

# Отчёт о работе Марии Дзюбы

Хисматуллин Владимир, 317 группа

17 мая 2022 г.

## 1 Введение

Работу можно разбить на код и эксперименты. Разобьём эксперименты на части:

1. С ретривером. На примере одной картинке применяются все методы для всех архитектур.
2. С бульдогом. Иллюстрация важных зон по каждому слою сети с использованием grad-cam.
3. С двумя классами. Визуализация внимания сети к двум объектам в одном изображении.
4. С церковью. Визуализация внимания сети при принятии решения.

## 2 Критика

### 2.1 Общая оценка

С одной стороны, проиллюстрирована работа разных алгоритмов на разных данных, показаны применения к конкретным задачам. Использованный код корректен.

С другой, в исследовании нет чёткой постановки задачи, мало выводов и анализа полученных результатов. В каждом эксперименте производится опыт на одной картинке, этого недостаточно, если автор хотел качественно рассказать о моделях.

Говоря вообще: задумка исследования правильная, отвечает на поставленный в задании вопрос, код написан верно, однако исполнение и анализ экспериментов оставляет желать лучшего. Всюду я опирался на постановку, как представлял её себе, так как автором она не приводится.

### 2.2 Понятность исследования

Не совсем ясна цель исследования. Вероятно, автор имел целью продемонстрировать разные методы и их приложения, показав, как можно интерпретировать работу нейросетей в разных ситуациях, для каких архитектур какие модели работают лучше. Однако о мыслях автора касательно цели исследования, ровно как и о большинстве экспериментов, можно лишь догадываться, а не судить. В отчёте не прописана постановка задачи и выводы из многих опытов.

## 2.3 Убедительность исследования

Разберём убедительность по экспериментам.

1. Saliency map - нет выводов. Grad-cam - нет выводов. Occlusion sensitivity - выводы неверные, об этом будет далее.
2. В данном случае всё хорошо, интересно смотреть, какие области считались важными на каждом слое. Одной картинке для этого опыта достаточно.
3. На примере одной картинке оба метода действительно хорошо отработали. Для убедительности можно было привести несколько примеров.
4. Не ясно, что именно на данном примере для нейросети отличает церковь от монастыря. Поясню: grad-cam для двух классов очень похожи. Разница между двумя классами в этом примере как раз не интерпретируется использованным методом. Обычно монастырь отличается от церкви тем, что он намного больше, ибо в нём проживают монахи, а не только проходят службы. На примере не понятно, как нейронка различает эти классы. Этот эксперимент скорее показывает неудачную интерпретацию изображения. Но о мыслях автора можно лишь догадываться, ибо выводов нет.

В итоге, данное исследование убедило меня в том, что можно использовать grad-cam для послойной интерпретации сетей и Occlusion Sensitivity для интерпретации внимания модели к разным классам. Однако я так и не узнал, какие архитектуры и какими алгоритмами интерпретируются лучше всего.

## 2.4 Корректность утверждений и выводов

Из утверждений, представленных в работе, есть одно некорректное:

Рассмотрим комментарий к опыту с Occlusion Sensitivity в эксперименте №1:

*... данные графики помогают нам понять, что для данных моделей лучше брать Patch=35 или Patch=45, т.к. в остальных случаях модель либо основывается на слишком маленькой, либо на слишком большой области изображения*

Утверждение однозначно некорректно. Параметр размера заплатки (patch) определяется для картинке или датасета, но не для архитектуры. Вся суть Occlusion Sensitivity в том, чтобы закрывать часть объекта в картинке, потому качество метода опирается именно на его размер. При этом я не утверждаю, что для разных моделей этот размер будет разным. Ключевой ошибкой является то, что patch априори подбирается под картинку. Если же нет - нужно было привести картинку на которых область внимания разная (морда собаки в  $\frac{1}{2}$  и в  $\frac{1}{20}$  экрана) и продемонстрировать, что на всех изображениях для данного патча качество интерпретации получилось хорошим.

## 2.5 Код

По первой части вопросов нет: реализации алгоритмов взяты из авторитетных источников. Комментарии по коду есть в источниках. В этом случае отсутствие комментариев не является недочётом. Код вывода картинок работает исправно и достаточно понятен.

Часть с записью и отображением рисунков следовало бы значительно упростить. Количество кода в этой части однозначно можно сократить в 5 раз. Блоки в 60+ строк кода можно было бы переписать в 20 и уместить в одну функцию. Понятно, что с такими блоками проще работать при отладке, но это значительно портит читаемость отчёта.

## **2.6 Рисунки**

Все рисунки понятны. Ясно, что изображено.

## **2.7 Требуют доработки**

В постановке задания говорится о качественных интерпретациях. Их не хватило. Были показаны интерпретации для разных картинок, но не ясно какие из них лучшие. Следовало приводить больше примеров. Или сделать отдельные эксперименты, где фиксированный алгоритм применяется к разным картинкам для разных архитектур. Потом подытожить какие архитектуры лучше всего интерпретируются и какими методами. Но больше всего не хватало выводов. Всё-таки в отчёте хочется видеть развитие мысли автора.