機械学習エンジニアコース
Sprint

－ CNN_コンボリューションNN －

DIVE INTO CODE

今回のモチベーション

目的はなにか

1. スクラッチを通して**CNN**を理解する
2. 線形モデルと異なる手法に触れる

このスライドは?

ここでは、CNNの基本的な知識を学びましょう

CNN とは何か

CNN（畳み込みニューラルネットワーク）とは、疎な（スパース）構造を持つ畳み込みレイヤーから成るネットワークのことである。
人間が視覚情報からパターンを認識するプロセスを模倣して考案された。

1958年ハーバード大学のヒューベルとウィーゼルが、猫の視覚野に特定の傾きをもつ線分を見せたときにだけ反応する細胞があることを発見した。このような細胞を単純細胞と呼ぶ。単純細胞は局所的なある有効範囲に特定のパターンが現れたときだけ発火する。

http://web2.chubu-gu.ac.jp/web_labo/mikami/brain/26/index-26.html

この後の流れ

① CNNの躍進を知る

② 畳み込み層と全結合層の違いを知る

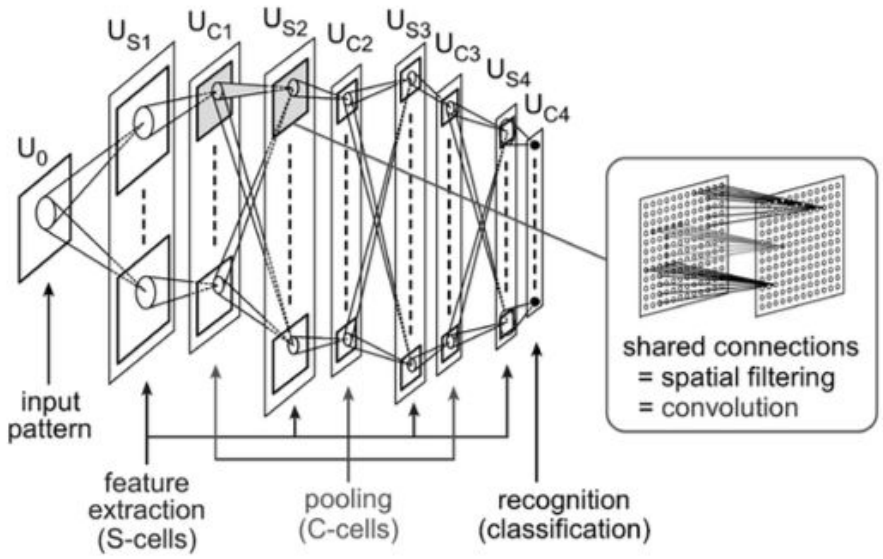
③ 畳み込み層とは何かを考える

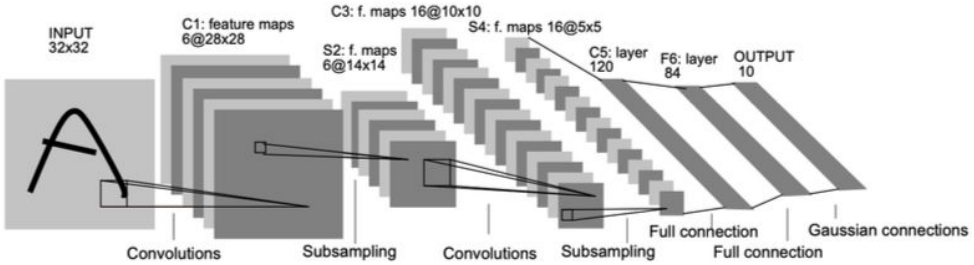
CNN の躍進

CNNの起源は、1982年に福島邦彦が発表したネオコグニトロン[1]に遡る。
現在のCNNもネオコグニトロンも同じ階層型の多層ネットワークで、局所的な特徴抽出の層と位置ズレ補正を行う層を持つ。
ネオコグニトロンはAdd-if-Silent則という学習則を用いるが、1998年Yann LeCun et al. によって発表されたCNN（LeNetと呼ばれる[2]）ではデルタ則の一般化（誤差逆伝播法）が採用された。

その後、2012年にAlexNetと呼ばれるCNNが、ImageNet LSVRC-2012というコンペティションで優勝し躍進を遂げた。

[1] http://www.visionsociety.jp/vision/vol29-1/29-1_1.pdf
[2]http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf





