

Deep Learning 輪読会 2017
第9章 畳み込みネットワーク

2017.11.27
東京大学
中村藤紀

第9章 畳み込みネットワーク

9.1 畳み込み処理

9.2 モチベーション

9.3 プーリング

9.4 無限に強い事前分布としての畳み込みとプーリング

9.5 基本的な畳み込み関数の変種

9.6 構造出力

9.7 データの種類

9.8 効率的な畳み込みアルゴリズム

9.9 ランダムあるいは教師なし特徴量

9.10 畳み込みネットワークの神経科学的基礎

9.11 畳み込みネットワークと深層学習の歴史

※ transposed convolution については扱われていない。

畳み込みネットワーク

- 畳み込みニューラルネットワーク
 - convolutional neural networks, CNNs
 - 格子状のトポロジを持つデータの処理に使われるニューラルネットワーク。
 - e.g. 時系列データ, 画像データ
 - 少なくとも1層で畳み込みを利用。

9.1 畳み込み処理

- 畳み込み (convolution)

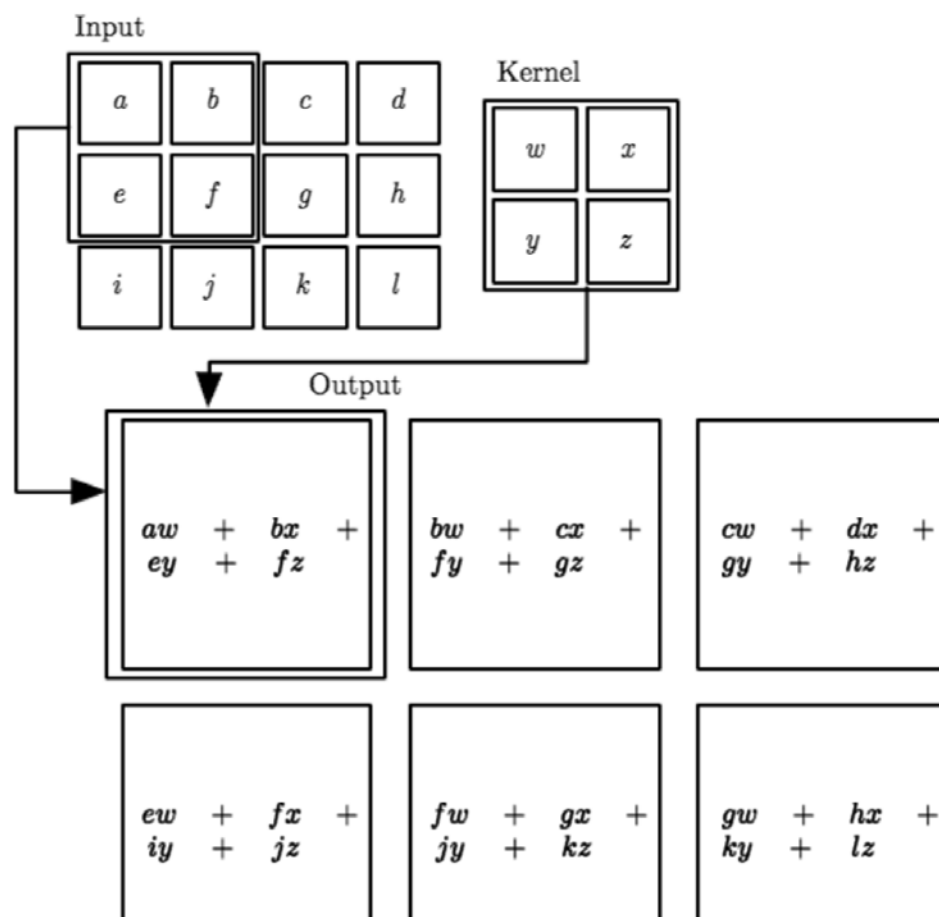
$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n). \quad (9.5)$$

- 相互相関 (cross-correlation)

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n). \quad (9.6)$$

– 本書ではどちらの処理も畳み込みと呼ぶ慣習に従う。

I : 入力
K : カーネル
出力 : 特徴マップ

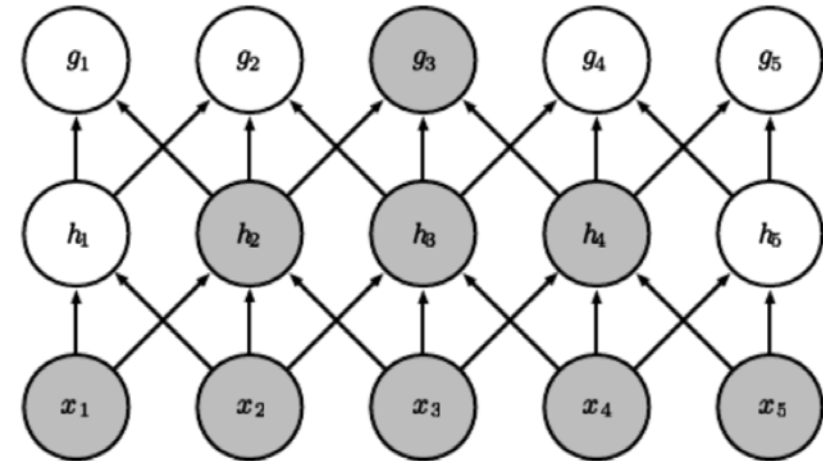
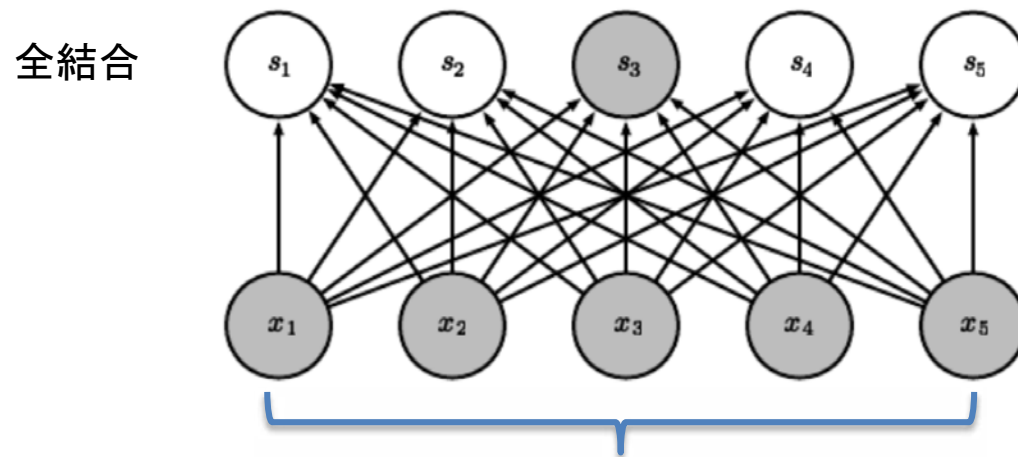
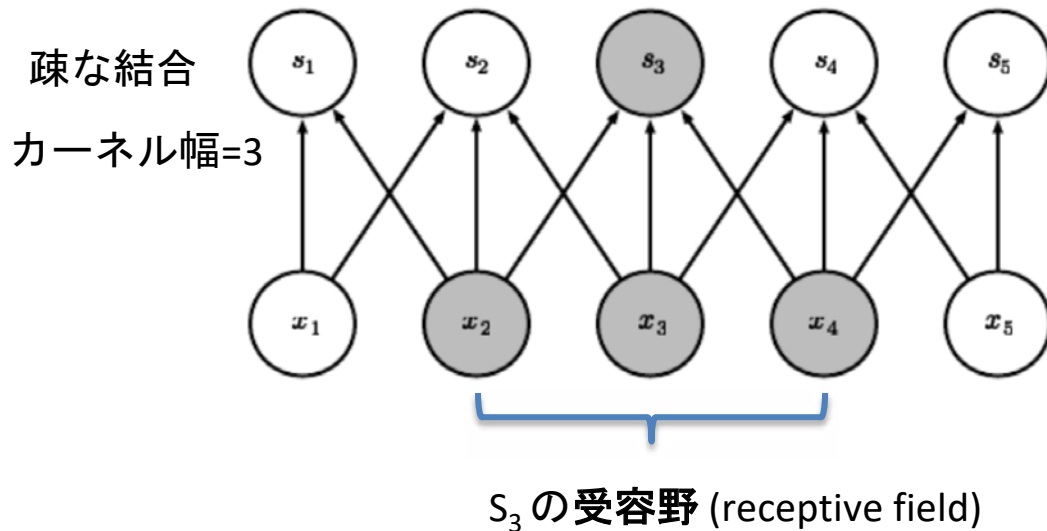


