

004.93'1

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТОПОЛОГИЙ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ КОНФОКАЛЬНОЙ МИКРОСКОПИИ ЩИТОВИДНОЙ И ПАРАЩИТОВИДНОЙ ЖЕЛЕЗ

Цалко И. Н.

*Гомельский государственный технический университет имени П.О. Сухого, г. Гомель
Научный руководитель: Курочка К. С., канд. техн. наук, доцент*

Осуществление автоматического анализа изображений конфокальной микроскопии с помощью нейронных сетей даст возможность уменьшить время и повысить точность диагностики, а также оценить результаты различных хирургических вмешательств и выбрать наиболее подходящее лечение для пациента.

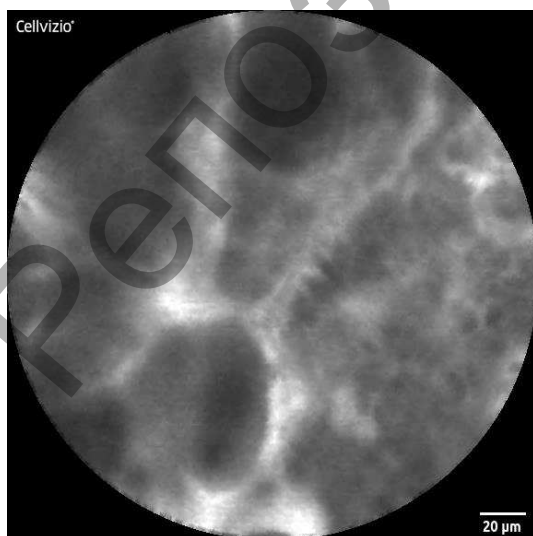
Для того, чтобы натренировать нейронную сеть требуемого качества, необходимо собрать достаточное количество данных и правильно их разметить при участии эксперта в предметной области.

Исходные данные конфокальной микроскопии находятся в видеофайлах формата avi с частотой кадров 12 кадров/сек, снятые с помощью микроскопа Cellvizio.

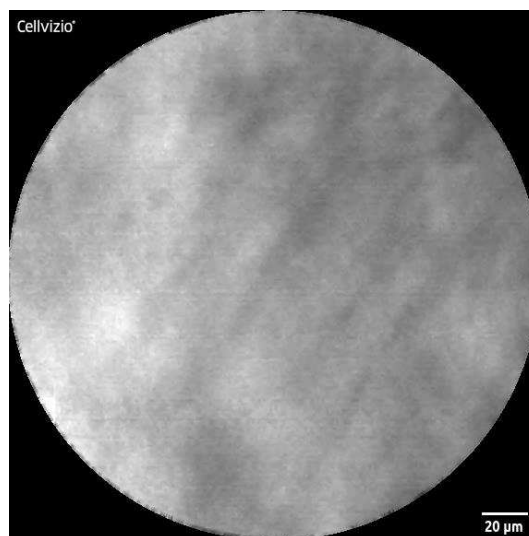
Цель исследования – выявление изображений, содержащих кровеносные сосуды, как предварительный этап анализа и предварительной фильтрации данных.

Для извлечения изображений из видео и их разметки была разработана программа, которая работает с видеофайлами любого размера и частотой кадров. С её помощью можно пометить некоторым классом как каждый кадр по отдельности, так и некоторые отрезки видео.

Разметка проводилась с разделением на два класса изображений "присутствие кровеносных сосудов" и "отсутствие кровеносных сосудов". Примеры данных изображений приведены на рисунках ниже.



Изображение сосудов



Изображение без сосудов

Всего было выбрано 5 видеофайлов для тренировочного набора данных и 1 видеофайл для тестового (каждый видеофайл был выбран из исследования отдельного пациента, чтобы исключить возможные пересечения в данных). В результате разметки было получено 2312 изображений с сосудами и 1665 изображений без сосудов в тренировочном наборе и 138 и 22 изображения в тестовом соответственно. Дисбаланс классов в наборах данных обусловлен содержанием исследуемых видеофайлов и, помимо процесса тренировки, это будет влиять на выбор метрики.

