

Красные смещения и машинное обучение

Позднухов Н.А.

208 группа

МГУ им. М.В. Ломоносова

Физический факультет

**Научные руководители: Троицкий Сергей Вадимович,
Рубцов Григорий Игоревич**

Постановка задачи

- Один из методов определения красного смещения объекта – идентификация спектральных линий.
- Блазары – подкласс квазаров, для которого характерна малая интенсивность эмиссионных линий. Из-за этого использование стандартных методов идентификации затруднено.

Примеры оптических спектров

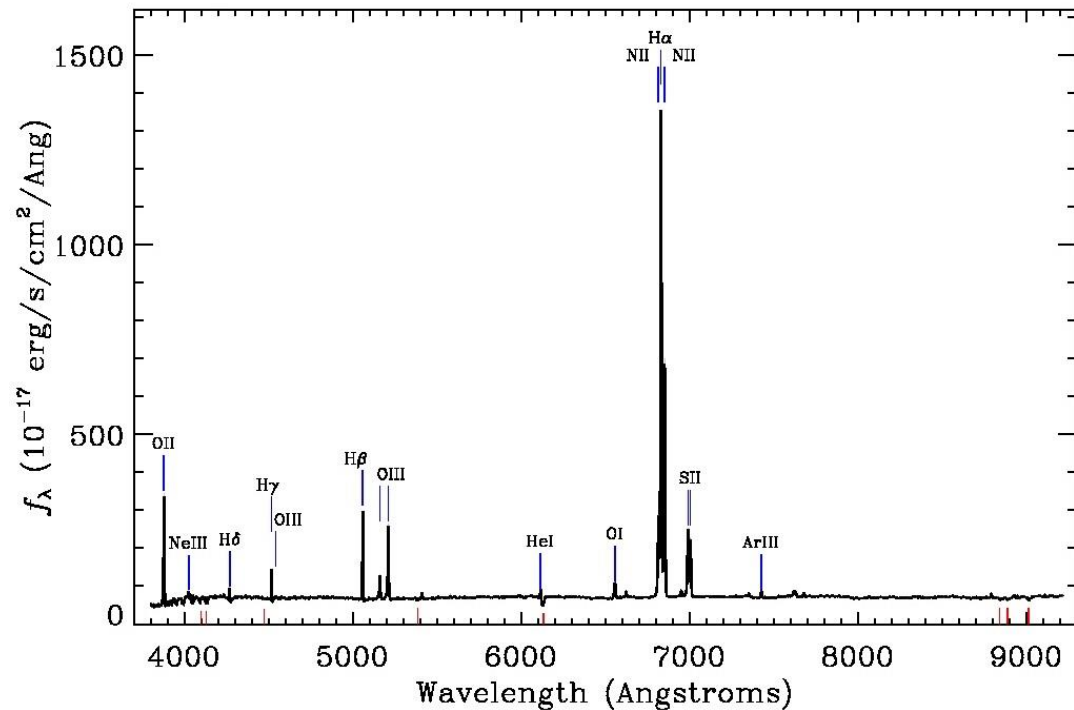


Рис. 1: Спектр галактики с явно выраженными линиями, $z = 0.036$

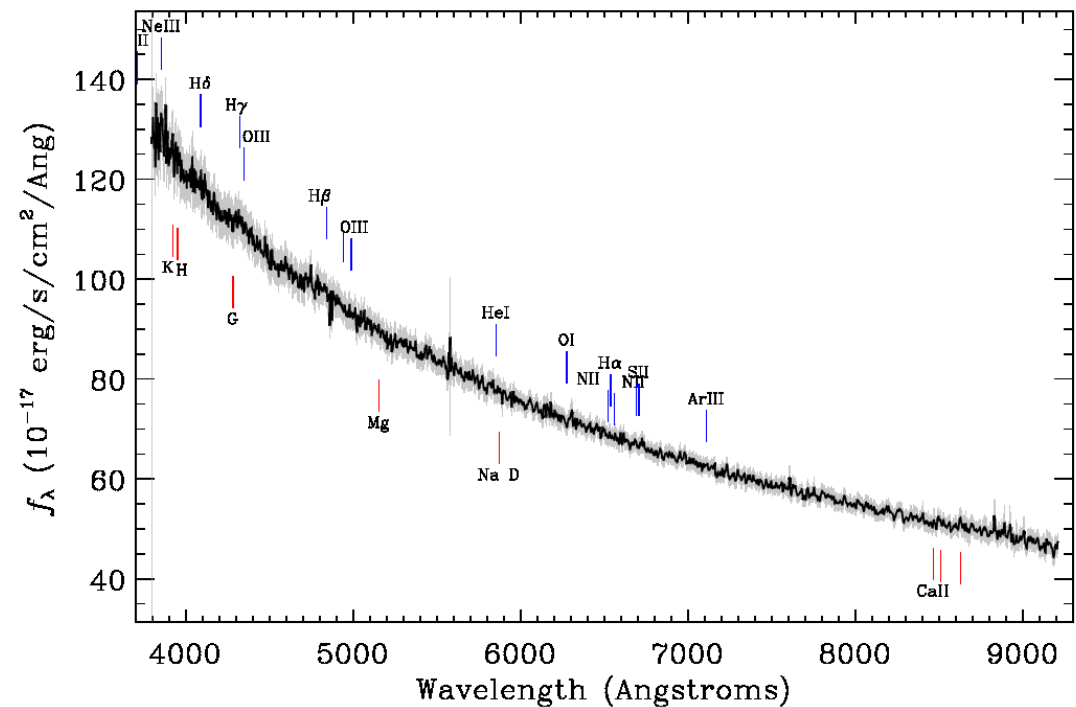


Рис. 2: Спектр блазара со слабыми линиями, $z = ???$

Постановка задачи

- Один из методов определения красного смещения объекта – идентификация спектральных линий.
- Блазары – подкласс квазаров, для которого характерна малая интенсивность эмиссионных линий. Из-за этого использование стандартных методов идентификации затруднено.
- Требуется новый метод отождествления спектральных линий блазаров.

Машинное обучение

- Имея выборку, для которой известен правильный ответ, можно “научить” алгоритм давать верные предсказания для других данных.
- Процесс обучения требует большого объёма выборки.
- Блэзаров с точно определённым красным смещением мало.
- Необходимо искусственно сгенерировать тренировочные данные.