カーネル反転を行わない2次元畳み込みの例

9.2 モチベーション

1. 疎な結合 (sparse interactions)

- 保管するパラメータ数を削減、メモリ削減、統計的な効率性を改善。

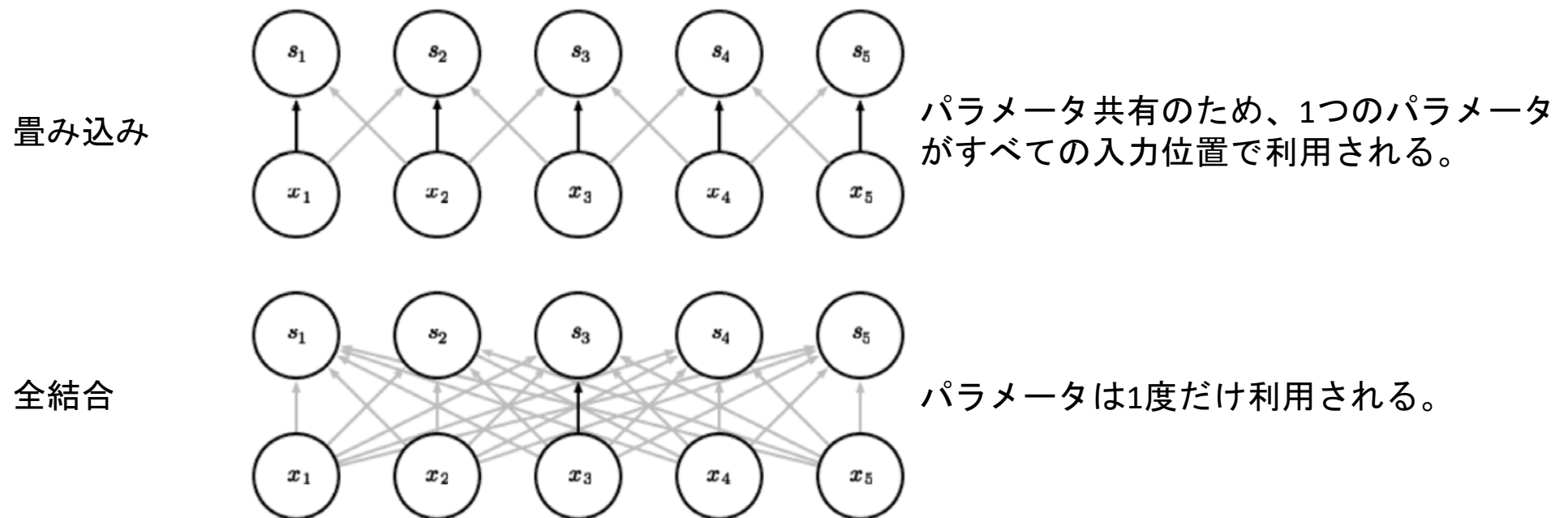


深い層のユニットの受容野の大きさ
> 浅い層のユニットの受容野の大きさ

深い層のユニットは、
入力のすべてあるいはほとんどと間接的に結合

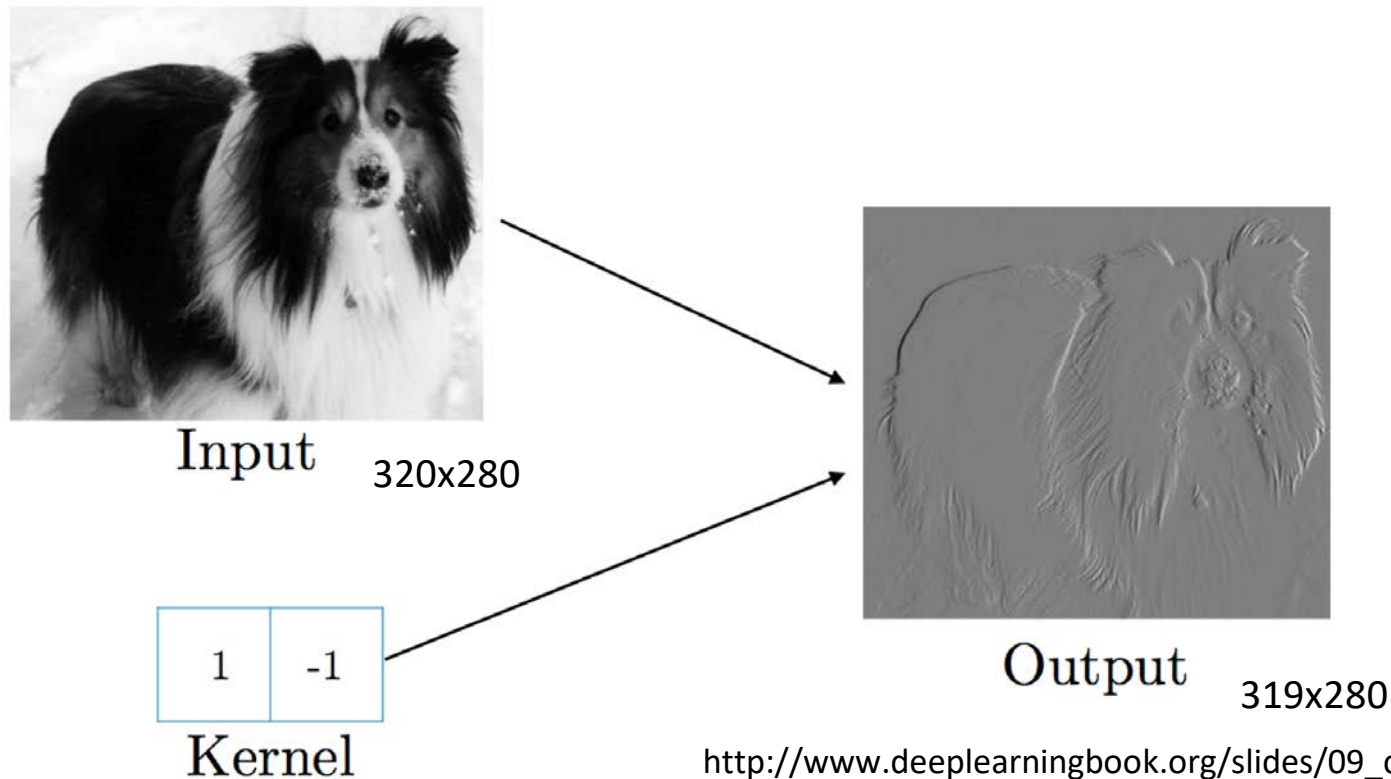
9.2 モチベーション

2. パラメータ共有 (parameter sharing)



9.2 モチベーション

- 疎な結合とパラメータ共有が画像中のエッジを検出する線形関数の効率を劇的に改善する例。



	畳み込み	全結合	畳み込みによる効率性
要素数	2	$319 \times 280 \times 320 \times 280 > 80\text{億}$	1 / 40億
演算数 (乗算, 加算)	$319 \times 280 \times 3 = 267,960$	$> 160\text{億}$	1 / 6万

9.2 モチベーション

3. 等価表現 (equivariant representation)

