

深層学習による異常検知手法の簡単な比較（第 6 報）

Fully Convolutional Siamese Network

廣川 勝久

Anomaly Detections using Deep Learning (VI)

Fully Convolutional Siamese Network

HIROKAWA Katsuhisa

Siamese Network を用いた異常検知手法を提案する。提案手法は fully convolutional 構造により position-invariant な特徴抽出を実現し、特徴ベクトルとそのユークリッド距離の連結処理によりクラスタリング性能の向上を図る。さらに、低解像度で学習したモデルを高解像度画像に直接適用することで、異常領域を位置情報とともに正確に検出できることを実験的に示す。また、未学習の画像に対しても異常検知が可能であることを確認する。

キーワード： Siamese Network、 fully convolutional、 position-invariant、 contrastive loss、 metric-based

1. 緒 言

異常検知における異常は、次の 2 種類に大別できる。

1. システムはほぼ正常に動作しているが、まれに予測不可能な異常が発生する場合。
2. 検出すべき異常の内容は判明しているが、正常状態そのものが変化する場合。

製造装置やネットワーク通信における異常は前者に属することが多く、製品や印刷物の検査における不良検出は後者に属することが多いと考えられる。とりわけ後者に属する広大な領域を対象とするリモートセンシングにおける物体検出や、製造品種が頻繁に切り替わる産業用検査画像における不良検出を想定した場合、ニューラルネットワークを用いた従来の異常検知手法にはいくつかの課題が存在する。すなわち、リモートセンシングや製品検査では検出対象の異常は特定可能である一方、それらを含む画像は広大かつ多様であり、学習に膨大なデータ量と計算資源を要する。さらに、製造品種の追加や変更が頻繁に生じるため、過学習や頻繁な再学習といった問題が避けられない。

これまで我々は、予測不可能な異常が発生するケースを対象に、正常状態のみを学習し、異常発生時に正常状態を生成してその差分から異常を検出する手法として AutoEncoder^{1,2)}、 Boltzmann Machine³⁾や Generative Adversarial Networks : GAN⁴⁾を用いた方法について報告を行った。その他に、One-Class Neural Network では、学習済み AutoEncoder の Encoder に Support Vector Machine: SVM を模したニューラルネットワークを接続し、正常な特徴ベクトルのクラスタに対する分離平面を学習して異常検知を行う手法を評価した⁵⁾。しかし、このアプローチでは AutoEncoder による特徴ベクトルのクラスタリングが十分

でなく、異常な特徴ベクトルが正常クラスタに混入する問題を確認した。

本稿では、何が異常か明確に定義されているケースに対応するため、Siamese Network⁶⁾を用いた異常検知手法を提案する。本手法は fully convolutional 構造に基づき、position-invariant な特徴抽出を実現する。この構造により、低解像度画像で学習したモデルを高解像度画像にそのまま適用可能とし、異常領域を位置情報とともに正確に検出できる。また、本手法では特徴ベクトルとそのユークリッド距離を連結する構造を導入し、クラスタリング性能の向上を実現している。さらに、本手法における特徴は、学習済みモデルが特徴量間の相対的な距離に基づいて異常を識別する点にある。このため、正常画像の特徴値分布を網羅的に学習する必要がなく、新たな正常画像が追加された場合でも再学習を行わずに適用が可能である。

2. Siamese Network

2.1 Siamese Network と異常検知

Siamese Network は、メタ学習における metric-based 手法に分類されるモデルである。Siamese Network では、構造および重みを共有する 2 つのニューラルネットワークを用いて入力データの特徴量を抽出し、特徴空間内においてその距離に基づく学習を行う。具体的には、ペアとして与えられた入力データをそれぞれネットワークに入力し、得られた特徴ベクトル間の距離が小さくなるよう類似ペアを、逆に距離が大きくなるよう非類似ペアを最適化する。この学習過程により、Siamese Network は特徴空間内において類似データをクラスタリングする能力を獲得する。本研究では、この性質を利用し、正常データと異常データ間の距離を学習させることで、異常検知への応用を試みる。