AlexNetとは

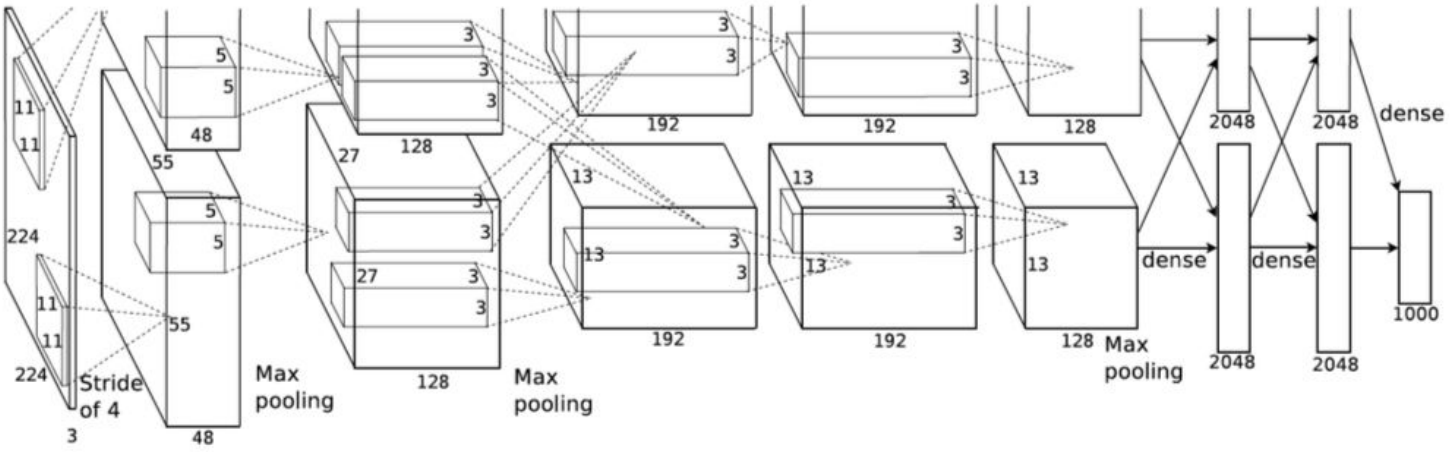
Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)

http://image-net.org/challenges/LSVRC/2012/results.html#abstract

8層からなるアーキテクチャで、ネットワークパラメータ数は6230万、フォワードパスには11億の計算ユニットを持つ。
畳み込み層の計算コストは90-95%を占め、すべてのパラメータの5%を持つ[3]。（全結合層は計算コストの5-10%、パラメータは全体の95%）

① 非線形性を追加するために、Tanhの代わりにReluを採用したことで、同じ精度で速度を6倍に加速した。
② 正規化の代わりにドロップアウトを使用し、過剰適合に対処した。ドロップアウト率が0.5の場合、トレーニング時間は2倍かかる。
③ 重複プーリングによるネットワークのサイズの削減。

[3] https://arxiv.org/pdf/1404.5997.pdf



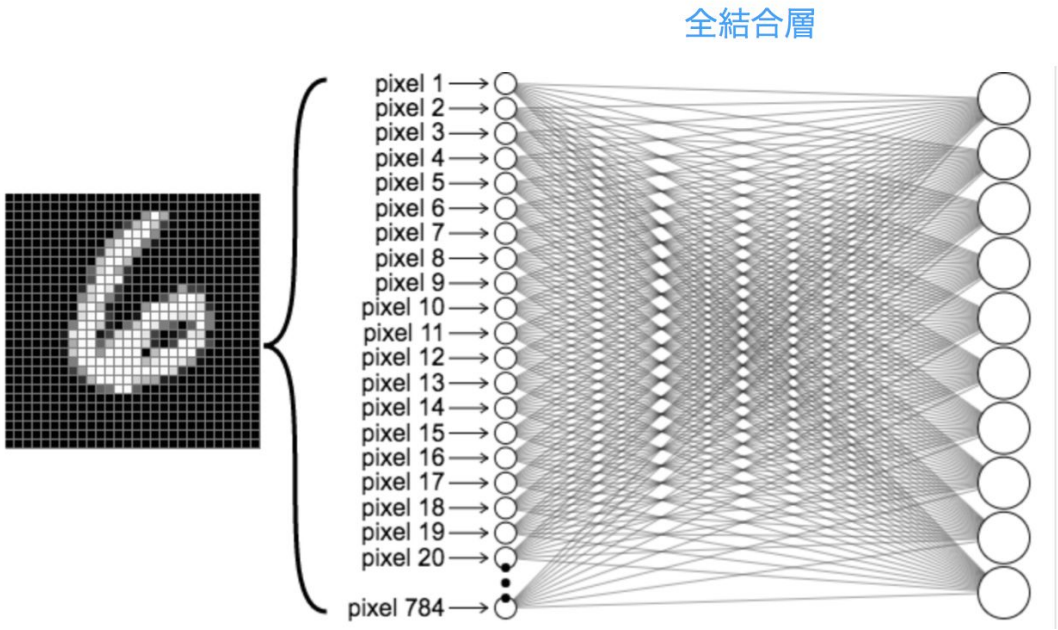
畳み込み層と全結合層の違い

全結合層

ニューラルネットワークでは、全結合層（affine層、Full Conected Layer、Dense Layerなどと呼ばれる）に2次元データ(sample数, feature数)を通していました。