- 移動に対する**等価性**

- ある関数が等価 = 入力に変化した場合に出力が同様に変化する。

- $f(g(x)) = g(f(x))$ ならば f と g は等価

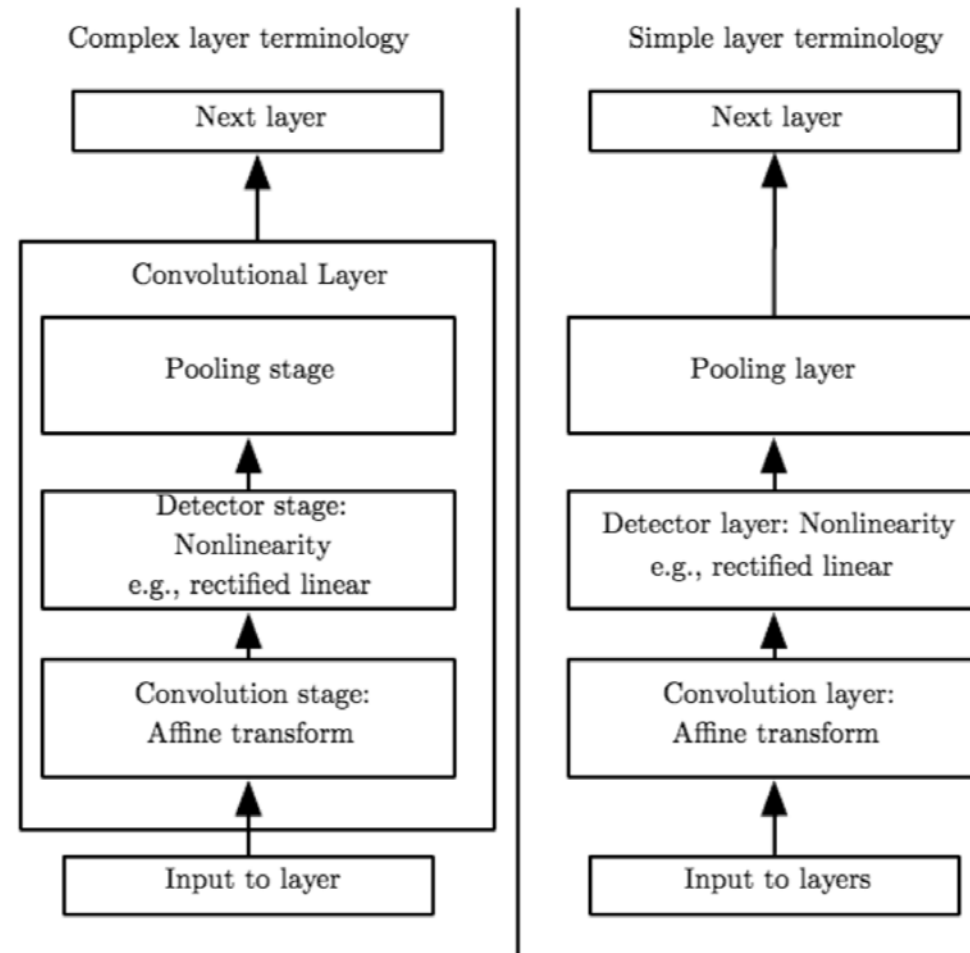
- e.g. f = 畳み込み, g = 入力をシフトする関数

- 入力中の物体を移動 → 出力の表現は同じ分だけ移動。

- 畳み込みは、画像のスケーリングや回転といった変換に対しては等価ではない。

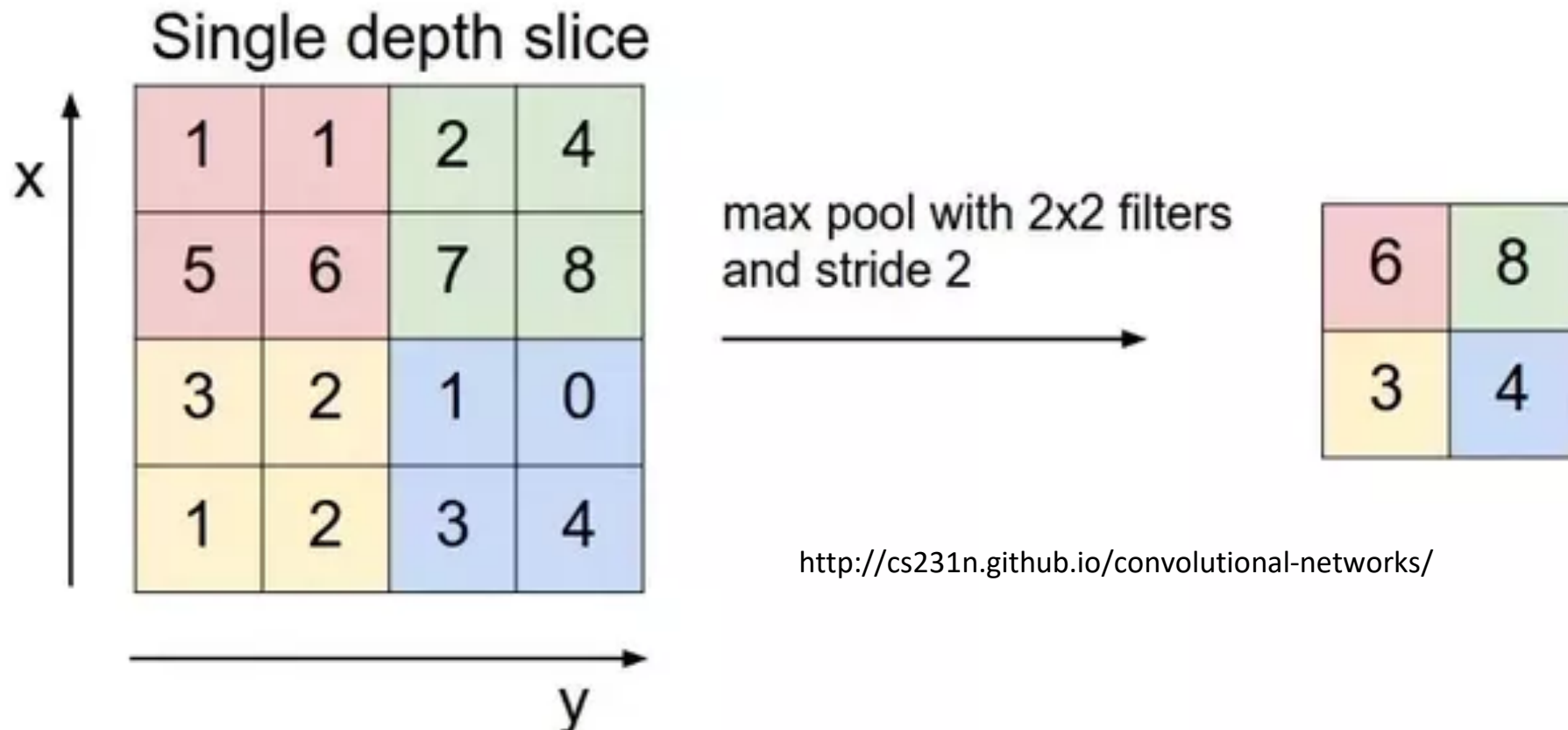
9.3 プーリング

- 典型的な CNN のアーキテクチャ
 - Conv -> ReLU -> Pool



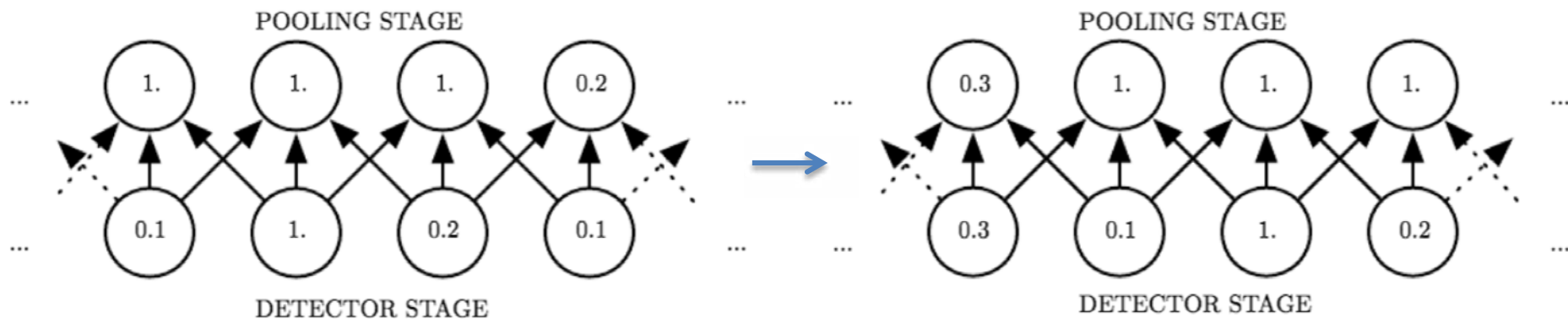