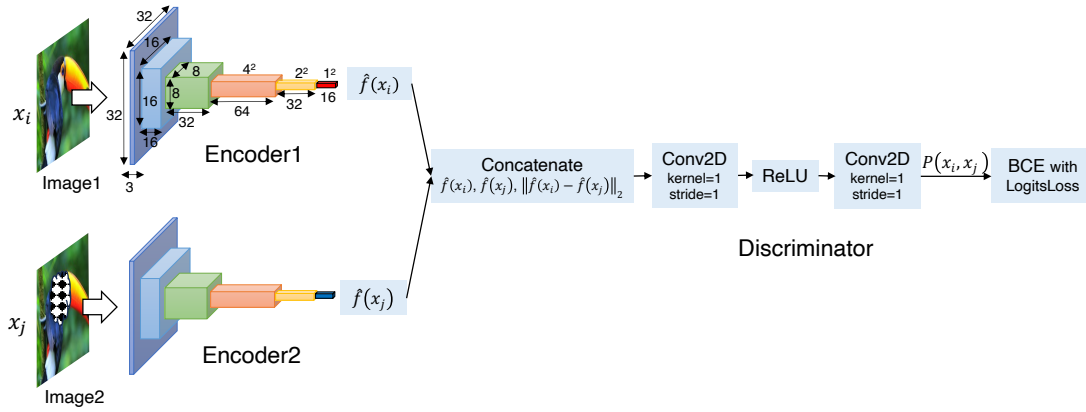


図 1 Fully Convolutional Siamese Network の構造

2.2 Fully Convolutional Siamese Network の構造

Siamese Network は、本来、入力されるデータペアが同一クラスに属するか否かを判定するために提案されたニューラルネットワークである。そのため、従来の構成では、入力画像ペアに対する類似性の判定は可能であるものの、類似領域の位置を特定することはできない。

本研究では、この制約を克服するため、図 1 に示すように、Encoder および Discriminator の双方を 2 次元畳み込み層のみで構成した Fully Convolutional Siamese Network を提案する。本モデルでは、低解像度画像で学習した重みパラメータを高解像度画像にそのまま適用可能であり、再学習を必要とせず、位置に対応した類似性の推定が実現される。ただし、得られる位置分解能は、学習時に用いた低解像度画像の画像サイズに依存する。

異常検知への応用においては、正常画像と異常画像を低解像度で学習させることで、実運用環境下では高解像度入力画像から異常領域を位置とともに特定できる。具体的には、Siamese Network の一方に高解像度正常画像をテンプレートとして入力し、他方に検査対象画像を入力することで、異常の有無およびその位置を同時に検出可能となる。さらに、本手法はテンプレート画像を未学習の正常画像に置き換えた場合であっても、異常領域を検出可能である。

2.3 Contrastive Loss

Contrastive Loss は、Siamese Network において特徴空間上で類似データを近づけ、非類似データを離すように学習させるための損失関数である⁷⁾。Siamese Network では、入力ペアデータ (x_i, x_j) をそれぞれのネットワークに入力し、対応する特徴ベクトル $\hat{f}(x_i)$ および $\hat{f}(x_j)$ を得る。このとき、両者のユークリッド距離 $d(x_i, x_j)$ は次式で定義される。

$$d(x_i, x_j) = \|\hat{f}(x_i) - \hat{f}(x_j)\|_2 \quad (1)$$

Contrastive Loss は、この距離を用いて次のように定義される。

$$\mathcal{L}_{CL} = \begin{cases} \frac{1}{2} d(x_i, x_j)^2, & \text{if } y = 1 \\ \frac{1}{2} \max(r - d(x_i, x_j), 0)^2, & \text{if } y = 0 \end{cases} \quad (2)$$

