【論　文】

# 深層学習における学習データクラスの t-SNE による解析

大関和夫・上條浩一

## Analysis of Train Data Class Using t-SNE in Deep Learning

Kazuo Ohzeki, Koichi Kamijo

**Abstract**：Although deep learning has progressed, it is still difficult to fully recognize the category of vehicles. In this study, we will investigate the relationship between the large category of vehicles and the small range of classes that are a subset of the vehicle type. In this study, we consider the distance of the set of vehicle classes and investigate the effect of merging the image sets of two vehicle types. Using t-SNE as a measure of distance, the recognition of CNN by the degree of similarity and merger of the pair of "taxi and sedan" and the pair of "1Box car and 1Box car with tilted front" which are subjectively similar. The change in the rate was investigated. Although it is a small number of cases, there was a discrepancy between the distance closeness of t-SNE and the recognition performance of CNN. Evaluation by t-SNE revealed that merging two classes with small distance into one class may form a better class. In the future, we will increase this verification quantitatively, and the class at a short distance indicated by t-SNE is a necessary condition for improving the recognition rate due to the merger, etc. recommend.

**Keywords**：artificial intelligence, train data, data class, dimension, reduction

## １．まえがき

自動運転では車両は周辺の状況を正確に把握する必要がある。車両に取り付けたセンサーは、前方、側方、後方の車両や歩行者を検知する。車両内のセンサーは物体があるとその先の隠れた部分は検知できない。例えば、前方の車両の前、後方の車両の後ろや路上の構造物の裏側などがある。そこで、最近では、車両内のセンサーのデータ処理に加え、路上に置いたカメラ等の情報を追加する「インフラ協調型」の自動運転の重要性が指摘され、そのための実験都市も開発されている [1][2]。インフラとなる道路側に置いたカメラ等は、障害物の裏に隠れた部分も見渡す事ができ、より安全な情報システムが構築できる。

深層学習を用いた車両等の検出の精度は大きく進歩したが、あらゆる条件下で完全であるわけではない。深層学習の認識精度は人間に比べ、平均値では、上回ってきており [3]、今後も向上が続くことが期待できる。しかし、比率は小さくても自動運転車が事故を起こせば、原因や責任の問題は複雑化する [4]。本研究では、道路上にあるカメラから車両を

認識し、道路内のすべての交通事象を完全に把握し、自動運転のインフラシステムの構築に貢献することを目指す。

道路上のカメラは、例えば図 1（a）のように各カメラの監視範囲が隙間なく繋がっていて、ある密度以上にあるものとする。また、各カメラは連携して認識動作を行うことができることにする。これにより、カーブでの車両認識は近接する直線道路で行う認識結果を引き継ぐことができることになる。車両の認識は、直線道路では、車両が平行移動に近い動きをするため、形状の変形が少なく、認識はやりやすい。一方、図 1（b）のようにカーブ上では、走行する車両は、回転して見えるため、形状の変化が大きく、認識にとっては負担が大きい。交差点における右折車が 45 度程度に曲がったところを学習データとして深層学習を行う実験を行ったが、撮影時間が少なく、図 1（b）のように同一車両の形状が少し異る画像を 10-20 枚取り込んだため、予想外に高精度で認識（識別）ができた [5]。そこで、データ採取の方法が結果に影響することがわかり、データの集め方や車種についてよく検討することを考えるようになった。

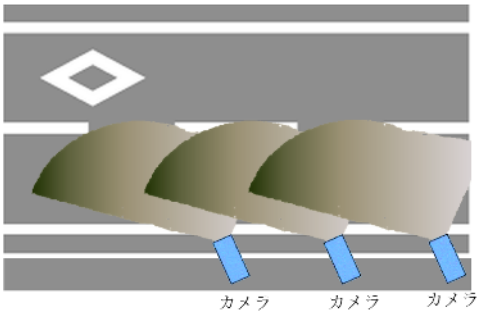


**図 1（a）　直線状の道路上のカメラ**
各カメラの監視範囲が隙間なく繋がっていて、道路全体を覆っている。道路は nuriyon.com のフリーイラストを使用禁転載。

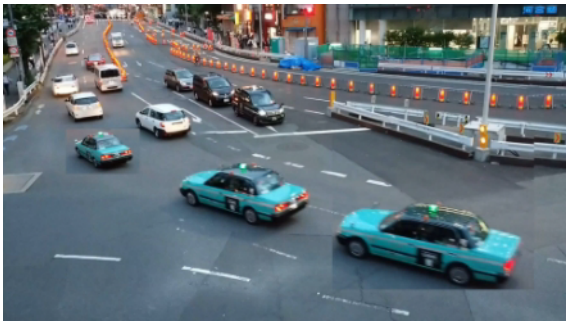

**図 1（b）　交差点などのカーブ状の道路と直線上の道路での車両形状の変化がある**
（東京都新宿区西新宿、西新宿交差点歩道橋上より撮影）
（同一車両を 3 つの位置で合成した画像）

しかし、図 1（b）は、撮影が右折時に限る、場所も交差点付近に限定されるなど制約が多く、多数の画像確保に適さなかった。そこで、本研究では、直線上の道路で、車両の進行方向にほぼ直交する真横から撮影した画像も加えて認識手法を開発していくことを目標とする。

## 2．認識システムの構成

車両の認識には、MNIST[6] や Cifar10[7] という画像データベースに対応した深層学習の認識ソフトウェアを用いる。これらは、物体認識の標準的なソフトウェアで現在でも参照ソフトとして広く使われている。このソフトは物体を 10 種類のクラスに分け、学習を行い、出力は 10 種類のクラスのうち最も類似度が高かったクラスになる。したがって、認識と言っても、10 種類の中での識別を行う機能を考えている。

自動運転の場合は、車両の認識を行うことを主眼としており、車両が入力されたにもかかわらず、車両でないと判定すれば、それに基づいた誤った判断がなされる可能性があり、次の行動に影響が出るため、事故が起こりかねない。また、もし車両でないものが入力さ

れた場合は、正しく拒絶する、すなわち車両でなかった、と出力する機能も必要となる。これは、車両でないもの例えば、動物や鳥が入力となったとき、車両と誤認識した場合は、急ブレーキや衝突回避の機能が働き、混乱を招く可能性があるためである。自動運転とは異なるが、現在でも「非接触事故」という事象が問題となっている [16]。これは、交通事故の当事者（二者）の他に、接触していない第三者が事故原因になっている事象である。この第三者の危険行動を避けようとして、回避行動を行った結果、この第三者には接触はなかったが、別の二者のみが事故になるような事象が起こっており、原因である第三者の情報や責任が追跡しにくいということ等が問題となっている。