9.3 プーリング

- **プーリング関数** (pooling function)
 - ある場所でのネットワークの出力を、周辺の出力の要約統計量で置き換える。
 - **max pooling** = 矩形の近傍中で最大出力を返す。



9.3 プーリング

- 移動に対する不変性
 - 入力を多少変化させたとしても、プーリング後のほとんどの出力値は変化しない。
 - 画像中のある特徴の厳密な位置より、画像中にその特徴が存在するかどうかに関心がある場合、局所的な移動に対する普遍性は有用。



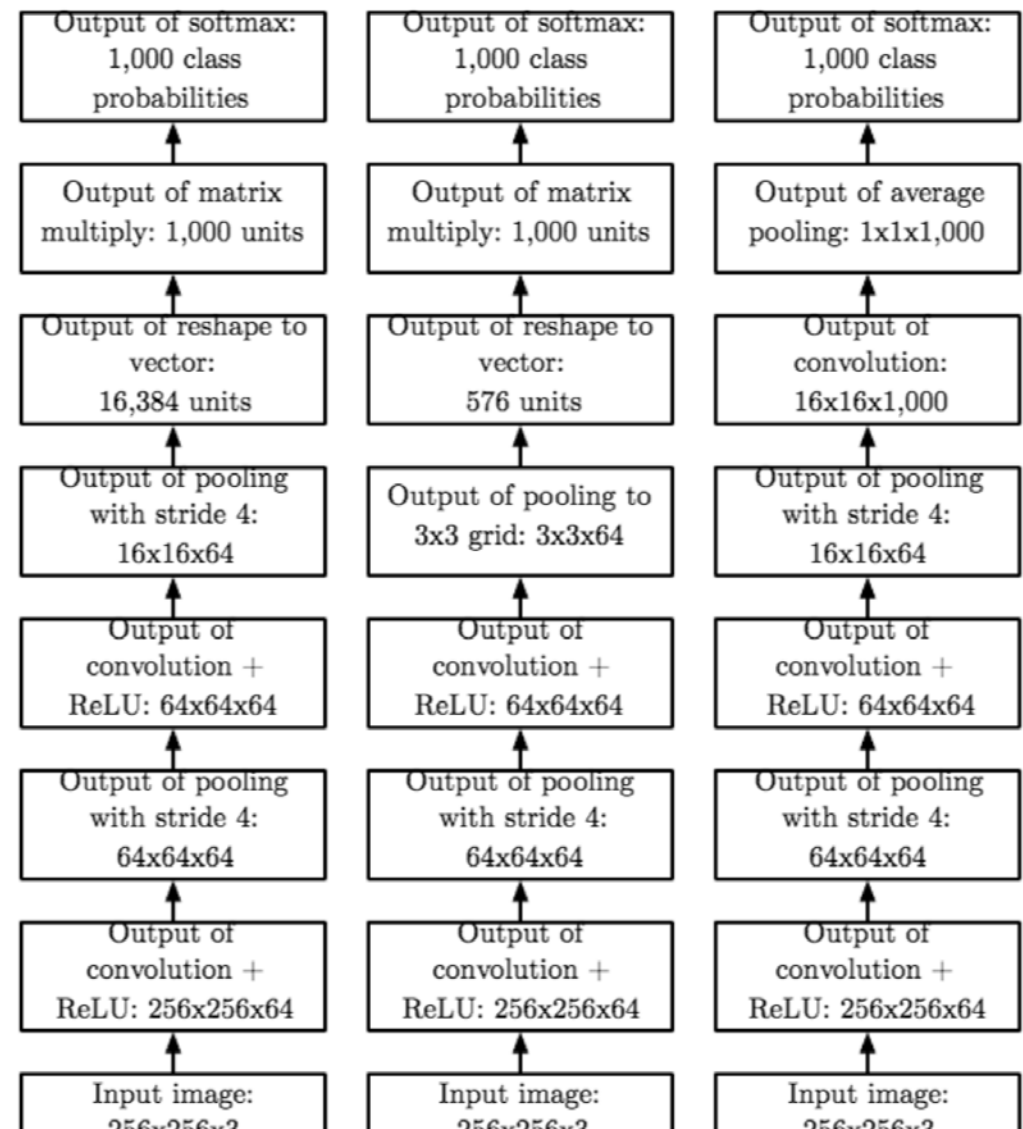
下部の値 (非線形処理の出力) がすべて変化しても、
上部のプーリング層の出力は半分の値だけが変化している。