これに対して、畳み込みニューラルネットワークでは4次元行列のデータ（縦幅, 横幅, チャネル数, バッチサイズ）(1)を畳み込み層へ通します。

(1)フレームワークによって順序が異なることがあります。



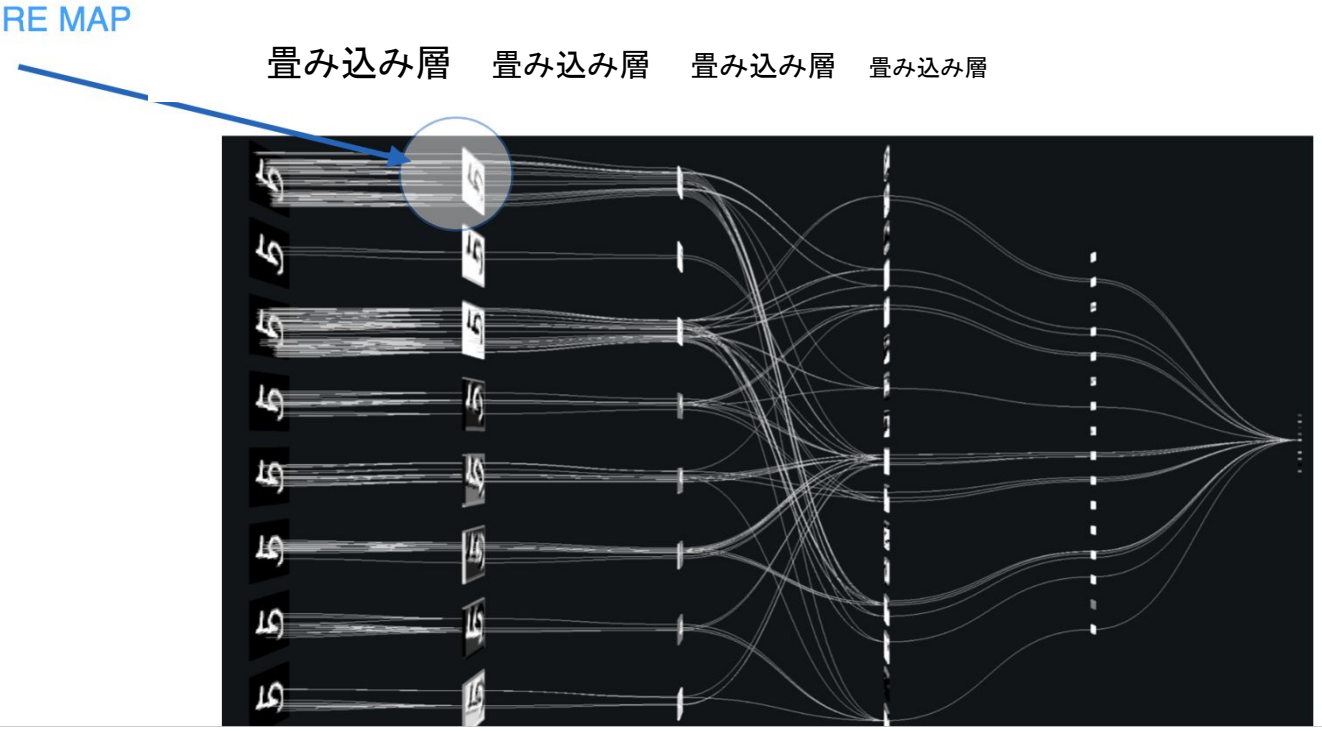
畳み込み層と全結合層の違い

畳み込み層

畳み込み層（convolution層、Conv2dなどと呼ばれる）からの出力は、一般に**Feature Map**と呼ばれる。

下のサイトでいろんな角度から眺めてみよう。4次元行列のchannelの軸はどれでしょう？
出力がだんだん小さくなっていくのはなぜだろう？

https://terencebroad.com/works/cnn-vis

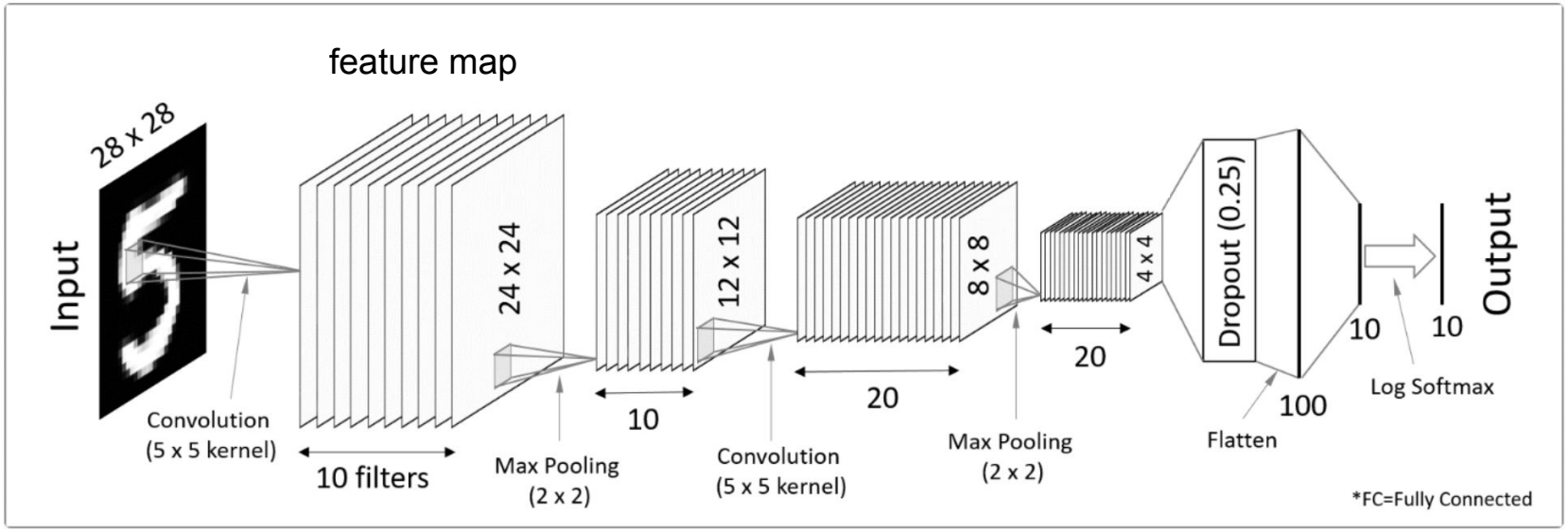


畳み込み層と全結合層の違い

畳み込み層

畳み込み層が入力と**局所的に結合する**重み行列のことを**フィルタ**または**カーネル**（Filter、kernel）と呼ぶ（厳密には、カーネルの集まりがフィルタと定義されている）。

このフィルタが入力の特徴（エッジやテクスチャと言われるパターン）を抽出する。
2次元畳み込みでは、一つのフィルタは入力の横縦方向にオフセットし、入力全体を網羅的に探索する。
