

УДК 004.853

**ТЕНДЕНЦИИ РАЗВИТИЯ
СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ****А.А. СОЛОВЬЁВ***(Представлено: канд. физ.-мат. наук, доц. Ю.Ф. ПАСТУХОВ)*

Рассматриваются и анализируются основные идеи и новшества в реализации свёрточных нейронных сетей, а также их результаты на основании ежегодного международного проекта по широкомасштабному распознаванию образов ILSVRC.

В последние несколько лет область машинного обучения добилась огромного прогресса в решении задач, связанных с компьютерным зрением. В частности, было обнаружено, что свёрточные нейронные сети могут достигать разумной производительности по задачам визуального распознавания, сопоставимой или превышающей производительность человека в некоторых случаях.

Исследователи продемонстрировали устойчивый прогресс в компьютерном зрении, подтвердив свою работу с ImageNet — проектом по созданию и сопровождению массивной базы данных аннотированных изображений, предназначенная для отработки и тестирования методов распознавания образов и машинного зрения. С 2010 года ведётся проект ILSVRC (ImageNet Large Scale Visual Recognition Challenge), в рамках которого различные программные продукты ежегодно соревнуются в классификации и распознавания объектов и сцен в базе данных ImageNet [1]. Эту кампанию можно рассматривать как ежегодную Олимпиаду компьютерного зрения, в которой команды со всего мира конкурируют, чтобы увидеть, кто имеет лучшую модель компьютерного зрения для таких задач, как классификация, локализация, обнаружение и многое другое.

Основной метрикой ошибки на этой базе является так называемая Топ-5 ошибка. Это когда берутся первые пять предсказаний сети, в которых она наиболее уверена, и если правильный класс из разметки попал в эти пять предсказаний, то можно сказать, что сеть права. И ошибка — это когда в первые пять предсказаний не попал правильный класс. Чтобы сравнить модели, будет рассмотрено, как часто модель не может предсказать правильный ответ в качестве одной из своих пяти догадок.

В 2012 году Алексом Крижевским была представлена глубокая свёрточная нейронная сеть, которая стала победителем проекта ILSVRC. В статье [2] группа создателей обсудила архитектуру сети (которая называлась AlexNet). Они использовали относительно простую компоновку по сравнению с современными архитектурами. Сеть состояла из 5 свёрточных (convolutional) слоёв, слоёв с максимальным пулом (max pooling), слоёв отсева (dropout) и 3 полностью связанных (fully connected) слоёв. Созданная ими сеть использовалась для классификации 1000 различных классов.

Данная сеть обучалась на выборке, содержащей более 15 миллионов аннотированных изображений, более чем из 22 тысяч классов. Новаторством данной модели было использование активационной функции ReLU, что позволило значительно уменьшить время обучения сети, в отличие от ранее широко используемой функции гиперболического тангенса. Также в данной работе были использованы методы аугментации (augmentation) данных, что значительно увеличило объём обучающей выборки и улучшило обучение. Нововведением в борьбе с проблемой переобучения сети стали слои отсева нейронов.

Нейронная сеть AlexNet стала первой моделью СНС, которая показала отличные результаты на исторически сложном наборе данных ImageNet. Используя методы, которые всё ещё используются сегодня, такие как увеличение данных и отсева, данная модель действительно иллюстрирует преимущества СНС над другими методами классификации изображений.

После победы AlexNet модели на конкурсе ImageNet, СНС широко вошли в решения задач, связанных с компьютерным зрением, и число свёрточных нейронных сетей, представленных на конкурсе ILSVRC в 2013 году быстро возросло. Победителем конкурса в этом году была сеть ZF Net, которая достигла 11,2% ошибок.

В статье [3] создатели архитектуры ZF Net обсуждают идеи о том, что новый интерес к СНС обусловлен доступностью больших обучающих наборов и увеличением вычислительной мощности с использованием графических процессоров. Они также говорят о ограниченном знании, которое исследователи использовали во внутренних механизмах этих моделей, заявив, что без этого понимания разработка лучших моделей сводится к методу проб и ошибок. Основным вкладом в эту работу являются детали немного модифицированной модели AlexNet и очень интересный способ визуализации карт функций.

Архитектура ZF Net очень похожа на AlexNet, за исключением нескольких незначительных изменений. Основными отличиями было то, что ZF Net обучалась всего на 1,3 миллиона изображений. Вме-

сто использования фильтров размером 11x11 в первом слое (что и было реализовано в AlexNet) использовались фильтры размером 7x7 и уменьшенное значение шага. Обоснование этой модификации заключается в том, что меньший размер фильтра в первом свёрточном слое помогает сохранить много исходной информации о пикселях во входном томе. Фильтр размером 11x11 оказался плох тем, что пропускал много информации, что особенно плохо сказывалось на обучении, поскольку это первый слой СНС. Также в их работе была разработана технология визуализации с именем Deconvolutional Network, которая помогает исследовать различные активизации функций и их отношение к входному пространству.