図 2 は入力の車両または非車両の例と、それをすべて識別しようとするシステムのブロック構成図である。深層学習は学習データによって学習し、一旦固定される。この学習済みのニューラルネットワーク構造データは HDF と言うファイルフォーマットのファイルとして、認識器のニューラルネットワーク「DNN 認識器」に送られる。DNN 認識器の動作は、学習データに対しては、ほぼ間違いなく認識がなされる。車両であっても学習時に使用されなかったデータ（非学習データ）は車両としての類似度があれば、車両として認識されるが、そうでない場合は、認識されないこともある。非車両のデータは、車両とは異なるため、本来は、車両ではないとの出力になるはずだが、形状の一部に車両の特徴を持つ場合は、車両と誤認識されることもありうる。

深層学習においては、このように学習したデータに関するものは、認識の対象になるが、そうでないものに対しては認識がうまく動作するかどうかはわからない、制御なしの状態になっている。このような範囲外のデータの扱いは、深層学習を用いた「異常検知」の分野で、活用されている。正常な認識状態と異なる場合は、異常である可能性があるため、一旦分離して、ある種の精密検査のようなことで、異常かどうかの判定を行うものである。工場の不良品の検査では、数が少なく形状が多岐にわたる不良品は学習用のデータを集めることが難しい。また、経年劣化したひび割れの検出などでも、ひびの形状は千変万化し、学習データとしての一般性が無い。そこで、ひびのない正常なデータを基準に、異常検出として、ひび割れの候補を見つける手法がある。

学習に使用するデータの範囲を超えて、認識を拡張しようとする手法に、学習データの入力のニューラルネットワークにクロス結合を入れる手法がある [8]。また、複数のカテゴリー間の混合により、より広い範囲の識別を求める方式もある [9]。これらは、学習データより関連する認識範囲を拡張しようとするものだが、学習しないデータに関する研究として、「out-of-distribution（OOD）[10]」や「Open Set Recognition [11]」がある。しかし、ODD では、確率分布から推定を向上させようとする試みであり、サーベイ論文 [11] では、種々の試みが紹介されているが、明確な方針や有効な手法が述べられていない。

本研究の最終形態では、学習に用いなかった、非車両のデータが入力されたときに、車両との明確な差異を判定できるように、車両データと、非車両のデータのニューラルネットワークにおける距離を検討し、識別性の向上を目指そうとするものである。今回は、そのうち車両のデータに対する、認識率を向上させるため、学習データの分布を距離の観点で検討することを試みる。

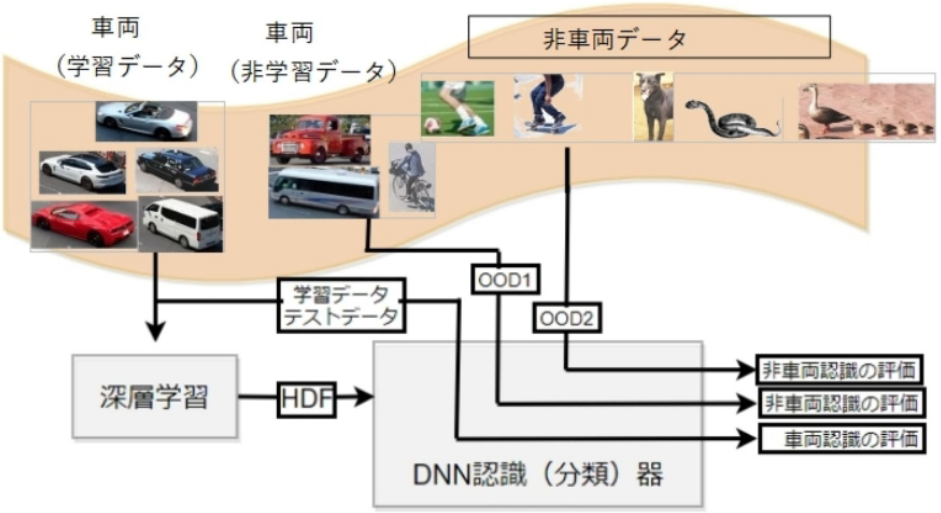


**図 2　入力画像データと、学習器、学習済み認識器の関係**

非車両は車両以外の物体で、道路上に現れる可能性のあるものすべてを想定する。今回は、そのうち、車両データの学習に関する検討に焦点を当てている。

（画像は足成、Pixabay（c）より使用した。（注）：足成は使用、改変が自由にできる、Pixabay は使用自由だが、再配布等は禁止されている。従って転載禁止）

### 2.1　車両画像について

車両の集合を識別のクラスとして設定し、類似の車種を同じクラスに統合していく事により、車両全体の車種の多大な種類を低減していく必要がある。車両のクラスを統合する際に、類似の 2 個のクラスを 1 個のクラスに統合をすれば、識別器の種類も減って、システムが構成しやすくなる。実際車両の種類は、大きくは、sedan 型乗用車、ワンボックスカー、トラック、バスなどが多いが、その他スポーツカー、消防車、救急車、特殊車などは、発生頻度が低いが、形状には変化が多い物がある。一方、非車両の方は、道路標識、信号機、歩行者、ベビーカー、サッカーボール、スケートボード、電動スクーター、ドローン飛行体、小動物（犬、猫、狸、イノシシ、鹿、猿、熊、蛇）、鳥類、などと種類が多い。認識器の仮の構成例を図 3 に示す。認識は多様な入力に対し、一段階では完了しないので、二段階構成になる。車両の確定的な識別を優先し、それ以外を非車両とする。次に非車両の集合から、人間を識別し、非車両として、信号機、小動物、サッカーボール等、ドローンなどを識別していく。本論文では、このうち、車両の識別の認識率の向上を目指すための車両クラスの分析について検討を行う。

図 3 の車両のクラスの例として taxi. sedan, van, one Box... などの形状は異なるが、車両として類似している車種がある。これらが真に類似しているかどうかは、形状の類似だけではなく、深層学習器の学習に於いて類似になりやすいかという点で考える必要がある。しかし、深層学習器で学習を行うことは、データのクラス調整やラベルの管理などの処理の手間や、学習における計算時間がかかるという問題がある。また、ある学習データを用いて学習した認識器はその学習データに固有の認識特性となり、別のデータを認識することにより、類似性を判別できるが、この別のデータを学習データに加えた場合は、別の認識器が生成される事になり、学習とクラスの形成が独立には行うことができないと言う問