以下は、クラシカルなCNNアーキテクチャである。



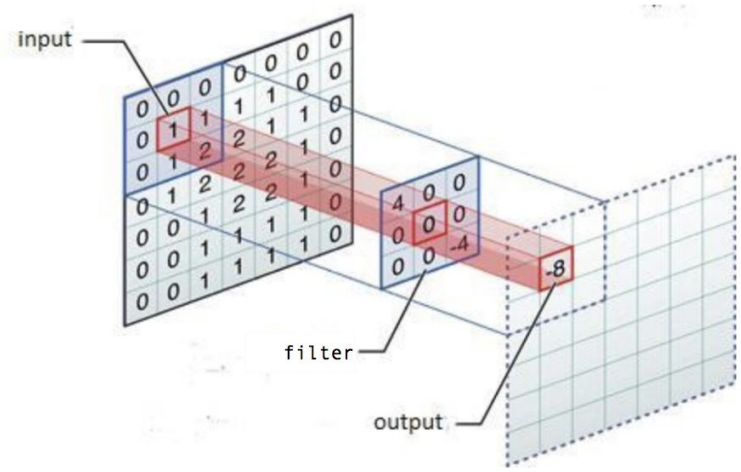
畳み込み層と全結合層の違い

畳み込み層

フィルタは何かしらの行列で、入力の部分と重なる箇所において積和演算を行い、特徴（パターン）を抽出する。一つのフィルタは横→縦に**オフセット**（画像ならばpixel単位で）**し**、各積和演算の結果を出力する（**局所ごとにフィルタを変えるのではなく、一つのフィルタが全体を網羅する**ため、**複数のpixelが同じ重みの値を共有する**ことになる）。フィルタサイズが1×1より大きければ、出力の縦横サイズは入力よりも小さくなる。

入力の縦横サイズを保って出力したい場合は、畳み込み演算をする前に入力の外周pixelを特定の値（通常はゼロ）で埋める（これをゼロパディングと呼ぶ）。フィルタは単体では3次元行列（filter縦幅,filter横幅,channel数）だが、一つの畳み込み層においては（filter縦幅,filter横幅,channel数,filter数）という4次元行列で用意する。つまり、一つの入力に対し複数のフィルタが適用されるということ。このことから、フィルタ数はNNのノード数のように捉えることができる。

