Глубокое обучение в компьютерном зрении

лекции - Иван Загайнов

семинары - Борис Зимка

1. Интро.

Что такое backpropagation и зачем он нужен.

Тренировка нейросетей. Back-propagation

Нейросеть – это дифференцируемая сложная функция, зависящая от параметров:

$y = f(\vec{x}, W_x)$

Тренировка нейросетей. Back-propagation

Нейросеть – это дифференцируемая сложная функция, зависящая от параметров:

$$y = f(\vec{x}, W_x) = f(g_x(\vec{v}, W_v), W_x)$$

Тренировка нейросетей. Back-propagation

Нейросеть – это дифференцируемая сложная функция, зависящая от параметров:

$$y = f(\vec{x}, W_x) = f(g_x(\vec{v}, W_v), W_x) = f(g_x(h_v(\vec{z}, W_z), W_v), W_x) = \cdots = M(I, \Theta)$$

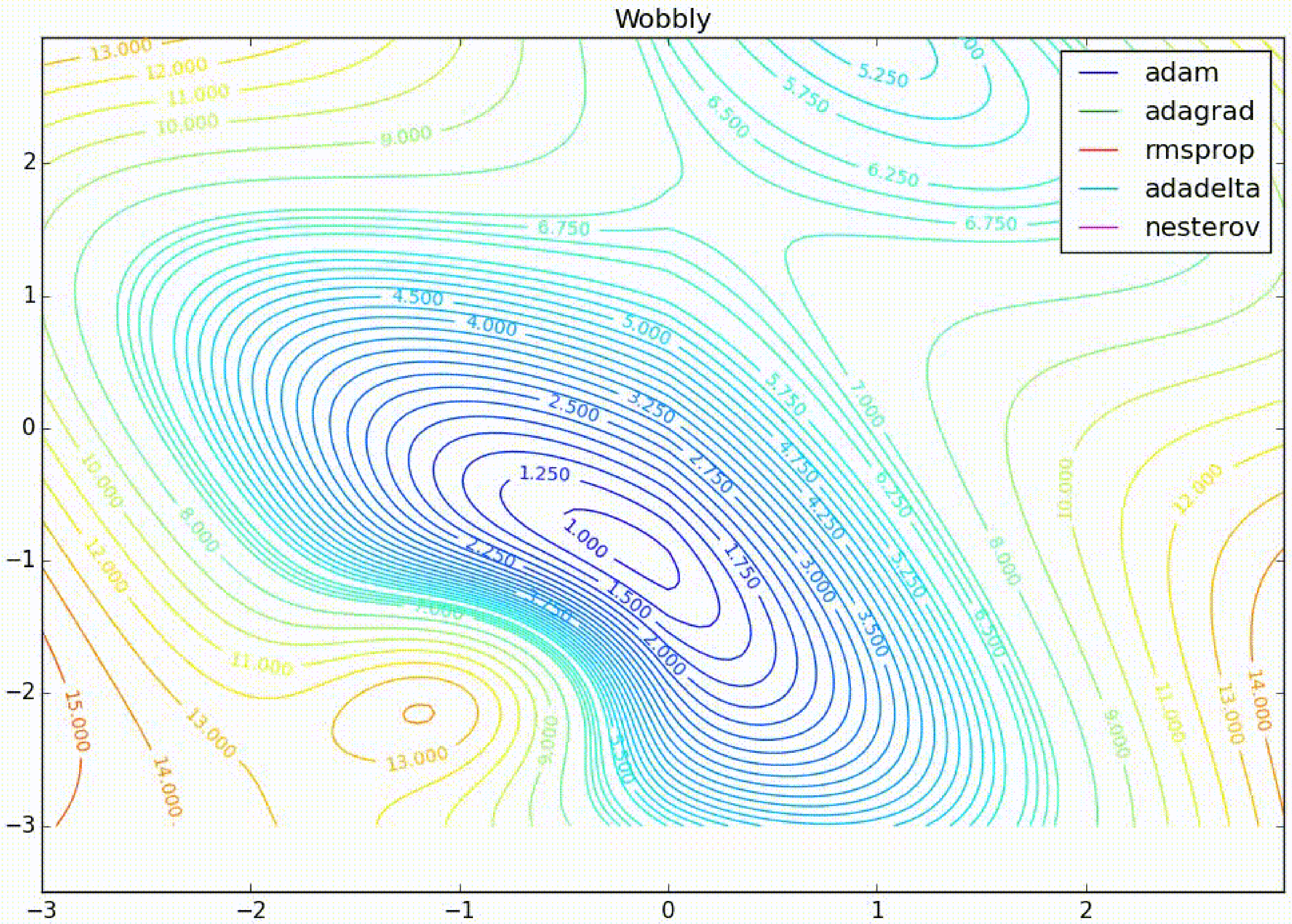
Тренировка нейросетей. Back-propagation

Нейросеть – это дифференцируемая сложная функция, зависящая от параметров:

$$y = f(\vec{x}, W_x) = f\big(g_x(\vec{y}, W_y), W_x\big) = f\big(g_x\big(h_y(\vec{z}, W_z), W_y\big), W_x\big) = \cdots = M(I, \Theta)$$

Ищем такие веса, что на данных будет минимальна ошибка L:

$$\Theta = argmin\, E_I\big[L\big(y_{true}, M(I, \Theta)\big)\big]$$



Градиентный спуск по L(w) для разных оптимизаторов (habr)

Тренировка нейросетей. Back-propagation

Нейросеть – это дифференцируемая сложная функция, зависящая от параметров:

$$y = f(\vec{x}, W_x) = f\left(g_x(\vec{y}, W_y), W_x\right) = f\left(g_x\left(h_y(\vec{z}, W_z), W_y\right), W_x\right) = \cdots = M(I, \Theta)$$

Ищем такие веса, что на данных будет минимальна ошибка L:

$$\Theta = argmin\, E_I[L(y_{true}, M(I, \Theta)]$$

Для этого используем градиентный спуск:

$$\Theta_{\mathrm{i+1}} = \Theta_{\mathrm{i}} - \alpha \frac{\partial L}{\partial \Theta}$$

Где взять градиент?

Производная сложной функции

$$L(y_1, y_2) = y_1^2 + y_2$$
$$y_1(\, x_1, x_2, x_3) = x_2 \cdot x_3$$
$$y_2(\, x_1, x_2, x_3) = x_1 + x_2 + x_3 + w$$

$$L(y_1, y_2) = L(y_1(\vec{x}), y_2(\vec{x})) = L(\vec{x}) = (x_2 \cdot x_3)^2 + (x_1 + x_2 + x_3 + w)$$

Производная сложной функции

$$L(y_1, y_2) = y_1^2 + y_2$$
$$y_1(\,x_1, x_2, x_3) = x_2 \cdot x_3$$
$$y_2(\,x_1, x_2, x_3) = x_1 + x_2 + x_3 + w$$

$$L(y_1, y_2) = L(y_1(\vec{x}), y_2(\vec{x})) = L(\vec{x}) = (x_2 \cdot x_3)^2 + (x_1 + x_2 + x_3 + w)$$

$$dL = \frac{\partial L}{\partial y_1} dy_1 + \frac{\partial L}{\partial y_2} dy_2 \equiv \frac{\partial L}{\partial x_1} dx_1 + \frac{\partial L}{\partial x_2} dx_2 + \frac{\partial L}{\partial x_3} dx_3 + \frac{\partial L}{\partial w} dw$$

Производная сложной функции

$$L(y_1, y_2) = y_1^2 + y_2$$
$$y_1(\, x_1, x_2, x_3) = x_2 \cdot x_3$$
$$y_2(\, x_1, x_2, x_3) = x_1 + x_2 + x_3 + w$$

$$L(y_1, y_2) = L(y_1(\vec{x}), y_2(\vec{x})) = L(\vec{x}) = (x_2 \cdot x_3)^2 + (x_1 + x_2 + x_3 + w)$$

$$dL = \frac{\partial L}{\partial y_1} dy_1 + \frac{\partial L}{\partial y_2} dy_2 \equiv \frac{\partial L}{\partial x_1} dx_1 + \frac{\partial L}{\partial x_2} dx_2 + \frac{\partial L}{\partial x_3} dx_3 + \frac{\partial L}{\partial w} dw$$

$$dL = \frac{\partial L}{\partial y_1} dy_1 + \frac{\partial L}{\partial y_2} dy_2 = \frac{\partial L}{\partial y_1} (2x_2^2 x_3 \cdot dx_3 + 2x_2 x_3^2 \cdot dx_2) + \frac{\partial L}{\partial y_2} (dx_1 + dx_2 + dx_3 + dw)$$