題を有している。

現在、データの類似性に関する尺度として UMAP [12]、t-SNE [17] と Siamese [13] 等がある。UMAP は Uniform Manifold Approximation and Projection for Dimension Reduction の略で、高次元の多様体の辺の関係性から順次縮約して、低次元化を図るものである。t-SNE は t-distributed Stochastic Neighbor Embedding の略で、同じく高次元の画像データを低次元に変換するものである。Siamese は学習データで学習した認識器を使用する距離の評価を行うもので、深層学習の性能の調整に使用されることがある。

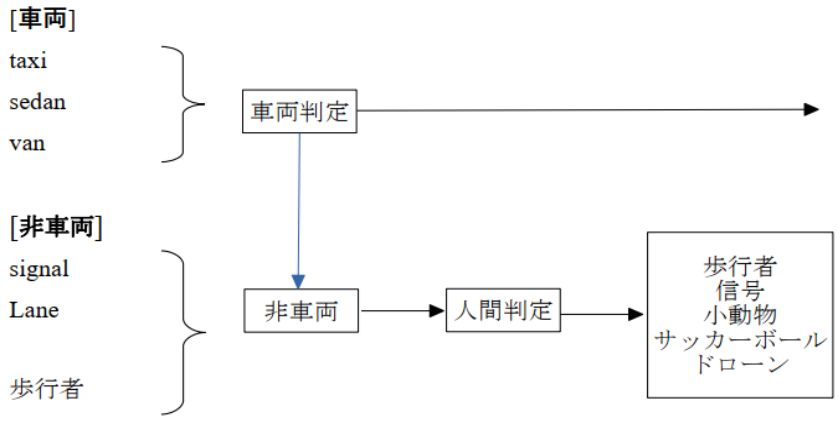


**図 3　認識器の構成例**

## 2.2　次元低減の方式と効果について

次元低減の手法の中で、t-SNE と UMAP について、その手法の概略と使用例と次元の低減後にクラス分けがどの程度正しく維持されるのかについて述べる。

UMAP, t-SNE,PCA などの次元低減の仕組みを文献により解説する。まず、古くからある PCA（Principal Component Analysis：主成分分析）は理論的な枠組みが明確であるが、非線形の次元圧縮を行えないので、数百次元という高次元から二次元という超低次元への低減では、特徴が保存されないという問題が強く、候補から外す。歴史的には、SNE[21], t-SNE[17], UMAP[12] の順に発表され、SNE や t-SNE はジャーナル論文などに述べられており、信頼性が高い。一方、UMAP は現時点では arXiv などに掲載されているだけで、正式ジャーナルには採択されていない。実際の使用例では、t-SNE に対し、UMAP は高速であり、次元削減後の状態も可視化した場合に優れているとの評価が多い。ここでは、web 上の UMAP や t-SNE についての動作の解説 [22,23] を引用して、簡単に解説する。大きい概念での動作は類似しているので、UMAP[22,23] の解説を引用する。UMAP の目的は、高次元データ X を低次元データ Y に変換することである。ただし X の局所構造と大域構造は保持したまま変換する。

UMAP の処理は、X と Y において距離に関するコスト関数を作り、このコスト関数を最小化する。コスト関数の導出は、Fuzzy topological expression のクロスエントロピーを使う。UMAP ではファジー関数を使うが、t-SNE では確率を使う。また、低次元化の初期値は、LaplacianEigenMap を使う。以上により低次元化したときに、相互に近い距離にあるものが保たれる作用が働く。（以上 ref [22,23] による）

この解釈をまとめると、t-SNE も UMAP も一定の定式化に基づく次元の低減を行う手法である、実際の計算は、望ましい低次元化を試みているが、確率やファジー関数を用いているため、低次元化により、距離の特徴が失われる可能性を含んいる、というところに集約できる。本研究では、このような次元低減手法を用いて、低次元化を行った結果において特徴が失われる場合の考察を行うことを目指す。

### 2.3　車両画像とクラス分けについて

車両の認識技術は、深層学習により大いに進歩したが、自動運転への適用には、完全性が求められるため、まだ完成したとは言えない。動画中から物体や輪郭を実時間で認識する YOLO や Mask-R-CNN 等のソフトウェアが公開されている [18][19]。2020 年に IEEE Transaction に論文として採択された Mask-R-CNN を用いた動作例を図 4 に示す。

車両（car）に加え、bus や truck などの詳細なクラスの認識を行うことができるが、シーンによっては、速度制限標識が検出できない場合（図 4 左）がある。入力画像を 2.75 倍まで拡大しても検出はしないが、2.95 倍まで、拡大した時、図 4 右のように検出できた。このような詳細な未検出を根絶するためには、学習データクラスの形成の手続きを見直しておく必要がある。Mask-R-CNN も例では、8 種の category を設定している。それらは、person, rider, car, truck, bus, train, mcycle, bicycle である。車両としては、バイク、乗用車、トラック、バス、自転車の 5 種であるが、これが選ばれたのは、得られている画像データベースにおける出現頻度の多いものを抽出した結果と推測される。車両という全体の中から、車種というような詳細なクラス分けを行うことが有効か、必要かなど明確な判定基準が無い。本研究では、このクラスの詳細化を自動車というカテゴリーの中で、クラス分けの判定基準を見出そうとしている。

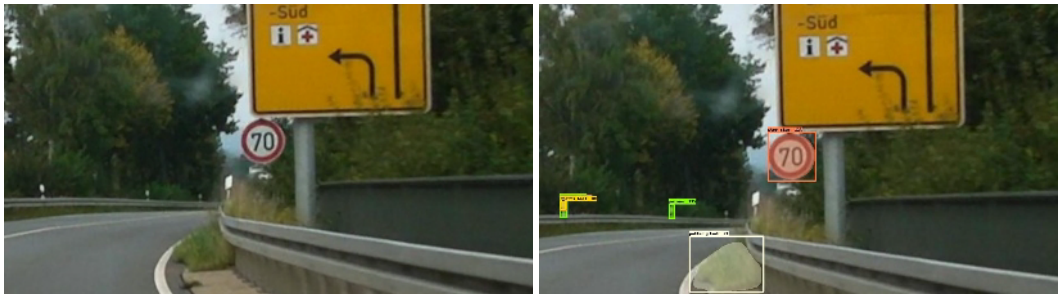


**図 4　Mask － R-CNN（Faster R-CNN）による動作例**

左：未検出。右：2.95 倍に拡大し検出できた例。1 ～ 2.75 倍まで左の未検出が続いた。各色のついた枠は、各種物体の検出を示している。

### 2.4　車両カテゴリーと車種の詳細クラス分けの関係

筆者らは、これまでクラスの分割や統合を試行することにより、影響を調べてきた [20]。表 1 は、車両（乗用車）（car）、バス（bus）、トラック（truck）の 3 車種に対する認識を行った予備実験の認識率のデータである。2 列目の結果は、乗用車については、高い認識率であるが、バスはやや低く、トラックは、0.25 と認識率が低い。これは、トラック専用の認識ではなく、トラックと車両の両方を対象としている認識器であったため、トラック単独の認識を取り出したときに認識率が低くなったと考えられる。次に、3 列目は 3 車種とも