畳み込み層



畳み込み層とは何か

「畳み込み」とはなにか？

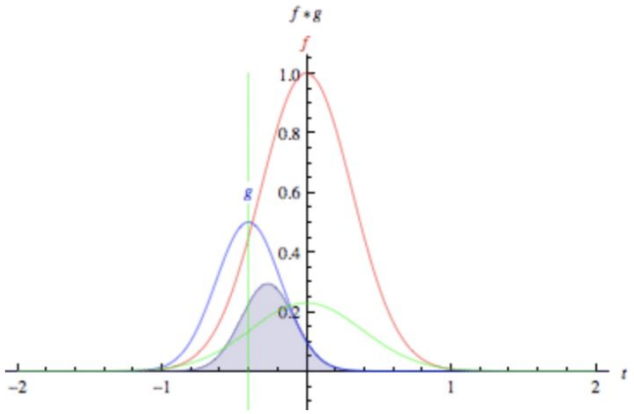
数学において連続値に対して**畳み込み積分**という演算があるが、CNNにおける畳み込みは、離散値に対する**畳み込み積和**に当たる。

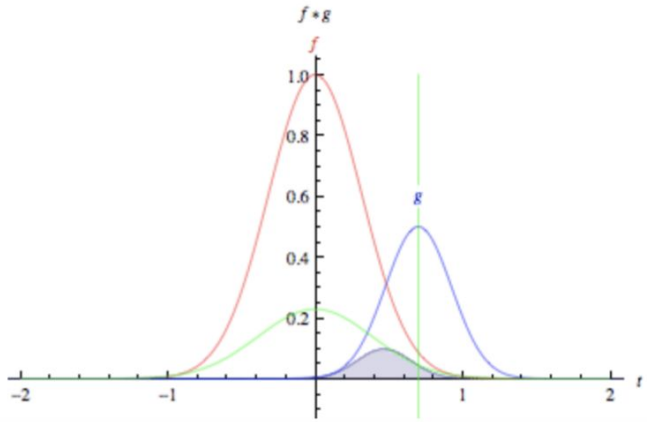
https://www.clg.niigata-u.ac.jp/~medimg/practice_medical_imaging/imgproc_scion/4filter/index.htm

http://tdual.hatenablog.com/entry/2018/05/02/113110

CNNの「畳み込み」は、以下のように捉えることができる

「積和」　　：特徴を抽出
「畳み込み」：総当たりする（重なりを測定する）





https://pathmind.com/wiki/convolutional-network

フィルターで特徴を抽出する

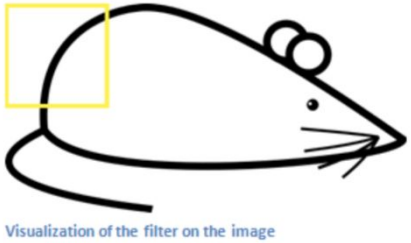
フィルタはいかにして特徴を検出するか。

ニューラルネットワーク以前には、フィルターの値は人間が計算して決めていた。

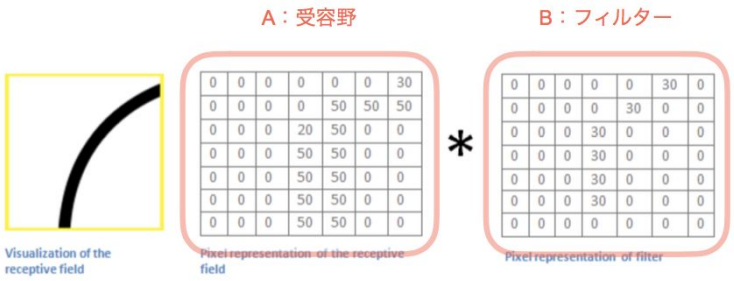
例えば下図のAのような受容野を検出するのに適したフィルターはBのような値を持つ。

Multiplication and Summation = (50*30)+(50*30)+(50*30)+(20*30)+(50*30) = 6600 (A large number!)

https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/



Visualization of the filter on the image



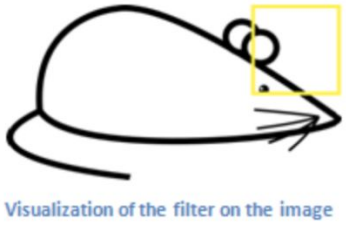
フィルターで特徴を抽出する

フィルタはいかにして特徴を検出するか。

以下のような受容野Aに対して、フィルターBが局所的に積和されるとき、計算結果は「0」となる。

（特徴量が抽出されない）

Multiplication and Summation = 0



Visualization of the filter on the image

A：受容野

0	0	0	0	0	0	0
0	40	0	0	0	0	0
40	0	40	0	0	0	0
40	20	0	0	0	0	0
0	50	0	0	0	0	0
0	0	50	0	0	0	0
25	25	0	50	0	0	0