ここで、 $y \in \{0,1\}$ はラベルを表し、 $y = 1$ は類似ペア、 $y = 0$ は非類似ペアを表す。また r は非類似ペア間の距離マージンである。この損失により特徴空間では類似ペアは近接し、非類似ペアは少なくともマージン r 以上隔てられるよう最適化される。

2.4 Binary Cross-Entropy Loss

Siamese Network における損失関数である Contrastive Loss は、特徴空間において入力データ間の高いクラスターリング分離能力を与える。一方で、図 1 に示すような Convolutional Neural Network: CNN を用いた Discriminator に直接適用することは困難である。本研究では、Encoder および Discriminator の双方を CNN により実装するために、Encoder からの出力データ間のユークリッド距離 $\|\hat{f}(x_i) - \hat{f}(x_j)\|_2$ に加え、各特徴ベクトル自体を連結し、Discriminator への入力とする手法を導入した。

$$d(x_i, x_j) = \text{Concat}[\hat{f}(x_i), \hat{f}(x_j), \|\hat{f}(x_i) - \hat{f}(x_j)\|_2] \quad (3)$$

この連結表現は、Graph Attention Network: GAT における処理と類似しており、ユークリッド距離に加えて各特徴ベクトル自体にも Attention が働くことを可能とする。その結果、類似データのクラスターリング性能が一層向上することを確認した。Discriminator は、連結されたベクトルを 1×1 の 2 次元畳み込み層を 2 層通過させる構造とし、その出力を次式で表す。

$$P(x_i, x_j) = \text{Conv}_{1 \times 1} \left(\text{ReLU} \left(\text{Conv}_{1 \times 1} (d(x_i, x_j)) \right) \right) \quad (4)$$

ここで、 $P(x_i, x_j) \in \{0,1\}$ は入力ペア (x_i, x_j) が類似ペアで

ある確率を表す。損失関数としては、二値分類の標準的な binary cross-entropy を用いた。

$$\mathcal{L}_{BCE} = y(x_i, x_j) \log P(x_i, x_j) + [1 - y(x_i, x_j)] \log [1 - P(x_i, x_j)] \quad (5)$$

これにより、Encoder は類似データごとに特徴空間内でクラスタを形成する。また、本研究で設計したネットワークは、全結合層を排除した fully convolutional 構造を採用しており、入力画像サイズに依存しない position-invariant な特性を有する。このため、低解像度画像で獲得した重みパラメータを高解像度画像に対しても再学習なしに適用可能であり、異なる解像度を扱う実環境での応用に有効である。

3. Siamese Network の実装

3.1 学習データ

Siamese Network のクラスタリング性能を評価するために、Modified National Institute of Standards and Technology : MNIST データセットを用いた⁸⁾。本データセットにより、10 種類の数字画像を対象として、Contrastive Loss および binary cross-entropy loss による損失関数の違いや、Discriminator への入力方法の差異が特徴空間における分布に与える影響を確認する。

さらに、Siamese Network の異常検知性能を評価するため、CIFAR-10 データセット⁹⁾を用いて画像中に意図的に異常領域を挿入し、異常画像の学習および検出能力を検証した。加えて、CIFAR-10 に基づき異常検知を学習したニューラルネットワークを、未学習の高解像度リモートセンシング画像¹⁰⁾へ適用し、異常領域およびその位置の検出性能を評価する。

3.2 Siamese Network の実装

binary cross-entropy loss との比較を行うため、(2)式の Contrastive Loss を以下のように実装した。

```
def forward(self, x1, x2, y):
    B=x1.size(0)
    x = (x1.view(B,-1) - x2.view(B,-1)) ** 2
    D2 = x.sum(dim=1)
```

```
D = D2**0.5
D = F.relu(self.margin - D)
loss = y * D2 + (1 - y) * D**2
loss = 0.5 * loss.mean()
return loss
```

一方、(5)式の binary cross-entropy loss については、PyTorch に実装されている log 機能を含む標準関数を用いた。

```
criterion = nn.BCEWithLogitsLoss()
```

