

【引用・参照】

- (1) 畳み込みニューラルネットワーク(CNN)をわかりやすく基本から実装まで解説
- <体験型>学習ブログ by zero to one (zero2one.jp)

<https://zero2one.jp/learningblog/cnn-for-beginners/>

- (2) CNN (畳み込みネットワーク) とは? 図や事例を用いながら分かりやすく解説! |
AI専門ニュースメディア AINOW

<https://ainow.ai/2021/09/16/258469/>

- (3) ニューラルネットワークとは? 仕組みや種類、学習手法や活用事例なども解説 |
EAGLYS株式会社

<https://www.eaglys.co.jp/news/column/neuralnetwork/>

CNNとは?

CNNについて勉強してみます。

これも、ネットでググると色々できますが、すんなりと分かるようなサイトはまずありません。

何でこうするの?とか、こんな面倒なことをする必要あるの?とかの説明がないまま、あれこれ解説が続くので、集中が続かないという人が殆どなのではないでしょうか?

そんなわけで、今回も私が大まかに理解できた道順に沿って、まず単語の解説からしていこうと思います。

まずはCNNの言葉の意味について 【CNN】とは **CNN(convolutional neural network, 畳み込みニューラルネットワーク)**のことですが、この単語のうち、**ニューラルネットワーク**から調べて行きます。

【ニューラルネットワーク】とは?⁽³⁾

ニューラルネットワークにはいくつかの種類があります。

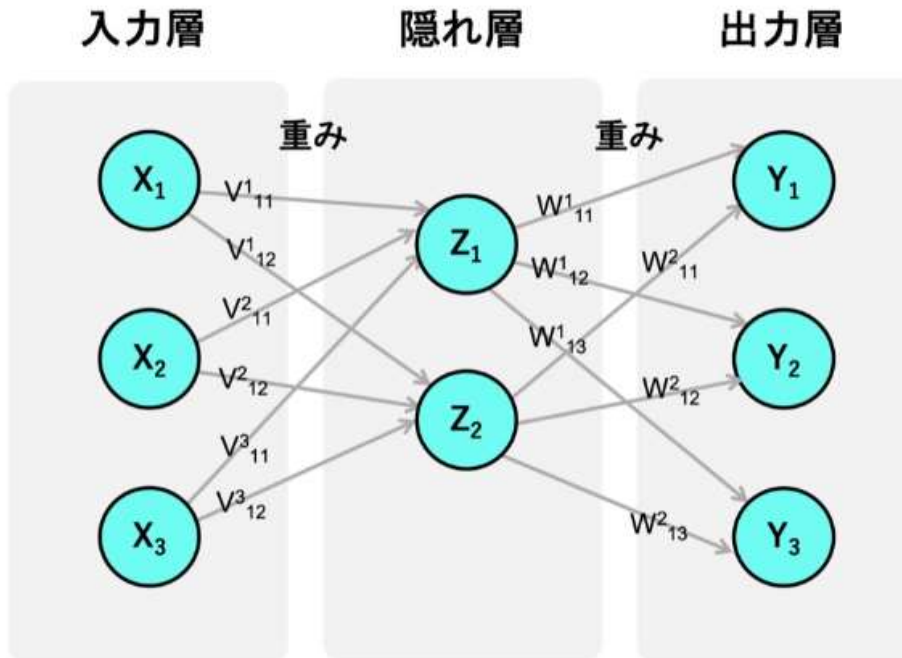
種類	特徴
ディープニューラルネットワーク (DNN)	ニューラルネットワークを多層化したモデル
畳み込みニューラルネットワーク (CNN)	いくつかの個性的な層を持ち、主に画像認識に用いられるモデル
再帰型ニューラルネットワーク (RNN)	時系列データなどを扱うことができ、自然言語処理の分野で期待されているモデル

上記のようなニューラルネットワークがありますが、これらに共通している特徴は入力層と、隠れ層と出力層があり、各層のノードに重みと呼ばれるパラメータをかけて、特徴を絞り込んでいくネットワークだということ。誤差逆伝搬法はこのパラメータを特徴に合わせてチューニングする手法です。

