

Рисунок 7 – График зависимости качества обучения от кол-ва слоёв

Литература

- 1 Image-net [Электронный ресурс]. – <http://www.image-net.org/challenges/LSVRC/>
- 2 LeCun Y. et al. Convolutional networks for images, speech, and time series //The handbook of brain theory and neural networks. – 1995. – Т. 3361. – №. 10. – С. 1995.
- 3 Осовский С. Нейронные сети для обработки информации / Пер. с пол. И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.: ил.
- 4 Ruder S. An overview of gradient descent optimization algorithms //arXiv preprint arXiv:1609.04747. – 2016.
- 5 Kaggle [Электронный ресурс]. – <http://www.kaggle.com>
- 6 Abadi M. et al. TensorFlow: A System for Large-Scale Machine Learning //OSDI. – 2016. – Т. 16. – С. 265-283.

Г.А. Альгашев, О.П. Солдатова

НЕЙРОПЛАСТИЧНОСТЬ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

(Самарский университет)

Цель данной научной работы состоит в исследовании эффективности переноса знаний (transfer learning) предварительно обученной свёрточной нейронной сети для решения других задач.

Для классификации и распознавания изображения большую популярность получили свёрточные нейронные сети. Такие сети состоят из нескольких свёрточных слоёв, которые в процессе обучения извлекают признаки подаваемых на вход изображений. Верхние слои сети занимаются извлечением низкоуровневых признаков, а глубокие слои занимаются извлечением высокоуровневых признаков. Чем ближе свёрточный слой находится к выходу сети, тем менее абстрактными становятся извлекаемые признаки (рисунок 1) [1].



В результате исследования работы свёрточных нейронных сетей был сделан вывод, что при обучении на различных наборах изображений, сети на верхних слоях формировали очень похожие низкоуровневые признаки. Таким образом, ядра свёртки разных свёрточных сетей на верхних слоях имеют похожие веса.

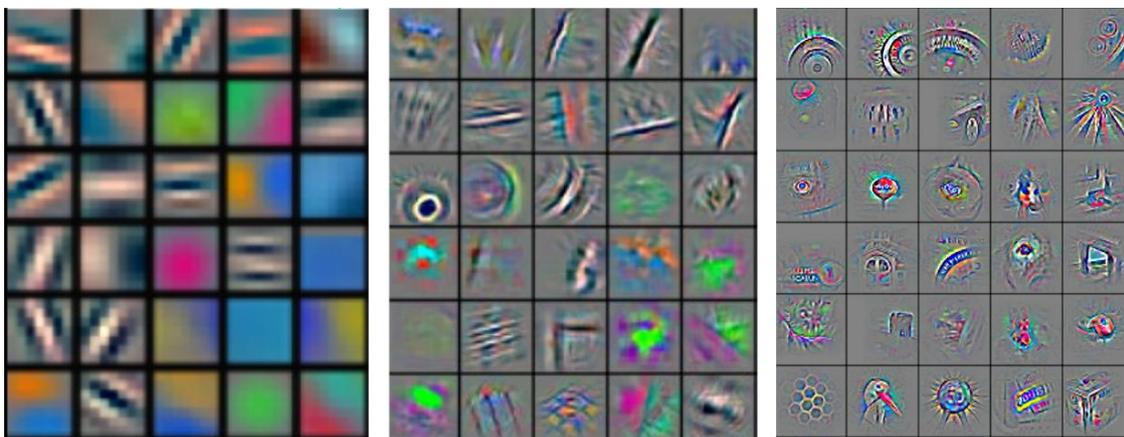


Рисунок 1 – Визуализация слоёв свёрточной сети

Сделанные выводы стали основополагающими в развитии метода переноса знаний или использования свойства нейропластичности, которое присуще головному мозгу. Основная идея метода состоит в достижении следующих целей: улучшение качества обучения уже на начальных итерациях, ускорение сходимости алгоритмов обучения, улучшение верхней достижимой границы качества [2].

Для применения метода переноса знаний, помимо обучающей выборки необходимо иметь хорошо обученную нейронную сеть. Такая сеть уже умеет хорошо извлекать низкоуровневые признаки, следовательно, в момент обучения новой сети веса на верхних слоях будут изменяться не значительно, по сравнению с глубокими слоями.

Для исследования поставленной задачи в качестве эталонной сети была взята сеть VGG-16. Эта сеть содержит 16 слоёв (рисунок 2) и состоит из двух частей: первая часть выделяет характерные признаки в изображении и состоит из 13 слоёв, вторая часть отвечает за классификацию объекта на изображения по выделенным на предыдущем этапе признакам и состоит из 3 слоёв [3].

