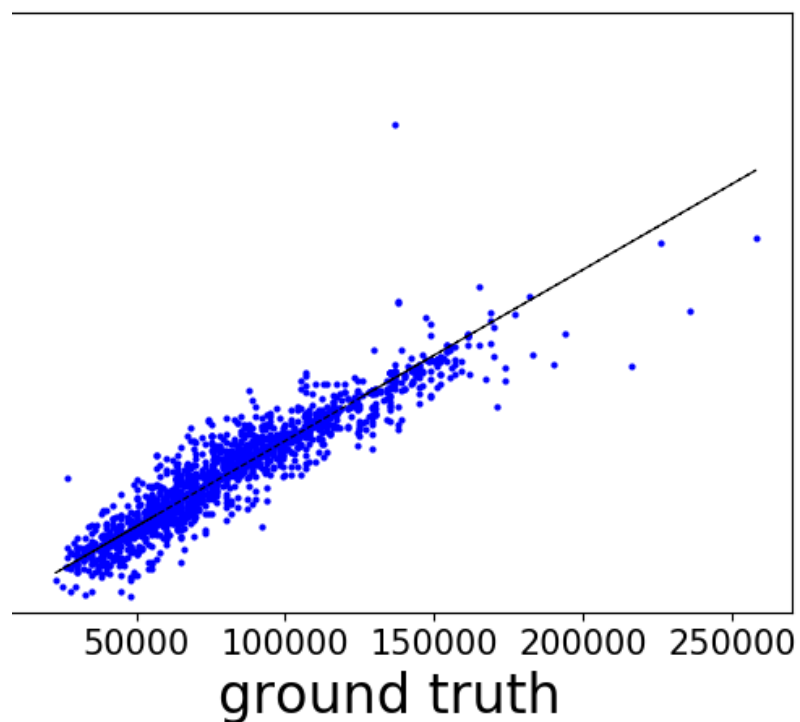
開発用データの学習時間[秒]

