機械学習エンジニアコース
Sprint

－ CNN2_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2DのCNNを理解する
2. 1DCNNとの異なる点に触れる

振り返り

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。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。

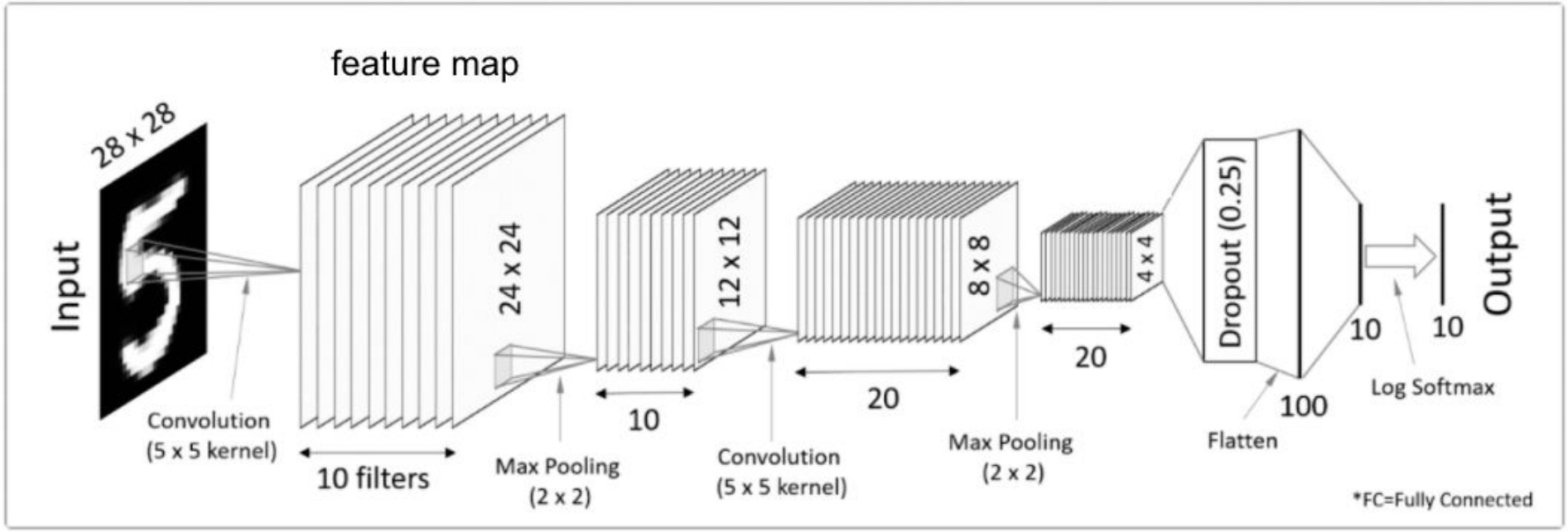
振り返り

畳み込みニューラルネットワークは以下を前提とする

①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）

CNN2 - CNNイメージ図（再掲）



CNN2 – マルチチャネル

マルチチャネル

マルチチャネル画像の典型的な事例はRGB画像です。各RGBチャネル（3チャネル）は**元の画像のさまざまな側面を強調**しています。

畳み込みニューラルネットワークの層においてもマルチチャネルデータを扱います。畳み込み層は通常、複数のチャネル（通常は数10〜数100チャネル）で構成されています。**各チャネルは、前の層のさまざまな側面について説明しています。**

Original image (RGB)