車両であることから、乗用車、またはバス、またはトラックという和集合（車両）とそれ以外（それ以外の車両や非車両）という補集合を用意しどちらに判定されるかの認識実験を示す。この場合は、乗用車、バス、トラックともに 90% 以上の比率で、車両として認識されている。この結果から、車種を細かく分解するのでは無く、車両という大枠のカテゴリーで学習や認識を進めれば良いという見解が生じる。一方、そのような大枠のカテゴリーにすると、形状の特徴が大量に含まれた集合になっているため、最終的な認識率が、100% に達しないで、停滞するという見解もある。二つの見解を表 2 にまとめた。

車両カテゴリーの単一クラスを使用する場合（見解 1）は、データ収集は、車両というものは何でも登録すればよいので、容易となる。一方、車両であっても極端に特殊な状況で撮影されたものは、異常データとして学習に悪影響がある場合がある。制限されたデータ数（例えば数万枚）などの状況では、認識率の向上を妨げる。また、車種クラスに細分化する見解 2 では、形状が異る車種に集合を分割するため、狭い範囲の認識として、認識率の向上が期待できる。各車両のクラスの中での形状の変動が小さいため、異常データの発見も容易であると考えられる。一方、データの収集は、種類ごとのラベルを合わせて収集していかないといけないので、見解 1 より困難になる。以上表 2 は、データの特徴を元に定性的な比較をしたものである。

定量的な実験例は、文献 [20] の表 6 にあり、もともと 5 種の車種（taxi,sedan,van,1-Box,1-Box_Slant）とそれ以外の 5 種のデータに対し、はじめの 5 種のうちから、2 種選んで、統合し、はじめと同じ条件に揃えるため、1 個のデータ（この場合は、交通標識データ）を追加し、認識率（accuracy）を調べている。結果は、統合した組み合わせにより、変動はあるが、いずれも認識率は向上した。これを単純に外挿すれば、全部統合して、車両 1 種にするのが最大の認識率を得られることになる。これは、見解 1 が良いことを示すことになる。しかし、この実験では、統合してクラスが不足した分を、交通標識を追加しているため、いずれも認識率が向上したとも考えられる。

そこで、交通標識が比較的同一の形状で背景も固定的であり、変化に乏しいデータであることから、それとは対照的な衣料データを追加した別の実験行ったところ、認識率が低下した。この衣料データは、形状の変化が多いことが多く、衣料データだけの実験でも認

**表 1　路上の固定カメラから撮影した道路画像の車種判別の結果**
[20] の Table3 より。Mask-R-CNN による認識結果、画像数（No. of images）は車種（Type of vehicle）が発生した回数。

| Type of vehicle（No. of images） | Recognition ratio | |
|---|---|---|
| | Type identification | Car or bus or truck |
| car（82） | 0.9756 | 0.9756 |
| bus（18） | 0.8889 | 1.0 |
| truck（28） | 0.25 | 0.9286 |

**表 2　車両カテゴリーを使用する場合と、車種に細分化する場合の定性的比較**

| 名称 | 集合の作り方 | 長所 | 短所 |
|---|---|---|---|
| 見解 1 | 車両カテゴリーを使用 | データ収集が容易 | 異常データの発見困難 |
| 見解 2 | 車種クラスに細分化する | 異常データ発見容易 | データ収集が困難 |

識率はかなり低くなるものである。追加した一つの衣料データの影響の方が統合したことによる変化よりも大きい可能性もある。そのため、追加するデータの特性の影響を除くことや、独立的な操作を考えておく必要がある。これらの実験から、単純に 2 クラスを統合してクラスの増減の影響を評価することは、正確では無く、全体のシステムに不動となる枠組みを設定するなどしないと、正確な比較は難しいことがわかった。

### 2.5　車両データの分布や相互距離の検討

そこで、車両データの分布を調べ、個別のデータ同士の距離や車種のクラスの重心（平均）の距離がどの程度かを探ることはできないかと考えた。そのような手法として、2.1 で触れた、UMAP、t-SNE などがある。また、Siamese は実際に深層学習を行って評価することまで含まれているので、除くことにする。UMAP と t-SNE では、学習データである画像データのサイズに起因した次元数があり、これが、はじめの空間となる。今回の例では、28 × 28 画素、32 × 32 画素等が使用されるので、これらが、原信号の次元数となる。具体的には、

28 × 28=784

32 × 32 ＝ 1024

次元の画像片である。ここで、使用している画像は、白黒画像であり、各画素は 8 ビットのグレースケールで記録されている。後出の図でカラー画像が表示されているのは、10 種のクラスを識別するために色付けされたものである。図 5 に、上記サイズに縮小する前の画像片の一部を示す。UMAP や t-SNE はこのような 784 次元× 1 バイトのデータをなるべく形状の特徴を保存するようにしながら、次元を低下させていくもので、最終的には、二次元にまで変換を行う。本研究では、このうち、t-SNE を使い、予め分類してある 10 種の路上物体がどのように分離されるか、各物体の重心（平均）の距離の有意性があるか、などを調べる。それにより、各 10 種のクラスの特徴を抽出できるか、どうかを調べることを目標とする。

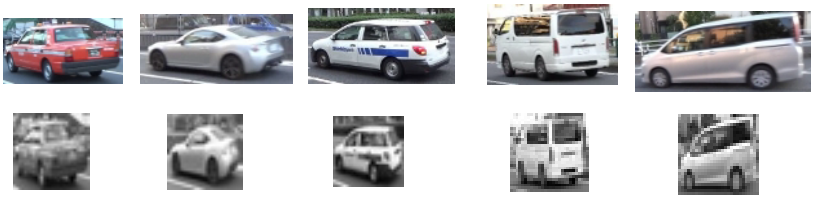

**図 5　画像データの例**

上段の左から、taxi, sedan, van, 1-Box, 1-Box_Slant、28 × 28 または 32 × 32 に縮小する前の画像、下段は各 28 × 28 画素のモノクロ画像に変換したもの、（名称：img_1_5_c_oz,img_1_5_d_28x2_jpg）。

## 3．実　　験

以上の検討のもとに、UMAP,t-SNE のプログラムを使用して、データ集合の分類の度合いを調べた。図 6 は文献 [12] による手書き数字 10 種の画像データの次元低減の例である。9 と 5 の領域の間に少しの混合が見られるが、二次元においても、各領域が分離されている。また、図 7 は衣服のデータ 10 種を同様に t-SNE で二次元に縮退したものである。衣服デー