Производная сложной функции

$$L(y_1, y_2) = y_1^2 + y_2$$
$$y_1(\,x_1, x_2, x_3) = x_2 \cdot x_3$$
$$y_2(\,x_1, x_2, x_3) = x_1 + x_2 + x_3 + w$$

$$L(y_1, y_2) = L(y_1(\vec{x}), y_2(\vec{x})) = L(\vec{x}) = (x_2 \cdot x_3)^2 + (x_1 + x_2 + x_3 + w)$$

$$dL = \frac{\partial L}{\partial y_1} dy_1 + \frac{\partial L}{\partial y_2} dy_2 \equiv \frac{\partial L}{\partial x_1} dx_1 + \frac{\partial L}{\partial x_2} dx_2 + \frac{\partial L}{\partial x_3} dx_3 + \frac{\partial L}{\partial w} dw$$

$$dL = \frac{\partial L}{\partial y_1} dy_1 + \frac{\partial L}{\partial y_2} dy_2 = \frac{\partial L}{\partial y_1}(2x_2^2 x_3 \cdot dx_3 + 2x_2 x_3^2 \cdot dx_2) + \frac{\partial L}{\partial y_2}(dx_1 + dx_2 + dx_3 + dw)$$

$$\frac{\partial L}{\partial x_1} dx_1 + \frac{\partial L}{\partial x_2} dx_2 + \frac{\partial L}{\partial x_3} dx_3 + \frac{\partial L}{\partial w} dw =$$
$$\frac{\partial L}{\partial y_2} dx_1 + \left(\frac{\partial L}{\partial y_1} 2x_2 \cdot x_3^2 + \frac{\partial L}{\partial y_2}\right) dx_2 + \left(\frac{\partial L}{\partial y_1} 2x_3 \cdot x_2^2 + \frac{\partial L}{\partial y_2}\right) dx_3 + \frac{\partial L}{\partial y_2} dw$$

Выводы

$$y = f(\vec{x}, W_x)$$

$$\mathrm{L} = \mathrm{L}(y)$$

1. (Forward) Если знаем x (и w) – можем вычислить $y(x)$
2. (Backward) Если знаем $\frac{\partial \mathrm{L}}{\partial \mathrm{y}}$ – можем вычислить $\frac{\partial \mathrm{L}}{\partial \mathrm{x}}, \frac{\partial \mathrm{L}}{\partial \mathrm{w}}$

Вопросы?

2. Практика.

Github repo. Домашка. Numpy.

Практика

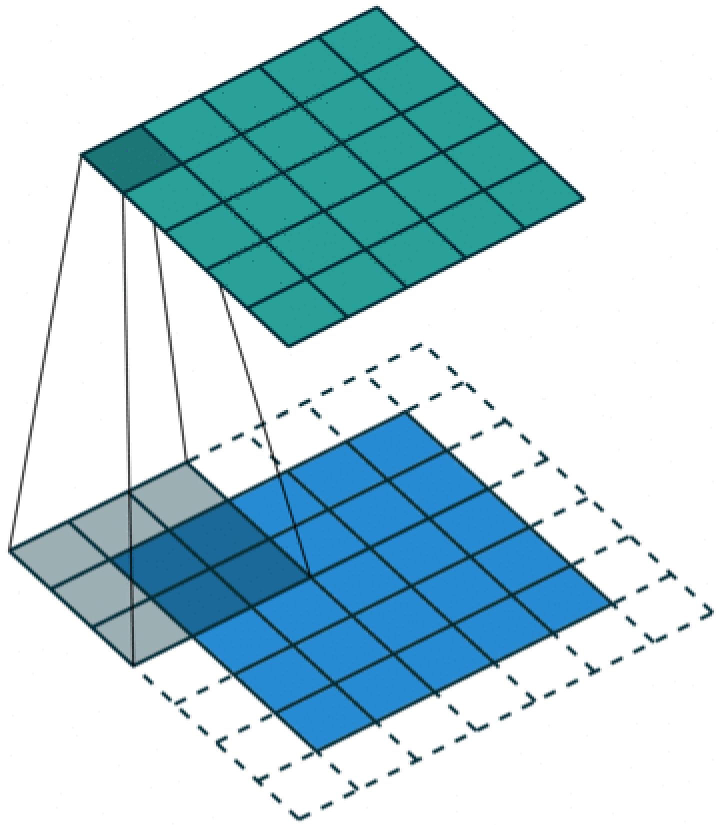
https://github.com/abbyy-edu/course_cvdl