R channel

G channel

B channel

CNN2 – フィルターとカーネル

フィルタとカーネルは同じようで、少し違う

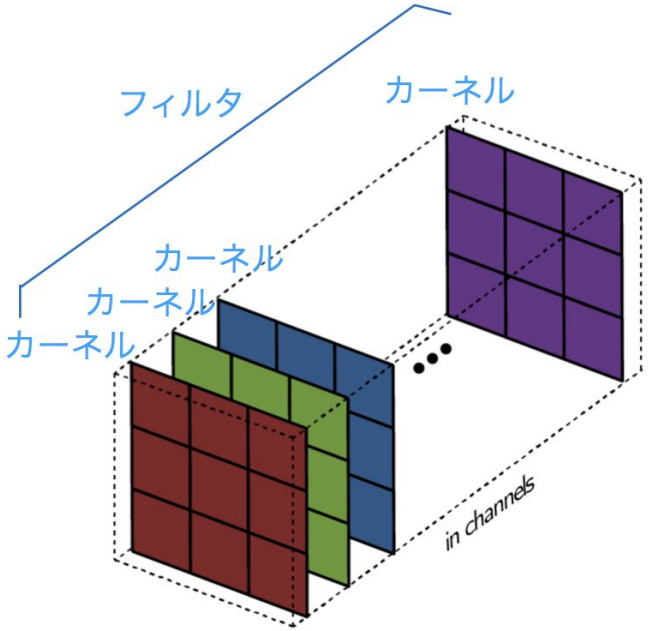
フィルタは**カーネルの集まり**と考えられます。

カーネルとは、2次元配列の重みを指します。一方フィルタは、重なった複数のカーネル（3D構造）を意味します。

入力が**1チャネル**の場合、カーネルは1枚になるので、フィルタとカーネルは同義語として扱われます。

入力が**マルチチャネル**（例えば3チャネル）の場合、カーネルは入力チャネルの数だけ用意され（3枚）、そのカーネルの集まりを１フィルタと数えます。

そのようなフィルタを4つ用意すれば、出力（フィーチャーマップ）のチャネル数は4チャネルになります。



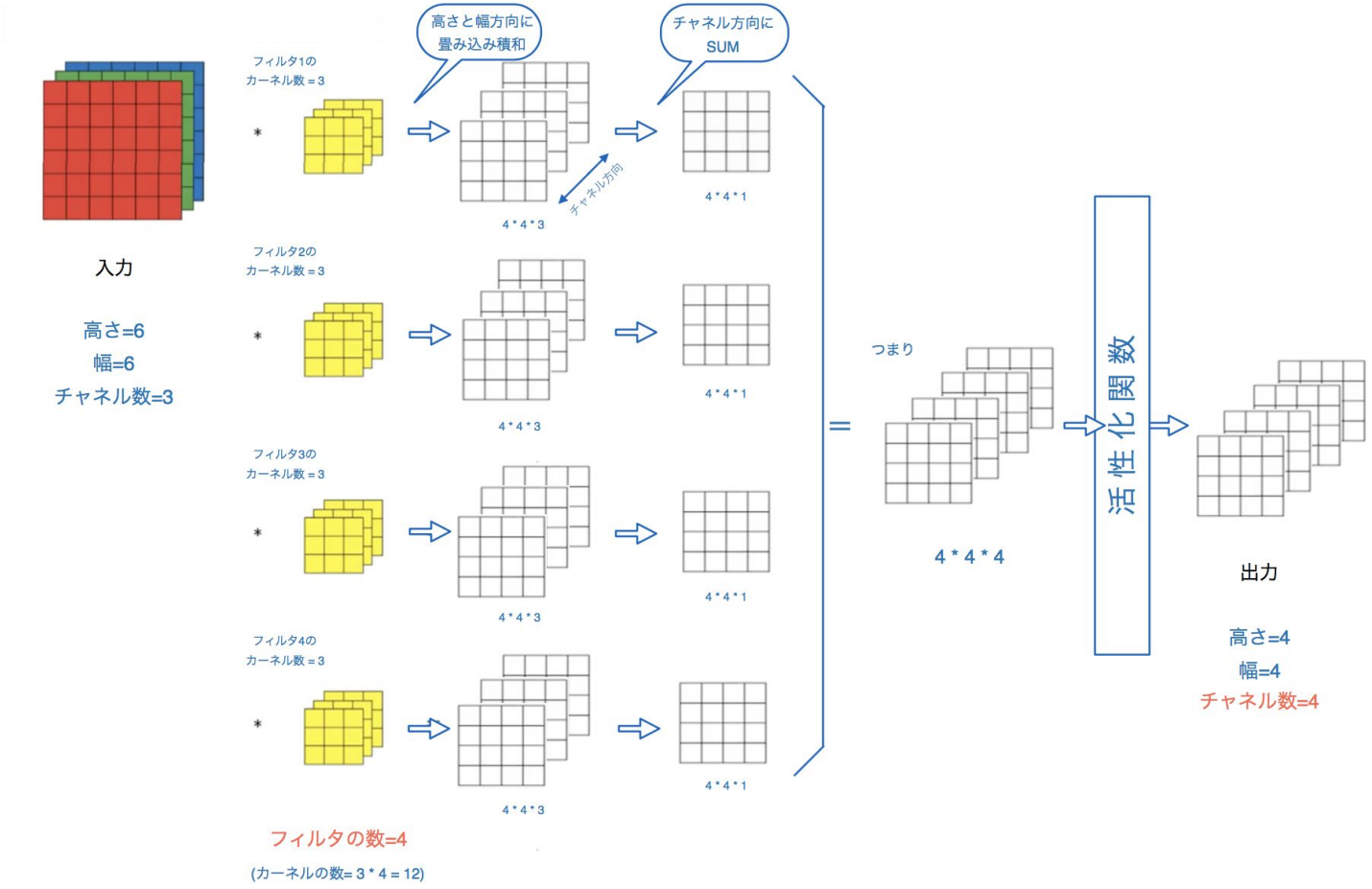
フィルタ：カーネルの集まり

CNN2 - 2DConvのイメージ

マルチチャネル

入力 → 2Dconv →

活性化関数 → 出力



CNN2 – フィルタ数の決め方

フィルタ数は

どのように決めるのか？

フィルタの数はハイパーパラメータである。
