



Title	ディープラーニングを用いた画像による屋内自己位置推定
Author(s)	
Citation	令和4（2022）年度学部学生による自主研究奨励事業研究成果報告書．2023
Version Type	VoR
URL	https://hdl.handle.net/11094/90994
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

令和4年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

ふりがな氏名	にしもと ようすけ 西本 遥裕	学部 学科	基礎工学部シス テム科学科	学年	3 年
ふりがな 共 同 研究者氏名	あかさき こうた 赤崎 功汰	学部 学科	基礎工学部シス テム科学科	学年	3 年
					年
					年
アドバイザー教員 氏名	松原 崇	所属	基礎工学研究科 システム創成専攻		
研究課題名	ディープラーニングを用いた画像による屋内自己位置推定				
研究成果の概要	研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。(先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。)				
1 はじめに 今日、我々の位置情報はナビゲーションアプリ、SNS やライフログのさまざまな場面で活躍している。屋外での位置情報は GPS (Global Positioning System) や GNSS (Global Navigation Satellite System) などにより取得されており、その精度やコスト面からもあらかた確立していると言えるだろう。[1]しかし、それらの屋外測位システムの屋内測位への利用は実用的とは言えない。屋内測位においては、近年のスマートフォンの普及や位置情報サービスの増加により、モバイル・ノードによる屋内測位に大きな関心が集まっている。特に Wi-Fi の信号強度を利用した測位方法は、その簡便性と必要機材の少なさから、屋内測位で広く利用されてきた。しかし、その測位精度はマルチマスや障害物により大きく左右される。[2] 本研究では、ディープラーニングを用いた画像認識での自己位置推定の手法を提案する。Convolutional Neural Network(CNN)によって、ある場所の画像の特徴量を抽出し、予めいくつかに分けられていた屋内での位置に分類し、自己位置推定を実現する。ディープラーニングでの画像認識での自己位置推定をマルチパスや障害物に影響されやすい信号強度に依存することなく実現でき、その結果、精度向上も期待できる。その他にも、マルチパスが起きやすい環境や人が密集していて電波状況が悪くなる場所での利用が期待される。また、この手法では、我々人間が見覚えのある景色から自分の位置を推定できるように深層学習でも人間と同じように推定をできるのではないかと(実験1、実験2)、というアイデアから、人間が周囲を見渡し自分の位置を推定するように、同じ位置の複数方向の画像から位置を推定する(実験3)。 2 実験準備 2.1 データ準備 図1のように大阪大学基礎工学部棟1Fを56箇所138方向に分類し、各方向で15枚ずつ(訓練用:検証用=12:3)高さや横方向の角度(±15度程度)をランダムに変えて撮影した。各撮影位置の間隔は、最大10m程(データ量がなるべく少ないことが理想であり、自分のある程度の位置が分かっている問題ないかと判断し、この間隔とした)である。また、ナビゲーションをする際、通路が交差している箇所は必要と考え、その箇所はすべて撮影した。					

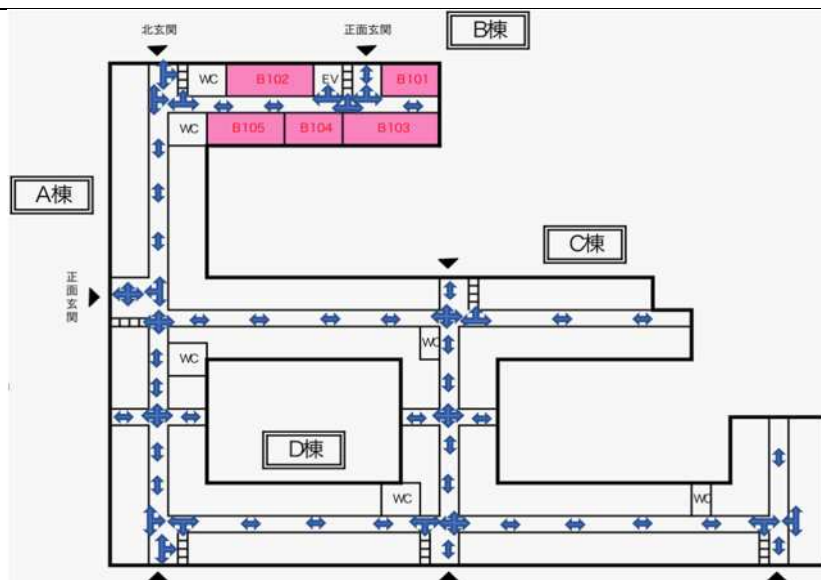


図 1：基礎工学部棟 1F

2.2 データ準備 2 (Transforms によるデータ前処理)