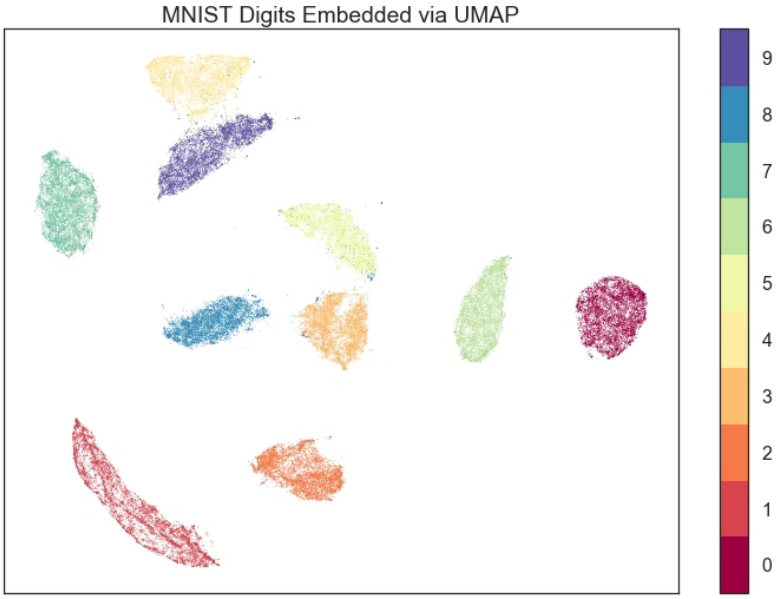


**図 6　手書き数字画像（28 × 28 画素）を UMAP で 2 次元にした図**
0 から 9 までの数字が色で示された領域に分布していることを示している。文献 [11] による。

タの方は、形状が不安定であり、領域ごとの境界も重なりが見られる。これから、UMAP は学習データの領域を示しているため、2 個のデータの遠近を示すものとして使用が可能なものであることがわかる。図 7 を見ると各領域の境界が不明確なところが見られるが、数字が幾何学模様に近いのに対し、衣服は、そうではないので、縮退時に混合が生じていると考えられる。

これに対し、実際の深層学習を行うのと同じ処理で、類似度を求める Siamese ネットワークというものがある。これは、各学習データの要素 A, B に対し距離

$$d(A,B) = ||f(A) - f(B)||^2 \text{ [14]} \tag{1}$$

を Siamese の計算により求め、これを誤差関数として認識対象である場合は、最小化を行い、非認識対象である場合は、最大化を図る学習を行うものである。しかし、深層学習の枠組みで距離を直接使用していくのが難しいと言われている [15]。そして深層距離学習と言う分野が盛んに研究され始めている。

このような背景の中で、学習データの設定にもっと注意を払う必要があると考えている。これは、学習データを基に学習が進められるため、対象の物体そのものが有する特徴は影や反射などを含めあってもよいが、他の物体がたまたま背景に混在している、などは、その物体固有の性質ではないため、学習に歪みが入ると考えられるからである。このような雑音成分は、無いのが基本的には望ましく、ある場合は、十分多数のサンプルがあって、雑音の特殊な影響が学習に入り込まないほうが望ましい。例えば、sedan と van は小型車と言う範囲で同じカテゴリーとみなせるが、別々に分けて 2 種の認識器を用意した方が良いのか、二者を混合し sedan と van を合併して形成される車種のクラスにした方が良いのかもはっきりしていない。本研究では、このようなクラスの区切りの変更が結果にどのような影響があるかを調べ、最も望ましいクラス分けを開発していくための指針を求めるこ

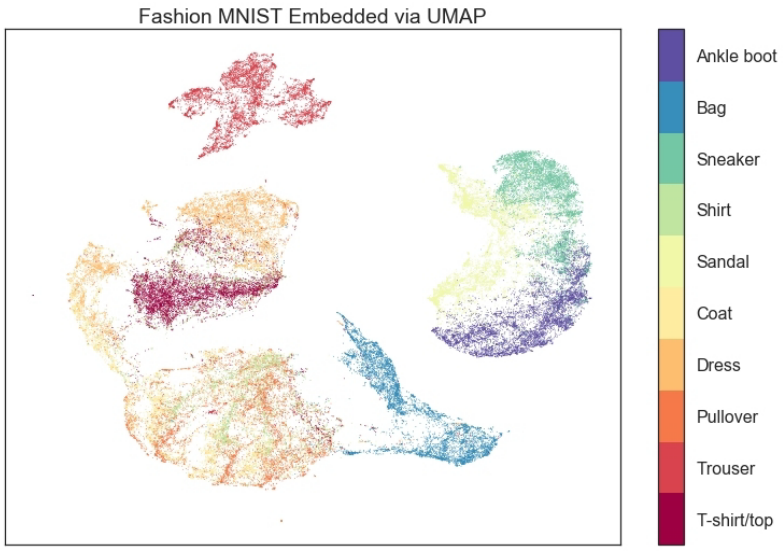


**図 7　衣服データ FashionMNIST（28x28 画素）の UMAP による 2 次元化**

とを将来的な目標とする。

図 8 は、深層学習用に集めた道路車両データ（45_2_d）の t-SNE による次元削減を行い二次元化したものである。データは各クラスごとに 100 枚あり、合計 1000 枚ある。この図をみると、各クラスはある程度の集まりがあるが、領域が広がっているクラスもあり、他のクラスと混合しているような領域も見られる。クラスの形状が類似しているものとして、taxi と sedan がある。実際の二次元上の分布も重なりが多い。また、ワンボックスカー（1-Box）とワンボックスカーで前部の傾斜が急なもの（1-Box_Slant）も形状に類似性があるが、二次元上の分布においても重なりがあり、また、広い領域に分散している。一方、静止物体である、信号機（signal）、パイロン（pylon）、車線のペイント（Lane）は比較的