このフィルタ数が、次の層の入力のチャネル数に対応するので、

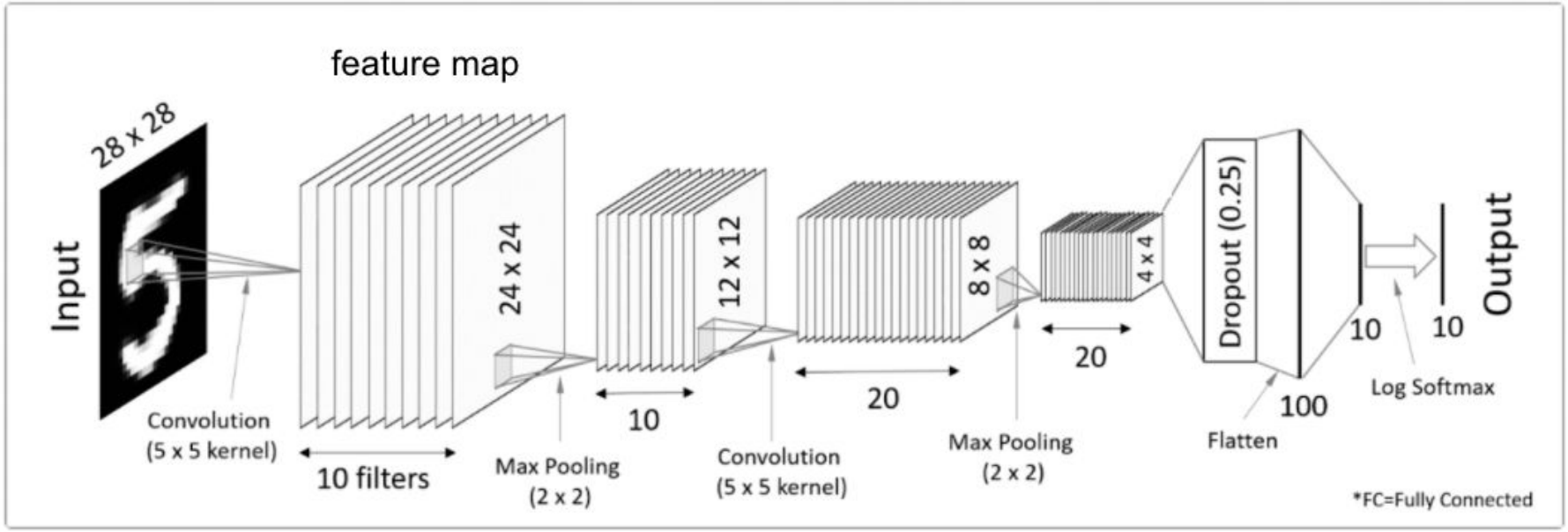
次の層は何チャネル入力にするか？を考えて今の層のフィルタ数を決める。（NNのノード数の決め方と同じ）

CNNレイヤー　　フィルタ数　　シェイプ

Layer	Number of Filters	Padding	Activation Shape	Activation Size
Input Image	-	-	(28,28,1)	784
Conv2d(f=3,s=1)	32	Valid	(26,26,32)	21,632
MaxPool(f=2,s=2)	-	Valid	(13,13,32)	5408
Conv2d(f=3,s=1)	64	Valid	(11,11,64)	7,744
MaxPool(f=2,s=2)	-	Valid	(5,5,64)	1600
Conv2d(f=3,s=1)	128	Valid	(3,3,128)	1,152
MaxPool(f=2,s=2)	-	Valid	(1,1,128)	128
Flatten	-	-	(128,1)	128
Dense	-	-	(64,1)	64
Softmax	-	-	(10,1)	10