このネットワークの構造は人間の神経構造を模したネットワークという事の様で、そのためにニューラルと呼ばれており、**こうした構造を持つネットワークがニューラルネットワークと呼ばれているようです。**

それぞれの役割を持った複数の層を構成する

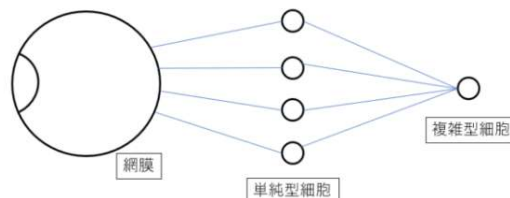
一般的なニューラルネットワークは、情報が入力される「入力層」、情報が発信される「出力層」、その中間にある「隠れ層」の3層で構成されています。



なかでもニューラルネットワークの肝となるのが、入力されたデータに対してさまざまな計算を行う隠れ層です。隠れ層をいくつも持つことで、より複雑な問題にも対処できるようになります。

それではCNNについて調べていきます。解説にもあるように中間に個性的な隠れ層を持ち、主に画像認識に使われます。このネットワークは人間の視覚神経を模したもので、下図のような説明がされています。(1)

以上の理由から、生物の脳の視覚野におけるニューロンの構造を模倣する方法が考案されました。我々生物の脳の視覚野には**単純型細胞**、**複雑型細胞**という二種類の細胞が存在し、それらが下の概略図のように目の網膜と接続しています。



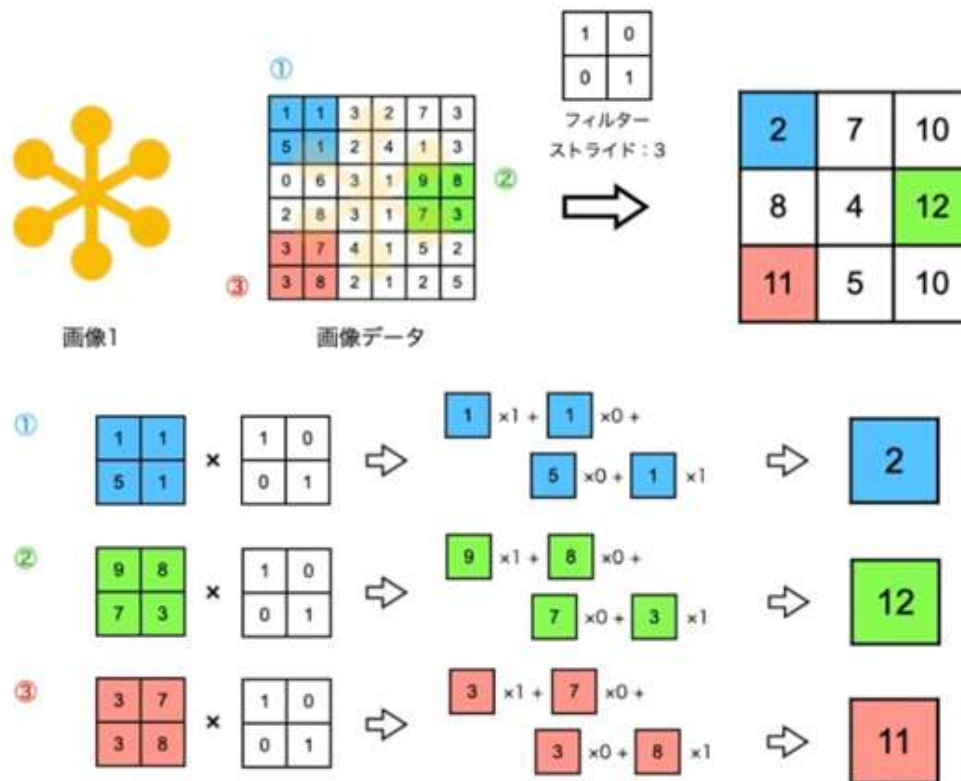
単純型細胞と複雑型細胞のイメージ

ここで重要なことは**部分的な特徴抽出と俯瞰的な特徴抽出の二つの機能がある**ということです。部分的な特徴抽出はどんな特徴があるかを抽出し、俯瞰的な特徴抽出はそれが全体画像のどこにあるかを判断します。二つが合わさって、全体の画像が何かを判断します。

