Abst

Artificial neural network (ANN) which has four or more layers structure is called deep NN (DNN) and is recognized as a promising machine learning technique.

4つ以上の層構造を持つ人工ニューラルネットワーク（ANN）はディープNN（DNN）と呼ばれ、有望な機械学習手法として認識されています。

Convolutional neural network (CNN) has the most used and powerful structure for image recognition.

畳み込みニューラルネットワーク（CNN）には、画像認識に最もよく使用される強力な構造があります。

It is also known that support vector machine (SVM) has a superior ability for binary classification in spite of only two layers.

また、サポートベクターマシン（SVM）には2つのレイヤーしかないにもかかわらず、バイナリ分類の優れた能力があることが知られています。

We have developed a CNN&SVM design and training tool for defect detection of resin molded articles, and the effectiveness and validity have been proved through several CNNs design, training and evaluation.

樹脂成形品の欠陥検出用のCNN＆SVM設計およびトレーニングツールを開発し、いくつかのCNNの設計、トレーニング、および評価を通じてその有効性と妥当性を証明しました。

The tool further enables to easily design a CNN model based on transfer learning concept.

このツールを使用すると、転移学習の概念に基づいてCNNモデルを簡単に設計できます。

In this paper, a CNN acquired by transfer learning of AlexNet, which is the winner of ImageNet LSVRC2012, is designed to recognize the orientation of objects.

この記事では、ImageNet LSVRC2012の勝者であるAlexNetの転移学習によって取得されたCNNは、オブジェクトの方向を認識するように設計されています。

The effectiveness of the transfer learning based CNN is evaluated through an actual pick and place test using an articulated robot named DOBOT.

転送学習ベースのCNNの有効性は、DOBOTという名前の多関節ロボットを使用した実際のピックアンドプレーステストを通じて評価されます。

Intro

Artificial neural network (ANN) which has four or more layers structure is called deep NN (DNN) and is recognized as a promising machine learning technique.

4つ以上の層構造を持つ人工ニューラルネットワーク（ANN）はディープNN（DNN）と呼ばれ、有望な機械学習手法として認識されています。

Convolutional neural network (CNN) has the most used and powerful structure for image recognition.

畳み込みニューラルネットワーク（CNN）には、画像認識に最もよく使用される強力な構造があります。

It is also known that support vector machine (SVM) has a superior ability for binary classification in spite of only two layers.

また、サポートベクターマシン（SVM）には2つのレイヤーしかないにもかかわらず、バイナリ分類の優れた能力があることが知られています。

Nagi et al. designed max-pooling convolutional neural networks (MPCNN) for vision-based hand gesture recognition [1].

ナギ等。視覚ベースのハンドジェスチャ認識用に設計された最大プーリ畳み込みニューラルネットワーク（MPCNN）[1]。

The MPCNN could classify six kinds of gestures with 96% accuracy and allow mobile robots to perform real-time gesture recognition.

MPCNNは6種類のジェスチャを96％の精度で分類し、モバイルロボットがリアルタイムのジェスチャ認識を実行できるようにします。

Weimer et al. also designed a deep CNN architectures for automated feature extraction in industrial inspection process [2].

また、ワイマー等は、産業検査プロセスで自動化された特徴抽出のための深いCNNアーキテクチャを設計しました[2]。

The CNN automatically generates features from massive amount of training image data and demonstrates excellent defect detection results with low false alarm rates.

CNNは、大量のトレーニング画像データから機能を自動的に生成し、低い誤警報率で優れた欠陥検出結果を示します。

Faghih-Roohi et al. presented a different type of deep CNN for automatic detection of rail surface defects [3].

Faghih-Roohi 等は、レール表面の欠陥を自動検出するための異なるタイプのディープCNNを提示しました[3]。

It was concluded that the large CNN model performed a better classification result than the small and medium CNN, although the training required a longer time.

大規模なCNNモデルは、中小規模のCNNよりも優れた分類結果を示したと結論付けられましたが、トレーニングにはより長い時間が必要でした。

Zhou et al. used a CNN to classify the surface defects of steel sheets [4].

周らはCNNを使用して、鋼板の表面欠陥を分類しました[4]。

The CNN could directly learn better representative features from labeled images of surface defects.

