

AIを活用した画像診断による橋脚の 地震時被災度判定の自動化に向けた検討

服部 匡洋¹・赤松 伸祐²・井藤 貴文³・茂呂 拓実⁴・金治 英貞⁵

¹正会員 博士(工学) 阪神高速先進技術研究所 調査研究部（〒541-0054 大阪市中央区南本町4-5-7）

²正会員 学士 阪神高速先進技術研究所 調査研究部（〒541-0054 大阪市中央区南本町4-5-7）

³正会員 修士 阪神高速道路 技術部（〒530-0005 大阪市北区中之島3-2-4）

⁴正会員 修士 阪神高速道路 技術部（〒530-0005 大阪市北区中之島3-2-4）

⁵正会員 工学博士 阪神高速道路 技術部（〒530-0005 大阪市北区中之島3-2-4）

1. はじめに

1995年1月17日の兵庫県南部地震では、阪神高速道路においても、3号神戸線や5号湾岸線で上部構造の落橋や支承の破壊、橋脚の倒壊など、甚大な被害を受けた。このとき、早急に構造物の損傷状況を把握し、迅速に応急復旧計画を策定することが望まれたが、実際には応急復旧計画を策定するまでにおよそ1ヵ月程度の時間を要した¹⁾。この一因として、職員が撮影した多数の被災構造物の写真を基に、複数の技術者が表-1に示す「被災度判定の基本方針」²⁾に基づき、図-1のとおり被災度判定を実施しており、判定のばらつきを担当者間で調整するのに時間を要したことが考えられる³⁾。

このような経験から、阪神高速では大規模地震後にも早期に道路サービスを再開し、道路機能を確保するために、事業継続計画（BCP）を定めるとともに、「阪神高速サイバーインフラマネジメント」を提案し、地震時の構造物の損傷写真の撮影・収集から、被災度の判定、被災写真及び被災度判定と橋脚の位置情報の紐づけ、復旧工期及び復旧工費の算出までを効率的・自動的に行い、3日以内に本復旧計画を策定することを目指している⁴⁾。（図-2）

本検討では、その要素技術の開発として、AIを活用した画像診断による地震時の被災度判定の自動化を試みた。本稿では、その第一段階として、RC単柱橋脚を対象に、学習の試行や学習用データの偏りの検証、最適な学習パラメータの検討を行った結果について示すものである。

2. 画像診断による被災度の自動判定方法

(1) 自動判定方法の検討

本検討では深層学習を用い、1枚の画像入力（写真）に対し単一のラベル（被災度ランク）を得ることを考える。深層学習のモデル構築にあたり、画像

表-1 被災度判定と被害状況の対応

ランク	被害状況
As	崩壊、倒壊した場合、またはそれに類するもの
A	耐力に著しい影響がある損傷を生じており、致命的な二次被害の可能性があるもの
B	耐力に影響のある損傷であるが、余震、活荷重による損傷の進行がなければ当面の利用が可能なもの
C	短期的には耐力に影響のないもの
D	耐力に関して特に異常が認められないもの

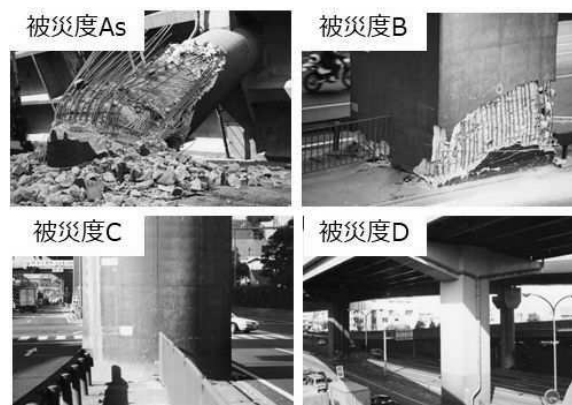


図-1 構造物の被災状況と被災度判定

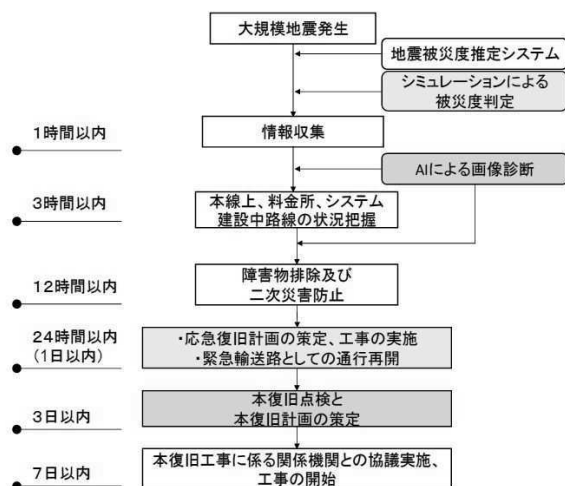


図-2 震災復旧の基本的な流れ
(参考文献4) 図-8より引用)