- パラメータ数の削減。
- 可変サイズの入力を扱うことができる。

9.3 プーリング

- CNN を用いて分類を行うアーキテクチャの例

- 左 : 固定サイズの入力
- 中央 : 可変サイズの入力→全結合
- 右 : 全結合層を含まない CNN

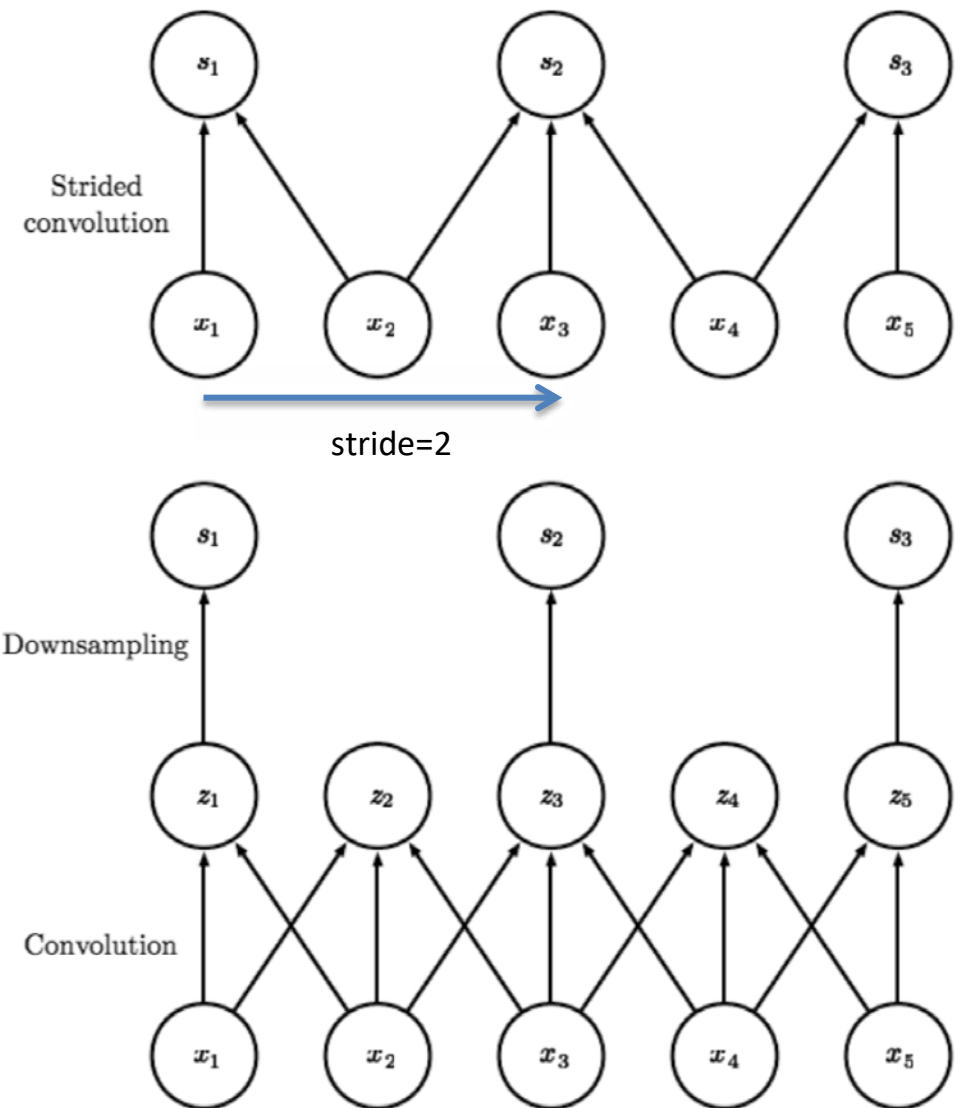


9.4 無限に強い事前分布としての畳み込みとプーリング

- モデルパラメータの事前分布上で、確率密度がどの程度密集しているかに応じて、事前分布の強さや弱さを考える。
 - 弱い事前分布 (e.g. 大きな分散を持つガウス分布, エントロピー大)
 - 強い事前分布 (e.g. 小さな分散を持つガウス分布, エントロピー小)
 - 無限に強い事前分布: パラメータがいくつかの値になる確率をゼロにして、完全に使用禁止とする事前分布。
- CNN を重みに対して無限に強い事前分布を持つ全結合ネットワークである
と考える。
 - 過少適合を引き起こす可能性がある。
 - 離れた場所の情報を入力に組み込む必要がある場合、畳み込みに基づく事前分布は不適切かもしれない。