パディングには
VALIDとSAMEの2種類がある。
調べてみよう。

CNN2 - CNNイメージ図（再掲）



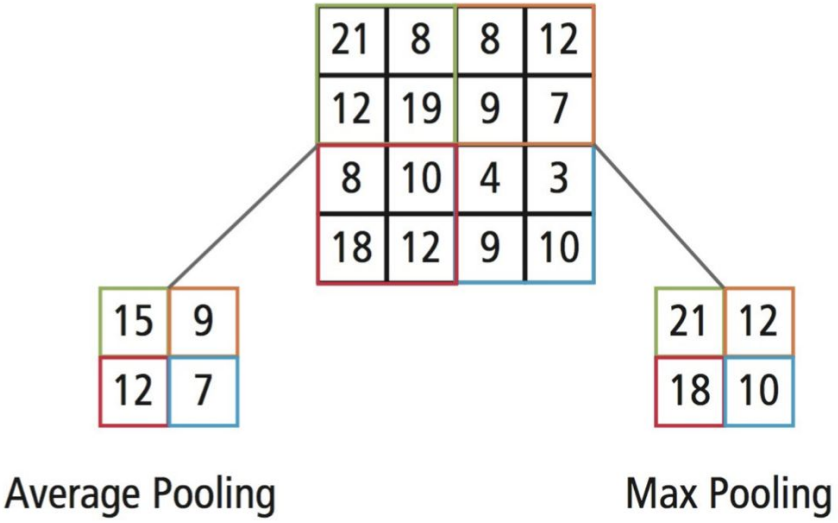
CNN2 – プーリング層の処理

プーリングの大事なこと

プーリングは最大値（あるいは平均値）をとる処理なので、**学習するパラメータはない。**また、**入力と出力においてチャネル数も変化しない。**

CNNのアイデアの起源は、1958年の猫の視覚野に関する実験で発見された単純型細胞と複雑型細胞に由来していた。
畳み込み層が単純型細胞の模倣である一方、プーリング層は複雑型細胞の模倣とされている。

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



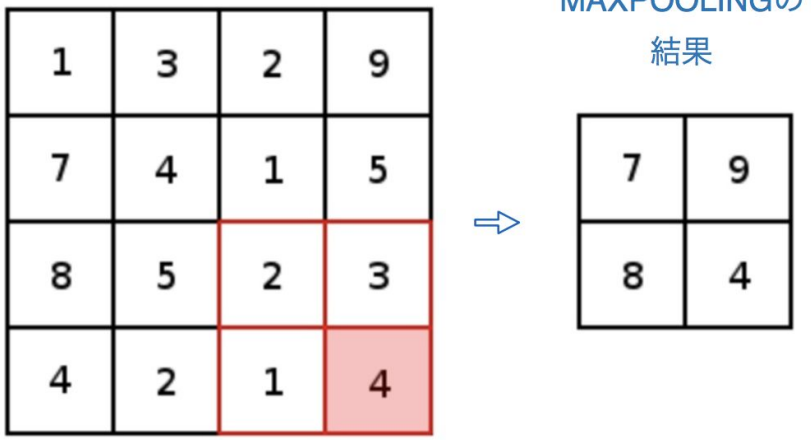
平均プールと最大プール

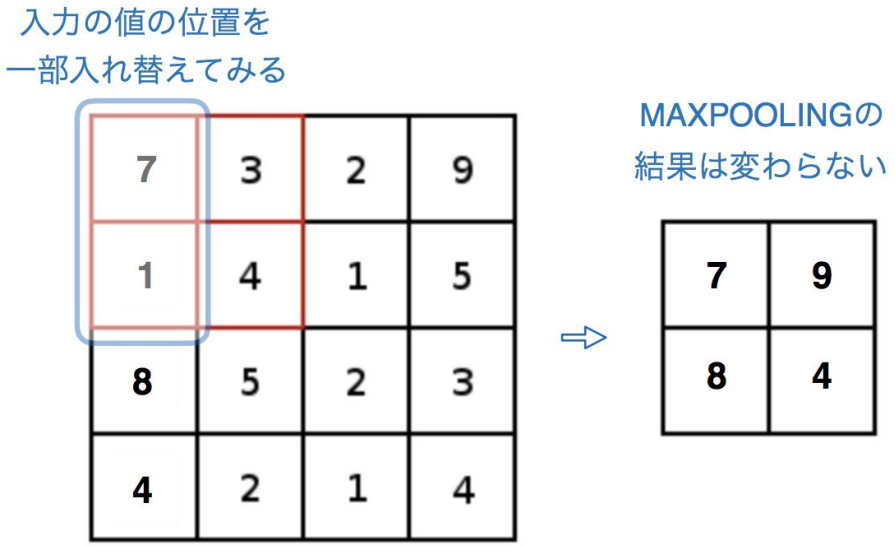
CNN2 – プーリングについて

プーリング層は何をしているか？

畳み込み層が局所領域ごとにパターンを抽出しているのに対し、プーリング層では、抽出された情報から局所領域ごとに代表値をとる。代表値に置き換えることで位置がずれたパターンも同定することが可能である。

下の左図は、2＊2の矩形フィルタをストライド2でずらすMAX POOLINGを施している図です。右図は入力の値を一部入れ替えても結果が左図と変わらないことを示している。
MNISTの場合、数字の「7」が画像の真ん中に書かれていても少し左や右にずれていても同じように「7」と判定することができる。