サイズは一定サイズにリサイズすることとした。これは、画像に写る橋脚の大きさが画像によって大きく異なるが、学習の計算の都合上画像サイズが一定である方が効率的に学習を進めることができるためである。また、深層学習の最終層の出力がラベルごとの確率となるよう構築し、最も高い確率を持つラベルをその橋脚の判別結果とした。

(2) 使用する画像データとラベル

本検討では、阪神高速道路3号神戸線に位置する橋脚のうち、RC単柱橋脚約450基を対象とする。写真は、被写体の橋脚と被災状況が視認できる1,108枚を使用した。

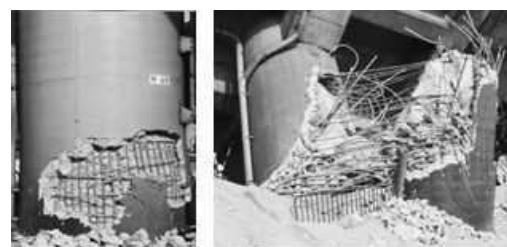
表-1に被災度ランクの定義を示す。表-1に示すとおり、被災度ランクには、被害が大きなものから、As, A, B, C, Dの5種類がある。しかし、兵庫県南部地震において、最も被害の大きいAsの画像枚数が非常に少なかったことからA判定に統合することとし、A, B, C, Dの4種類のラベルを用いて検討することとした。その結果、A, B, C, Dの画像はそれぞれ494枚、155枚、114枚、345枚であった。

(3) サンプル画像の準備

サンプル画像を準備するため、まず橋脚番号と外観評価のリストを照らし合わせ、各画像に被災度A, B, C, Dをラベルとして付与した。また、図-3に示すように、画像編集ツールを用いて橋脚部の矩形領域を手動で設定し、橋脚部画像の切り抜きを行った。さらに、図-4に示すとおり切り出した画像はすべて128×128 (pixel) にリサイズした。



図-3 橋脚部の切り出し



切り出し画像のリサイズ



図-4 切り出した画像のリサイズ

(4) 切り出し画像の学習

図-5にニューラルネットワークを用いた学習イメージをに示す。本検討では入力画像サイズを一定とし、最終層の出力を被災度としてニューラルネットワークを構築する。構築したネットワークは畳み込みニューラルネットワーク (CNN) の1種であるResidual Network (ResNet)⁵⁾と呼ばれるネットワークの小規模版を採用した。

3. 学習の試行

本検討の実現可能性を確認するため、ニューラルネットワークによる学習の試行を実施した。この試行では、画像10枚ごとに1枚を検証用画像として振り分け、学習と判別を実施しF値により評価した。ここでの最適化アルゴリズムはAdam、学習率の係数は0.001、バッチサイズは32とした。表-2に学習及び検証に使用した画像枚数を示す。

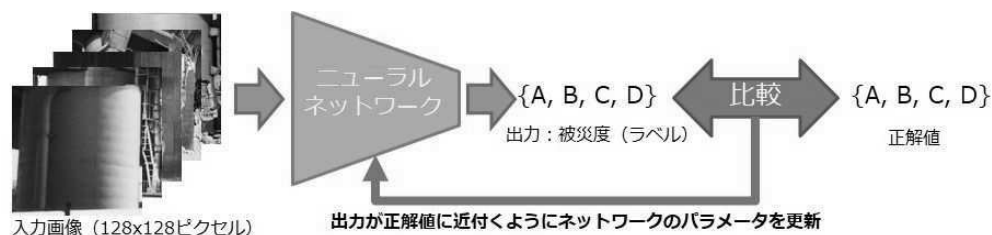


図-5 深層学習のイメージ

表-2 学習用及び検証用の画像枚数 (学習の試行)