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

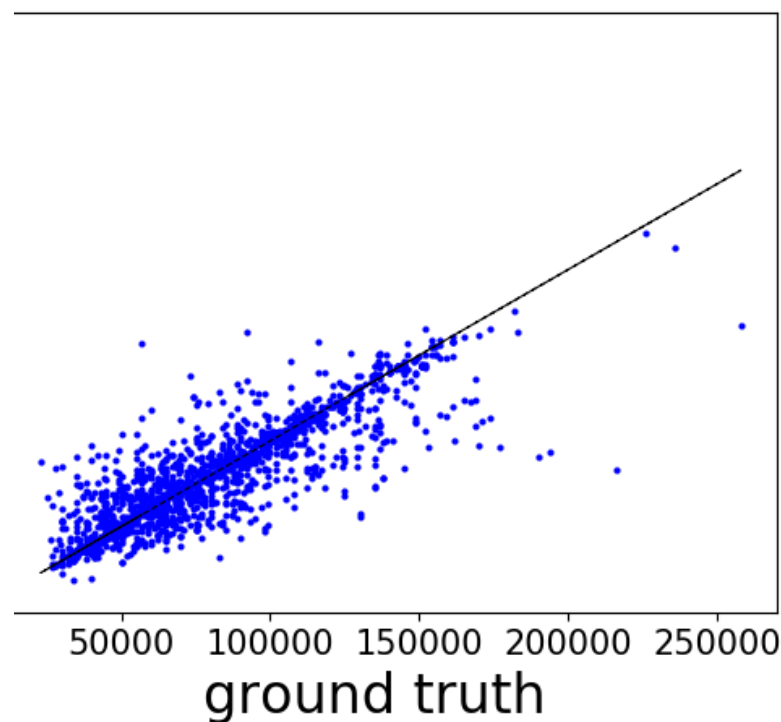
2-2. 予測値の可視化



LR