1. ДЗ-1
2. Numpy-туториал

3.Свертка

Вывод backpropagation для свертки.

Визуализация



https://github.com/vdumoulin/conv_arithmetic

«Свертка» (кросс-корреляция)

$I[C, H, W], K[1, C, D, D], O[1, H, W]$ – input, kernel, output

Введём симметричный индекс kernel $d = \frac{D-1}{2}$, размер kernel - всегда нечетный

$$O[h, w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} I[c, h + y, w + x] \cdot K[c, y, x]$$

«Свертка» (кросс-корреляция)

$I[C, H, W], K[1, C, D, D], O[1, H, W]$ – input, kernel, output

$d = \frac{D-1}{2}$, размер kernel всегда нечетный (пусть D=3, d=1)

$$O[h, w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} I[c, h+y, w+x] \cdot K[c, y, x]$$

$$O[h, w] = \sum_{c=0}^{C-1}$$

$$I[c, h-1, w-1] \cdot K[c, -1, -1] \quad + \quad I[c, h, -1, w+0] \cdot K[c, -1, 0] \quad + \quad I[c, h-1, w-1] \cdot K[c, -1, -1]$$
$$I[c, h+0, w-1] \cdot K[c, +0, -1] \quad + \quad I[c, h+0, w+0] \cdot K[c, +0, 0] \quad + \quad I[c, h, +0\ w-1] \cdot K[c, +0, -1]$$
$$I[c, h+1, w-1] \cdot K[c, +1, -1] \quad + \quad I[c, h+1, w+0] \cdot K[c, +1, 0] \quad + \quad I[c, h, +1, w-1] \cdot K[c, -1, -1]$$

Q: Между вычислением $O[0, 2]$ и, например, $O[1, 3]$ будет небольшое различие. В чём?

«Свертка» (кросс-корреляция)

$I[C,H,W], K[1,C,D,D], O[1,H,W]$ – input, kernel, output

$d = \frac{D-1}{2}$, размер kernel всегда нечетный

$$O[h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} I[c, h+y, w+x] \cdot K[c,y,x]$$ - ломается на краях!

У input получаются негативные индексы – добавим zero-padding

$\hat{I}[c, -d\colon H-1+d, -d\colon W-1+d] = Pad(I,d)$

$$O[h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \hat{I}[c, h+y, w+x] \cdot K[c,y,x] = CC(\hat{I}, K)$$

Градиенты «свертки»

$$O[h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \hat{I}[c,h+y,w+x] \cdot K[c,y,x] = f_K(I)$$

$$dO[1,h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c,h+y,w+x] \cdot K[c,y,x] + \hat{I}[c,h+y,w+x] \cdot dK[c,y,x]$$

Нам нужны градиенты:

- $\frac{\partial L}{\partial K[c,y,x]}$ - чтобы «подправить» K
- $\frac{\partial L}{\partial I[c,h,w]}$ - чтобы передать дальше по графу вычислений

Градиенты «свертки»: $\frac{\partial L}{\partial K[c,y,x]}$

$$dL = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} dO[h,w] = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O}[h,w] dO[h,w]$$

$$dO[1,h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c,h+y,w+x] \cdot K[c,y,x] + \hat{I}[c,h+y,w+x] \cdot dK[c,y,x]$$

$$dO[1,h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \cancel{d\hat{I}[c,h+y,w+x] \cdot K[c,y,x]} + \hat{I}[c,h+y,w+x] \cdot dK[c,y,x]$$

$$dL = \cdots + \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \hat{I}[c,h+y,w+x] \cdot dK[c,y,x]$$