被災度(ラベル)	学習用	検証用
A	444	50
B	139	16
C	102	12
D	310	35
合計	995	113

表-4 ラベルごとの F 値 (学習の試行)

F 値			
D	C	B	A
0.81	0.57	0.39	0.82

表-3 混同行列とラベルごとの再現率, 適合率 (学習の試行)

		推定値				再現率
		D	C	B	A	
真値	D	30	0	5	0	0.86
	C	3	6	0	3	0.50
	B	3	3	7	3	0.44
	A	3	0	8	39	0.78
適合率		0.77	0.67	0.35	0.87	

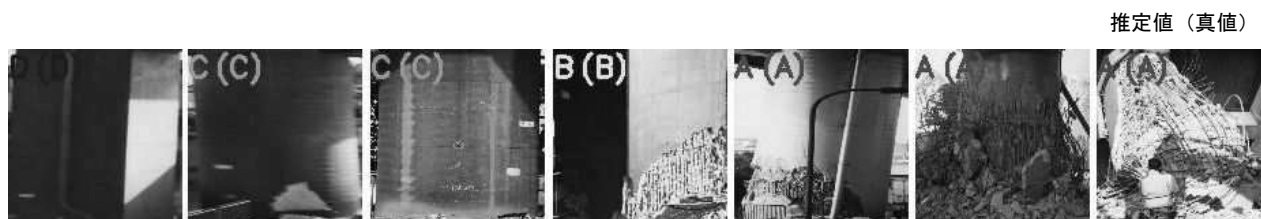


図-5 認識成功例 (試行)



図-6 認識失敗例 (試行)

表-3に試行により得られた混同行列とラベルごとの再現率と適合率を, 表-4にラベルごとのF値を示す. ここで, 再現率は実際にあるラベルであるもののうち, そのラベルであると予測されたものの割合を, 適合率はあるラベルと予測したデータのうち, 実際にそのラベルであるものの割合を示す. また, F値とは正確性と網羅性の総合的な評価の際に利用される尺度であり, 再現率と適合率の調和平均で表される. これらの値は大きいほど精度が高いことを示す. 試行の結果, 平均適合率は0.66, 平均再現率は0.64, 平均F値は0.65であった. サンプルが比較的多いAとDでは再現率, 適合率ともに高い傾向が見

られ, サンプルの少ないCとBでは再現率適合率ともに低い傾向が見られた.

また, 図-5に試行での認識の成功例を, 図-6に失敗例を示す. 劣化部位が目視で分かりやすい場合は正しく判別できるものの, 地表面付近に大きな損傷がある場合や, 解像度の制約上ひび割れが目視でも確認できない場合などに誤判別が多く確認された. また, BとCにおいてはサンプル数が少ないことに加え, 正解データとしている人の判別結果にもばらつきがあることが正解率が低い原因と考えられる. 以上の結果, 誤認識も多く見られるものの, 正解データにばらつきが含まれることを考慮すると, ニュ

ーラルネットワークを用いることで、一定精度の被災度判定を得ることができるといえる。

4. データの偏りの検証（交差検証）

3.で実施したように、データセットをあるルールによって学習用と検証用に振り分けた場合、特異なサンプルがことごとく検証用のデータに含まれるなど、データの性質的な偏りがある可能性がある。そのため、振り分けのパターンを複数回試すことにより、データセットの違いでどの程度モデルの性能が変化するかの確認を行った。

(1) 検証手順

まず、全サンプルから1割ランダムに抽出し、これを検証用に、残りを学習用とする。続けて、学習と検証を行い、各ラベルのF値、および平均F値を計算する。以上の手順を繰り返し行う。ここでは、10回の検証を実施した。

(2) 検証結果

表-5に各試行の結果（F値）に示す。平均F値は最小で0.56、最大で0.67となった。ラベルごとのF値を見ると、BとCは試行ごとに精度が大きく変わっている。これは、D以外の各ラベルのデータのばらつきが大きく疎になっており、近い特徴を持つデータが学習用と検証用の両方に入りづらくなっていると考えられる。また、学習時のランダム要素とGPU計算による誤差があり、学習を実行するたび結果が変わることが課題として明らかになった。