CNNは、表面欠陥のラベル付き画像からより良い代表的な特徴を直接学習できます。

Further, Ferguson et al. presented a system to identify casting defects in X-ray images based on the Mask Region-based CNN architecture [5], [6].

さらに、ファーガソン等は, マスク領域ベースのCNNアーキテクチャ[5]、[6]に基づいて、X線画像の鋳造欠陥を識別するシステムを発表しました。

It is reported that the proposed system simultaneously performed defect detection and segmentation on input images making it suitable for a range of defect detection tasks.

提案されたシステムは、入力画像に対して欠陥検出とセグメンテーションを同時に実行し、一連の欠陥検出タスクに適していることが報告されています。

We have developed a CNN&SVM design and training tool for defect detection of resin molded articles and the effectiveness and validity have been proved through several CNNs design, training and evaluation [7–9].

樹脂成形品の欠陥検出用のCNN＆SVM設計およびトレーニングツールを開発し、いくつかのCNNの設計、トレーニング、評価を通じて有効性と有効性を証明しました[7–9]。

The tool further enables to easily design a CNN model based on transfer learning concept.

このツールを使用すると、転移学習の概念に基づいてCNNモデルを簡単に設計できます。

When industrial robots are applied to pick and place tasks of resin molded articles, information of each object ’s position and orientation is essential.

産業用ロボットを樹脂成形品のピックアンドプレースタスクに適用する場合、各オブジェクトの位置と方向の情報が不可欠です。

Recognition and extraction of object position in an image are not so difficult if using image processing technique, however, that of orientation is not easy due to the variety in shape.

画像処理技術を使用する場合、画像内のオブジェクト位置の認識と抽出はそれほど難しくありませんが、方向の認識は、形状の多様性のために容易ではありません。

In this paper, a CNN acquired by transfer learning of AlexNet, which is the winner of ImageNet LSVRC2012, is introduced to recognize the orientation of objects in images.

この論文では、ImageNet LSVRC2012の勝者であるAlexNetの転移学習によって取得したCNNを導入して、画像内のオブジェクトの方向を認識します。

The effectiveness of the CNN is evaluated using test image data set of thin resin mold articles.

CNNの有効性は、薄い樹脂成形品のテストイメージデータセットを使用して評価されます。

**Design & training tool for CNN and SVM**

Figure 1 shows the main dialogue of the developed CNN&SVM design tool. In training of CNN, pre-training using randomly initialize weights and additional (successive) training with once trained weights can be selected.

図1は、開発されたCNN＆SVM設計ツールの主要なダイアログを示しています。 CNNのトレーニングでは、ランダムに初期化される重みを使用した事前トレーニングと、一度トレーニングされた重みを使用した追加（連続）トレーニングを選択できます。

As for SVM, one-class unsupervised learning and two class supervised learning can be selectively executed.

SVMについては、1クラスの教師なし学習と2クラスの教師あり学習を選択的に実行できます。

Also, favorite CNN, which is used for a feature extractor, and Kernel function are selected.

また、機能抽出に使用されるお気に入りのCNNとカーネル関数が選択されます。

The tool has another promising function to design original CNNs based on transfer learning.

このツールには、転移学習に基づいて元のCNNを設計する別の有望な機能があります。

For example, the following main items can be set for the operation of transfer learning through the dialogue.

例えば、以下の主な項目は、ダイアログを介した転移学習の操作のために設定できます。

• Folders for training and test images.

トレーニングおよびテスト画像用のフォルダー。

• Base CNNs used for transfer learning such as AlexNet, VGG16, VGG19, GoogleNet and Inception-V3.

AlexNet、VGG16、VGG19、GoogleNet、Inception-V3などの転送学習に使用されるベースCNN.

• Learning parameters such as max epochs, mini batch size, desired accuracy and loss, learning rates for convolution layers and fully connected layers.

最大エポック、ミニバッチサイズ、必要な精度と損失、畳み込み層と完全に接続された層の学習率などの学習パラメーター。

The software shown in Fig. 1 is developed on MATLAB system optionally installed with Neural Network Toolbox, Parallel Computing Toolbox for GPU, Deep Learning Toolbox, Statistics and Machine Learning Toolbox.

