

Об алгебраических свойствах операций, используемых при построении современных свёрточных нейронных сетей

Фадеев Е. П., Зубюк А. В.
<http://NeuroFuzzy.phys.msu.ru>,
fadeevegor@yandex.ru,
zubuk@cmpd2.phys.msu.ru

5 декабря 2019 г.

Непредсказуемость поведения искусственных нейронных сетей



Перс. кот



Карл. пудель



Карл. пудель

S. Dodge and L. Karam, "Understanding How Image Quality Affects Deep Neural Networks",
ArXiv e-prints, 2016. arXiv: 1604.04004 [cs.CV]

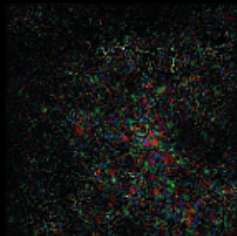
Непредсказуемость поведения искусственных нейронных сетей



Кит



Черепаха



S. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, “Deepfool: A simple and accurate method to fool deep neural networks”, in **CVPR**, 2016, pp. 2574–2582

Непредсказуемость поведения искусственных нейронных сетей



Корабль
Автомобиль



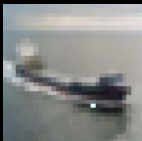
Лошадь
Лягушка



Автомобиль
Самолёт



Птица
Лягушка



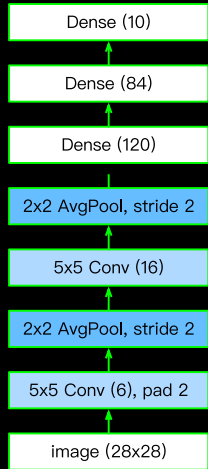
Корабль
Самолёт



Кошка
Собака

J. Su, D. Vasconcellos Vargas, and S. Kouichi, "One pixel attack for fooling deep neural networks", [ArXiv e-prints](#), Oct. 2017. arXiv: 1710.08864 [cs.LG]

Архитектура свёрточных нейронных сетей



► Свёртка

Выделение признаков (выделение краёв объектов и др.) на изображении на разных масштабах, одинаковое внимание к всем областям изображения.



0	-1	0
-1	4	-1
0	-1	0



► Активация

Нелинейность.

► Пулинг

Уменьшение размерности, выделение полезной и отсеивание лишней информации.

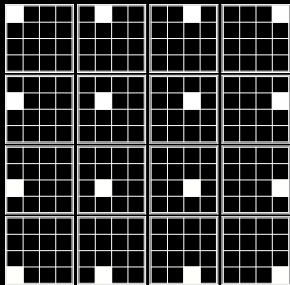
Y. Lecun **et al.**, “Gradient-based learning applied to document recognition”, **Proceedings of the IEEE**, vol. 86, pp. 2278–2324, Dec. 1998. DOI: 10.1109/5.726791

Алгебраические свойства операций

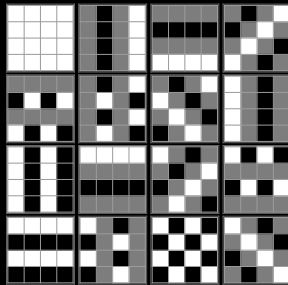
Предлагается рассматривать изображения как элементы эвклидова пространства $\mathbb{R}^n$ со скалярным произведением

$$(X, Y) = \sum_{i=1}^h \sum_{j=1}^w X_{ij} Y_{ij},$$

где X, Y — изображения, h и w — высота и ширина изображений соответственно.



Исходный базис



Базис Фурье

Сингулярное разложение линейного оператора

Пусть $A : \mathbb{R}^N \rightarrow \mathbb{R}^M$ — линейный оператор ($N \geq M$). Тогда $\exists(U, S, V)$, где $U = \{u_i\}_{i=\overline{1, N}}$ — базис в $\mathbb{R}^N$, $S = \{s_i\}_{i=\overline{1, M}}$ — набор неотрицательных чисел, $\{v_i\}_{i=\overline{1, M}}$ — базис в $\mathbb{R}^M$:

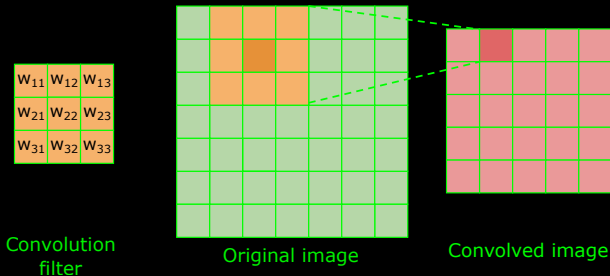
$$Au_i = s_i v_i, \quad i = 1, \dots, M,$$

$$Au_i = 0, \quad i = M + 1, \dots, N.$$

При этом U и V называются сингулярными базисами оператора A , а S его сингулярными значениями.