5. 学習パラメータの検討

本検討では、使用できるサンプルが限られているため、現状のサンプルで学習を繰り返した場合にできるだけ結果が安定するよう学習の仕方を工夫する必要がある。

(1) ニューラルネットワークの基本構造

図-7にニューラルネットワークの基本構造であるパーセプトロンの構造を示す。パーセプトロンは入力 x に対し重み w をかけ、バイアス b を足した結果 u に活性化関数 ϕ を通したものを出力 z とする構造になっている。これを多層化したものをニューラルネットワーク、層数の多いものをディープニューラルネッ

表-5 試行回数ごとのF値（交差検証）

		試行回数									
		1	2	3	4	5	6	7	8	9	10
ラ ベ ル	D	0.75	0.68	0.86	0.80	0.67	0.83	0.85	0.75	0.80	0.76
	C	0.53	0.21	0.38	0.53	0.47	0.33	0.40	0.24	0.53	0.33
	B	0.55	0.42	0.48	0.40	0.30	0.55	0.40	0.44	0.39	0.43
	A	0.85	0.92	0.83	0.88	0.88	0.87	0.86	0.85	0.88	0.87
平均		0.67	0.56	0.64	0.65	0.58	0.64	0.63	0.57	0.65	0.60

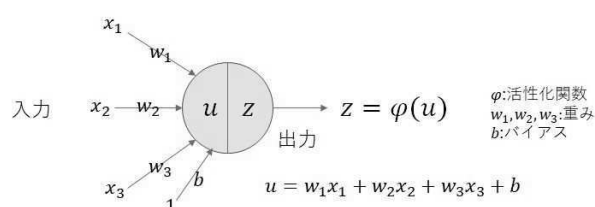


図-7 ニューラルネットワークの基本構造

トワークと呼び、深層学習はこのディープニューラルネットワークのパラメータ w を学習（最適化）することを指す。深層学習モデルの性能は、学習データセットをはじめ、ネットワークの構造、最適化アルゴリズム、損失関数、各種パラメータ等様々な要素によって決まる。

(2) 検討ケース

本検討では、最適化アルゴリズムと学習パラメータについて検討する。このうち、最適化アルゴリズムには最良と考えられるAdam、スパースなデータを扱うのに適するAdaGradの2種類について検討する。また、学習パラメータは、バッチサイズと学習率の係数について試みる。ここで、学習率の係数とは、誤差逆伝播によるネットワークパラメータ更新時の係数である。これは、AdamとAdaGradに共通するパラメータである。また、バッチサイズとは繰り返し学習時のデータサンプリング数である。これは、最適化アルゴリズムによらないパラメータであり、小さいほど1つ1つのデータに敏感に反応する。本検討では、以下の計8パターンについてそれぞれ10回の試行を実施し、学習の安定性と精度を検証した。

最適化アルゴリズム：Adam, AdaGrad

学習率の係数：0.001, 0.0001

バッチサイズ：32, 16

また、データセットの学習用・検証用の振り分けによる違いも考慮するため、表に示した交差検証の際の1回目および2回目試行のデータセットを用いた。

表-6 各パターンでの平均F値（学習パラメータの検討）

(a) データセット 1

最適化 アルゴリズム	学習率の 係数	バッチ サイズ	D の 平均 F 値	C の 平均 F 値	B の 平均 F 値	A の 平均 F 値	平均 F 値の 平均
Adam	0.001	32	0.769	0.248	0.294	0.879	0.547
		16	0.776	0.170	0.410	0.910	0.567
	0.0001	32	0.786	0.387	0.470	0.899	0.636
		16	0.790	0.360	0.456	0.905	0.628
AdaGrad	0.001	32	0.793	0.332	0.477	0.902	0.626
		16	0.785	0.308	0.489	0.894	0.619
	0.0001	32	0.739	0.157	0.089	0.871	0.475
		16	0.765	0.201	0.163	0.881	0.503

(b) データセット 2

最適化 アルゴリズム	学習率の 係数	バッチ サイズ	D の 平均 F 値	C の 平均 F 値	B の 平均 F 値	A の 平均 F 値	平均 F 値の 平均
Adam	0.001	32	0.838	0.356	0.515	0.896	0.651
		16	0.849	0.402	0.577	0.899	0.682
	0.0001	32	0.856	0.560	0.563	0.912	0.723
		16	0.846	0.444	0.644	0.930	0.716
AdaGrad	0.001	32	0.845	0.455	0.603	0.908	0.703
		16	0.845	0.538	0.612	0.913	0.727
	0.0001	32	0.745	0.304	0.265	0.825	0.535
		16	0.726	0.199	0.268	0.818	0.503