図1に示すソフトウェアは、オプションでNeural Network Toolbox、GPU用Parallel Computing Toolbox、Deep Learning Toolbox、Statistics and Machine Learning ToolboxをインストールしたMATLABシステムで開発されています。

IMAGES FOR TRAINING AND TEST

Training image generator was already proposed to efficiently augment limited number of training images [7].

トレーニング画像ジェネレーターは、限られた数のトレーニング画像を効率的に増強するためにすでに提案されました[7]。

By using the generator, images for training are prepared considering typical twelve orientations, i.e., 0◦, 15◦, 30◦, 45◦, 60◦, 75◦, 90◦, 105◦, 120◦, 135◦, 150◦ and 165◦.

ジェネレーターを使用することにより、0◦、15◦、30◦、45◦、60◦、75◦、90◦、105◦、120◦、135◦、150◦、および165◦。

Figures 2 and 3 show examples of the training images for the categories of 45◦ and 165◦, respectively.

図2と3は、それぞれ45◦と165◦のカテゴリのトレーニング画像の例を示しています。

The resolution and channel are 200×200 and 1, respectively.

解像度とチャネルは、それぞれ200×200と1です。

**IV. TRANSFER LEARNING BASED CNN**

1. **Design and Training**
2. 設計とトレーニング

In this section, a transfer learning based CNN is designed to learn the feature of orientation included in images as shown in Figs. 2 and 3.

このセクションでは、転送学習ベースのCNNを設計して、図4および図5に示すように、画像に含まれる方向の特徴を学習します。 2および3。

Figure 4 illustrates the structure of the original AleXnet consisting of 25 layers, which can classify input images into one of 1,000 categories.

図4は、25層で構成される元のAleXnetの構造を示しており、入力画像を1,000のカテゴリのいずれかに分類できます。

In order to make the CNN have an ability to classify input images into 12 categories as 0◦, 15◦, 30◦, 45◦, 60◦, 75◦, 90◦, 105◦, 120◦, 135◦, 150◦ and 165◦, the fully connected layers are replaced as shown in Fig. 5 before executing transfer learning.

CNNに入力画像を0◦、15◦、30◦、45◦、60◦、75◦、90◦、105◦、120◦、135◦、150◦、165◦の12のカテゴリに分類する機能を持たせるため、転送学習を実行する前に、図5に示すように完全に接続されたレイヤーが置き換えられます。

6,889 images consisting of 12 categories are used for the transfer learning.

12のカテゴリで構成される6,889枚の画像は、転移学習に使用されます。

As for training parameters, mini batch size is given 50.

トレーニングパラメーターについては、ミニバッチサイズは50になります。

Iteration is the number of mini batches needed to complete one epoch, so that one epoch in this transfer learning is composed of 6,889/50 ≒ 137 iterations.

反復は、1つのエポックを完了するために必要なミニバッチの数であるため、この転送学習の1つのエポックは6,889 / 50≒137の反復で構成されます。

Desired accuracy and loss are set to 1 and 0, respectively.

必要な精度と損失は、それぞれ1と0に設定されます。

Besides, learning rates of convolutional layers and fully connected ones are set to 0.0001 and 0.001, respectively.

また、畳み込み層と完全に接続された層の学習率は、それぞれ0.0001と0.001に設定されます。

It is important for fast and stable convergence in transfer learning to set the learning rate in convolutional layers smaller than that of fully connected layers.

転移学習の高速で安定した収束には、完全に接続された層よりも畳み込み層の学習率を小さく設定することが重要です。

If the nth image for training is given to the input layer of the transfered CNN, then the softmax layer produces the probability pni(i = 1, 2, · · · , 12) called the score for twelve categories, which is given by

トレーニング用のn番目の画像が転送されたCNNの入力層に与えられた場合、softmax層は、12個のカテゴリのスコアと呼ばれる確率pni（i = 1、2、・・・、12）を生成します。

where yn = [yn1 yn2 · · · yn12]T is the output from the last fully connected layer corresponding to the nth image.

ここで、yn = [yn1 yn2···yn12] Tは、n番目のイメージに対応する最後の完全に接続されたレイヤーからの出力です。

The transfered CNN is trained based on the back propagation algorithm using the loss function called cross entropy.