Для создания классификатора изображений были использованы следующие архитектуры:

- VGG [1]
- Inception [1]
- ResNet [1]
- DenseNet [1]

Выбор был обусловлен следующими причинами:

- VGG – является одной из первых по-настоящему глубоких сетей (имеется два варианта сети: 16 и 19 слоев), добившейся большой точности на наборе данных ImageNet

- Inception – является примером сети, в которой в отличие от сети VGG используются не просто свертки, а целые блоки, состоящие из набора сверток 1x1 и 3x3, соединенных определенным образом (такие блоки получили название Inception block). В отличие от VGG, она имеет большую точность, несмотря на то, что параметров в сети примерно в 5 раз меньше.

- ResNet – одна из первых архитектур, которая популяризовала так называемые shortcut connections или residuals, которые позволили замедлить процесс затухания градиентов при обучении сети и создать сеть с большим количеством слоев (в статье предлагается сеть с количеством слоев вплоть до 152).

- DenseNet – сеть с очень малым количеством параметров (около 7 миллионов), что в несколько раз меньше, чем у вышеупомянутых сетей, но при этом точность классификации является сопоставимой.

Тренировка сетей осуществляется не с нуля, так как набор изображений недостаточен по сравнению с количеством параметров сети. Именно поэтому используется техника переноса обучения (transfer learning). Вместо полного обучения всей сети обучается только последний слой, отвечающий за классификацию. Такая техника позволяет, во-первых, предотвратить переобучение, а во-вторых, использует признаки из более крупного набора данных ImageNet.

Несмотря на то, что обучается лишь только последний слой нейронной сети, увеличение данных может также привести к улучшению результатов классификации. Для этого используется техника раздувания данных (augmentation). Исходя из представленного набора изображений и информации из предметной области, можно заметить следующее:

- Поворот изображения на произвольный угол не изменяет его класс (то есть сосуды не могут ни появиться, ни исчезнуть)

- Горизонтальное отражение не меняет класс
- Вертикальное отражение не меняет класс
- Добавление нормально шума с небольшой дисперсией также не меняет класс

Комбинируя данные преобразования, можно увеличить набор данных для обучения и тестирования. Однако, если использовать все возможные комбинации трансформаций, время обучения увеличится экспоненциально, что является неоправданным.

Таким образом, проведя аугментацию набор аданных с двумя видами отражений, поворотами на углы, кратные 45 градусам, и добавив шум с $\mu = 0.0$ и $\sigma = 0.01$, можно увеличить набор данных в 72 раза. Тренировочный набор увеличивается примерно до 286 тысяч изображений, а тестовый – до 11 тысяч, что является достаточным для эксперимента [2].

Так, сети VGG, Resnet и DenseNet требуют на вход трехканальные изображения размером 224x224 пикселя, а Inception - 299x299, то изображения из набора данных масштабируются до соответствующих размеров.

В качестве метрики был выбран показатель F1, так как он является сбалансированным показателем между точностью (precision) и откликом (recall), а также лучше отражает работу классификатора на несбалансированных наборах данных, в отличие от точности (accuracy).

F1 вычисляется по формуле:

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

В качестве функции потерь была выбрана функция кросс-энтропии [3], как самая наиболее используемая в задачах классификации.

В качестве алгоритма оптимизации был выбран стохастический градиентный спуск с моментом инерции. Параметры оптимизатора следующие $\alpha = 0.01$ и $\beta = 0.99$. Кроме того, темп обучения α снижается в 2 раза, если на протяжении 2 эпох обучения не наблюдается улучшения метрики на валидационном наборе данных.

Также, если метрика на валидационном наборе не улучшается в течение 5 эпох, обучение останавливается, вне зависимости от того, достигло ли оно максимального количества эпох (по-умолчанию =20).

Результаты обучения представлены в таблице.

Архитектура сети	F1 (тренировочный набор)	F1 (тестовый набор)
VGG11	0.8507	0.8623
Inception V4	0.8405	0.8307
Resnet18	0.8622	0.8648
DenseNet121	0.8878	0.8729

В ходе исследования все архитектуры показали схожие результаты, однако так как сеть DenseNet121 имеет гораздо меньшее количество параметров, то для дальнейших исследований и апробации сети на встраиваемом устройстве была выбрана данная архитектура.

Список цитированных источников

1. Николенко, С. Глубокое обучение / С. Николенко, А. Кадури, Е. Архангельская. – СПб.: Питер, 2019 – 480 с.: ил.
2. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – 3-е издание, исправленное и дополненное. — М.: Техносфера, 2012. — 1104 с.
3. Нейросетевые технологии обработки данных : учеб. пособие/В. А. Головкин, В. В. Краснопрошин. – Минск : БГУ, 2017. – 263 с. – (Классическое университетское издание).