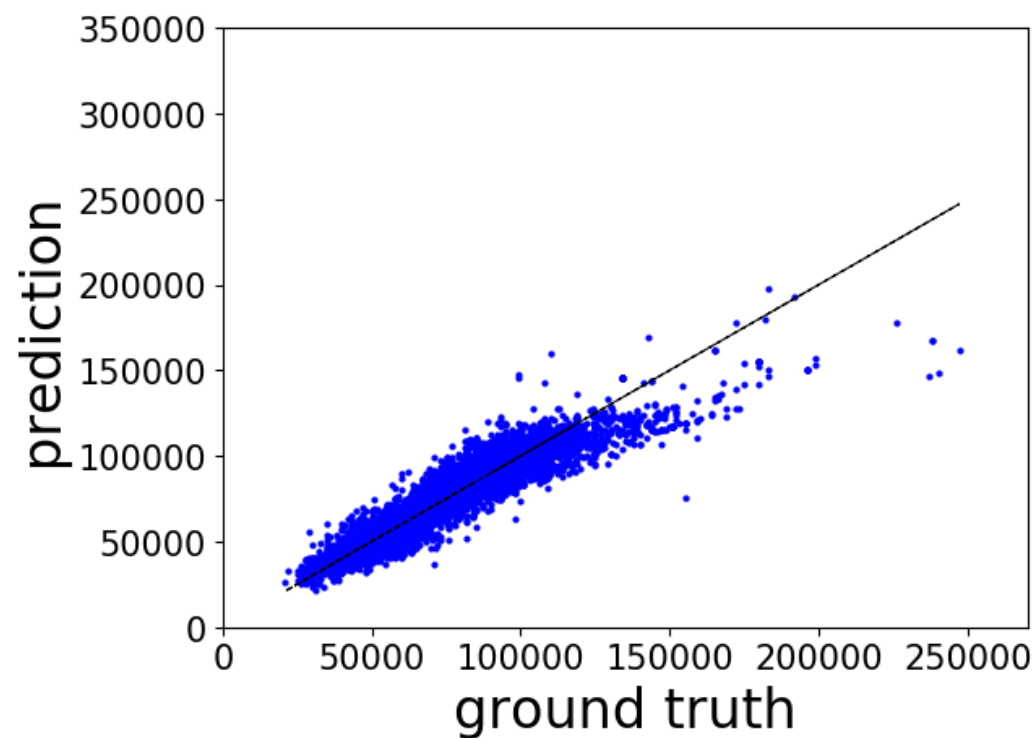


PCA-LR

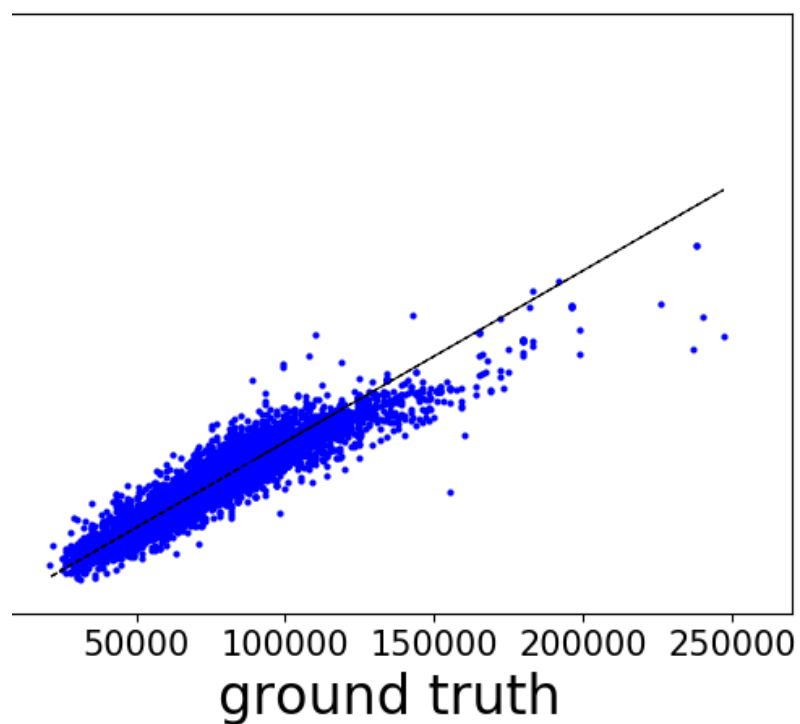


VGG-LR

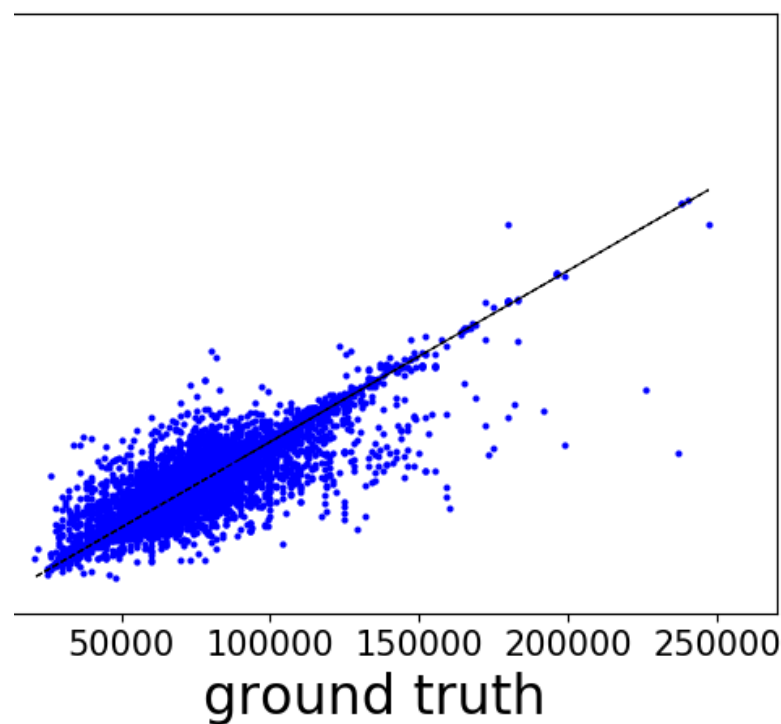
間取り規格K