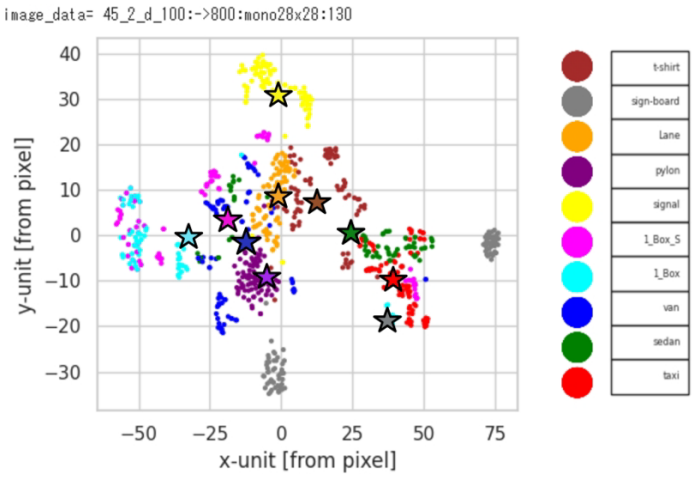


**図 8　道路車両データ、45_2_d の t-SNE による二次元化**

10 種のクラス（taxi, sedan, van, 1Box, 1Box_Slant, signal, pylon, Lane, sign-board, t-shirt）に対し、可視化のため、色付け（赤緑、青、水色、桃色、黄色、小豆色、橙色、灰色、茶色）してある。また、星印☆は各クラスのデータ（点）の重心位置である。（以下同様）

かたまって分布している。

t-SNE の次元低減の性能を調べるため、上記のデータに対し、ワンボックスカー（1-Box）を廃止し、ワンボックスカーで前部の傾斜が急なもの（1-Box_Slant）のみとした。この廃止により一つクラスが減少した変化を補うため、上着服の画像（jaket）を追加したクラス（各 100 枚）で構成されたデータ（mono28x28_6_d）を作成した。このデータに対して、t-SNE を適用したのが、図 9 である。赤い点がワンボックスカー前部傾斜であるが、単独の集まりにやや近い分布になってはいるが、水色の点が「ワンボックスカー前部傾斜」であるが、重なりは少ないが、3 ヶ所（左下、中央、上部）に分散している。上部では taxi の重心の近くに分布しているのが見られる。また、追加したジャケットは T シャツと重なるように分布している部分が見られる。

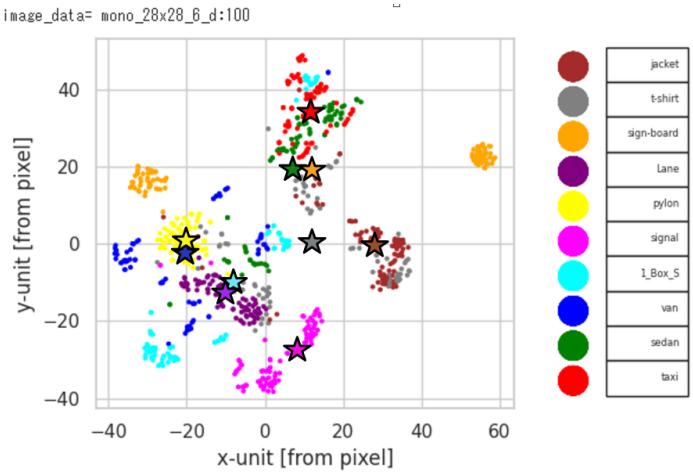


**図 9　道路車両データ、mono28x28_6 の t-SNE による 2 次元化**

図 8 のデータの 1-Box を廃止し、f_sign_board を追加した。

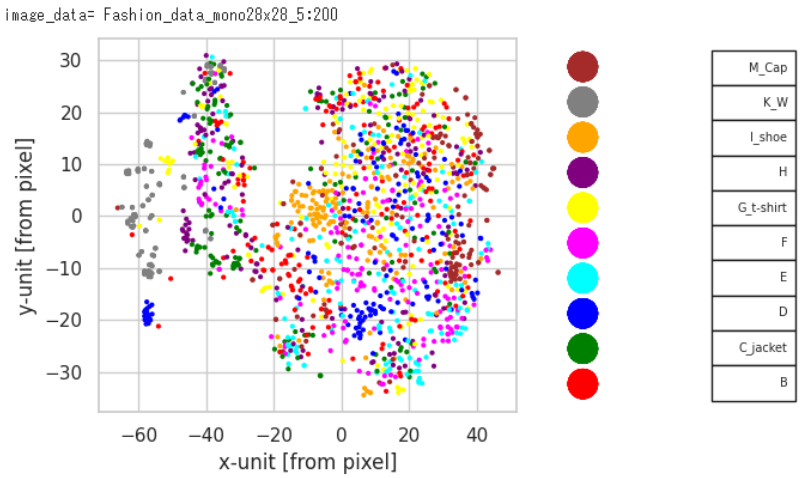


**図 10　衣服画像データ、mono28x28 の t-SNE による 2 次元化**

0：上下一体型の Long コート等、1：上着のみ（jacket）、2：G パン、3：スカート、4：短パン、5：T シャツ、6：シャツ、丸首、7：靴、8：水着、9：帽子、データは、DeepFashion という公開された衣服のデータベースから抽出した。[24]

図 10 は、別のデータとして、衣服データとして収集した 10 クラスのデータを t-SNE で次元低減を行い、二次元表示したものである。衣服データを実験に用いたのは、車両のように形状が確定した物体に対し、形状が不安定な物体での状況を見ておくためである。10 クラスは、0：上下一体型の Long コート等、1：上着のみ（jacket）、2：G パン、3：スカート、4：短パン、5：T シャツ、6：シャツ、丸首、7：靴、8：水着、9：帽子である。全体的に分離が明確ではないが、例えば 2 ～ 5 が何箇所かに分散していることが見られる。

以上、二次元まで次元削減した図での視察によると、車両等の路上の物体からなる 10 クラスや衣服データから成る 10 クラスは、ある程度のまとまりはあるが、そのうちのいくつかのクラスは分散していたり、他クラスと混在しており、二次元上では、クラスの明確な識別ができない。これが、元の画像データが明確なクラス分けされていないためなのか、t-SNE の処理のためなのかは、不明である。

t-SNE や UMAP はクラスの分離を保証するものでは無いが、その点を明確化するため、更に次の実験を行った。画像の枚数を 104 枚に増やし、次元低減後の各クラスの重心（平均）や分散、標準偏差を求めた。

表 3 は、10 クラスの重心（平均）の座標、分散と標準偏差を示す。画像は 130 枚に増加したが､そのうち学習データとして用いる 80% のデータ 104 枚を使用している。重心(平均）は各二次元 X,Y の成分ごとの平均を用いた。また、図 11 は二次元まで、t-SNE で低次元化したあと表示するとともに､重心（平均）を星印で示した。重心の値は､視察により、各クラスの平均を表していることが確認できる。各クラスの点は集まっていると言うよりは、分散しているように見え、各クラスの重心が各クラスの要素で囲まれていたり、埋もれている例が少ない。紫色の Lane、黄色の signal はよく集積している。また、灰色の sign-board も離れており､混合はなく比較的集積している。標準偏差では､class 番号で 5,6,7 の 3 つが小さい。これらは、5：signal（黄色）、6：Lane（紫色）、7：truck（肌色）であり、二次元にまで次元削減した図 11 の分布から見られる分離の程度の印象と同じであった。