Градиенты «свертки»: $\frac{\partial L}{\partial K[c,y,x]}$

$$dL = \dots + \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} \hat{I}[c, h+y, w+x] \cdot dK[c,y,x]$$

$$dL = \dots + \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} dK[c,y,x] \sum_{\boldsymbol{h},\boldsymbol{w}=\boldsymbol{0}}^{\boldsymbol{H-1},\boldsymbol{W-1}} \hat{\boldsymbol{I}}[\boldsymbol{c}, \boldsymbol{h}+\boldsymbol{y}, \boldsymbol{w}+\boldsymbol{x}] \frac{\boldsymbol{\partial L}}{\boldsymbol{\partial O}[\boldsymbol{h},\boldsymbol{w}]}$$

<u>Q:</u> Через какую операцию можно записать выделенную жирным часть формулы?

Градиенты «свертки»: $\frac{\partial L}{\partial K[c,y,x]}$

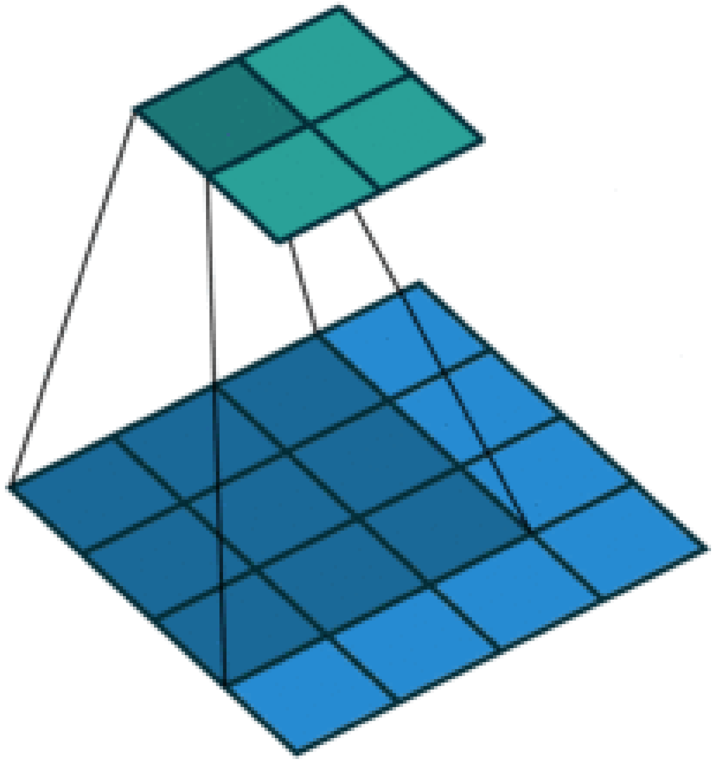
$$dL = \cdots + \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} \hat{I}[c, h+y, w+x] \cdot dK[c,y,x]$$

$$dL = \cdots + \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} dK[c,y,x] \sum_{h,w=0}^{H-1,W-1} \hat{I}[c, h+y, w+x] \frac{\partial L}{\partial O[h,w]}$$

$$dL = \cdots + \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} dK[c,y,x] \cdot CC\left(\hat{I}[c], \frac{\partial L}{\partial O}\right)$$

$$\frac{\partial L}{\partial K[c]} = CC\left(\hat{I}[c], \frac{\partial L}{\partial O}\right)$$

<u>Градиент K: «свертка» Input и градиента O.</u>



https://github.com/vdumoulin/conv_arithmetic

Градиенты «свертки»: $\frac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} dO[h,w]$$

$$dO[1,h,w] = \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c,h+y,w+x] \cdot K[c,y,x] + \cancel{\hat{I}[c,h+y,w+x] \cdot dK[c,y,x]}$$

$$dL = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c,h+y,w+x] \cdot K[c,y,x] + \cdots$$

Градиенты «свертки»: $\frac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c, h + y, w + x] \cdot K[c, y, x] + \cdots$$

Q: Индексы в dI варьируются по x и y. Что делать?