LR



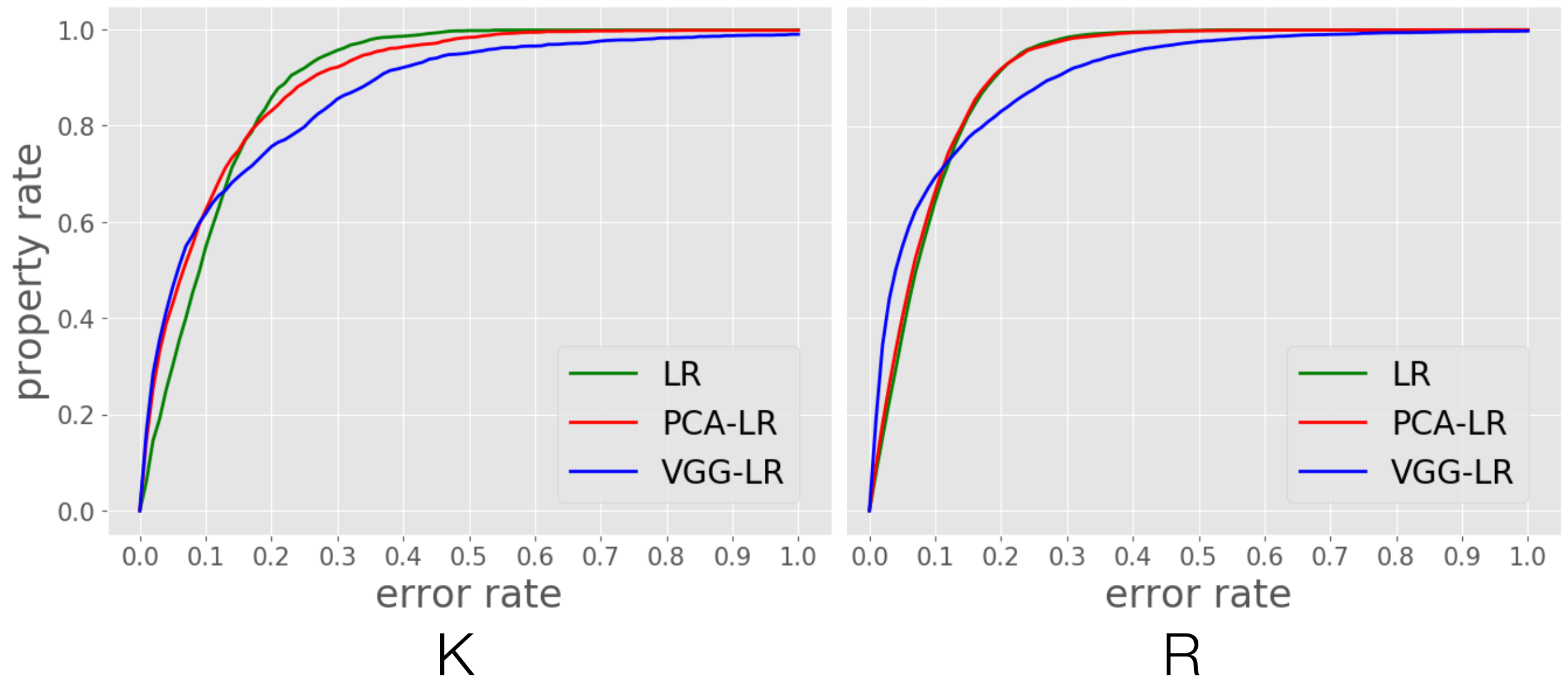
PCA-LR



VGG-LR

間取り規格R

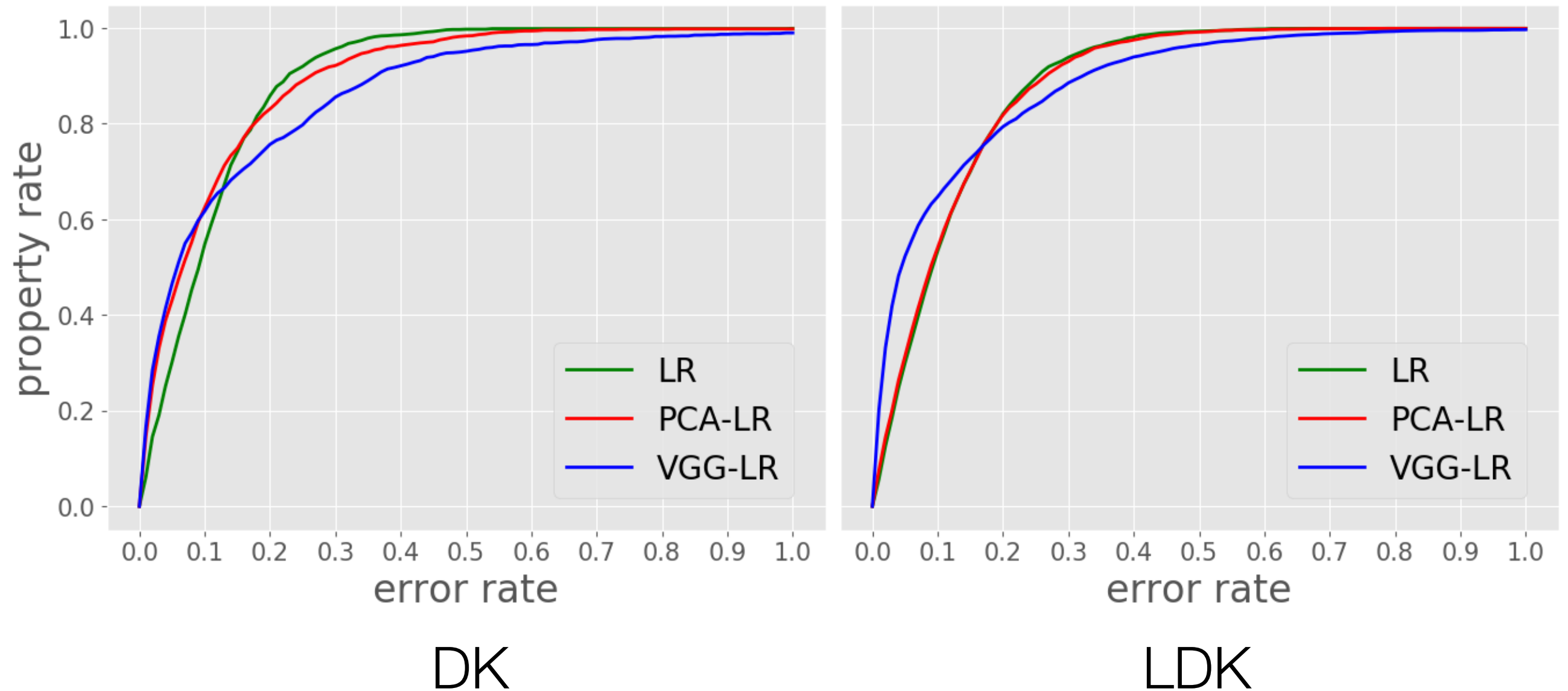
2-2. 累積密度関数 (1/2)