表 4 に　道路画像（img_1_5_d_d_28x2_jpg：104 枚× 10）を t-SNE で二次元まで、低減化したときの各クラスの識別率を accuracy で求めたものを示す。各クラスの重心から最短の他の重心までの距離の 1/2 をしきい値として、すべての点をクラスの内部か、外部かの判定を行っている。TP（True Positive）は少ないものも多いが、FP, FN も一定に押さえ

**表 3　道 路 画 像（img_1_5_d_d_28x2_jpg：104 枚 ） を t-SNE で 二次元まで低減化したときの重心（平均）と分散、標準偏差**

| 4_Fig11 | mean | | var | stdvar |
|---|---|---|---|---|
| class | X | Y | | |
| 0 | 11.734 | 29.437 | 245.060 | 15.654 |
| 1 | 7.592 | 14.573 | 559.295 | 23.649 |
| 2 | 4.651 | −9.015 | 373.391 | 19.323 |
| 3 | 2.674 | −23.691 | 1199.939 | 34.640 |
| 4 | −1.613 | −15.226 | 1108.063 | 33.288 |
| 5 | −33.847 | −10.290 | 224.123 | 14.971 |
| 6 | −8.926 | −3.902 | 92.834 | 9.635 |
| 7 | −8.677 | 11.615 | 228.277 | 15.109 |
| 8 | 45.409 | 6.364 | 310.347 | 17.617 |
| 9 | −14.102 | 18.670 | 617.000 | 24.839 |

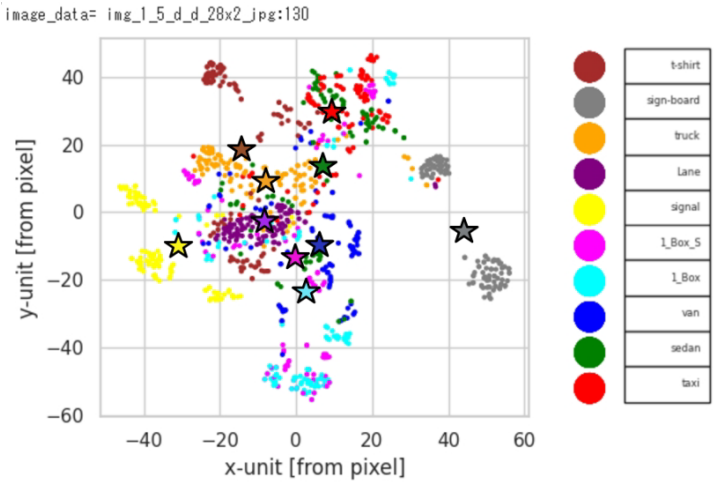


**図 11　道路画像（img_1_5_d_d_28x2_jpg：104 枚× 10）を t-SNE で二次元まで低減化したときの各点の分布**
重心（平均）を☆で示している。

**表 4　道路画像（img_1_5_d_d_28x2_jpg：104 枚× 10）を t-SNE で二次元まで、低減化したときの各クラスの識別率（accuracy）**

4_Fig11

| TP | FN | FP | TN | total | accuracy |
|---|---|---|---|---|---|
| 52 | 50 | 69 | 869 | 1040 | 0.886 |
| 4 | 96 | 58 | 882 | 1040 | 0.852 |
| 4 | 93 | 19 | 924 | 1040 | 0.892 |
| 0 | 112 | 16 | 912 | 1040 | 0.877 |
| 0 | 104 | 13 | 923 | 1040 | 0.888 |
| 76 | 30 | 56 | 878 | 1040 | 0.917 |
| 81 | 17 | 42 | 900 | 1040 | 0.943 |
| 21 | 87 | 5 | 927 | 1040 | 0.912 |
| 111 | 0 | 10 | 919 | 1040 | 0.990 |
| 3 | 99 | 11 | 927 | 1040 | 0.894 |

られているので、TN（True Negative）が多い結果となっている。図 11 のデータ（道路画像 img_1_5_d_d_28x2_jpg：104 枚× 10）を深層学習した場合の accuracy は 0.912 であった。t-SNE の二次元分布において、距離の近い重心に割り当てる識別を行った時は、accuracy は 0.886 である。

## 4．考　　察

以上の実験結果から、考察を行う。784 次元などの高次元から二次元まで次元低減を行うときの、相互距離の近いという特徴点の集まりを維持できるか、どうかについては、手法として近い状態を維持しようとしているだけで、二次元まで低減した時の結果には保証があるものではない。文献 [17] では、低減した結果のクラス分けのエラー率を quality の基準として使っている。MNIST という手書き数字データ 60000 枚に対し、784 次元（28 × 28）の原データにおけるエラー率は 5.75% であり、t-SNE で二次元にまで低減した時

のエラー率は 5.13% に低下したと述べられている。なお、MNIST の手書き数字データを深層学習した場合の accuracy は 98% 近くの例もあり、t-SNE の距離による識別性が十分では無いことも確認できる。

t-SNE と UMAP の比較としては、文献 [12] では、UMAP の方が処理速度が数倍速いことが強調されている。質的な面では、詳しく述べておらず、エラー率などの評価も無い。他には、同程度や、UMAP のほうがやや良い例の報告が見られた。他にも UMAP の論文があったが、データが Telecom データや一次元の時系列データという画像データでないデータであったため、本研究では参考とし実験としては取り上げなかった。

以上より、t-SNE 等では、形状が類似しているというクラスの性質は、主に距離で評価されていることがわかった。距離が近い点として類似しているという性質は、次元低減でかなり保たれるという例が示されている。一方､距離による識別によるエラー率の評価と、次元削減しない原データの深層学習による認識率（accuracy）はいくつかの例で、全て、深層学習のほうが、低いエラー率となっている

同じ形状の車種クラスでも二次元に低減した時、重心（平均）に集まらず、重心から一定距離離れた位置に分布する場合がある。[25] では、MNIST などのデータを可視化するための新手法 GNNis を開発し、UMAP などと比較している。しかし比較の基準は処理速度であり、可視化の結果を比較していない。次元の削減は UMAP や t-SNE で可視化できても、可視化後の特徴の維持については、本研究の重心（平均）と分散が客観評価として有効と考えられる。t-SNE の特徴的な問題点として、二次元まで次元削減後に、クラス間距離よりもクラス内の距離が大きくなる例があったことである（signal、signboard など）。つまり、t-SNE はクラス間の距離を保って次元削減を行えない場合が多いことがわかる。一方、t-SNE が示すクラス間距離の近いクラスは統合によりまとまりの高いクラスになる可能性があるが、実際にそのような例が得られ、クラス統合の指針の一つとして有力な候補となった。少なくとも全探索でクラスの統合の優劣を調べるよりも大幅に少ない試行で有効なクラス統合を探し出すことができると考えられる。