転送されたCNNは、クロスエントロピーと呼ばれる損失関数を使用した逆伝播アルゴリズムに基づいてトレーニングされます。

The cross entropy is calculated by

クロスエントロピーは次によって計算されます.

where tn = [tn1 tn2 · · · tn12]T means the nth desired output for twelve categories, i.e., only one element in tn has 1, remained elements have 0.

ここで、tn = [tn1 tn2···tn12] Tは、12個のカテゴリのn番目の望ましい出力を意味します。つまり、tnの1つの要素のみが1で、残りの要素は0です。

N is the total number of samples in the training set.

Nは、トレーニングセット内のサンプルの総数です。

The training was conducted using a single PC with a Core i7 CPU and a GPU (NVIDIA GeForce GTX 1060, 6GB).

トレーニングは、Core i7 CPUとGPU（NVIDIA GeForce GTX 1060、6GB）を搭載した単一のPCを使用して実施されました。

The training progress is shown in Fig. 6, in which both the training accuracy and loss seem to well converge to desired values.

トレーニングの進捗状況を図6に示します。この図では、トレーニングの精度と損失の両方が目的の値によく収束しているようです。

It actually took about 40 minutes until the learning was stopped since both the accuracy and loss had not been improved during 10 consecutive iterations or more.

実際には、10回以上の反復で精度と損失の両方が改善されなかったため、学習が停止されるまで約40分かかりました。

Note that this training could be completed within one epoch by severally giving different learning rates in convolutional layers and fully connected layers.

このトレーニングは、畳み込み層と完全に接続された層で異なる学習率をいくつか与えることにより、1つのエポック内で完了することができます。

Through the process explained above, an original CNN model acquired by transfer learning of AlexNet, which is the winner of ImageNet LSVRC2012, is presented to recognize the orientation of objects.

上記のプロセスを通じて、ImageNet LSVRC2012の勝者であるAlexNetの転送学習によって取得された元のCNNモデルが提示され、オブジェクトの向きが認識されます。

1. **Generalization Ability**

After the training, the generalization ability of the transfer learning based CNN is checked using 15 test images imitating resin molded articles which have not been included in the training data set.

トレーニング後、トレーニング学習セットに含まれていない樹脂成形品を模した15個のテスト画像を使用して、転送学習ベースのCNNの一般化能力を確認します。

Figure 7 shows the photos and their classification results, i.e., the angles shown in the JPEG images are the outputs from the CNN.

図7は写真とその分類結果を示しています。つまり、JPEG画像に示されている角度はCNNからの出力です。

It is observed from the results that the obtained CNN has a promising generalization ability that can recognize the orientations of objects in the images.

得られたCNNには、画像内のオブジェクトの向きを認識できる有望な一般化能力があることが結果から観察されます。

However, some visual inconsistencies, e.g., between “test7.jpg” (75◦) and “test12.jpg” (60◦); “test.jpg” (150◦) and “test3.jpg” (150◦) are observed.

ただし、「test7.jpg」（75◦）と「test12.jpg」（60◦）の間など、視覚的な矛盾がいくつかあります。

As can be clearly seen, some images in Fig. 7 are not complete square.

明らかなように、図7の一部の画像は完全な正方形ではありません。

That is the reason why the main cause of these results seems to be the conversion of resolution before classification.

これが、これらの結果の主な原因が分類前の解像度の変換であると思われる理由です。

The resolution of images given to the input layer is forced to be converted to 227×227×3 fixed according to the input layer of the AlexNet, which brings out some undesirable deformation of images and the resultant ambiguities in classification.

入力層に与えられた画像の解像度は、AlexNetの入力層に従って固定された227×227×3に強制的に変換されます。これにより、画像の望ましくない変形とその結果の分類のあいまいさが生じます。

V. **EXPERIMENT OF RICK AND PLACE**

An actual pick and place experiment is conducted using a small articulated robot named DOBOT.

実際のピックアンドプレース実験は、DOBOTという名前の小型多関節ロボットを使用して実施されます。

The experimental setup is shown in Fig. 8.

実験のセットアップを図8に示します。

Position [x y z]T and yaw angle R of the gripper in robot coordinate system can be controlled by an API function SetPTPCmd(xd, yd, zd,Rd).

ロボット座標系でのグリッパーの位置[x y z] Tおよびヨー角Rは、API関数SetPTPCmd（xd、yd、zd、Rd）によって制御できます。

Note that the yaw angle R is dealt with the orientation of a target object in this experiment.