データ数が少ないため、訓練前に以下の Data Augmentation で前処理をすることにより学習効率を高める。

○訓練にのみ使用

- Random Affine

ランダムにアフィン変換する。今回は、回転のみ設定。

- Random Perspective

ランダムに射影変換を行う。

- Grid Mask[3]

ランダムに一部の画素を削除する。

○訓練、検証に使用

- Normalize

データを正規化する。 $\mu = \sigma = 0.5$ とする。

2.3 事前学習済みモデルの利用

本研究では、VGG-19 with batch normalization という事前学習済みモデルを使用する。また、データ数が少ない場合に向いていると言われる転移学習で学習する。最適化関数には、Adam (Adaptive moment)、損失関数には、交差エントロピー関数 CrossEntropyLoss を使用する。

3. 実験

3.1 実験 1

3.1.1 実験 1 の手法

138 方向の画像の 1 つを入力とし、138 クラスに分類し自己位置とその方向を出力するよう VGG-19 with batch normalization を訓練し、その訓練モデルで自己位置を推定する。この手法は、自己位置推定をマルチパスや障害物に影響されやすい信号強度に依存することなく実現できることが期待できる。訓練用画像データ：12 * 138 = 1656 枚、検証用画像データ：3 * 138 = 414 枚を用意し、各エポックの損失と精度の推移を観察する。これは、ディープラーニング (CNN) を用いて、画像による屋内自己位置推定の可能性を示すことを目的とする。

3.1.2 実験 1 結果

訓練、検証の損失の推移を図 2 に、精度の推移を図 3 に示す。検証精度は最高 94.203% であった。このことから、画像による屋内自己位置推定は十分可能であることが示された。

検証用の間違えた画像の一部を図 4 に示す。予測の間違い方は 3 パターンあった。「正解と予測が隣同士で方向は同じ」(図 4 の 4801:4901、2501:2401)、「正解と予測が隣同士で方向が反対」(図 4 の 0402:0501、3302:3204 など)、「正解と予測が全く違う」(図 4 の 3002:2003 など)、と

いう 3 つである。

また図 5 には、「正解と予測が全く違う」場合の、間違えて予測してしまった位置と方向の実際の画像を示す。「正解と予測が全く違う」間違え方で間違えた画像は、特徴の少ない通路の画像や、他にとても似た風景の位置画像がある画像の 2 パターンであった。(表 1)