Машинное обучение

- Для задач распознавания используются свёрточные нейронные сети

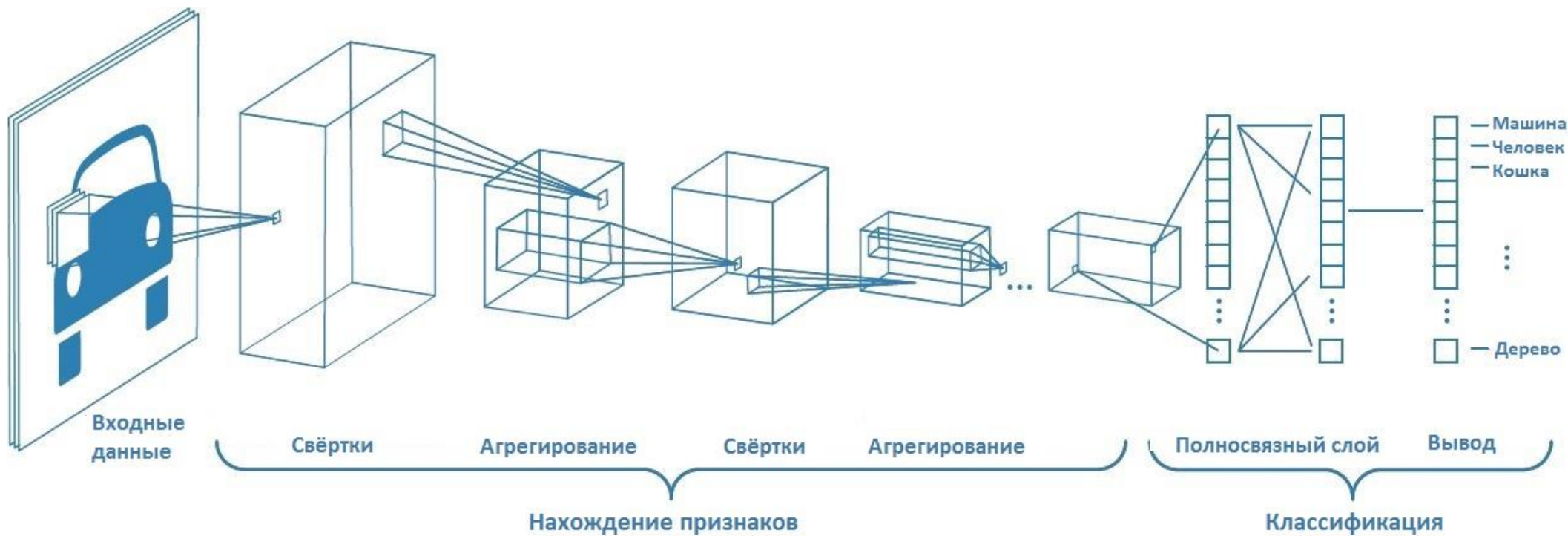


Рис. 3: Структура свёрточных сетей

Результаты

- Модель обучается отличать линии от шума.
- Точность распознавания незначительно зависит от интенсивности линий.
- При обучении на линиях малой интенсивности, модель даёт почти 100% точность на ярко выраженных линиях.