ZF Net была не только победителем конкурса в 2013 году, но также продемонстрировала больше способов повышения производительности. Описанный ими подход к визуализации признаков помогает не только объяснить внутреннюю работу СНС, но также даёт представление об её улучшении.

Простота и глубина — это то, что модель, созданная в 2014 году, наилучшим образом демонстрировала с её частотой ошибок 7,3%. Карен Симонян и Эндрю Зиссерман из Оксфордского университета создали 19-слойную СНС, названную VGG Net [4], в которой использовались фильтры 3x3 с шагом и 1, что существенно отличается от фильтров 11x11 сети AlexNet в первом слое и фильтров 7x7 сети ZF Net. Авторы утверждают, что комбинация двух 3x3 свёрточных слоёв имеет эффективное восприимчивое поле 5x5. Это, в свою очередь, имитирует больший фильтр, сохраняя при этом преимущества меньшего размера фильтра. Одним из преимуществ их сети является уменьшение количества параметров. Кроме того, с двумя свёрточными слоями можно использовать два слоя ReLU вместо одного.

По мере уменьшения пространственного размера входных данных на каждом уровне глубина данных увеличивается из-за увеличения количества фильтров прохода по сети. Стоит заметить, что количество фильтров удваивается после каждого слоя максимального пулинга, что усиливает идею сокращения пространственных размеров и роста глубины сети.

VGG Net архитектура укрепила представление о том, что свёрточные нейронные сети должны иметь глубокую сеть слоёв, чтобы иерархическое представление визуальных данных работало.

Победителем в 2014 году с коэффициентом ошибок 6,7% стала 22-уровневая СНС сеть GoogLeNet [5]. Это была одна из первых архитектур СНС, которые действительно отклонялись от общего подхода простой укладки свёрточных слоёв и слоёв пулинга в последовательной структуре. В данной модели не использовались полносвязные слои, вместо этого они использовали средний пулинг, чтобы перейти от объёма 7x7x1024 к тому 1x1x1024, что сэкономило огромное количество параметров (в 12 раз меньше, чем в AlexNet).

Данная сеть была одной из первых моделей, которые представили идею о том, что слои СНС не всегда должны быть сложены последовательно. Авторы показали, что творческое структурирование слоев может привести к повышению производительности и эффективности вычислений.

Относительно новая 152-слойная сетевая архитектура ResNet выиграла ILSVRC 2015 с невероятной частотой ошибок 3,6% (в зависимости от их квалификации и опыта люди обычно колеблются с частотой ошибок 5-10%). Данная модель является по настоящему глубокой, однако авторы утверждают, что наивное увеличение уровней приводит к более высокой тренировочной и тестовой ошибкам.

Модель ResNet — это одна из лучших архитектур СНС, которую мы имеем в настоящее время.

Проанализировав рассмотренные СНС можно провести параллель между количеством слоёв и ошибкой сетей, представленных в таблице.

Таблица. — Показатели свёрточных нейронных сетей

Архитектура	Количество слоёв	Значение ошибки
AlexNet	8	15.4
ZF Net	8	11.2
VGG Net	19	7.3
GoogLeNet	22	6.7
ResNet	152	3.6

Проанализировав таблицу, можно сказать, что имеет место тенденция углубления архитектур СНС, что позволяет получать лучшие результаты по классификации изображений. Однако в новых статьях, опубликованных на научных конкурсах было описано использование моделей с намного большим количеством слоёв, которые показали результаты заметно хуже, чем та же модель ResNet, что говорит о том, что развитие СНС должно продолжаться не только в сторону увеличения глубины сети, но в улучшении старых и создании новых методов обучения.

ЛИТЕРАТУРА

1. ImageNet [Электронный ресурс]. — Режим доступа: <https://ru.wikipedia.org/wiki/ImageNet>. — Дата доступа: 24.09.2018г.
2. Krizhevsky, A. Imagenet classification with deep convolutional neural networks / Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton // NIPS. — 2012.
3. Zeiler, M. Visualizing and Understanding CNN / Matthew D. Zeiler, Rob Fergus // ECCV. — 2014.
4. Simonyan, K. Very Deep Convolutional Networks for Large-Scale Image Recognition / Karen Simonyan, Andrew Zisserman // arXiv technical report. — 2014.
5. Wei, L. Going deeper with convolutions / Christian Szegedy, Wei Liu, Pierre Sermanet, Scott Reed, Vincent Vanhoucke, Dragomir Anguelov, Andrew Rabinovich, Yangqing Jia, Dumitru Erhan // CVPR. — 2015.
6. Kaiming, H. Deep Residual Learning for Image Recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // CVPR. — 2016.