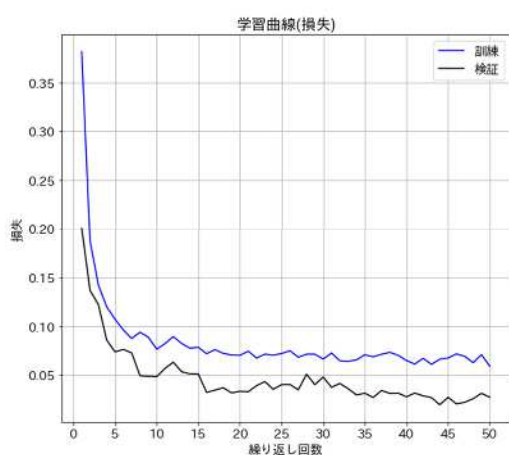


図 2：実験 1 損失曲線

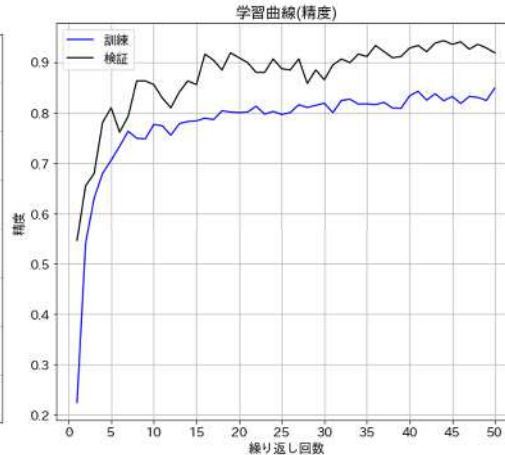


図 3：実験 1 精度曲線

4503:2404 2004:3704 2501:2401 1901:2901 4801:4901 0402:0501 3903:4104



2501:2401 5603:5304 3302:3201 2501:2602 3002:2003



図 4：間違えた画像の一部（正解：間違えた予測）



図 5：間違えて予測した位置と方向の実際の画像

表 1：間違え方

間違え方	画像
正解と予測が隣同士で方向は同じ	図 4 の 4801:4901 や 2501:2401 など
正解と予測が隣同士で方向が反対	図 4 の 0402:0501、3302:3204 など
正解と予測が全く違う	2 パターン：特徴の少ない通路の画像（図 4 の 3002:2003 など） or 他の位置にとっても似た風景がある画像（図 4 の 5603:5304 など）

3.1.3 実験 1 考察

予測の間違え方、「正解と予測が隣同士で方向は同じ」、「正解と予測が隣同士で方向が反対」、「正解と予測が全く違う」、という 3 つのパターンの前 2 つは、撮影した場所が隣同士であるので、その通路の特徴のあるものが同じで、それが各画像内に写ってしまい、その特徴から同じ場所の方向を撮影したと判断したと考えられる。3 つ目は、前 2 つとは違い、異なるものの特徴を抽出し、その特徴から同一物として、同じ場所の方向を撮影したと判断したと考えられる。つまり、前 2 つは、同じ通路の特徴が同一と判断されているのに対し、3 つ目は、異なる通路の特徴を同一と判断しているということである。3 つ目の間違え方は、深刻な問題と考える。

3 つ目の間違え方で間違えた画像は、特徴の少ない通路の画像や、他の位置にとっても似た風景が

ある画像の 2 パターンと述べたが、この画像は我々人間でも判断が難しいものがほとんどであったが、逆に、他の位置によく似ているが判断するのにあまり困難でない画像もあった（図 4 の 4503 : 2404、図 5 の 2404 の画像）。この画像をどうにかして正しく推定できるようにしたい。

3.2 実験 2

実験 1 で画像による屋内自己位置推定の可能性を示すことができた。しかし、今基礎工学部棟を 56 箇所に分割しているが、その位置と位置の間が最大 10m 程離れていることや撮影するときのノイズなどを考えると、実際に撮影するとき、特に撮影位置同士の間で撮影したときに精度が下がることを想定し、精度をもう少し改善しておきたい。

3.2.1 実験 2 の手法

実験 1 の結果から間違いやすい場所を調べ、その特定場所の分類を得意としたディープラーニングモデルを新しく作り、複数のモデルで 138 分類し、自己位置とその方向を推定する。新しいディープラーニングモデルを訓練・検証するためのデータは、実験 1 で使ったデータをそのまま利用した。

実験 1 の他の位置によく似ているが判断するのにあまり困難でない画像同士の間違えを防止するために、ここでは図 6 の画像分類が得意なモデル net1 と図 7 の画像分類が得意なモデル net2 を新しく作る。net1、net2 の学習には、事前学習済みモデル AlexNet を使用。（分類先が少なく、あまり深くしなかったため）