Результаты обучения

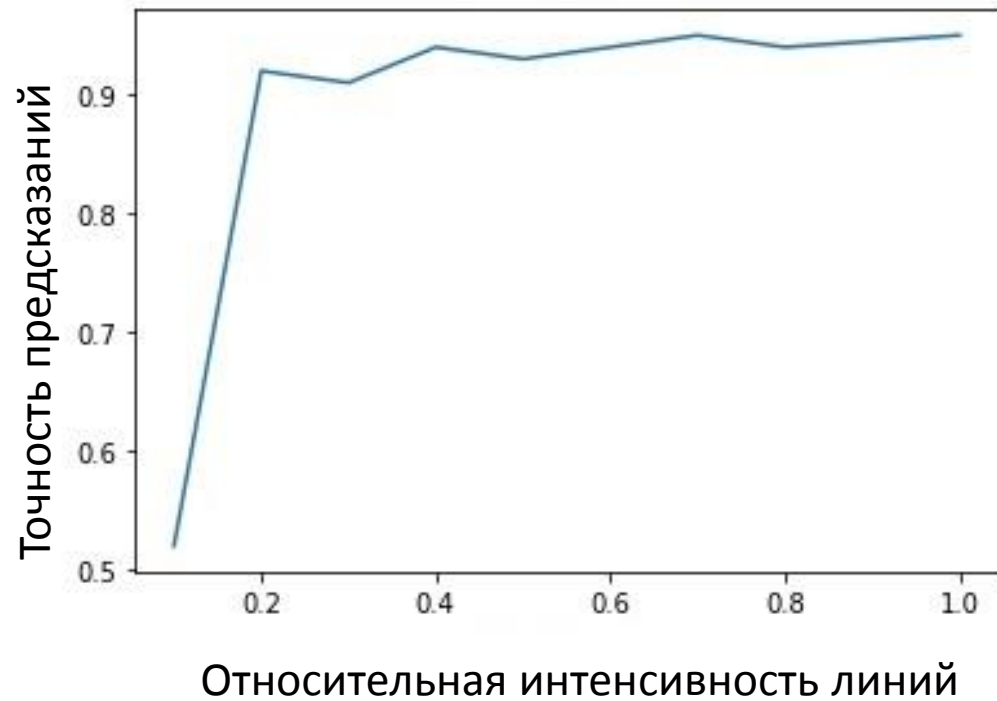


Рис. 4: Зависимость точности модели от интенсивности линий

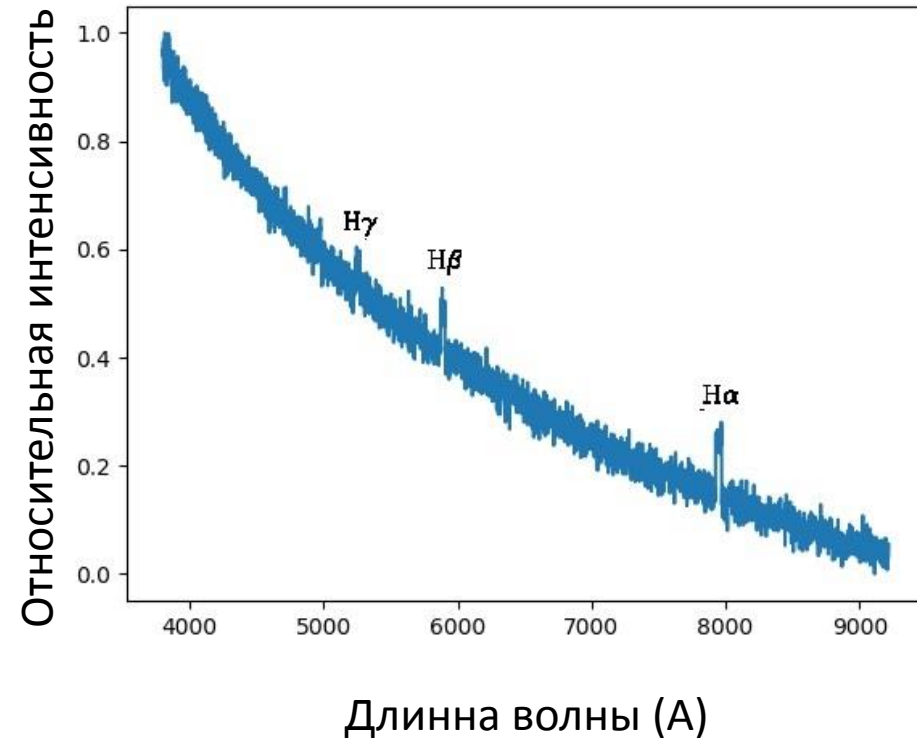


Рис. 5: Типичный искусственный спектр, $z = 0.22$

Результаты-2

- Нейронная сеть отличает разные серии линий друг от друга
