Глубокое обучение и вообще

Шигапова Фирюза и Соловей Влад

21 января 2021 г.

Посиделка 13: Свёрточные сетки

Risk
modelling.research

Agenda

- Как видит компьютер
- Свёртка
- Пулинг
- Собираем свою свёрточную сетку
- Data augmentation

Затравка (2006)

Затравка (2006)

- Капча портит вид сайтов, довольно бесполезная, в Microsoft придумывают в 2006 году новый вид капчи. Надо отличать котов от собак.
- В то время разделение собак от кошек было очень сложной задачей, лучшая точность была 0.6.
- У нас ест 12 картинок, робот нагнёт нас с вероятностью 0.6^{12}.
- Пул картинок пополнялся фотографиями из приютов.

В 2014 проект закрыли

Dogs vs. Cats

Create an algorithm to distinguish dogs from cats

215 teams · 6 years ago

Overview | Data | Notebooks | Discussion | Leaderboard | Rules

Public Leaderboard | **Private Leaderboard**

The private leaderboard is calculated with approximately 70% of the test data.
This competition has completed. This leaderboard reflects the final standings.

Refresh

Gold | Silver | Bronze

#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	—	Pierre Sermanet			0.98914	5	6y
2	▲ 4	orchid			0.98308	17	6y
3	—	Owen			0.98171	15	6y
4	—	Paul Covington			0.98171	3	6y

Risk
modelling.research

Как видит компьютер

Картинка — тензор

- Каждая картинка — это матрица из пикселей
- Каждый пиксель обладает яркостью по шкале от 0 до 255

Картинка — тензор

- Цветное изображение имеет три канала пикселей: красный, зелёный и синий (rgb), размерность изображения $5 \times 5 \times 3$

Картинка — тензор

Картинка — тензор

5 x 5 x 3

Обычная сетка

- Очень много весов
- Теряется информация о взаимном расположении пикселей
- Изображение в разных местах картинки даёт разные веса

Обычная сетка

- Хотим, чтобы информация не терялась
- Хотим, чтобы сетка была устойчива к положению картинки
- Хотим одинаковые веса $\Rightarrow$ свёртка

Свёртка

Свёртка

Risk
modelling.research

```
[[62, 36,  9],
 [62, 36,  9],
 [61, 35,  8],
 ...,
 [58, 43, 20],
 [57, 45, 21],
 [57, 45, 21]]
```

Красный

```
[[62, 36,  9],
 [61, 35,  8],
 [59, 33,  6],
 ...,
 [53, 41, 17],
 [53, 41, 17],
 [52, 40, 16]]
```

Зеленый

```
[[62, 36,  9],
 [60, 34,  7],
 [57, 31,  4],
 ...,
 [50, 38, 14],
 [49, 37, 13],
 [48, 36, 12]]
```

Синий

$$\frac{1}{9}\cdot\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}$$

$$\begin{pmatrix}0.1 & 0.1 & 0.1\\ 0.1 & 0.1 & 0.1\\ 0.1 & 0.1 & 0.1\end{pmatrix}$$

$$\begin{pmatrix}0.1 & 0.1 & 0.1\\ 0.1 & 2 & 0.1\\ 0.1 & 0.1 & 0.1\end{pmatrix}$$

Свёртка

- Разные ядра помогают накладывать на кртинку различные эффекты
- Какие-то ядра помогают искать границы
- Идея: возможно, с помощью некоторых ядер можно искать разные образы...

Классификатор слэшей

Risk
modelling.research

Классификатор слэшей

Risk
modelling.research

Свёртка инвариантна к расположению

0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	1

Input

$*$

1	0
0	1

Kernel

$=$

0	0	0
0	1	0
0	0	2

Output

1	0	0	0
0	1	0	0
0	0	0	0
0	0	0	0

Input

$*$

1	0
0	1

Kernel

$=$

2	0	0
0	1	0
0	0	0

Output

Свёртка инвариантна к расположению

Дополнение (Padding)

Padding используют, чтобы пространственная размерность картинки не уменьшалась после применения свёртки, это помогает не терять информацию на краях и избежать проблем с размерностями

Дополнение (Padding)

Свёрточный слой

Shared bias:
b

Shared kernel:

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

0	0	0	0	0
0	0	1	0	0
0	1	1	0	0
0	1	0	1	0
0	0	0	0	0

Input 3x3 image with zero **padding** (grey area)

$\sigma(w_6 + \boldsymbol{w_8} + w_9 + b)$	…	…
…	…	…
…	…	…

9 output neurons (**feature map**) with only 10 parameters

Risk
modelling.research

Свёрточный слой

Risk modelling.research

Backpropagation для свёрточного слоя

$$b = w_1 x_{11} + w_2 x_{12} + w_3 x_{21} + w_4 x_{22}$$

$$a = w_1 x_{12} + w_2 x_{13} + w_3 x_{22} + w_4 x_{23}$$

Risk
modelling.research

Backpropagation для свёрточного слоя

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a} \cdot x_{11} + \frac{\partial L}{\partial b} \cdot x_{12} + \frac{\partial L}{\partial c} \cdot x_{22} + \frac{\partial L}{\partial d} \cdot x_{12}$$

Risk
modelling.research

Свёрточный слой

- Слой действует одинаково для каждого участка картинки, в отличие от полносвязного
- Нужно оценивать меньшее количество параметров
- Слой учитывает взаимное расположение пикселей
- Можно учить тем же самым backpropagation
- Свёрточный слой - это полносвязный слой с ограничениями, попробуйте нарисовать свёрточный слой также как мы рисовали полносвязный

Одного ядра недостаточно

Risk
modelling.research

Одного свёрточного слоя недостаточно

- Нейроны первого слоя смотрят на поле 3×3
- Если интересующий нас объект больше, нам нужна вторая свёртка

Одного свёрточного слоя недостаточно

- N слоёв со свёртками 3×3
- На N-ом слое поле обзора $(2N+1) \times (2N+1)$
- Если наш объект размера 300, надо 150 слоёв... $\Rightarrow$ нужно растить поле обзора быстрее

Risk modelling.research

Сдвиги и Пулинг

Нужно растить поле обзора быстрее!

- Пиксели локально скоррелированы — соседние пиксели, как правило, не сильно отличаются друг от друга

- Если будем делать свёртку с каким-то шагом, сэкономим мощности компьютера и не потеряем в информации

- Очень агрессивная стратегия снижения размерности картинки

Сдвиг (Stride)

Risk
modelling.research

Пулинг слой (Pooling)

- Будем считать внутри какого-то окна максимум или среднее и сворачивать размерность, пользуясь локальной скоррелированностью

Простейшая CNN

Нейросетка LeNet-5 (1998) для распознавания рукописных цифр

http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

Risk modelling.research

Что выучивают нейросети

Risk
modelling.research

Что выучивают нейросети

Layer 1

Layer 2

Layer 3

Layer 4

Layer 5

https://arxiv.org/pdf/1311.2901.pdf

Risk
modelling.research

Data augmentation

Data augmentation

- В сетке может быть миллионы параметров!
- Естественная регуляризация, дополнительная регуляризация, генерация новых данных (data augmentation)
- Генерируем новые цвета, сдвигаем, искажаем и тп

https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html

Data augmentation

- Сдвиги (вместо них лучше пулинг)
- Увеличение, уменьшение
- Повороты
- Искажение
- Затенение
- Смена стиля (красок)

Делает модель более устойчивой, полезна при маленьких выборках. На больших датасетах также улучшает результаты.

Risk
modelling.research

Собираем свою собственную CNN