また、Fully Convolutional Siamese Network のメインプログラムは、1 つの CNN Encoder に対して 2 つの入力データを与え、それぞれの出力を用いて勾配計算を行う構造とした。ここで使用した CNN モジュールは、AutoEncoder の性能検証に用いた Encoder 部を転用した²⁾。実装例を以下に示す。

```
def forward(self, x1, x2):
    x1 = self.cnnnet(x1)
    x2 = self.cnnnet(x2)
    x = torch.abs(x1-x2)
    x = torch.cat([x1,x2,x],dim=1)
    x = self.conv(x)
    x = self.relu(x)
    x = self.conv2(x)
    return x
```

(3)式および(4)式についても同様に容易に実装可能である。

4. Siamese Network の性能評価

4.1 クラスタリング性能評価

Siamese Network の異常検知への応用に先立ち、本節では MNIST データセットを用いて、Siamese Network のクラスタリング性能を比較・検証した。具体的には、contrastive loss を用いた場合、および Discriminator の入力方法を変えた場合における特徴空間の分布の違いを評価した。

図 2 に Siamese Network からの特徴ベクトル次元を 3 次元に設定した場合のクラスタ分布を示す。図 2(a)は Contrastive Loss を用いた場合であり、10 種類の数字が特

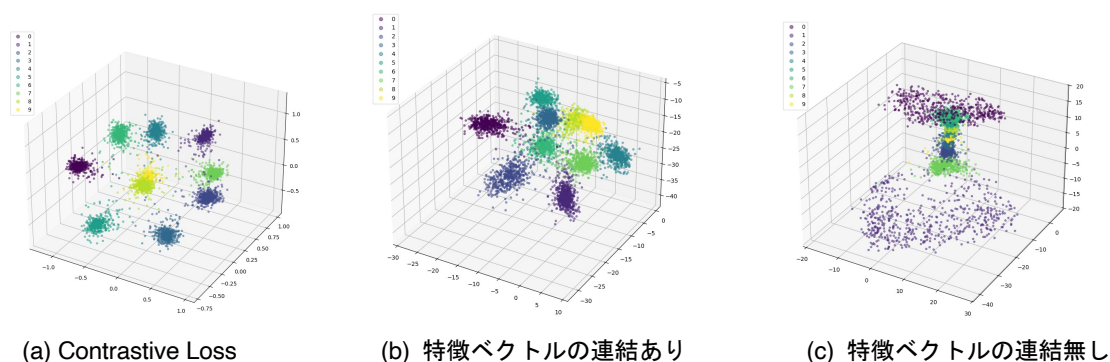


図 2 特徴空間内のベクトル分布

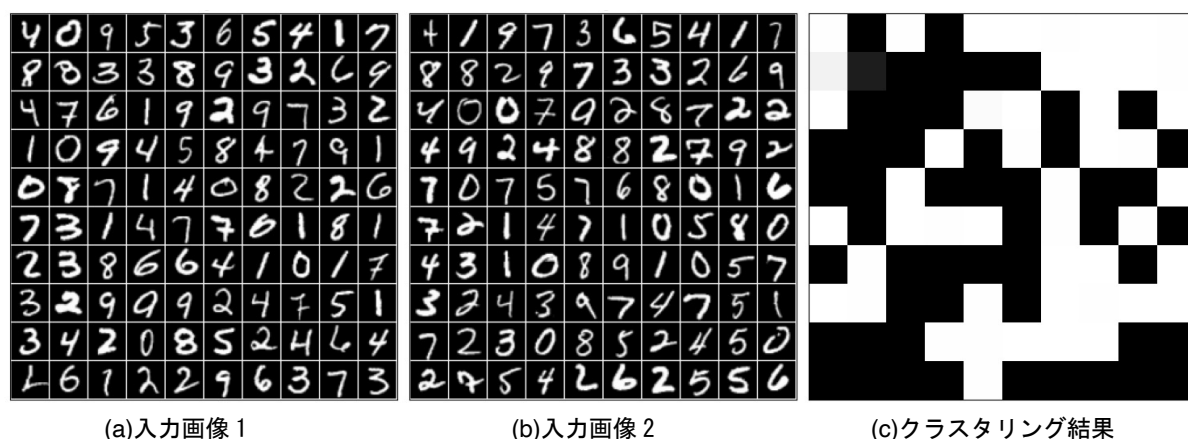


図3 100枚のMNIST未学習画像を使ったクラスタリング結果