Pixel representation of receptive field

*

B：フィルター

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

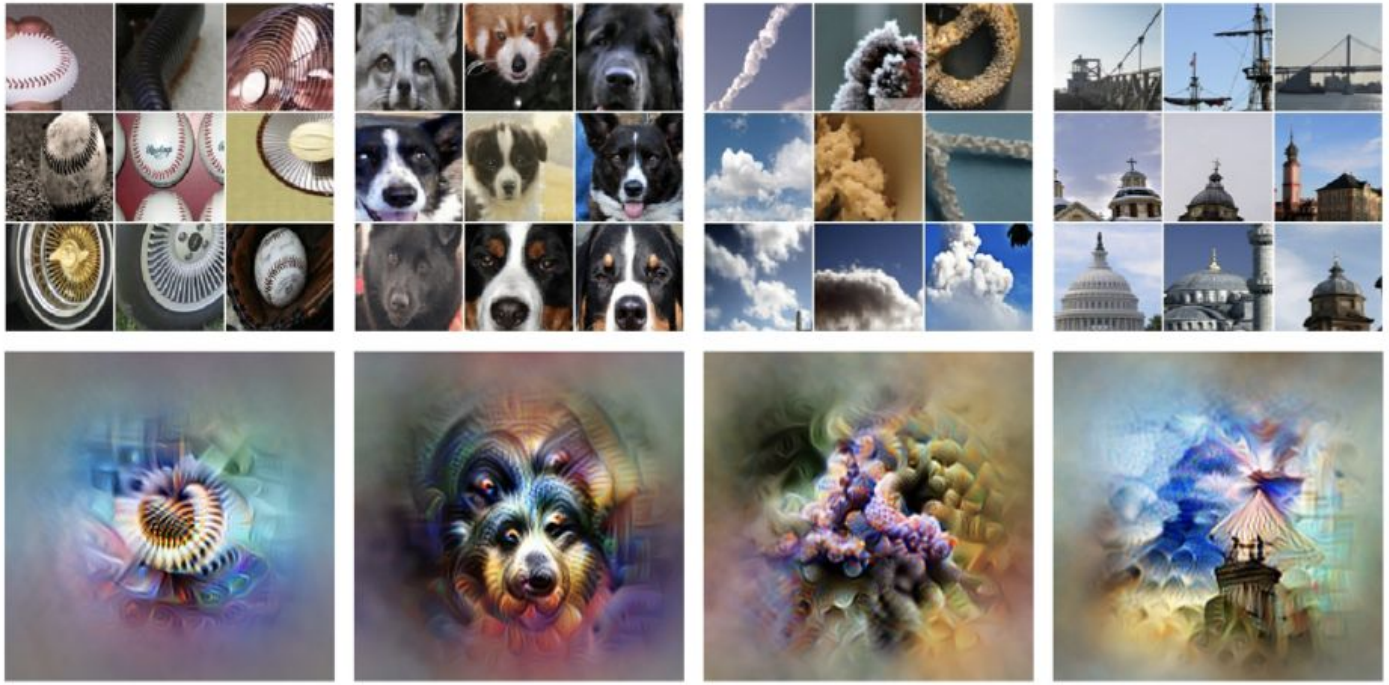
Pixel representation of filter

Feature Map

CNNは何を見ているか

CNNの最適化されたFeature Mapをのぞいてみよう。

https://distill.pub/2017/feature-visualization/



