

ADL—CNN

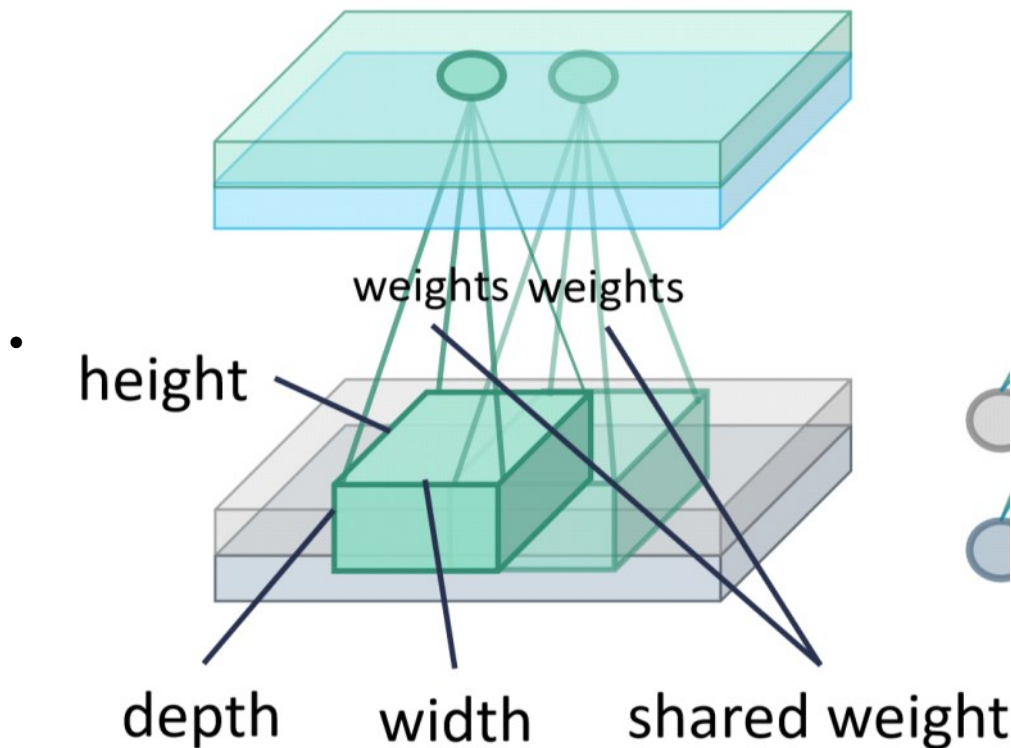
2019年4月30日 上午 11:09

特性

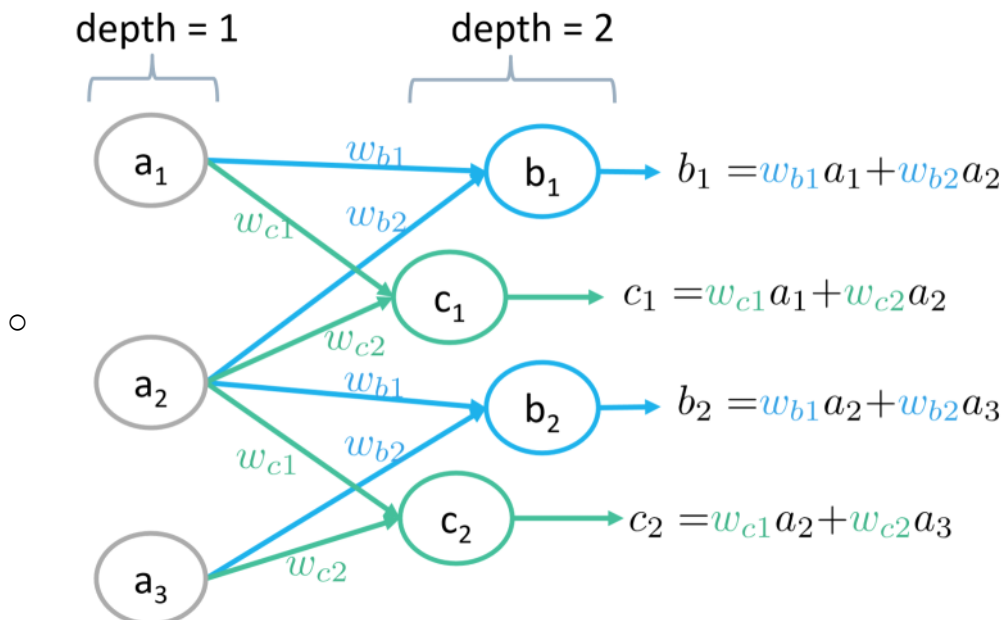
- 特徵位置
 - 有一些特徵不用看整張圖
 - 相連在一起的region
 - pixel跟pixel之間連接的特性，不需要每個pixel都一個neuron (減少參數)
- 同樣pattern可能會出現在不同位置
 - 既然是要做一樣的事情detect同一件事情，所以應該要可以shared weight
 - 不同位置處理相同特徵
- subsampling壓縮不會改變圖，原本的鳥還是鳥 (四個pixel只取左上角)
 - 變小的image代表input進來的參數比較少，要學的參數就會比較少
 - 希望訓練過程的參數可以少一點
- local connectivity
 - 希望可以考慮原本input的位置關係，要跟旁邊region關聯
 - 把pixel放進去
 - 文字間的context放進去，為了理解中間字，會把周圍的字放進去
 - 希望進入neuron是有小塊相連，把整個小區的input都放進去
 - 把數個small region input進去
- Parameter sharing
 - 同一個位置會有不同組參數抽取不同feature
 - 相同feature會有相同參數在不同位置去抽取