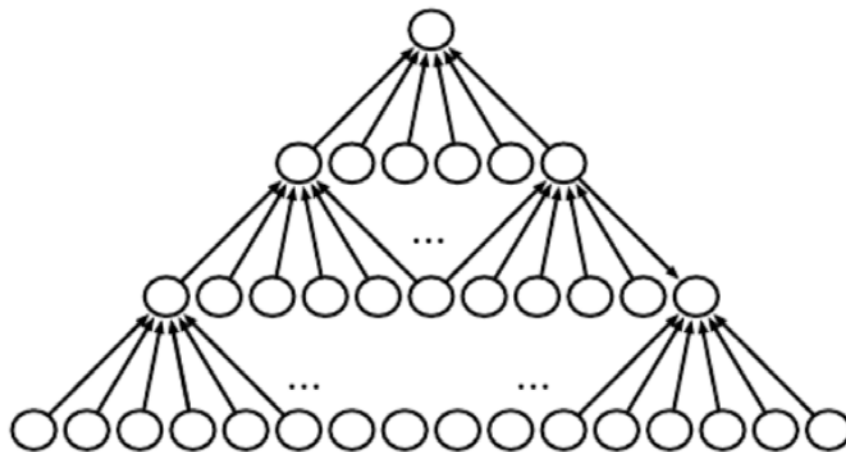
9.5 基本的な畳み込み関数の変種

- ストライド (stride) を利用した畳み込み
 - 精緻な特徴の獲得を犠牲にして、
計算コストを削減。
- = 完全な畳み込み関数の出力を
ダウンサンプリング



9.5 基本的な畳み込み関数の変種

- ゼロパディング (zero padding)
 - 入力 \mathbf{V} の幅を広げる。valid, same, full



valid: ゼロパディングを利用しない
各層で表現が縮小している。

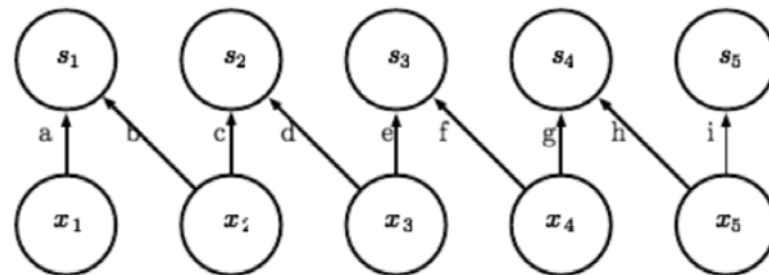


same: ゼロパディングを利用
深層において表現が縮小することを回避。

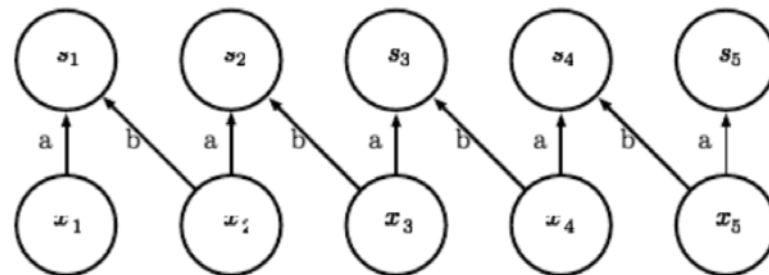
9.5 基本的な畳み込み関数の変種

- 局所結合層 (locally connected layers)
 - a.k.a. **非共有畳み込み** (unshared convolution)
 - 各特徴が空間の小さな一部分の関数であるが、同じ特徴が空間全体で発生すると考える理由がない場合 (e.g. 顔画像)

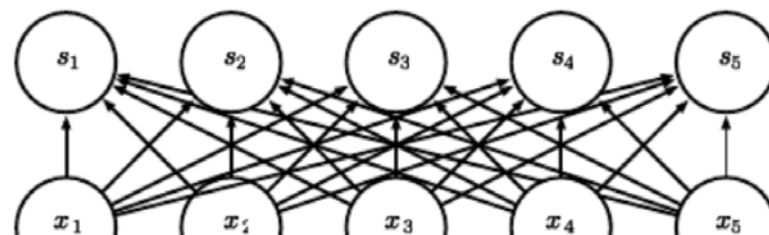
局所結合, パッチサイズ=2ピクセル
- 結合は局所的
- 重みは共有されない
(異なるアルファベットで表現)



畳み込み, カーネル幅=2ピクセル

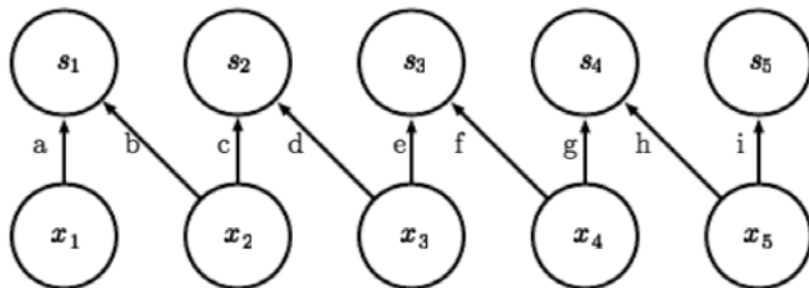


全結合

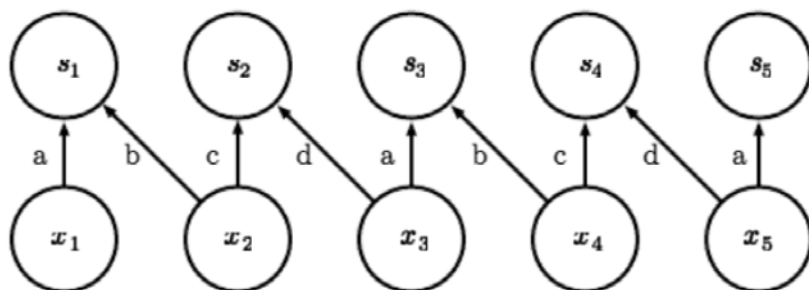


9.5 基本的な畳み込み関数の変種

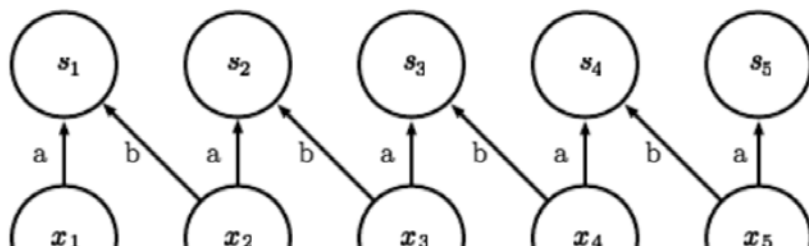
- **タイル型畳み込み** (tiled convolution)
 - 空間上のすぐ近傍の場所のみ、局所結合層のように異なるフィルタを持つ。
 - t 種類の異なるカーネル