CNNの畳込み = convolutionalの意味は何でしょうか？

これは文字通り、特徴を折り紙の様に畳込んで、大きい色紙を小さく織り込んでいくようなイメージです。分かりやすい画像がないので、畳込み計算の図面を流用しますと、以下の様です。(2)

【畳込み層の図解】

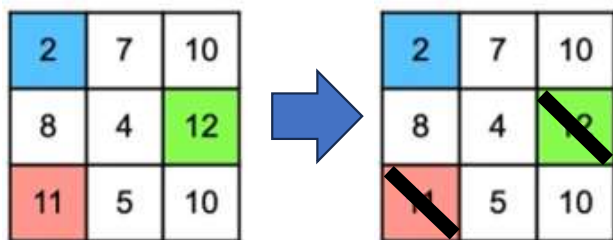


上の図では6×6の画像データを3×3の画像データに畳み込んでいます。畳込みのルールは①～③の例の様に、元の画像データの一部とフィルターで行列演算し、右の3×3の画像データに置き換える操作です。**でも、何でこんなことをするのか？何の意味があるのか？分からなければ続きを読む気にはなりません。**

この説明は一切ないので、あちこち調べた結果を補足として追加しておきます。

この演算は行列演算の内積と呼ばれる演算で、フィルターの行列と同じ性質を持つ画像データをあぶりだす機能を持っています。だからさらっと説明なく書かれているフィルターがとても重要な意味を持っており、このフィルターの場合、フィルターを掛けた部分画像に、フィルターと同じ左上から右下向きの部分画像がある場合には、この内積した結果の数値が最大になります。

畳み込んだ画像データを見てみると、右の列の真ん中と左下に大きな数値が現れています。これはこの辺りに左上から右下に向かう特徴をもった画像があるらしいと言う事になります。



ですから、このフィルターの行列を変化させれば、色々な特徴をあぶりだせることになります。

例えば、以下の表の例のようになります。

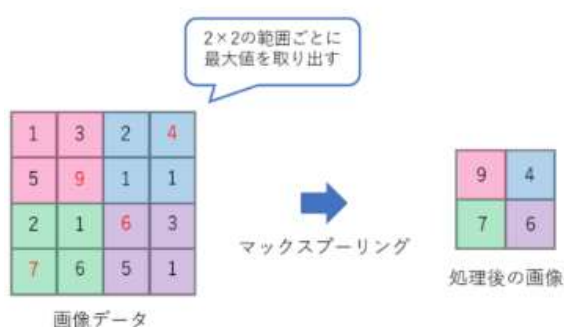
横上棒		横下棒		右棒		左棒	
1	1	0	0	0	1	1	0
0	0	1	1	0	1	1	0
ブロック		かぎ型1		かぎ型2		逆かぎ型	
1	1	1	1	1	1	0	1
1	0	0	1	1	0	1	1

このようなフィルターをたくさん作り、順次、元の画像データに適用すれば、各ブロック毎の画像の特徴が浮かび上がってきます。例えば、横上棒、横下棒、右棒、左棒どのフィルターも数値が大きければ、これは黒いブロックではないかと判断します。

これは簡単な例ですが、複雑なフィルターを作って内積を取り、各ブロックの特徴量を計算することで、もっと詳細な画像の特徴抽出をすることができ、実用的な画像認識ツールが出来上がることになります。

内積の計算は一次関数の計算なので、GPUを使うと高速に計算ができ、詳細なフィルターを作って計算量が大きくなっても計算時間にはさほど影響を与えません。

さらに判定を素早くするために、フィルタごとの特徴がどこにあるか大まかな略図を示す数値もあります。それがプーリング値です。荒い特徴だけで描いた画像になります。⁽¹⁾



マックスプーリングの例

これまでの説明をまとめると、CNNは巨大なジグソーパズルのデータベースを持っており、画像のどの位置にどんな特徴があるかを判断して、画像を再構成する画像処理装置という事になります。この特徴量から構成されたジグソーパズルの絵のデータから、入力された画像がどういう画像なのかを最終的に認識することができます。

以上