横軸：error rate = $|1 - \frac{\hat{y}}{y}|$ y ：真値 \hat{y} ：予測値

縦軸：error rate 以下の誤差に収まっているデータ数の割合

2-2. 累積密度関数 (2/2)



LIFULL HOME'Sデータセット外の物件予測

	間取り図	構造	間取り 規格	賃料	LR	PCA-LR	VGG-LR
自宅		RC	K	65,000	61,361	62,545	83,695
学生寮		RC	K	47,700	56,931	49,716	64,526
A宅		木造	K	74,000	66,695	68,701	58,561
B宅		鉄筋	LDK	128,000	130,988	122,887	181,913

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

間取り図を考慮することで予測誤差が改善する

- ・ 特徴量抽出に主成分分析を用いることでRMSEが改善する
 - ・ 間取り図の特徴量は1024, 2048次元の高次元が望ましい
 - ・ 1024, 2048次元の間取り図の特徴量で間取り図の大部分の特徴量を抽出できる
- ・ 特徴量抽出にVGGを用いることで, error rateが0.1以下の物件数が増える
 - ・ 予測誤差のばらつきが課題
 - ・ 計算コストが課題

PCAによる次元別の間取り図の特徴量

検証データ（LDK）に対する各次元数からの復元図

入力画像

64次元

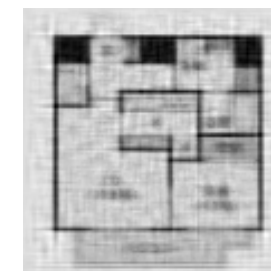
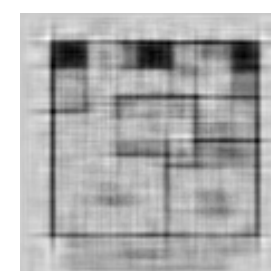
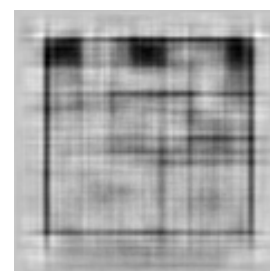
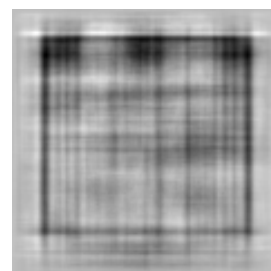
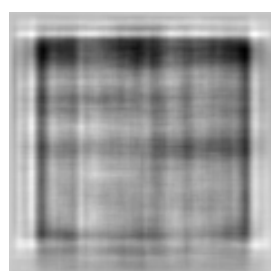
128次元