この実験では、ヨー角Rがターゲットオブジェクトの方向に対応していることに注意してください。

Figure 9 shows the developed control dialogue for the robot.

図9は、ロボット用に開発された制御ダイアログを示しています。

Pick and place task while recognizing the orientations of target objects can be executed through the dialogue.

ターゲットオブジェクトの向きを認識しながらタスクをピックアンドプレースすることは、ダイアログを介して実行できます。

Figure 10 illustrates the flowchart of the pick and place task, which is implemented in a timer interrupt routine.

図10は、タイマー割り込みルーチンで実装されるピックアンドプレースタスクのフローチャートを示しています。

In the timer interrupt, first of all, a snapshot of 1600×1200 resolution is captured.

タイマー割り込みでは、まず、解像度1600×1200のスナップショットがキャプチャされます。

After binarized into black and white, a connected component with the largest area is found as a target object and the COG position [Ix Iy]T (1 ≤ Ix ≤ 1600, 1 ≤ Iy ≤ 1200) in image coordinate system is extracted.

白黒に二値化された後、最大面積を持つ連結成分が対象オブジェクトとして検出され、画像座標系のCOG位置[Ix Iy] T（1≤Ix≤1600、1≤Iy≤1200）が抽出されます。

Consequently, desired position [xd yd]T in robot coordinate system to move the gripper to the COG position can be obtained by

したがって、グリッパーをCOG位置に移動するためのロボット座標系での望ましい位置[xd yd] Tは、次のようにして取得できます。

where [X1 Y1]T and [X2 Y2]T in robot coordinate system are the positions of left upper and right bottom of the snapshot as shown in Fig. 11.

ここで、ロボット座標系の[X1 Y1] Tおよび[X2 Y2] Tは、図11に示すように、スナップショットの左上および右下の位置です。

The part of the connected component is further cropped centering the COG from the original snapshot as shown in Fig. 11.

接続されたコンポーネントの一部は、図11に示すように、元のスナップショットからCOGを中心にさらにトリミングされます。

The cropped image is resized into 227×227 and given to the input layer of the transfer learningbased CNN designed in the previous section.

トリミングされた画像は227×227にサイズ変更され、前のセクションで設計された転送学習ベースのCNNの入力層に与えられます。

Finally, the orientation of the object can be estimated by the CNN, which is used for the desired yaw angle Rd so that the robot can successfully grasp the object with a long-axis shape.

最後に、CNNによってオブジェクトの向きを推定できます。CNNは、ロボットが長軸形状のオブジェクトを正常に把握できるように、望ましいヨー角Rdに使用されます。

VI. CONCLUSION

In this paper, a CNN acquired by transfer learning of AlexNet, which is the winner of ImageNet LSVRC2012, is introduced to recognize the orientations of objects.

この記事では、ImageNet LSVRC2012の勝者であるAlexNetの転移学習によって取得したCNNを導入して、オブジェクトの方向を認識します。

Originally, the AlexNet had been able to classify input images into one of 1,000 kinds of objects, however the transfered CNN has been able to recognize the orientation of an object in images with 12 kinds of degrees.

もともと、AlexNetは入力画像を1,000種類のオブジェクトの1つに分類できましたが、転送されたCNNは12種類の角度の画像でオブジェクトの向きを認識できました。

The effectiveness and promise of the CNN are evaluated using test images imitating thin resin mold articles.

CNNの有効性と見込みは、薄い樹脂成形品を模したテスト画像を使用して評価されます。

Then, we actually applied this transfer learningbased CNN to a small articulated robot to demonstrate a pick and place task which required orientation information of target objects.

次に、実際にこの転送学習ベースのCNNを小型の関節ロボットに適用して、ターゲットオブジェクトの方向情報を必要とするピックアンドプレースタスクを実証しました。

In future work, we are planning to apply the proposed system to an actual production line of resin molded articles with various shapes.

今後の作業では、提案システムをさまざまな形状の樹脂成形品の実際の生産ラインに適用する予定です。

Because orientation information is essential for industrial robots to successfully play a pick and place task.

産業用ロボットがピックアンドプレースタスクを正常にプレイするには、方向情報が不可欠です。