表-7 各パターンでのF値の標準偏差（学習パラメータの検討）

(a) データセット 1

最適化 アルゴリズム	学習率の 係数	バッチサイズ	D の F 値の 標準偏差	C の F 値の 標準偏差	B の F 値の 標準偏差	A の F 値の 標準偏差	平均 F 値の 標準偏差
Adam	0.001	32	0.0500	0.1679	0.1415	0.0226	0.0736
		16	0.0297	0.1122	0.1325	0.0276	0.0575
	0.0001	32	0.0310	0.1621	0.1264	0.0259	0.0718
		16	0.0290	0.1500	0.2109	0.0226	0.0821
AdaGrad	0.001	32	0.0317	0.1099	0.1044	0.0165	0.0440
		16	0.0283	0.1102	0.0916	0.0151	0.0395
	0.0001	32	0.0191	0.0755	0.0670	0.0148	0.0415
		16	0.0250	0.0636	0.0953	0.0173	0.0375

(b) データセット 2

最適化 アルゴリズム	学習率の係数	バッチサイズ	D の F 値の 標準偏差	C の F 値の 標準偏差	B の F 値の 標準偏差	A の F 値の 標準偏差	平均 F 値の 標準偏差
Adam	0.001	32	0.0354	0.1988	0.1142	0.0267	0.0755
		16	0.0362	0.1821	0.0780	0.0202	0.0550
	0.0001	32	0.0378	0.1281	0.0904	0.0171	0.0471
		16	0.0298	0.1209	0.0614	0.0185	0.0372
AdaGrad	0.001	32	0.0300	0.1272	0.0939	0.0102	0.0481
		16	0.0334	0.0932	0.0752	0.0161	0.0289
	0.0001	32	0.0213	0.0662	0.0703	0.0146	0.0298
		16	0.0255	0.0936	0.0811	0.0202	0.0355

(3) 検討結果

表-6、表-7に各パターンにおける検討結果を示す。ここで、ラベルごとのF値と平均F値、および平均F値の平均により学習の安定性を、ラベルごとのF値および平均F値の標準偏差により精度を表す。

精度については、全体の傾向として、データセット2の方が平均F値の平均が高くなる傾向があった。また、どちらのデータセットでも、Adamでは学習率の係数0.0001、AdaGradでは学習率の係数0.001の場合にそれぞれのF値が高くなる傾向がみられた。バッチサイズによる違いはほとんど現れなかった。データセット2を用い、AdaGradの学習係数を0.001、バッチサイズを16とした場合に平均F値の平均が最大となった。

また、安定性については、Adamに比べAdaGradの方がF値の標準偏差が小さくなる傾向にあり、今回のデータに対してはAdaGradの方が安定した結果が得られるといえる。また、AdaGradでは学習率の係数が小さい方が安定する傾向が見られるが、安定性が高い場合の精度（平均F値）は全体的に低くなっていることから、学習中に最適解に到達する前に、学習率が減衰し、学習不足の状態で収束してしまっていると考えられる。

以上を踏まえ、本条件における最適なパラメータの組み合わせは、AdaGradの学習率の係数0.001といえる。なお、バッチサイズの違いによる影響は小さく、いずれでもよい。

(4) 最適パラメータでの検討結果

表-8に平均F値が最大となった最適パラメータにおける10回の試行結果を示す。また、表-9、表-10には10回の試行のうち、最も平均F値が高くなった9回目の試行の混同行列とF値を示す。適合率の平均値、再現率の平均値、F値の平均値はそれぞれ0.82、0.74、0.77となり、3.で実施した試行と比べどの数値も0.1程度向上したことから、最適パラメータを選定したことにより精度が向上したといえる。

図-8に最適パラメータを用いた検討における誤認識例を示す。図-8(a)では外観上大きな損傷が生じているように見えるにもかかわらず正解ラベルはCとなっており、正解ラベルのばらつきが影響したものと考えられる。また、図-8(b)は損傷部位が白飛びしており損傷状況が見えづらい、図-8(c)は損傷部位が確認できない、図-8(d)は解像度の制約上ひび割れが確認できない、図-8(e)は損傷部位が画像右端しか確認できないという特徴がある。このよう