256次元

512次元

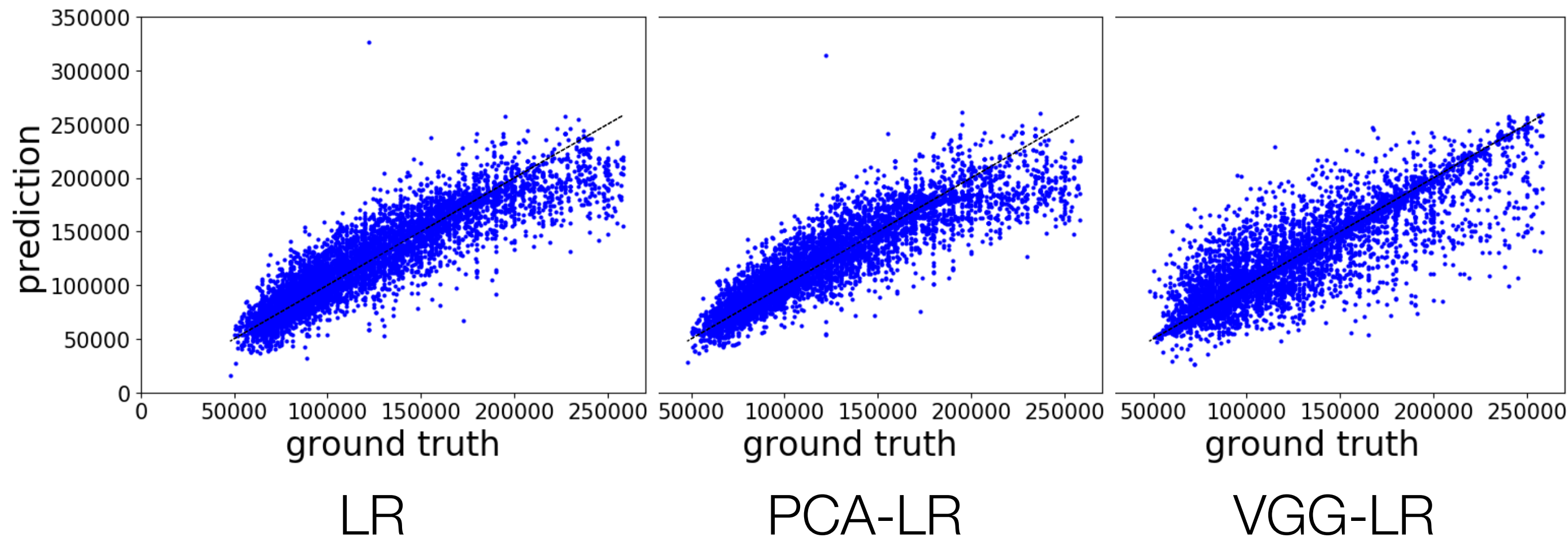
1024次元

2048次元



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

LDKのテストデータに対する予測値の可視化



LDKにおけるVGG16のみで予測精度が高い 物件事例

テストデータ数の31.6%がVGG16のみで高い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		146,000	木造	13	1	練馬区
事例 2		140,000	木造	8	1	世田谷区
事例 3		237,000	SRC	17	4	千代田区

LDKにおけるPCA-LRのみで予測精度が高い 物件事例

テストデータ数の10.1%がPCA-LRのみで高い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		236,000	SRC	4	10	新宿区
事例 2		120,000	軽量鉄骨	4	1	府中市
事例 3		128,000	軽量鉄骨	0	2	小金井市

LDKにおける両方のモデルで予測精度が低い 物件事例

テストデータ数の36.9%が両方のモデルで低い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		114,000	RC	20	2	江戸川区
事例 2		171,000	RC	0	2	品川区
事例 3		82,000	木造	28	2	三鷹市