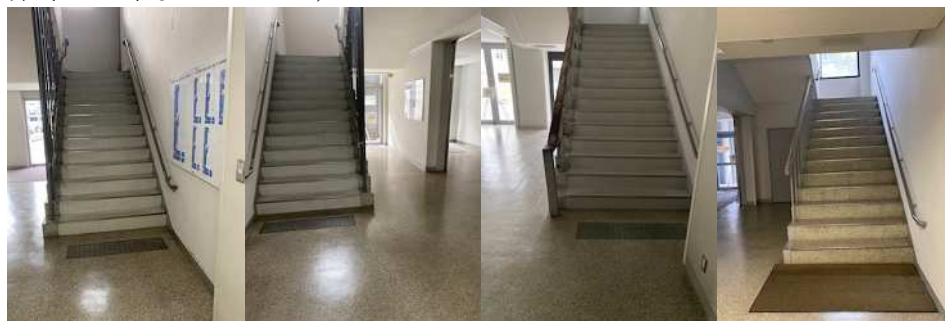


図 6 : 左から 2804,2404,4503,0703

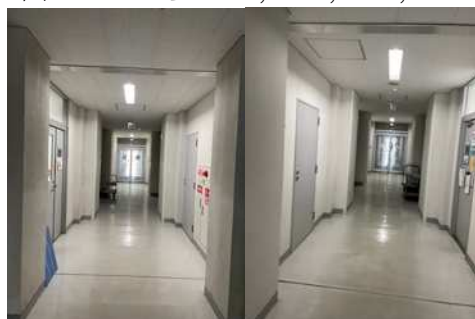


図 7 : 左から 4104,3903

推定は以下のようなアルゴリズムで行われる。

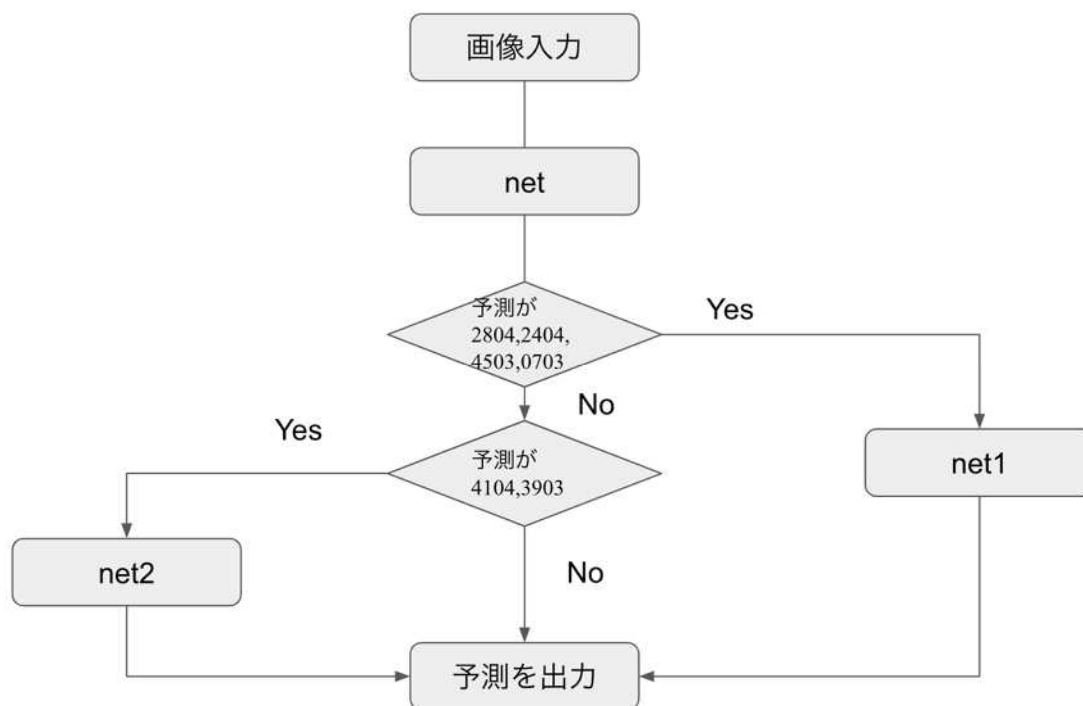


図 8：実験 2 の推定アルゴリズム

3.2.2 実験 2 結果

net1 と net2 の最高精度はどちらも 100%となった。

上のような手法で検証用データを推定した結果、もともと 91.787%だったものが、92.512%となり約 0.7 ポイント精度が改善した。このことから、図 6、図 7 の画像を実験 1 より正しく分類できたことがわかる。

3.2.3 実験 2 考察

この手法は、実験 1 のモデルの弱点に、新しいモデルを追加することによって、直接手を加えている。これが効果的であることは実験 2 で示すことができたが、実際に撮影した画像に対して実験 2 と同じような挙動をするか調べる必要がある。また、この手法は、間違えた画像のパターンを調べ、正解できるはずと考えた画像の組みを作り新しいモデルを作っているため、非効率的な改善方法なのかもしれない。