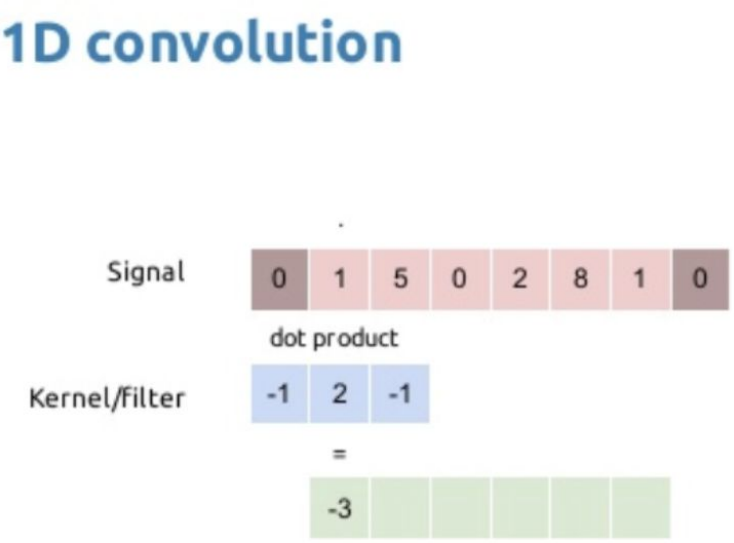
1Dconvolutionとは

2Dconvolutionが基本だけれど、

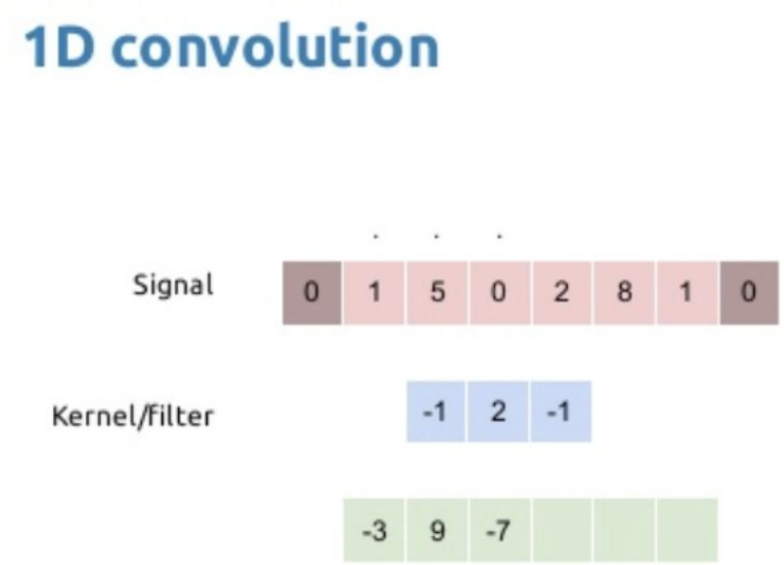
1Dconvolutionから始めよう。

フィルタの動きを追ってみよう。

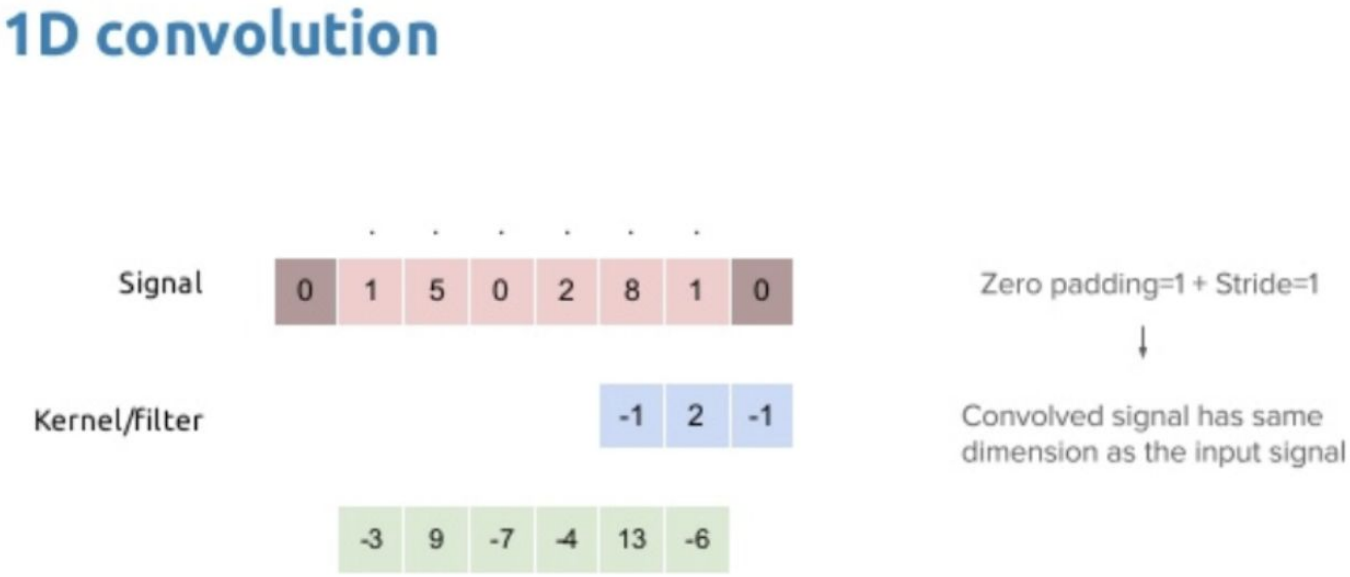
設定は、zero padding = 1, stride = 1 で行う。



1Dconvolutionとは



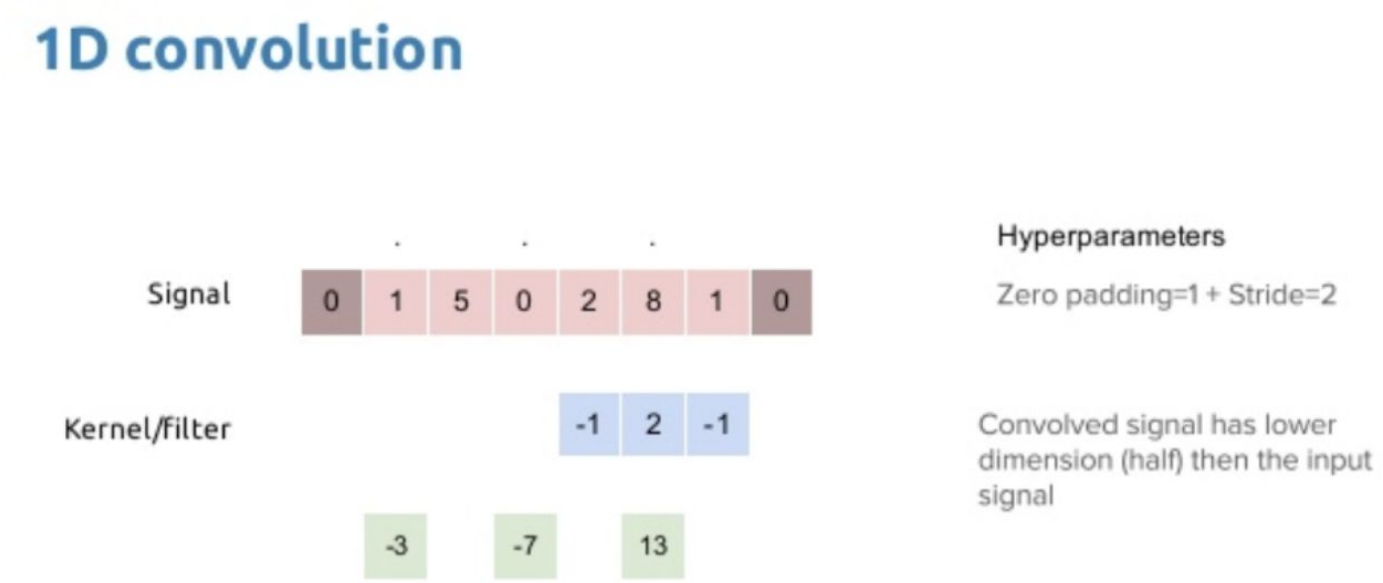