微空間内で一定の距離を保ちつつ明確にクラスタリングされている。図2(b)には(3)式で示した連結ベクトルを入力とした場合の結果を示す。Contrastive Lossの場合ほどではないが、各クラスは一定の大きさをもつクラスタを形成しており、クラス間の分離も保持されている。これに対し、図2(c)は(1)式のユークリッド距離のみをDiscriminatorに入力した場合の結果を示す。図2(b)と比較すると、クラスによってクラスタサイズに差が生じており、類似データのクラスタリング性能が低下していることがわかる。

以上の結果から、特徴ベクトルの連結処理は類似データのクラスタリング性能を向上させる有効な手法であることを確認した。なお、図2はMNISTデータセットの学習用画像からランダムに選んだ2枚の画像からなる5万ペアを用い、100エポック学習後に同データを入力した場合の特徴ベクトル分布を示している。

連結処理を用いて学習したSiamese Networkに対し、MNISTデータセットからランダムに抽出した未学習画像を入力した際のクラスタリング結果を図3に示す。図3(a)および図3(b)は、同一座標に配置された画像ペアをSiamese Networkに入力したものであり、それらの出力結果を図3(c)の同一座標に示している。ここで、黒から白への明度はSiamese Networkが出力した類似度の確率的指標を表しており、白いほど同一クラスタの確度が高いことを示す。図3(c)に示すように、全ての数字においてほぼ白と黒に分離されておりクラスタリングが正確に行われていることが確認できる。

4.2 低解像度学習に基づくSiamese Networkのスケラブル異常検知

本節では、Siamese Networkを異常検知へ応用するため、CIFAR-10データセットを用いた学習を行った。具体的には、画像中に意図的に異常領域を挿入した異常画像と正常画像を用いて学習を実施した。CIFAR-10は32×32画素の低解像度カラー画像から構成され、本研究では学習用画像5万枚を用いて20エポックの学習を行った。学習の際には、まず学習用画像からランダムに2枚を選びペ

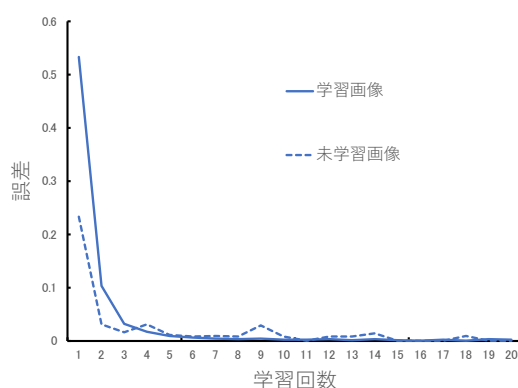
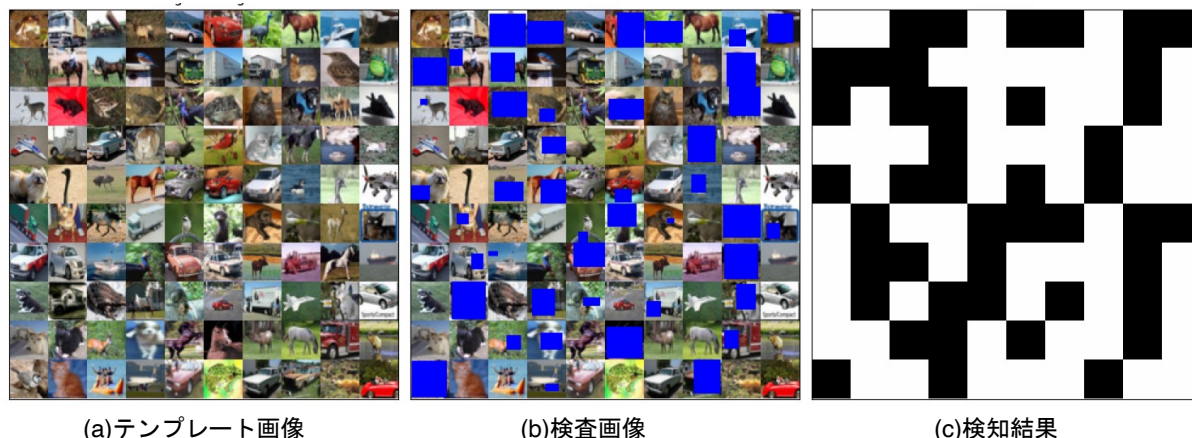