Conv layer

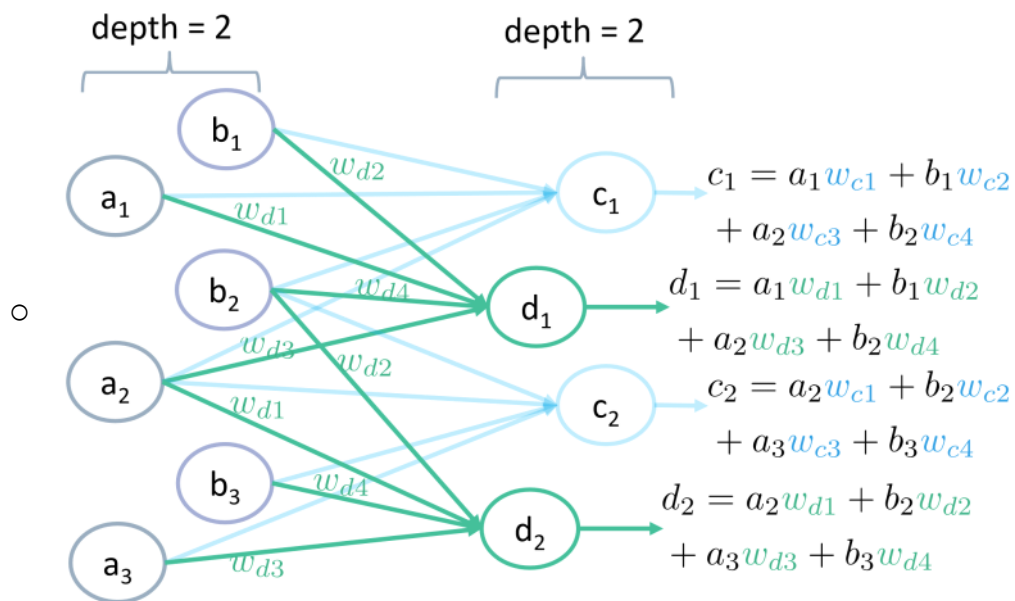
- 把抽取的feature匯集到某一個位置上



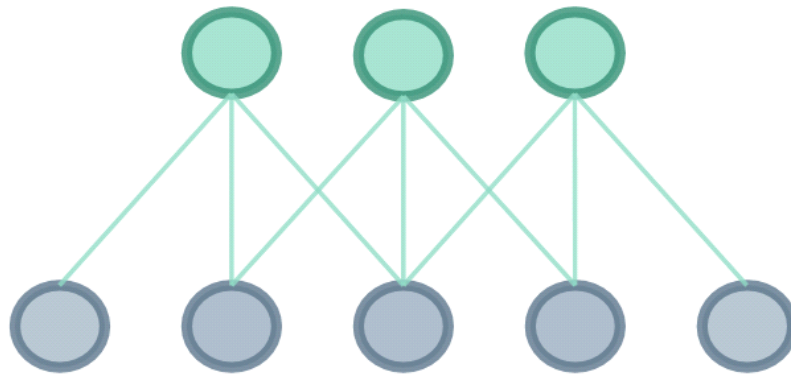
- 抽feature也會把不同的height深度資訊拉進來
- b1、b2都要抽取鳥嘴特徵，只是想看不同地方的位置，所以既然他們要抽取一樣特徵，則weight應該要一樣:
 - input進來的會是不一樣的region
 - c則是另外一個深度的資訊，要做的事情跟b不一樣，所以會有c自己的weight，雖然input都是從a來，但看的重點不同