Свёртка (Convolution)

Свёртка — линейное преобразование с сингулярным базисом Фурье (с последующим добавлением смещения в исходном базисе).



Теорема о образе свёртки

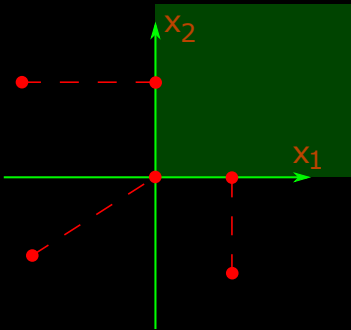
$$\mathcal{F}(f * g) = \mathcal{F}(f) \cdot \mathcal{F}(g),$$

Здесь $\mathcal{F}(\cdot)$ — дискретное преобразование Фурье, $*$ — свёртка двух изображений, $\cdot$ — покомпонентное произведение в пространстве образов.

Активация

Активация — Покомпонентное нелинейное преобразование в исходном базисе.

$\text{ReLU}(x) = \max\{0, x\}$ — кусочно-линейное преобразование, которое можно рассматривать как проекцию на не отрицательный ортант.



Пулинг

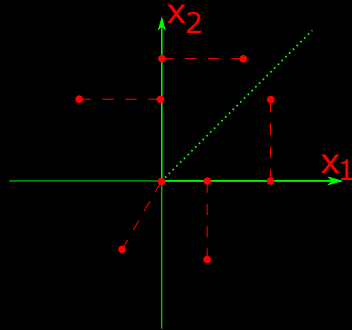
Max-pooling (P_{max}) — Нелинейное (кусочно-линейное) преобразование, которое применяется к группе компонент в исходном базисе.

- Операции P_{max} и ReLU можно менять местами, а именно пусть $P_{max} : \mathbb{R}^{H,W} \rightarrow \mathbb{R}^{h,w}$ — max-pooling понижающий размер изображения с (H, W) до (h, w) , тогда

$$P_{max} \circ \text{ReLU}^{H,W} = \text{ReLU}^{h,w} \circ P_{max},$$

где $\text{ReLU}^{H,W}$ и $\text{ReLU}^{h,w}$ — операторы ReLU действующие в пространствах изображений размерами (H, W) и (h, w) соответственно.

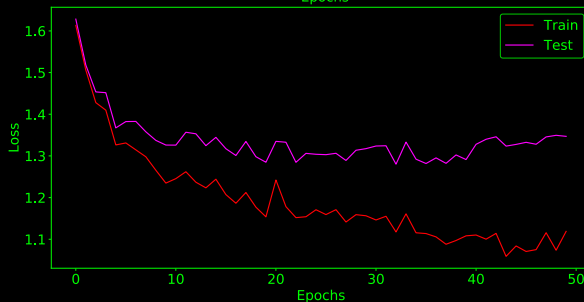
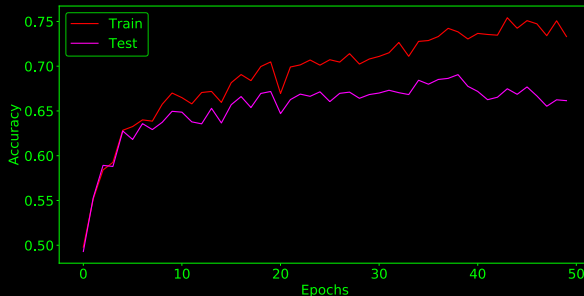
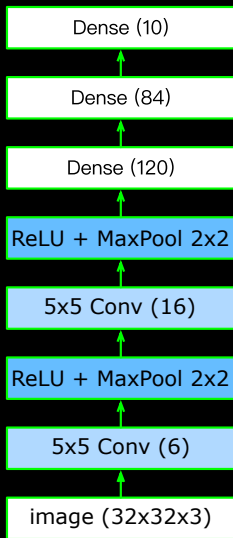
- Композицию ReLU и P_{max} можно рассматривать как длину проекции на множество, состоящее из положительных лучей осей.