CNN2 - CNNの歴史

state-of-the-art (CNNs)の変遷

CNNのstate-of-the-art（最先端の）アーキテクチャの変遷

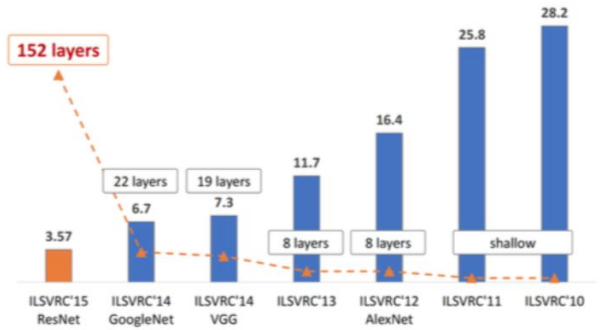


Fig. 1. The evolution of the winning entries on the ImageNet Large Scale Visual Recognition Challenge from 2010 to 2015. Since 2012, CNNs have outperformed hand-crafted descriptors and shallow networks by a large margin. Image re-printed with permission [36].

AlexNet を始めとする過去の典型的な CNN では、後段に数層の全結合層をもち、これが最後の畳込み層と出力層の間に入る。

全体の重みの 9 割以上がこれら全結合層に偏り、その規模がモデルのパラメータ(重み)数の大半を占めていた。

GoogLeNetやXception、ResNetでは層数が増えたにも関わらず、大きな全結合層を持たないため（代わりにプーリング層で取りまとめて出力層へ繋ぐ）、全体のパラメータ数が減少している。

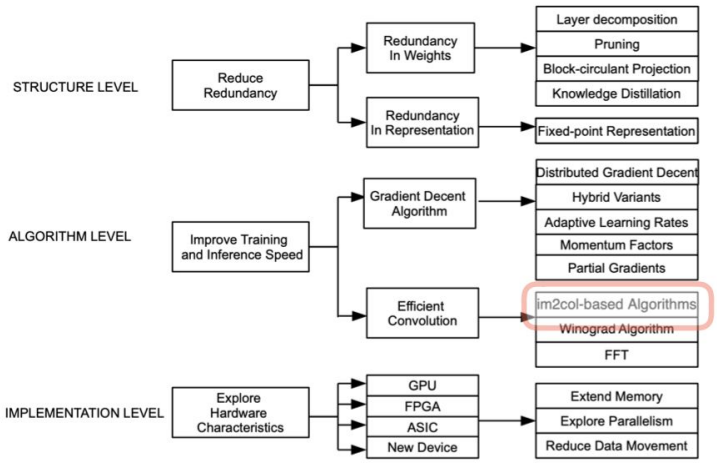
Model	Layer Size	Configuration	Feature	Parameter Size	Application
LeNet [22]	7 layers	3C-2S-1F-RBF output layer		60,000	Document recognition
AlexNet [23]	8 layers	5C-3S-3F	Local response normalization	60,000,000	Image classification
NIN [24]	-	3mlpconv-global average pooling (S can be added in between the mlpconv)	mlpconv layer: 1C-3MLP; global average pooling	-	Image classification
VGG [25]	11-19 layers	VGG-16: 13C-5S-3F	Increased depth with stacked 3 × 3 kernels	133,000,000 to 144,000,000	Image classification and localization
ResNet [26]	Can be very deep (152 layers)	ResNet-152: 151C-2S-1F	Residual module	ResNet-20: 270,000; ResNet-1202: 19,400,000	Image classification, object detection
GoogLeNet [27]	22 layers	3C-9Inception-5S-1F	Inception module	6,797,700	Image classification, object detection
Xception [28]	37 layers	36C-5S-1F	Depth-wise separable convolutions	22,855,952	Image classification

Table 1: CNN model summary.
C: convolutional layer, S: subsampling layer, F: fully-connected layer

CNN2 - im2colの紹介

CNNをアクセラレート（高速化）する方法

- 冗長さの削減
- 学習・推論の高速化
- ハードウェア特性の探索



efficientなconvolutionとして、im2col[1]という手法が有名である。
この手法では、入力データとフィルタの次元を落とすことで1回の行列積演算でconvolutionを行う。
im2colで得られた2次元行列に対しさらに特異値分解を行い低ランク近似によって1%程度の精度の低下と引き換えに、2倍程度の高速化を得る手法もある。