1Dconvolutionとは



1Dconvolutionとは

ストライド = 2 だと、

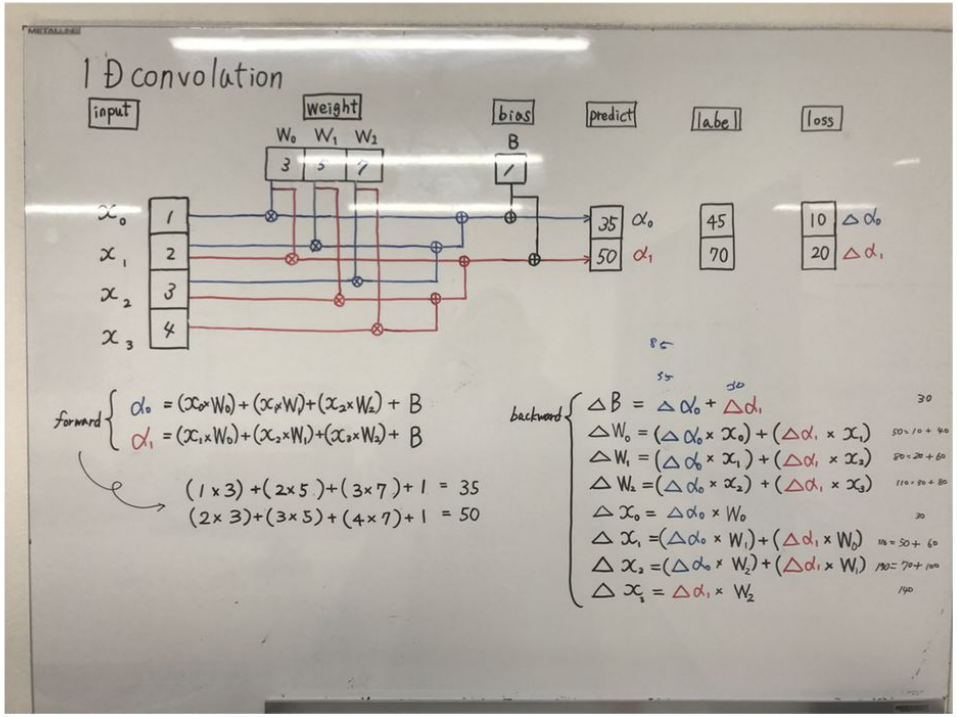
アウトプットサイズが小さくなる。



conv1Dのヒント

今日のDiver

単体テストデータをconv1Dに通してみよう



まとめ

① 畳み込み層は、局所的な受容野（有効範囲）からパターンを抽出する

② パターンを抽出する一つのフィルタは入力データに対して局所的な結合をオフセットすることで全体を網羅的に探索する。（重み共有）

CNN_コンボリューションNN
完