図4 CIFAR-10画像の学習曲線

アを構成した。各画像には50%の確率で異常領域を挿入し、一方のみに異常領域が存在するペアを非類似ペア、いずれも正常であるか、あるいは双方に異常領域を有するペアを類似ペアとしてラベル付けした。挿入する異常領域は、画像全体に対する面積比1%~99%、縦横比0.5~1.5の矩形とし、矩形の色は青とした。学習の収束速度は速く、図4に示す学習曲線からも明らかのように、およそ5エポック程度で収束域に達した。

さらに、低解像度画像で学習した重みパラメータを高解像度画像に直接適用し、異常領域を位置とともに特定する能力を検証した。検証には、CIFAR-10データセットの学習用画像100枚を縦横10枚に配置し、320×320画素からなる高解像度テンプレート画像を生成した。また、異常画像として、テンプレートに用いた100枚のうち50%に学習時と同様の異常領域を挿入し、1枚の高解像度検査画像を作成した。これら2枚の画像を低解像度画像で学習したモデルにそのまま適用した結果を図5に示す。図5(c)から明らかのように、高解像度画像に対して領域単位で明瞭に白黒が分離しており、異常領域の位置特定が正確に行われていることが確認できた。

この結果は、低解像度画像で学習されたSiamese Networkが、高解像度画像においても特徴量の検出能力を維持し、スケラブルに適用可能であることを示している。すな



(a)テンプレート画像

(b)検査画像

(c)検知結果

図5 100枚のCIFAR-10画像を使った異常検知結果

わち、本手法は高解像度リモートセンシング画像や産業用検査画像といった大規模データに対しても有効であり、学習データの解像度制約を受けずに異常領域の検出を実現できる点で、実運用上の有用性が高い。

4.3 高解像度未学習リモートセンシング画像における異常検知

図4に示した未学習画像の学習曲線からも推察されるように、テンプレートが未学習であっても異常領域の検出は十分に可能であると考えられる。ただし、図4で用いた未学習画像は、学習画像に類似したパターンを含む可能性も否定できない。そこで本節では、学習時に一切使用していない高解像度リモートセンシング画像をテンプレートとし、異常領域の検出を試みた。実験には、CIFAR-10データセットを用いて異常検知を学習したモデルをそのまま適用した。テンプレート画像として使用した高解像度リモートセンシング画像は解像度1260×720画素である。異常画像は、テンプレート画像に縦横5～15画素からなる矩形をランダムな位置に100個挿入することで作成した。図6に検出結果を示す。結果より、低解像度のCIFAR-10で学習したモデルであっても、全く未学習の高解像度リモートセンシング画像に対して異常領域の検出が可能であることが確認された。