Идея исследования

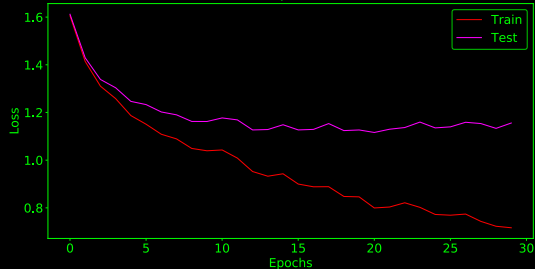
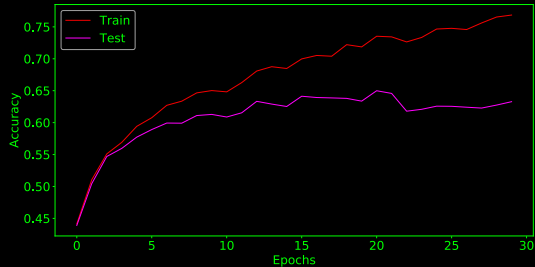
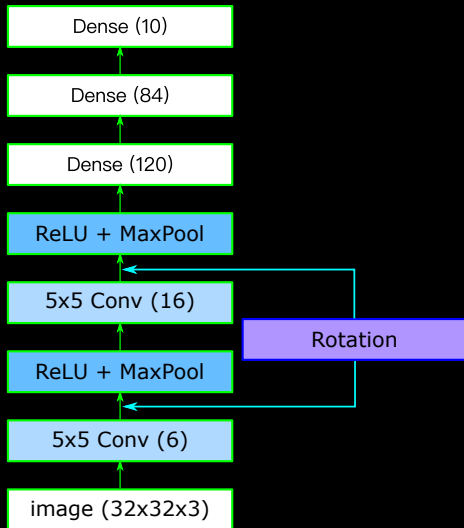
Действие свёрточной нейронной сети можно рассматривать, как композицию растяжений в одном базисе и нелинейных преобразований в другом базисе. Успех нейронных сетей обеспечивается поочередным применением этих операций.

Стартовая архитектура



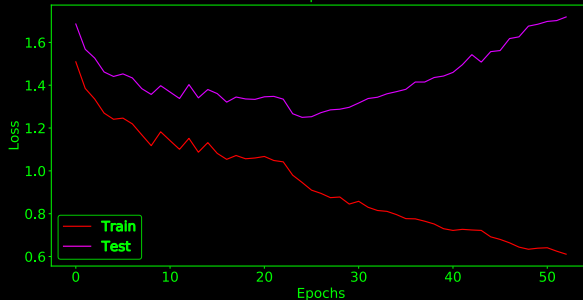
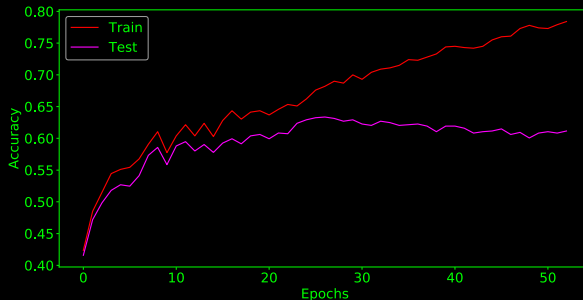
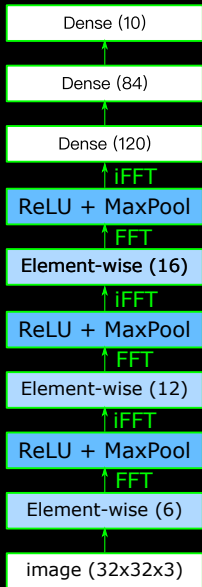
Количество параметров 62006
Пиковая точность 0.68

Добавление поворота



M. Erdoğan and M. Özdemir, **Generating four dimensional rotation matrices**, 2015. DOI: 10.13140/RG.2.1.4118.3442

Базисы наоборот



Количество параметров 48844

Пиковая точность 0.63

Спасибо за внимание!