Градиенты «свертки»: $\frac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{h,w=0}^{H-1,W-1} \frac{\partial L}{\partial O[h,w]} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} d\hat{I}[c, h + y, w + x] \cdot K[c, y, x] + \cdots$$

$$\hat{h} = h + y$$
$$\hat{w} = w + x$$

$$dL = \sum_{\hat{h},\hat{w}=-d}^{H-1+d,W-1+d} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \frac{\partial L}{\partial O[\hat{h} - y, \hat{w} - x]} d\hat{I}[c, \hat{h}, \hat{w}] K[c, y, x] + \cdots$$

Градиенты «свертки»: $\dfrac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{\hat{h},\hat{w}=-d}^{H-1+d,W-1+d} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \frac{\partial L}{\partial O[\hat{h}-y,\hat{w}-x]} K[c,y,x]\, d\hat{I}[c,\hat{h},\hat{w}] + \cdots$$

Градиенты «свертки»: $\frac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{\hat{h},\hat{w}=-d}^{H-1+d,W-1+d} \sum_{\substack{c=0 \\ x,y=-d}}^{\substack{C-1 \\ d,d}} \frac{\partial L}{\partial O[\hat{h} - y, \hat{w} - x]} K[c,y,x]\, d\hat{I}[c,\hat{h},\hat{w}] + \cdots$$

$$\hat{x} = -x$$
$$\hat{y} = -y$$

Градиенты «свертки»: $\frac{\partial L}{\partial I[c,h,w]}$

$$dL = \sum_{\hat{h},\hat{w}=-d}^{H-1+d,W-1+d} \sum_{\substack{c=0\\x,y=-d}}^{\substack{C-1\\d,d}} \frac{\partial L}{\partial O[\hat{h}-y,\hat{w}-x]} K[c,y,x]\, d\hat{I}[c,\hat{h},\hat{w}] + \cdots$$

$$\hat{x} = -x$$
$$\hat{y} = -y$$
$$K[c,x,y] = K[c,-\hat{x},-\hat{y}] = \widehat{K}[c,\hat{x},\hat{y}]$$

$$dL = \sum_{\hat{h},\hat{w}=-d}^{H-1+d,W-1+d} \sum_{\substack{c=0\\\hat{x},\hat{y}=-d,-d}}^{\substack{C-1\\d,d}} \frac{\partial L}{\partial O[\hat{h}+\hat{y},\hat{w}+\hat{x}]} \widehat{K}[c,\hat{y},\hat{x}]\, d\hat{I}[c,\hat{h},\hat{w}] + \cdots$$

$$\frac{\partial L}{\partial I[c]} = CC(\frac{\partial L}{\partial O}, \widehat{K}[c])$$

Выводы

$$O = CC(\hat{I}, K)$$

$$\frac{\partial L}{\partial \hat{I}[c]} = CC\left(\frac{\partial L}{\partial O}, \widehat{K}[c]\right)$$

$$\frac{\partial L}{\partial K[c]} = CC\left(\hat{I}[c], \frac{\partial L}{\partial O}\right)$$

Обозначения:

- CC – это кросс-корреляция
- $\hat{I}$ - это I с паддингами
- $\widehat{K}$ - это K с обратной индексацией (повернутое на 180 градусов)

Выводы

$$O = CC(\hat{I}, K)$$

$$\frac{\partial L}{\partial \hat{I}[c]} = CC\left(\frac{\partial L}{\partial O}, \widehat{K}[c]\right)$$

$$\frac{\partial L}{\partial K[c]} = CC\left(\hat{I}[c], \frac{\partial L}{\partial O}\right)$$

Заметки:

- Будьте внимательны с индексами и паддингами
- Помните, что в «свертке» также есть добавление bias: O = CC(I, K) + b
- Помните, что Kernel имеет размер [**<u>G</u>**, C, D, D], а Output[**<u>G</u>**, H, W]
- Используйте для дебага результаты torch

Выводы

$$O = CC(\hat{I}, K)$$

$$\frac{\partial L}{\partial \hat{I}[c]} = CC\left(\frac{\partial L}{\partial O}, \widehat{K}[c]\right)$$

$$\frac{\partial L}{\partial K[c]} = CC\left(\hat{I}[c], \frac{\partial L}{\partial O}\right)$$

В домашке нет:

- Dilation («разреженных сверток»)
- Четных размеров Kernel **И** Input
- Stride (шаг) > 1