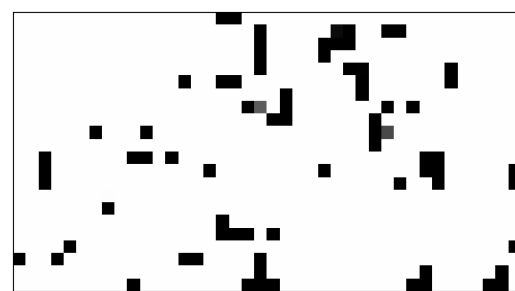
本手法における再学習不要性は、学習済みモデルが画像の絶対的な特徴量ではなく、特徴量の相対的な距離に基づいて異常を識別している点に起因する。すなわち、モデルは正常・異常といった画像のクラス分類を行うのではなく、入力ペア間の特徴量の距離を評価する。一般的なニューラルネットワークを用いた異常検知手法では、正常データの分布をモデルが直接学習するため、新たな正常画像が追加されると、正常画像、追加画像、異常画像すべての再学習が必要となる。一方、本手法は特徴量の距離に基づいているため、正常データの分布を網羅的に学習する必要がなく、追加データに対しても再学習を行わずに適用可能である。この性質は、特に全画像を学習することが困難なりリモートセンシング画像や、製造品種が頻繁に変更される産業用検査画像のように、多様か



(a)テンプレート画像



(b)検査画像



(c)検知結果

図6 高解像度リモートセンシング画像を使った異常検知結果

つ膨大なデータが取り扱われる領域において、実用上大きな利点となる。

5. 結 言

本研究では、Siamese Network を用いた異常検知手法

を提案した。各種データセットを用いた評価により、本手法は **position-invariant** な構造と特徴ベクトルの連結手法により、高いクラスタリング性能を実現することが確認された。さらに、低解像度で学習したモデルを高解像度画像にそのまま適用することで、異常領域を位置情報とともに正確に検出可能であり、追加データに対する再学習も不要である。これにより、本手法はリモートセンシング画像や産業用検査画像など、実運用における多様かつ大規模なデータ環境に対しても、効率的かつスケラブルに適用可能であることを示した。

文 献

- 1) 廣川 勝久：広島県立総合技術研究所東部工業技術センター研究報告，深層学習による異常検知手法の簡単な比較（第1報），**35**（2022）。
- 2) 廣川 勝久：広島県立総合技術研究所東部工業技術センター研究報告，深層学習による異常検知手法の簡単な比較（第2報），**36**（2023）。
- 3) 廣川 勝久：広島県立総合技術研究所東部工業技術センター研究報告，深層学習による異常検知手法の簡単な比較（第4報），**36**（2023）。
- 4) 廣川 勝久，佐々木 秀和，押野 純大，山形 亮太：広島県立総合技術研究所東部工業技術センター研究報告，深層学習による異常検知手法の簡単な比較（第5報），**37**（2024）。
- 5) 廣川 勝久：広島県立総合技術研究所東部工業技術センター研究報告，深層学習による異常検知手法の簡単な比較（第3報），**36**（2023）。
- 6) Jane Bromley, James W. Bentz, Léon Bottou, Isabelle M Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, Roopak Shah: Signature Verification using a "Siamese" Time Delay Neural Network, *International Journal of Pattern Recognition and Artificial Intelligence* **7** (1993).
- 7) María Leyva-Vallina, Nicola Strisciuglio, Nicolai Petkov: Generalized Contrastive Optimization of Siamese Networks for Place Recognition, *arXiv:2103.06638* (2021).
- 8) THE MNIST DATABASE of handwritten digits : <http://yann.lecun.com/exdb/mnist/>
- 9) The CIFAR-10 dataset: <https://www.cs.toronto.edu/~kriz/cifar.html>
- 10) Simon Speth, Artur Gonçalves, Bastien Rigault, Satoshi Suzuki, Mondher Bouazizi, Yutaka Matsuo, Helmut Prendinger: Deep Learning with RGB and Thermal Images onboard a Drone for Monitoring Operations, *Journal of Field Robotics* **39** (2022).