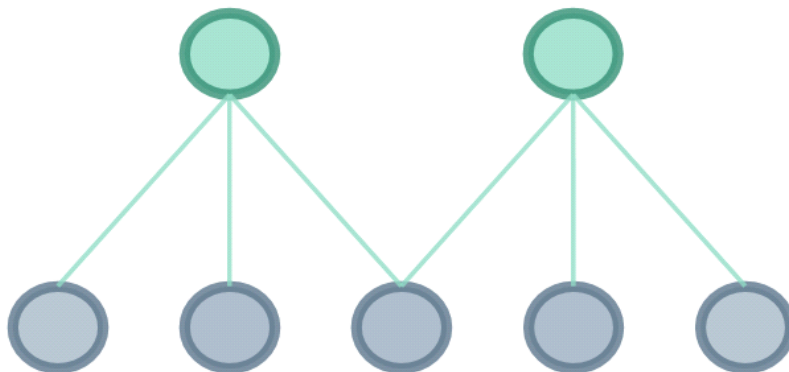
- 同樣位置不同feature，同樣feature不同位置
- b可能是抽取鳥嘴，a可能辨識出羽毛，把小feature組合以後丟給c跟d，從不同feature來越後面level月高階，所以c跟d都可以detect不同位置的鳥嘴根羽毛



- Stride
 - 下一個neuron要跳幾格



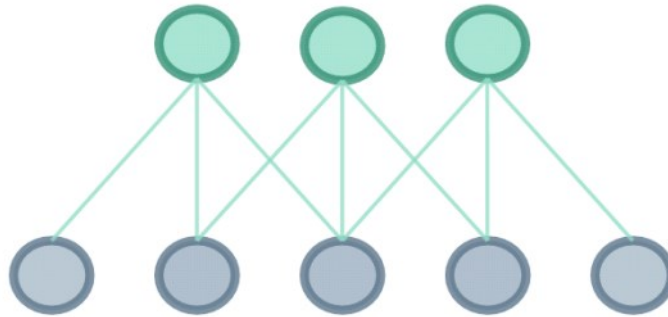
Stride = 1



Stride = 2

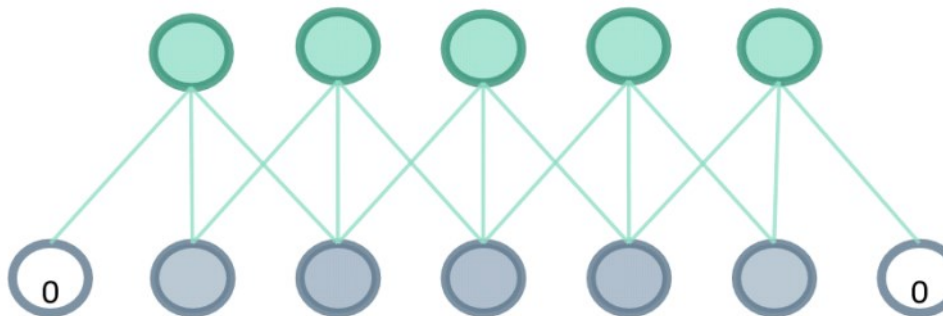
- 圖的細節多，上面比較好，但如果不要學那麼多參數疊那麼多 layer 可以下面就好了

- stride越小layer要越多
- Padding:
 - 上面沒pad，下面有
 - padding是希望input維度要可以控制或是一樣的，控制dim
 - 主要是try error，看方便為主



Padding = 0

○



Padding = 1

Relationship with conv

- 跟signal processing有關係，不同的input會有不一樣的weight往後移動，去看pattern

Nonlinarity (activation function)

- ReLU沒有上限輸入
 - sigmoid如果疊很多次，都要一直壓到1裡面，越多層就會變成越小的數值，這樣signal會很難propagate到最後做出貢獻
 - sigmoid到後面就會飽和了，很難調整到最前面的input
- ReLU可以對輸出造成影響
 - 疊很多層是比較適合的
 - 希望可以都大於0，可以對後面做出影響
 - ReLU高的值保留刺激，比較容易對output造成影響
- 避免gradient vanishing

Pooling Layer

- max: 四格裡面選最大的
- avg: 四格裡面取平均

- 沒有東西要訓練(no weights)，region no overlap
- 把某這塊feature貢獻最高的保存下來
- EX: 怎麼辨識5或7，兩個組合起來會是某一個數值，希望保留最重要的feature當成下一層input，最一開始把最基本feature抽出來，月high level需要看的東西只要比較少就好了
- 抽出pooling的資訊再去conv，所以前面線條，後面形狀，越後面越具有semantic資訊

FC + softmax:

- 把pooling資訊壓平，餵進去DNN過softmax，end2end訓練
- 前面conv只是去抽取重要特徵(local feature)，DNN是要去學特徵怎麼組合(global aggregate)

Update:

- max pooling: BP只會對max的那個path的weight造成影響。其他不更新
- avg pooling: BP會大家的 BP weight平分gradient