[1] https://hal.inria.fr/inria-00112631/document

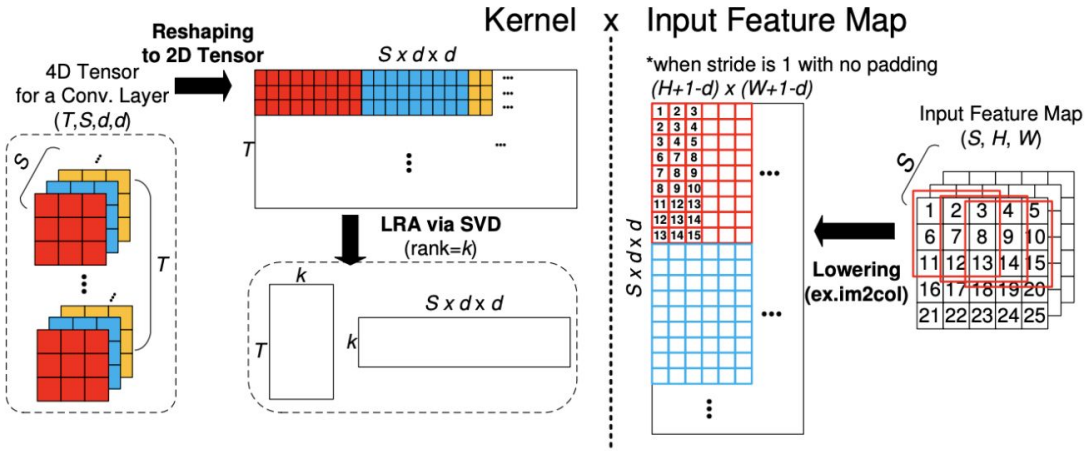
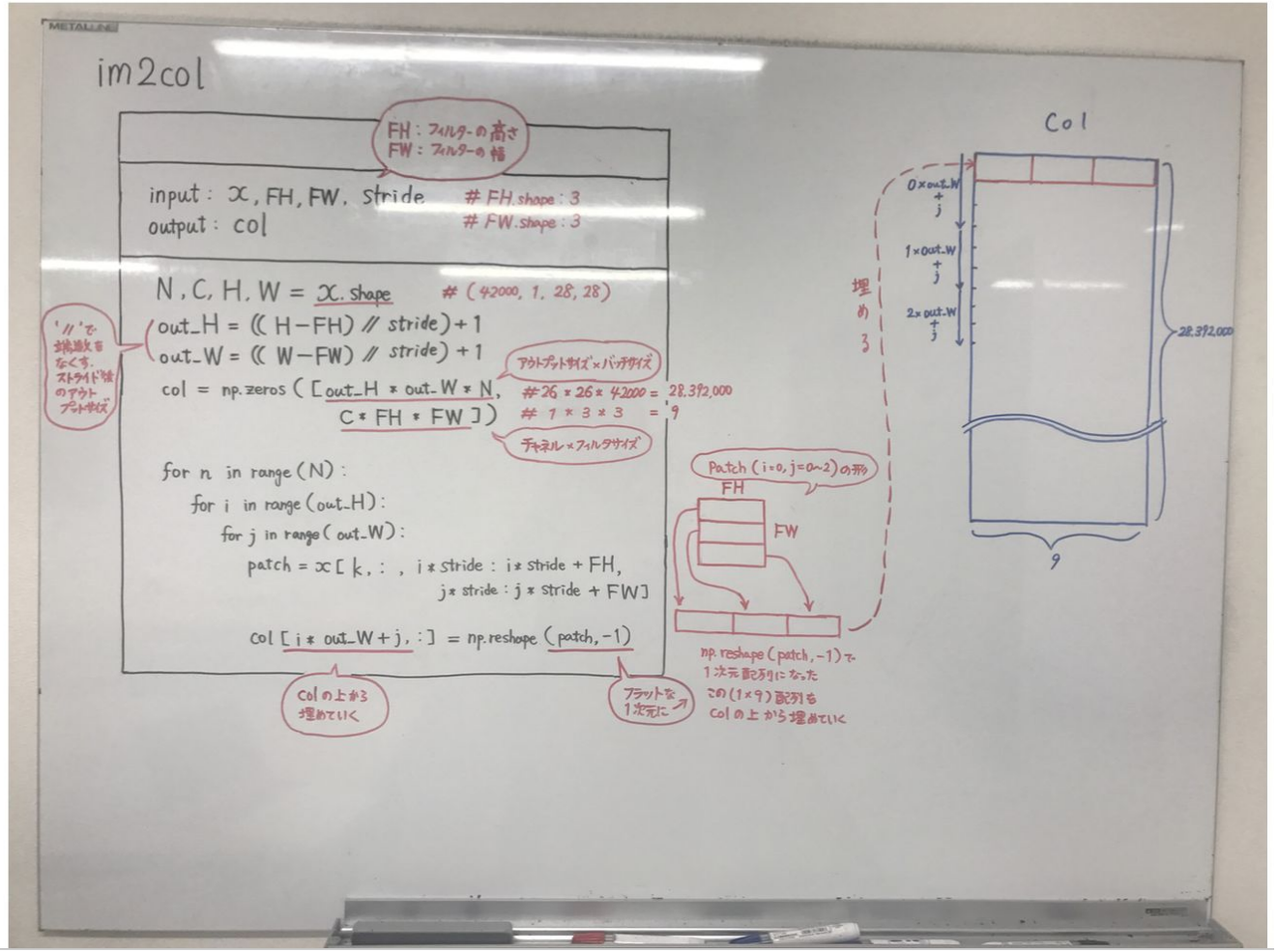