局所結合



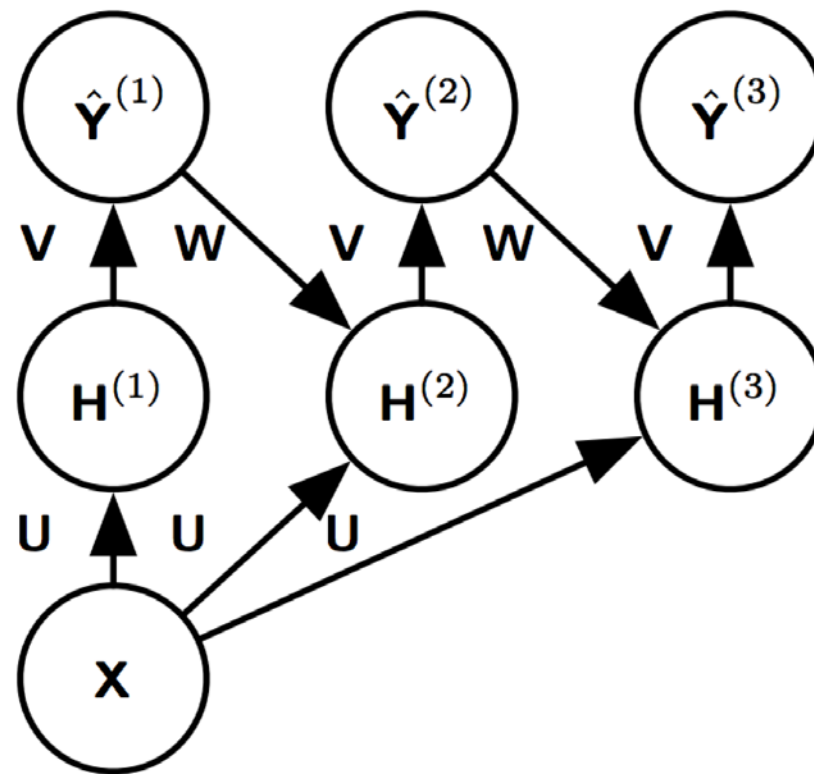
タイル型畳み込み, 2種類のカーネル
- 数種類の重みが共有される。



畳み込み

9.6 構造出力

- CNN は分類タスク、回帰タスク以外にも用いられる。
 - segmentation
 - 出力テンソルは (画像の行, 画像の列, ピクセルが各クラスに属する確率)



回帰結合型畳み込みネットワークの構造
各ステップで同じパラメータが利用される。

、4.6章 回帰結合型畳み込みネットワークの構造

9.7 データの種類

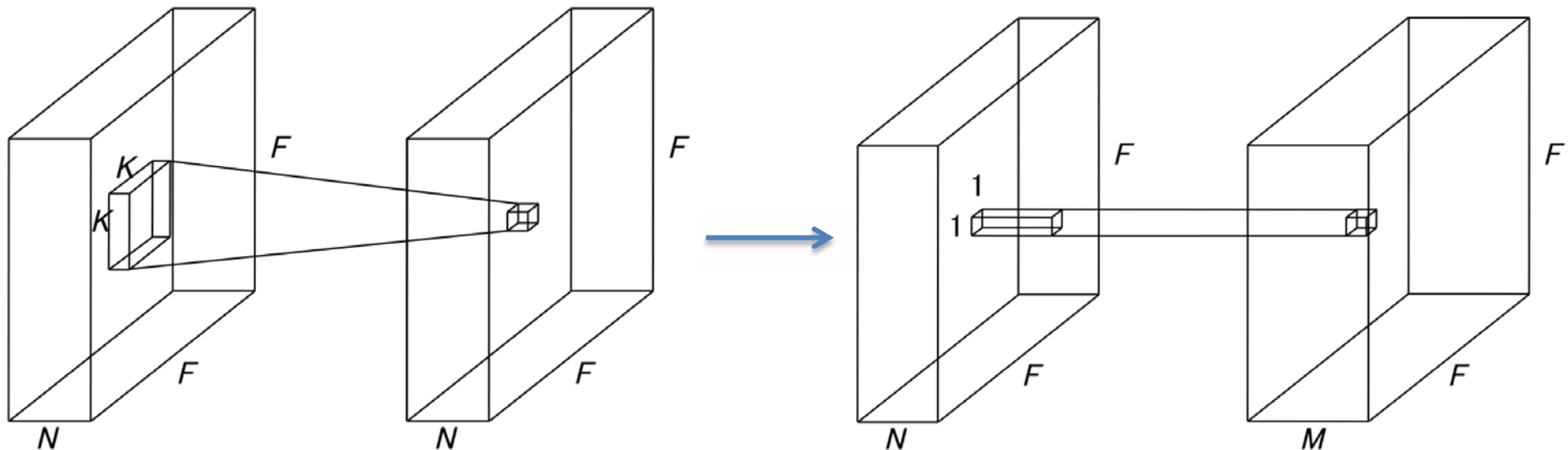
- 次元とチャンネル数に基づくデータの種類

	単一チャンネル	複数チャンネル
1次元	音声波形	スケルトンのアニメーションデータ
2次元	フーリエ変換で事前処理された音声データ	カラー画像データ
3次元	体積データ	カラー動画データ

- CNN の利点: 空間的大きさの異なる入力も処理できる。
 - MLP は異なる大きさの入力に対応できない。
 - NLP のタスクでは入力 (= 文章) が可変長のため、入力サイズに比例したプーリングが有用。

9.8 効率的な畳み込みアルゴリズム

- separable convolution
 - パラメータ数の削減。
 - Xception, MobileNet
 - Depthwise convolution → Pointwise convolution



チャンネルごとに空間方向の畳み込みをした後、チャンネル方向の畳み込みを行い、目的の出力テンソルを得る (<https://qiita.com/yu4u/items/34cd33b944d8bdca142d>)

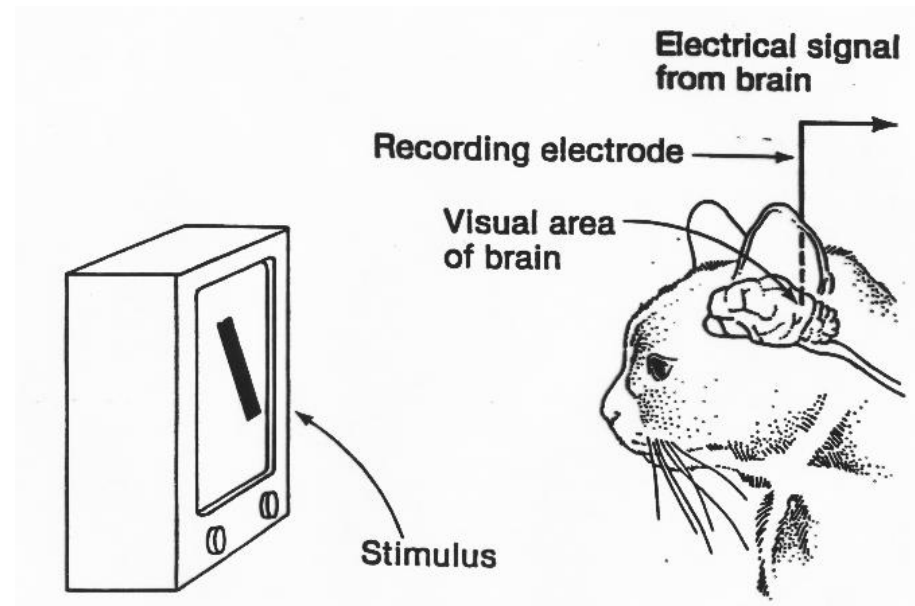
- モデルの精度を損なうことなく畳み込み処理を高速化する方法の開発は、活発な研究領域。