=

ImageNet致敬WordNet

VGG

- filter小一點可以抽一些細緻資訊，performance比較好，所以前幾層的filter要小
- 大filter要aggregate的資訊很多可能學不好，小filter的performance比較好

GoogLeNet

- Inception module
 - 平行的做，一邊做1*1、3*3、5*5再去做pooling
 - 把各類型不同aggregate的資訊合在一起做
 - 多多利用1*1 convolution先讓dim變小，不看local其他的資訊
 - 長寬都是1，指示去做dimension reduction，不同類型feature稍微互相組合，但沒有考慮local資訊
 - 所以可以先做完1*1再去做3*3或5*5

ResNet-152 (Microsoft)

- 非常多層，設計residual function (skip-connection)
 - 經過太多層資訊會loss，好feature會被後面沒用的feature影響，所以讓他可以彈性決定要否保留
 - 會有兩種資訊輸出去

Open Image Extended - Crowdsourcing (google)

- 有標上某個tag的某些圖，非常複雜背景

Why pretrained model:

- imageNet太大，自己train很麻煩，期望用人家的pretrained可以當成representation，就不用管image 處理
- 如果資料跟image Net很像，可以transfer learning，可以把中間的conv都

fix住，去fine-tune FC layer

- 如果data跟image Net不像，只會fix最前面幾層的基本layer，因為high level feature可能會不適用在情境，basic feature如線條可以沿用，所需資料也不用太多，大概600~1000張就足夠了

Visualize CNN Features (看每一層CNN學到甚麼，可以知道接幾層比較好，重要特徵在哪裡)

- 第一層可能是邊
 - 中間可能是材質
 - 後面可能是object part
 - 最後可能是class資訊
- 調整輸入的noise要可以接近filter response，做gradient ascend
 - 調整input可以跟output接近
 - 把weight固定下來
 - 希望filter跟某個花比較像，希望某些feature數值比較高
 - 可以顯現一些比較重要的feature，改變上升的數值
 - 從noise變成類似花的圖案
 - 不同層的response filter不一樣
 - 越後面層資訊越抽象
 - multiscale (resize=>visualize=>...)
 - 網站: deepdreamgenerator，加上machine看到的，因為CNN是image訓練的，所以會對這些東西重要特徵產生反應，調整怎樣會比較高的response

CNN as Artist (現在都用GAN):

- 希望encode content保留，希望encode style保留下來
- 藉由修改input產生loss，希望activate neuron
- 希望最後產生的內容可以讓activate 的feature接近，所以minimize loss
- 改變一張圖讓他某一個接近
- 為了不失敗，所以會在原本的圖加上noise，再去做
- style
 - 每張圖的各個部分都可以裝作是style的一部分
 - 希望可以學習怎麼樣叫做style
 - 遮不同地方，希望出來的matrix都一樣
 - 畫布經過一連串就可以跟原本的gram matrix一樣
 - 內容不管
- 合在一起
 - 希望中間的content跟原本接近，gram matrix跟原本接近，可以依照比例調整

Why CNN for Go?

- 有些pattern看local的region就夠了

- 同樣pattern會出現在不同位置，同樣參數 (5*5去捕捉local region feature)
- pooling? =subsampling，會讓圍棋盤面跟原本不同，影像做subsampling還是影像，但是圍棋盤面都保留左上角很怪，不符合模型假設
- 5*5 filter，zero padding，stride=1，relu，filter 3*3 stride=1
- 要考慮每一層對input的假設才可以拿進來使用

CNN Sentence encoding:

- 希望可以找出local重要的，找出重要向量
- 語言上會用1D conv，只考慮左右
- 好幾層的1D conv weight，組合不同level的信息
- max pooling，文字上的假設，有些字比較重要，有些字不重要，文法結構比較弱的關係不重要，stop word可能可以拿掉，可以變成對句子結構的理解
- 學習句子組合關係
- sentiment analysis，可以考慮負負得正的情況，最後才會知道是怎樣組合的結構
 - 經過max pooling希望weight可以不同高
 - 同一層可以用不同類型不同size的傳上去
- 句子比較難resize，影像可以四個挑一個，但是文字不一定可以兩個字挑一個字
 - 希望不管多少長度都有相同vector，但如果長度不一會不知道怎樣得到相同size vector
 - 如果長度不一樣，最後output vector也會不一樣
 - 處理不同長度? padding?
 - 透過pooling到相同長度
 - K-max pooling，不管怎樣最後都可以得到相同的K維度
 - wide convolution & K-max pooling
- 利用不同size conv

=

acGAN的clf，會跟D shared weight

WGAN -GP

spectral norm

衡量embed space

FID score

inception score