3.3 実験 3

実験 1、実験 2 で一方向の画像からある程度自己位置を推定でき、精度 92.512%とかなり良い精度が出ることがわかった。しかし、間違えた画像の中では、一方向の画像では位置を推定することはかなり難しい画像も多数見られた。そこで、我々は初心に帰り、人があたりを見渡してその景色の情報から自己位置を推定すること、をヒントに以下のような手法を検証した。

3.3.1 実験 3 の手法

ある一つの地点で撮った異なる二つの方向の画像を入力として、56 箇所の中から一つ選びその位置を出力として、自己位置を推定する。

人があたりを見渡して自分が今いる位置を推定するように、ひとつの方向からではなく、複数方向（二方向）の画像情報から自分の位置を推定する。具体的には、以下のようなアルゴリズムである。

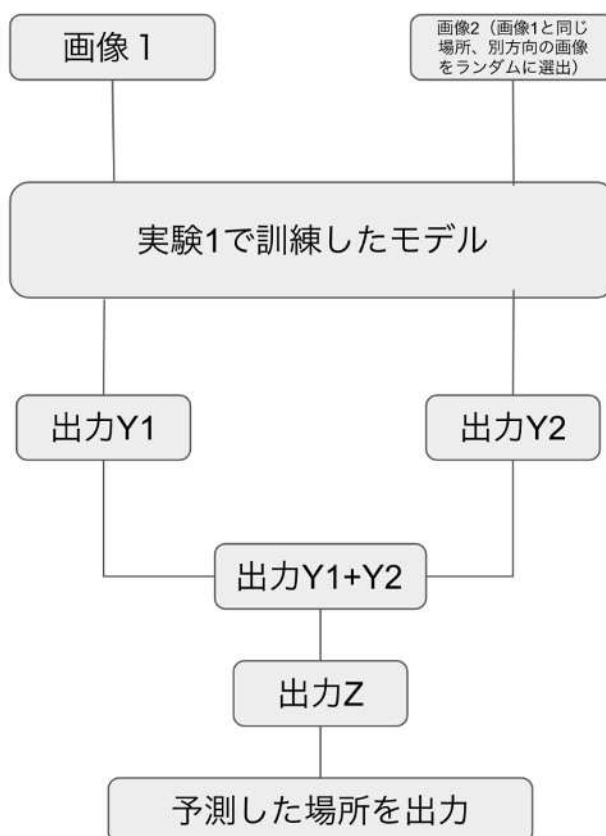


図 9 : 実験 3 の推定アルゴリズム

$$\text{ただし、} Y1 = \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,138} \end{bmatrix}, Y2 = \begin{bmatrix} y_{2,1} \\ y_{2,2} \\ \vdots \\ y_{2,138} \end{bmatrix}, Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{56} \end{bmatrix}$$

$$z_i = \frac{1}{n_i} \sum_{j=\sum_{k=1}^{i-1} n_k}^{n_i + \sum_{k=1}^{i-1} n_k} (y_{1,j} + y_{2,j}), n_i: \text{場所 } i \text{ で撮影した方向の個数}$$

Zの中で最大のクラスを予測した場所として出力する。

3.3.2 実験 3 結果

上のような手法で検証用データを推定した結果、実験 2 では、92.512%だった精度は、99.517%と約 7 ポイント上昇した。入力画像が 1 つの場合に比べ、入力画像が 2 つの場合は、情報量が 2 倍になっており、やはり精度が上がることを確認することができた。検証用データ 414 枚中推定位置を間違えたのは、2 枚のみであり、その 2 枚のうち 1 枚は推定位置に隣接している場所で撮影したもので、もう一枚も推定位置が 3 つ隣の場所で撮影されたものだった。

3.3.3 実験 3 考察

この実験から 2 つの方向の画像からは、1 つの方向の画像と比べ情報が 2 倍となり、他の場所と区別するだけの情報を入手出来たと言えるだろう。そして精度は 99.512%と、とても高い精度であり、これは既存の技術と同等もしくはそれ以上の精度を持つ推定を実現することができたことを示すと考える。