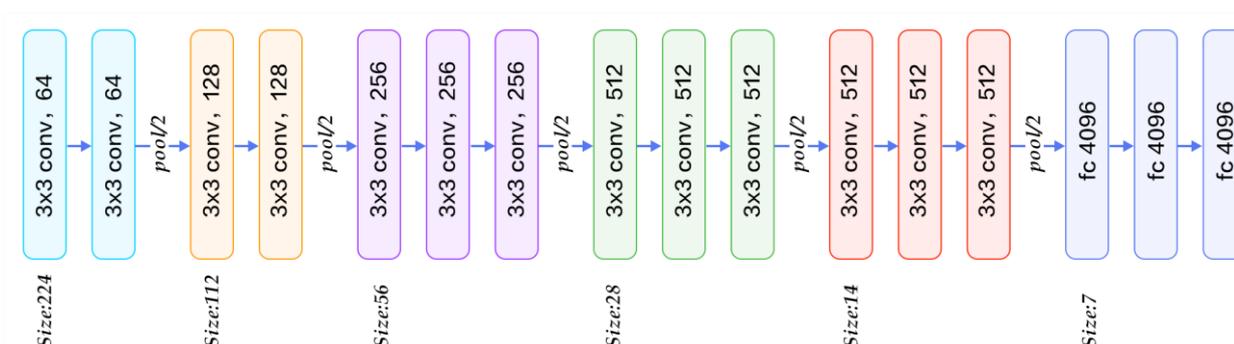


Рисунок 2 – Структура свёрточной сети VGG-16



В качестве обучающей выборки был взят набор изображений маркированных клеток в крови (рисунок 3). В набор входят клетки 4 видов: эозинофилы, лимфоциты, моноциты и нейтрофилы. Каждый класс состоит из 800 обучающих изображений, 100 изображения для промежуточной проверки и 100 изображения для тестовой проверки.

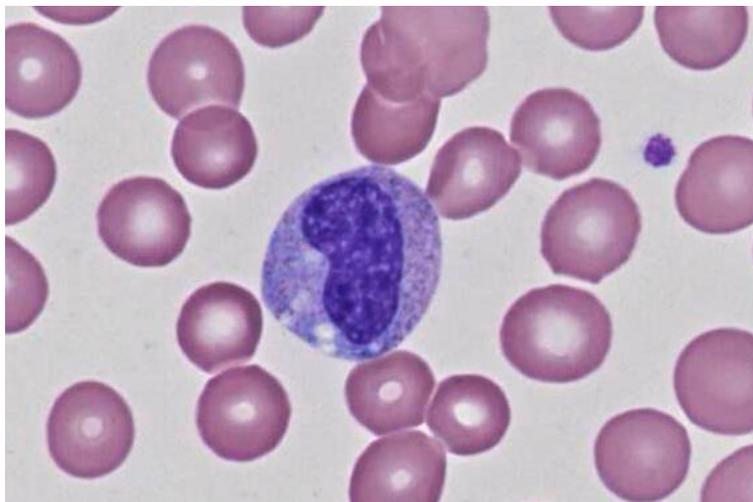


Рисунок 3 – Маркированная клетка для классификации

Важным фактором в проведении исследования является тот факт, что обучающая выборка не входит в число классов, на основе которых обучали эталонную сеть. Таким образом, мы имеем дело с сетью, которая была обучена на наборе данных, сильно отличающихся от нашей обучающей выборки.

Для исследования метода переноса знания было принято решение использовать архитектуру сети VGG-16. Первые 13 свёрточных слоёв остались без изменения структуры, а последние три слоя были заменены на 2 полносвязных слоя, состоящих из 256 нейронов и 4 нейронов (в соответствии с количеством классов в обучающей выборке).

В процессе исследования было решено провести две серии экспериментов. При этом в обоих случаях архитектура сети, её настройки и количество эпох обучения остаются одинаковыми. В первой серии эксперимента все слои нейронной сети были инициализированы случайными весами. Во второй серии эксперимента первые 13 слоёв инициализировались весами сети VGG-16, натренированными на наборе данных ImageNet, а последние слои, где происходит классификация изображений, были инициализированы случайными весами. Обе сети обучались в течении 15 эпох. В таблице 1 приведены результаты двух экспериментов.

На рисунке 4 приведены зависимости точности классификации от числа эпох обучения.

Первая сеть на тестовых данных, которые она не видела в процессе обучения распознала их с точностью в 61,5%, а вторая сеть распознала с точностью в 98,50%. Исходя из выше приведённых результатов, мы видим, что сеть, в которой мы использовали уже натренированные веса показала лучший результат, чем сеть, в которой все веса были случайно инициализированы.