Figure 1: An example of lowering technique using im2col.

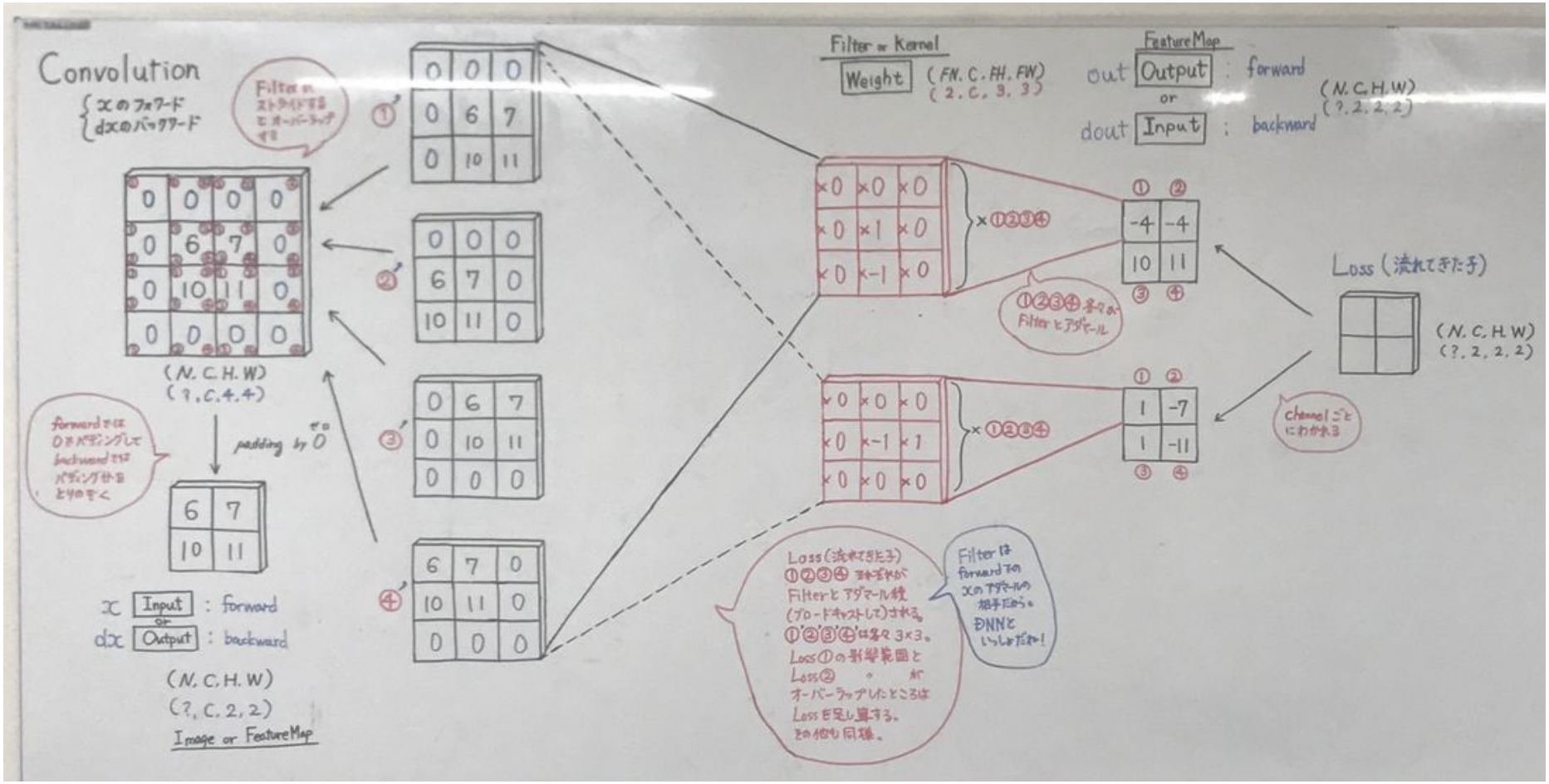
CNN2

im2colの擬似コード



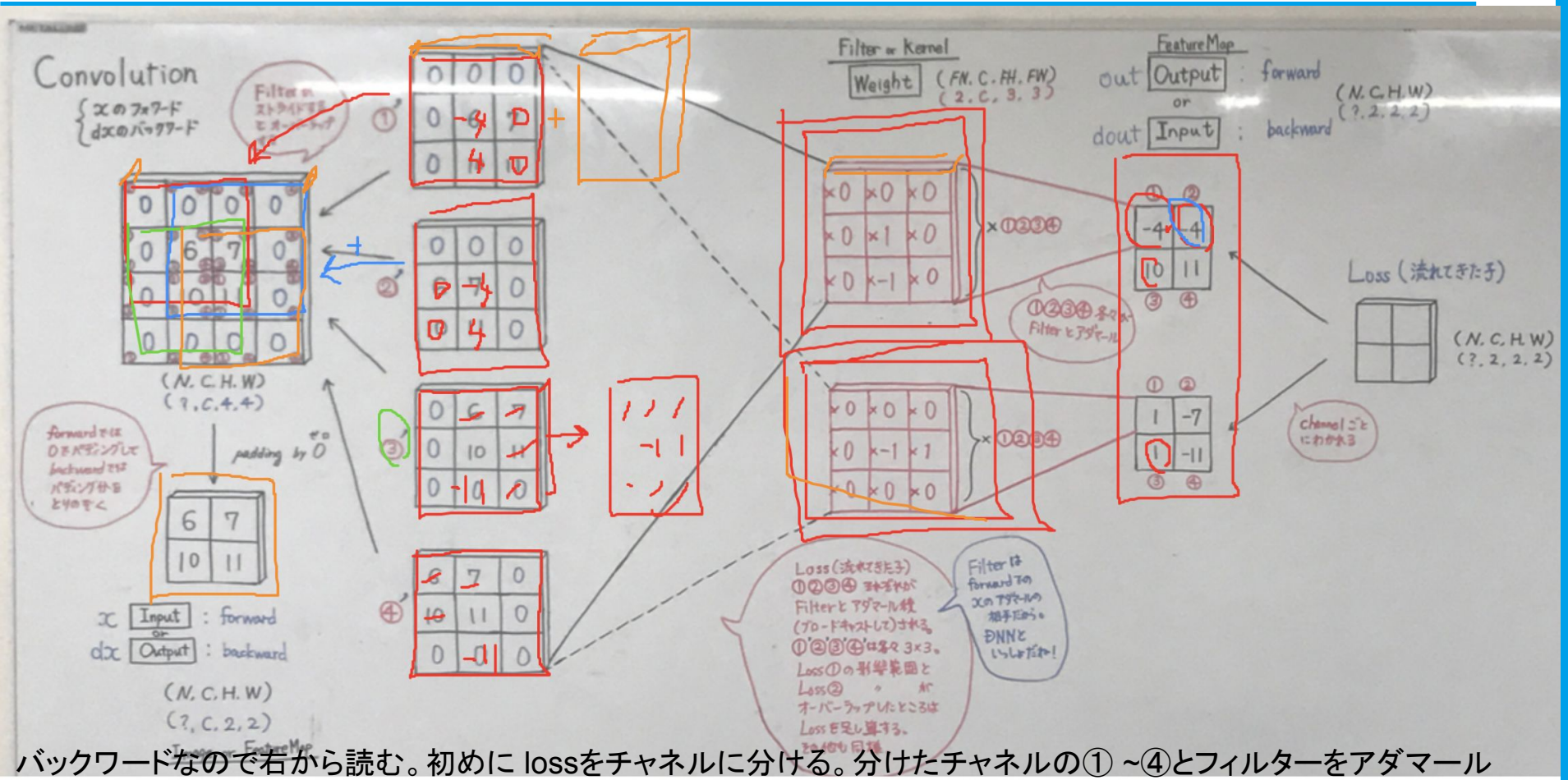
CNN2 - 2DConvの参考資料

今日のDiver Conv2Dのバックワード(右から左へ)



CNN2_コンボリューションNN
完

畳み込み層と全結合層の違い



バックワードなので右から読む。初めに lossをチャネルに分ける。分けたチャネルの① ~④とフィルターをアダマールして①'~④'を作成する。①'~④'をそれぞれチャネル方向で sumする。オフセットするので① '~④'を該当の場所に足していく。cnnの逆をすると周りにゼロが生まれるので省いた物がバックワードの最終地点。
ちなみに値はフォワードで書いてあるのでバックの時は数値が合わない。