9.9 ランダムあるいは教師なし特徴量

- CNN の特徴量の学習は訓練のコストが高い。
- 教師あり学習を利用せずに畳み込みカーネルを求める。
 1. handcraft な特徴量の設計
 2. ランダムに初期化されたフィルタ
 3. 教師なし学習によりカーネルを学習
 - layer-wise な貪欲事前学習
 - e.g. 深層畳み込み信念ネットワーク (convolutional deep belief net)

9.10 畳み込みネットワークの神経科学的基礎

- Hubel と Wiesel による実験
- CNN の設計が捉える V1 (1次視覚皮質) の特徴
 1. 特徴マップ → V1 は網膜上の画像の構成を反映する2次元構造
 2. 検出ユニット → V1 における **単純細胞** (simple cells)
 3. プーリング → V1 における **複雑細胞** (complex cells)→ “Deep” にする = **おばあさん細胞** (@内側側頭葉)
- CNN の最終層 ≈ 下側頭葉 (IT)
 - CNN が IT の発火頻度を予測することができる



9.10 畳み込みネットワークの神経科学的基礎

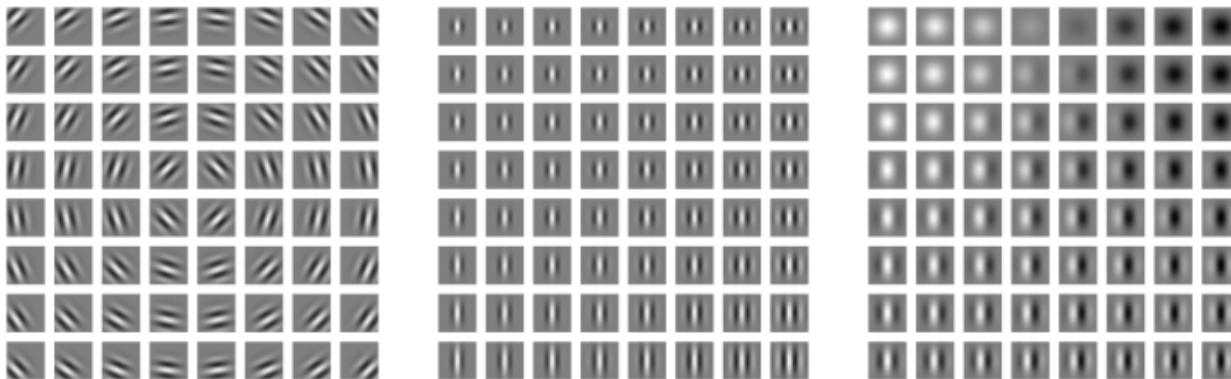
CNN と哺乳類の視覚システムとの差異

- 人間の視覚システムは、
 - 中心窩 (fovea) : 中心窩以外では低解像度。
 - 聴覚のような他の多くの感覚等と統合されている。
 - 単に物体を認識する以上のことをおこなう。
 - 高次領域から V1 へのフィードバック。
 - IT の発火頻度は CNN とほぼ同じ情報を捉えているものの、中間の計算の類似性は不明。
- 神経科学からは CNN の訓練方法については、ほとんど知見が得られていない。
- FYI
 - PredNet: 新皮質の局所回路を定性的によく反映していると思われる ANN モデル。
 - <https://wba-meetup.doorkeeper.jp/events/46611>

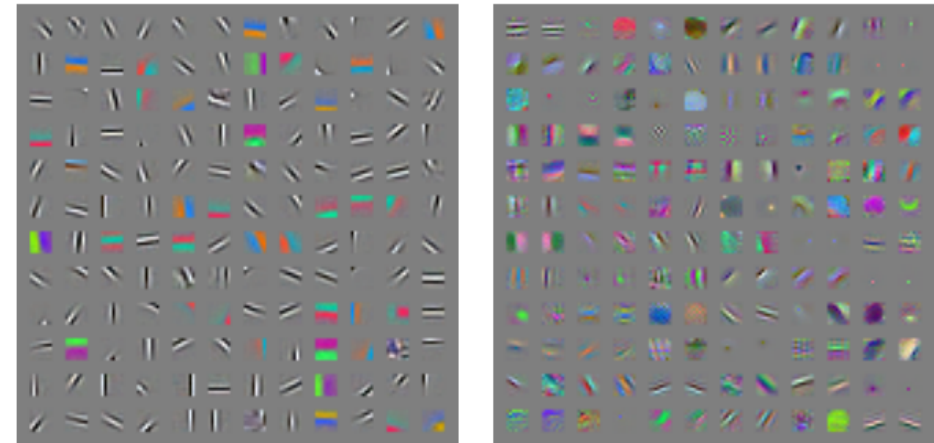
9.10 畳み込みネットワークの神経科学的基礎

- **逆相関法** (reverse correlation)
 - 視覚刺激に対するニューロンの活性化を、取り付けた電極で記録。
 - ニューロンの重みの近似を求める。
 - V1 にある細胞の持つ重み = **ガボール関数** (Gabor functions)
- 機械学習モデルが学習した特徴と V1 で使われる特徴は視覚的に類似。
 - 逆に、こうした特徴が得られていないなら、そのモデル (e.g. CNN) は学習できていないと考えられる。

ガボール関数



学習された特徴



9.11 畳み込みネットワークと深層学習の歴史

- CNN は、脳の研究から得られた洞察を機械学習に応用した重要な成功事例。
- また、商用利用の問題を解決した最初のニューラルネットワーク。
 - 小切手読み取り, OCR, 手書き文字認識, etc
- ImageNet Large Scale Visual Recognition Competition (ILSVRC)
 - AlexNet (Krizhevsky et al., 2012)
 - 商用利用での強い関心。そして今に至る。
- CNN は誤差逆伝播法で訓練して機能した初期の深層ネットワークの1つ。
 - CNN が ニューラルネットワークが一般的に受け入れられる道を切り開いたと言える。
 - Deep Learning の代名詞的存在 (とも言える)

参考文献

- Deep Learning
 - Ian Goodfellow, Yoshua Bengio, Aaron Courville
 - 日本語版
<https://www.amazon.co.jp/%E6%B7%B1%E5%B1%A4%E5%AD%A6%E7%BF%92-Ian-Goodfellow/dp/4048930621>
- CS231n Convolutional Neural Networks for Visual Recognition
 - Spring 2017 の講義では本章で扱われなかった transposed convolution の解説もある。