Таблица 1 – Результаты экспериментов

Номер эпохи	Без переноса знания		С переносом знаний	
	ошибка	точность	ошибка	точность
1	1.3863	0.2500	0.9689	0.6800
2	1.3863	0.2625	0.4277	0.8425
3	1.3863	0.2500	0.2819	0.8825
4	1.3863	0.2500	0.1647	0.9350
5	1.3863	0.2500	0.1366	0.9400
6	1.3863	0.2750	0.1135	0.9525
7	1.3862	0.2950	0.1916	0.9275
8	1.3746	0.2525	0.1554	0.9450
9	1.2085	0.4025	0.1080	0.9675
10	1.0993	0.4550	0.0948	0.9725
11	1.0924	0.4575	0.0989	0.9650
12	1.0946	0.4275	0.1439	0.9525
13	0.9956	0.4875	0.0933	0.9700
14	0.9393	0.4775	0.0831	0.9800
15	0.7962	0.6300	0.0795	0.9775

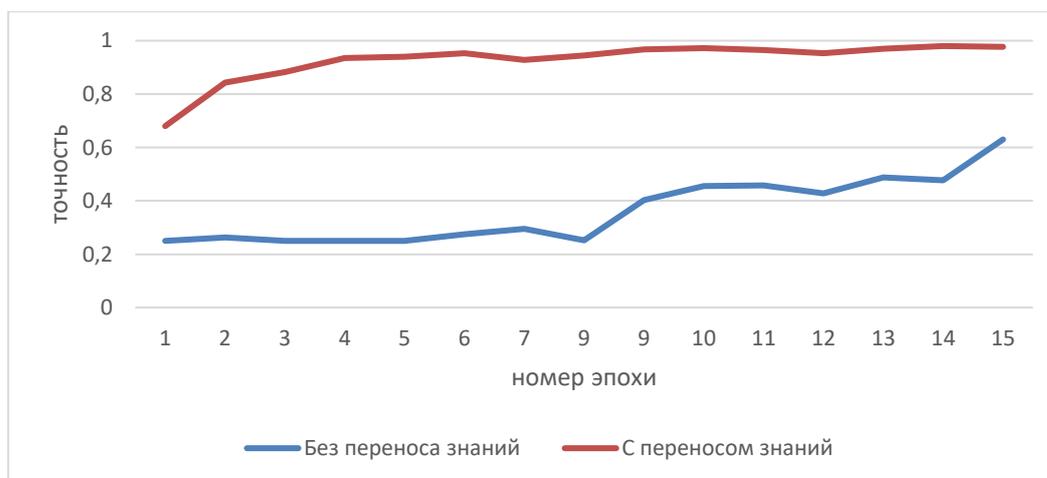


Рисунок 4 – Графики изменения точности классификации в зависимости от числа эпох обучения

Задача классификации изображения требует больших вычислительных мощностей, в результате чего обучение сети может занимать от нескольких часов до нескольких дней. Использование метода переноса знаний позволяет сократить время обучение сети и повысить точность её обучения, что и было показано в данном исследовании.

Литература

1 Very Deep Convolutional Networks for Large-Scale Visual Recognition [Электронный ресурс] //Visual Geometry Group [сайт] URL: http://www.robots.ox.ac.uk/~vgg/research/very_deep/ (дата обращения: 25.02.2018)



2 Torrey L. Transfer Learning. Handbook of Research on Machine Learning Applications / L.Torrey, J.Shavlik // IGI Global 2009, 22 с.

3 Zeiler M.D. Dept. of Computer Science / M.D.Zeiler, R.Fergus. // New York University, USA 2013, 16 с.

О.В. Антошина

ПРИМЕНЕНИЕ АЛГОРИТМОВ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТОВОЙ ИНФОРМАЦИИ ДЛЯ ОПРЕДЕЛЕНИЯ ЗАИМСТВОВАНИЙ

(Самарский университет)

Введение

Свободный доступ к огромным массивам информации в сети Интернет способствует развитию плагиата в различных сферах человеческого общества. Под плагиатом понимают незаконное использование чужого изобретения или произведения без указания источника заимствования [1]. Для выявления заимствований и оценки самостоятельности автора применяют специализированные системы [2]. Главным недостатком некоторых автоматизированных систем поиска плагиата является отсутствие методов лингвистического анализа, в том числе семантического и морфологического.

Методы анализа текстовой информации

Задачи семантического и морфологического анализа способствуют классификации текстов, определению предметной области, а также более точному обнаружению плагиата.

Одной из задач морфологического анализа является определение нормальной формы искомого слова (процесс лемматизации), от которого наследуется данная словоформа, и набора параметров этой словоформы [3]. Таким образом, морфологический анализ позволяет избежать дополнительной обработки всех словоформ.

Другой тип морфологического анализа построен на основании правил с использованием процедурного подхода и позволяет привести словоформу к единой форме с помощью стемминга – выделения основы слова. Рассмотрим оба подхода [4].

1. Лемматизация – процесс приведения словоформ с одинаковым понятием к единой нормальной форме для увеличения релевантности поиска [5] и уменьшения количества анализируемых слов. Операция лемматизации может быть представлена в виде отображения (1):

$$T \rightarrow L, \quad (1)$$

где T – множество всех терминов, L – множество всех лемм. Данное преобразование позволяет уменьшить размер индексной информации и ускорить обработку текстового документа.

2. Стемминг – приближенный эвристический процесс, который находит основу слова [6]. В отличие от лемматизации стеммер не требует наличия