表-8 最適パラメータでの試行

		試行回数									
		1	2	3	4	5	6	7	8	9	10
ラベル	D	0.86	0.87	0.89	0.87	0.82	0.79	0.83	0.79	0.87	0.86
	C	0.50	0.47	0.70	0.53	0.56	0.40	0.59	0.44	0.67	0.53
	B	0.71	0.61	0.50	0.67	0.55	0.65	0.52	0.56	0.64	0.71
	A	0.92	0.92	0.91	0.90	0.92	0.94	0.88	0.92	0.89	0.92
平均		0.75	0.72	0.75	0.74	0.71	0.70	0.71	0.68	0.77	0.76

表-9 最適パラメータにおける混同行列

		推定値				再現率
		D	C	B	A	
真値	D	31	1	1	1	0.91
	C	3	6	1	1	0.55
	B	1	0	9	6	0.56
	A	2	0	1	46	0.94
適合率		0.84	0.86	0.75	0.85	

表-10 最適パラメータにおけるラベルごとのF値

F 値			
D	C	B	A
0.87	0.67	0.64	0.89

推定値（真値）

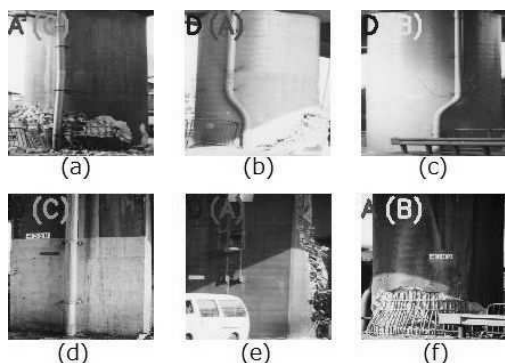


図-8 最適パラメータでの誤認識例

な事例は学習データにそれほど多く含まれていないため、誤認識したと考えられる。さらに、図-8(f)では損傷部位にガードレールおよび柵が重なるとともに、撮影面の外観損傷ではAのようにも見えるため、Aと判定されたと考えられる。その他の誤認識についても確認すると、図-8(f)と同様の誤認識が散見された。ニューラルネットワークは抽出すべき特徴を自動的に学習するため、人間の考える判断基準とはしばしば異なる特徴、すなわち今回の場合ではガードレールや柵より判別していることが原因と考えられる。このような現象は、学習用画像の中に同様の特徴を含む画像が多い場合に起こりうることから、より正確に判定を行うためには、損傷写真を撮影するためのルールを定めるのが望ましいといえる。

6. まとめ

本検討により得られた知見を以下にまとめる。

- ・ 深層学習を用いた被災度自動判定の試行を行った結果、誤認識するものも見られるが、一定精度で被災度判定を得ることができた。
- ・ データのばらつきを交差検証により確認した結果、BやCでは試行ごとに大きく精度が変化したことから、各ラベルのデータに偏りがあり、近いデータが学習用と検証用の両方に入りづらい傾向にあると考えられる。また、学習時のランダム要素とGPU計算による誤差のため、学習を実行するたび判別結果が変わるという計算上の課題が明らかになった。
- ・ 可能な限り学習結果を安定させつつより精度の高いモデルを作成するため、学習パラメータの検討を行った結果、本条件では最適化アルゴリズムにAdaGradを用い、学習率の係数を0.001とするのがよい。

今後は、まず正解ラベルやそのラベルと橋脚番号の紐づけについて精査した上で、ある面にのみ損傷が集中しているような橋脚に対しても精度よく判別できるよう、同一の橋脚を撮影した複数の画像から判別された結果を統合する処理を検討することが必要と考える。また、正解率を高めるためには学習データを増やすことも重要と考える。

謝辞：本研究を進めるにあたっては、(株)構造計画研究所関係各位にご協力いただいた。ここに記して謝意を表する。

参考文献

- 1) 阪神高速道路公団：大震災を乗り越えて—震災復旧工事誌一，1997.
- 2) 阪神高速道路株式会社：高速道路の被害分類，<https://hanshin-exp.co.jp/earthquake/classifying1.html>（2021年7月1日閲覧）
- 3) 阪神高速道路公団：阪神高速神戸線復旧への軌跡—新技術と使命感でクリア！—，1997.
- 4) 伊佐 政晃，茂呂 拓実，金治 英貞：阪神高速サイバーインフラマネジメントの取り組みと価値創造の可能性，AIデータサイエンス論文集，2020年1巻J1号pp.252-260，2020.
- 5) He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.