

Рецензия на работу *Борисова Алексея*

Рецензент: *Суглобов Кирилл*

## 1 Насколько хорошо выполнено задание?

В целом, задание выполнено хорошо, но есть ряд замечаний:

1. Автор пишет: "ширина соответствует числу каналов... параметризуется числом  $w$ ... Если в слое было  $C_i$  каналов, то станет  $w \cdot C_i$ ".  
Далее, при описании параметров «по умолчанию» в модели указывается: "Ширина — 1.0;".  
Понятно, что под этим значением подразумевается не число каналов, а коэффициент  $w$ , что входные изображения датасета SVHN имеют три канала цвета.  
Но на мой взгляд стоило отдельно написать, что входное число каналов равно 3 и заменить "Ширина — 1.0;" на "Масштаб ширины — 1.0;".
2. При исследовании ширины автор отмечает, что "с ростом ширины сети точность на обучающем множестве возрастает, а на тестовом уменьшается, что говорит о том, что сеть начинает сильнее переобучаться". Это не соответствует результатам эксперимента: при увеличении ширины качество на обучении не убывает, а качество на тесте уменьшается с 0.811 до 0.770, далее до 0.690, а потом увеличивается до 0.744 (при коэффициенте ширины  $w = 2.6$ ). Я думаю, что по данной неточности стоило бы написать пояснение, например, что дело в погрешности. Или, если не так, то не делать вывода об усилении переобучения.
3. В экспериментах с dropout, с ненулевыми инициализациями весов и с оптимизаторами присутствуют выводы о качестве модели, а выводы по поводу распределений градиентов, как мне кажется, несодержательны, либо размыты: автору не хватило рассмотренной выборки графиков распределений градиентов для выявления каких-либо закономерностей на соответствующих экспериментах.
4. Графики распределений градиента имеют разный масштаб, из-за чего достаточно сложно сравнивать их. К тому же, я полагаю, что автору стоило пояснить, что подразумевается под "абсолютной величиной градиента" на этих графиках: дисперсия, размах или какая-либо другая характеристика распределения.
5. Нет общего вывода по проделанной работе. Я считаю, следовало бы собрать и подытожить результаты на всех экспериментах.

## 2 Понятно ли Вам исследование?

Исследование понятное, работа читается легко. В начале ставится задача, а далее перед каждым экспериментом и после него автор оставляет пояснения: мотивацию эксперимента и его итог соответственно. Однако по трём последним пунктам исследования автор оставляет достаточно размытые выводы (см. п.1.3).

## 3 Убеждает ли оно Вас в чём-то?

Да. Например, в том, что skip-connection помогает решить проблему затухания градиентов в глубокой сети, это видно из проделанного эксперимента. Но работа не убеждает в усилении переобучения при увеличении ширины сети (см. п.1.2). Наоборот, исходя из небольшой выборки слоёв кажется, что только некоторый диапазон ширины подвержен переобучению, после чего оно резко уменьшается.

## 4 Корректны ли утверждения и выводы?

Большинство утверждений корректны, а сомнение в корректности вызывает лишь утверждение об усилении переобучения с ростом ширины сети (см. п.1.2). При этом утверждений делается не так много: выводы по экспериментам с dropout, с ненулевыми инициализациями весов и с оптимизаторами кажутся недостаточными (см. п.1.3), так что не могут быть признаны корректными или некорректными.

## 5 Понятен ли Вам код, есть ли в нём ошибки?

Здесь надо отдать должное: модель, функции обучения и оценки модели, визуализация результатов экспериментов - всё вынесено в самостоятельные модули, код которых понятен, а в самой работе место не отнимается на побочный, неинформативный или лишний вывод при обучении модели / оценке модели / визуализации. Ошибок в коде не замечено.

## 6 Понятны ли Вам рисунки и таблицы?

Графики понятны и достаточно аккуратны, оси и заголовки подписаны. Разве что не очень удобно сравнивать графики между собой из-за различного, а часто и сильно различного масштаба (см. п.1.4). Для лучшего понимания можно было бы организовать дополнительный вывод рассматриваемых графиков в единой шкале.

## 7 Что на Ваш взгляд можно ещё сделать? (но это пока не сделано)

Ещё можно было бы:

1. Детальнее исследовать вопрос усиления переобучения при увеличении ширины сети.
2. Попытаться найти закономерности в экспериментах, где из-за небольшой выборки графиков и малого числа рассматриваемых параметров не удалось сделать конкретные выводы по распределению градиента.
3. Для более удобного сравнения распределений градиентов, можно было бы попробовать отобразить нужные графики рядом на единой логарифмической шкале.
4. Сделать общий вывод по работе.