4. 評価

実験 1、実験 2、実験 3 の手法を実際に使ってみると、どの程度の精度になるのか検証する。

4.1 評価方法

訓練、検証用データとは別に基礎工学部棟から適当に 20 箇所を選び、各箇所で 2 方向の画像を 1 枚ずつ撮影し、合計 40 枚を準備する。その画像を入力として、1 枚入力の場合、2 枚入力の場合の自己位置推定の精度を検証する。検証には、実験 1、実験 2、実験 3 の提案手法を使う。ただし、基礎工学部棟を 56 箇所に分割しているが、その箇所同士の間で撮影した画像は、その撮影箇

所から一番近い場所を推定すると正解ということとする。正解箇所より 1 ずれると誤差 1、2 ずれると誤差 2 とする。

4.2 評価結果

実験 1 の手法と実験 2 の手法は、同じ結果となった。

表 2：実験 1、実験 2 手法の検証

	(誤差) 0	1	2	不正解
方向一致	21	0	0	15
方向不一致	3	0	1	

表 2 のようになり、予測のうち場所、方向ともに正解した個数は、21 となり、正解率は、52.5%となった。また方向は違うが場所は一致したものを合わせると、精度は 60%となった。

実験 3 の手法は以下の通りである。

表 3：実験 3 の手法の検証

(誤差) 0	1	2	不正解
13	1	0	6

表 3 のようになり、誤差 0 の正解率は、65%となり、誤差 1 を合わせると、70%となった。

6. 考察

実験では、期待通り以上の精度を出すことができたが、実際に撮影した画像を入力とすると、精度 70%となり、実験ほどの高精度な結果とはならなかった。これは、やはり基礎工学部棟を 56 箇所に分けて撮影した画像から訓練しているため、その箇所と箇所の間で撮影した画像を入力すると、どうしても精度が落ちてしまうと考えられる。本研究での手法は、自己位置推定を分類問題として扱った手法であり、準備された問題（本研究では、56 箇所 138 方向の画像入力）の推定は得意であるが、準備されていない問題（本研究では、56 箇所 138 方向以外の箇所と箇所の間で撮影した画像）での推定が苦手であることがわかった。この屋内自己位置推定を分類問題でより高精度に実現するためには、本研究では基礎工学部棟 56 箇所に分けたが、それより多くの箇所に分割して撮影し、データ量を増やす必要があると考える。しかし、データ量を増やすと、既存の屋内測位システムよりはるかに手間のかかる作業が必要となってしまうことが問題となる。

7. 結論

本研究において、ディープラーニングを用いた画像による屋内自己位置推定の手法の可能性を示すことができた。実験 2 より似通った景色の画像の分類を得意とするモデルで推定精度を改善することができ、実験 3 より入力画像を 1 から 2 に増やすことで、精度は大きく改善することができた。これは、データ量を増やし、画像による屋内推定をする場合でも同じような効果が期待できるだろう。

また、本研究での手法での推定が、準備されていない問題での推定が苦手であることやデータ量を増やすことによる手間を考慮すると、少ないデータ量でより柔軟な推定のできるようなシステムへの新たなアプローチが必要であると考ええる。

8. 謝辞

本研究を行うにあたり、ご多忙の中ご指導くださったアドバイザー教員の松原崇先生に深く感謝申し上げます。

9. 参考文献

- [1] 海老澤颯, 田島孝治, 出口利憲, ” CNN を用いた画像による屋内測位”, 電気・電子・情報関係学会東海支部連合大会講演論文集(CD-ROM) (電気・電子・情報関係学会東海支部連合大会(CD-ROM)),2019, ROMBUNNO.F3-1
- [2] Hirokazu Miura, Kazuhiko Hirano, Noriyuki Matsuda, Hirokazu Taki, Norihiro Abe, Satoshi Hori, ” Indoor Localization for Mobile Node Based on RSSI”, Knowledge-Based Intelligent Information and Engineering Systems, 2007, Volume 4694
- [3] Pengguang Chen, Shu Liu, Hengshuang Zhao, Jiaya Jia, ” Grid Mask Augmentation”, arXiv:2001.04086,2020
- [4] 赤石雅典, ” 最短コースでわかる PyTorch & 深層学習プログラミング”, 日経 BP,2021