## 5．結　　論

本論文では、道路画像の深層学習を完全化するため、学習画像を車両という大きいカテゴリーに対し、車種という小さいクラスに詳細化して確実な認識を行うための、クラス分けの基準を模索してきた。t-SNE や UMAP により、次元削減し可視化した図から得られる距離関係は、確実な根拠は示されていないが、エラー率などからは、ある程度の信頼性があることも示された。そこで、t-SNE で分布が近かった、taxi と sedan、またワンボックスカー（1_Box）と前部が傾斜したワンボックスカー（1_Box_Slant）は統合して一つの車種クラスにすることが有効であるという指針が得られた。また、signal や signboard は人間が見れば、各々信号機と道路標識の特徴を有する単純な画像であるが、t-SNE の二次元分布では、数個の領域に分散しており（図 11）、signal や sign-board のデータから形状を詳細に検討する必要性などの問題点が得られた。

今後は、車種のクラスの統合を試行し、t-SNE の低次元での分布の観察や、深層学習を

行った時の認識率（accuracy）の向上を調べていく。

**参考文献**

[1] 熊小敏、杨　鑫、刘兆杨璘、朱雪田、「车路协同的云管边端架构及服务研究」2019 年电子技术应用第 8 期（Xiong Xiaomin、Yang Xin、Liu Zhaolin、Zhu Xuetian,「クラウドネットワークエッジターミナルアーキテクチャと車両と道路のコラボレーションのサービスに関する研究」電子技術の応用、2019、45（8）：14-18, 31.）

[2] MDOT staff , "Michigan Avenue Planning and Environment Linkages（PEL）Study", March 3, 2021 http：//www.michiganpel.com/media/ygrhu0aq/2021-03-03_presentation.pdf

[3] Kaiming He; Xiangyu Zhang; Shaoqing Ren; Jian Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE CVPR June 2016

[4] 谷辺 哲史 , 唐沢 かおり ,「自動運転による事故とメーカー、ユーザーに対する責任帰属」実験社会心理学研究、資料論文、日本グループ・ダイナミックス学会、2021 年 61 巻 1 号 p. 10-21

[5] 大関和夫、上條浩一、「 インフラ協調型自動運転における 認識処理の比較」電子情報通信学会、ソサイエティ大会、A-13-6　2021 年 9 月

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86（11）：2278-2324, Nov. 1998. http：//yann.lecun.com/exdb/mnist/

[7] Alex's CIFAR-10 tutorial, Caffe style https：//caffe.berkeleyvision.org/gathered/examples/cifar10.html

[8] Cuiping Shi, Cong Tan, and Liguo Wang, "A Facial Expression Recognition Method Based on a Multibranch Cross-Connection Convolutional Neural Network" IEEE ACCESS, VOLUME 9, 2021 March 16, 2021.

[9] Yun Liang; Keisuke Maeda; Takahiro Ogawa; Miki Haseyama, "CROSS-DOMAIN SEMI-SUPERVISED DEEP METRIC LEARNING", IEEE International Conference on Acoustics, Speech and Signal Processing（ICASSP）MMSP-2.1 June 2021.

[10] Yen-Chang Hsu1, Yilin Shen2, Hongxia Jin2, Zsolt Kira, "Generalized ODIN：Detecting Out-of-distribution Image without Learning from Out-of-distribution Data", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）, 2020, pp. 10951-10960.

[11] Chuanxing Geng, Sheng-Jun Huang and Songcan Chen, “Recent Advances in Open Set Recognition：A Survey”, IEEE Trans PAMI pp.1-18, March 2020.

[12] Leland McInnes John Healy and James Melville, "UMAP：Uniform Mnifold Approximation and Projection for Dimension Reduction", arXiv：1802.03426v3 , Feb, 2018.

[13] Soumava Kumar Roy, Mehrtash Harandi, Richard Nock, Richard Hartley, "Siamese Networks：The Tale of Two Manifolds", IEEE/CVF International Conference on Computer Vision（ICCV）, 27th Oct.-2nd Nov. 2019.

[14] TUM, “Siamese Neural Networks and Similarity Learning” Lecture Note , Advanced Deep Learning for Computer vision（ADL4CV）（IN2389）TUM

[15] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, "Deep Metric Learning with Angular Loss" , IEEE International Conference on Computer Vision（ICCV）, Volume：1, Pages：2612-2620, 22-29 Oct. 2017.

[16] 一般財団法人 東京都交通安全協会 > 活動の内容 > 誘因事故（非接触事故） https：//www.tou-an-kyo.or.jp/soudanjirei/4_list_detail.html

[17] L. v. d. Maaten and G. Hinton. “Visualizing data using t-sne”, Journal of machine learning research, 9（Nov）：2579–2605, 2008.
[18] J. Redmon et al, "You Only Look Once：Unified, Real-Time Object Detection" IEEE CVPR June. 2016.
[19] Kaiming He, et al, "Mask R-CNN", IEEE trans PAMI Vol 42, pp386-397, Feb. 2020.
[20] Kazuo Ohzeki, Koichi Kamij and Stefan A. Schneider, “Vehicle Recognition in an Autonomous Driving System for Road and Vehicle Cooperation”, Proceedings of FastZero ‘21 Sept. 2021.
[21] G.E. Hinton and S.T. Roweis. Stochastic Neighbor Embedding. "Advances in Neural Information", Processing Systems, volume 15, pages 833–840, Cambridge, MA, USA, 2002. The MIT Press.
[22] @odanny、「t-SNE より強い UMAP を（工学的に）理解したい」
https：//qiita.com/odanny/items/06ab88353bcee7bf6aa7
投稿日 2020 年 02 月 18 日
[23] kntty.hateblo.jp、「UMAP の仕組み —— 低次元化の理屈を理解してみる」
https：//kntty.hateblo.jp/entry/2020/12/14/070022
[24] Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou, "DeepFashion：Powering Robust Clothes Recognition and Retrieval with Rich Annotations", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition（CVPR）}, June, 2016.
[25] Yajun Huang Jingbin Zhang Yiyang Yang Zhiguo Gong Zhifeng Hao GNNVis：Visualize Large-Scale Data by Learning a Graph Neural Network Representation CIKM '20, October 19–23, 2020, Virtual Event, Ireland, Association for Computing Machinery.

大関和夫　東京国際工科専門職大学　工科学部　情報工学科　教授
上條浩一　東京国際工科専門職大学　工科学部　情報工学科　教授