- Точность достигает 92% - это хороший результат для машинного обучения

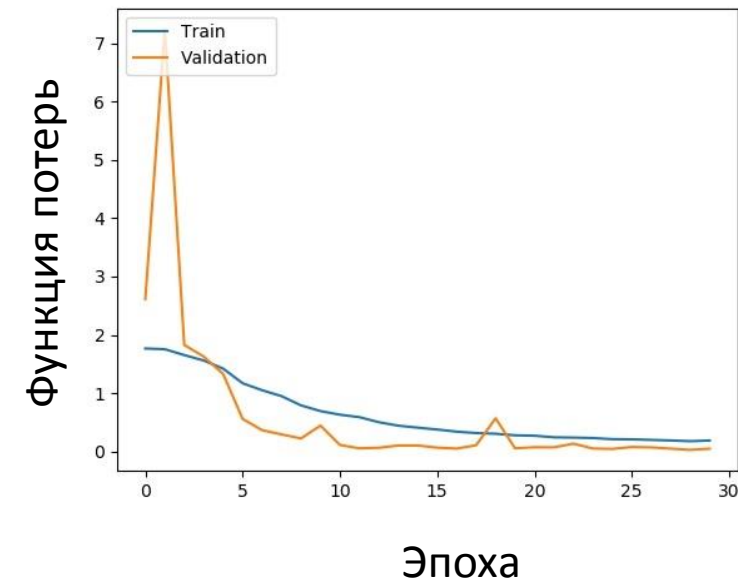
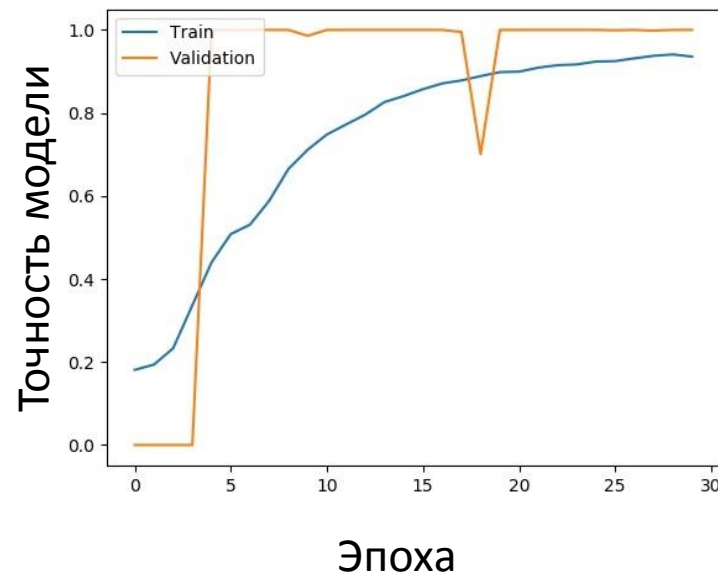


Рис 6: Точность и функция потерь при обучении

Результаты-3

- На наборе данных с комбинациями линий точность модели значительно ниже
- Модель переобучается – этот результат нельзя считать настоящим

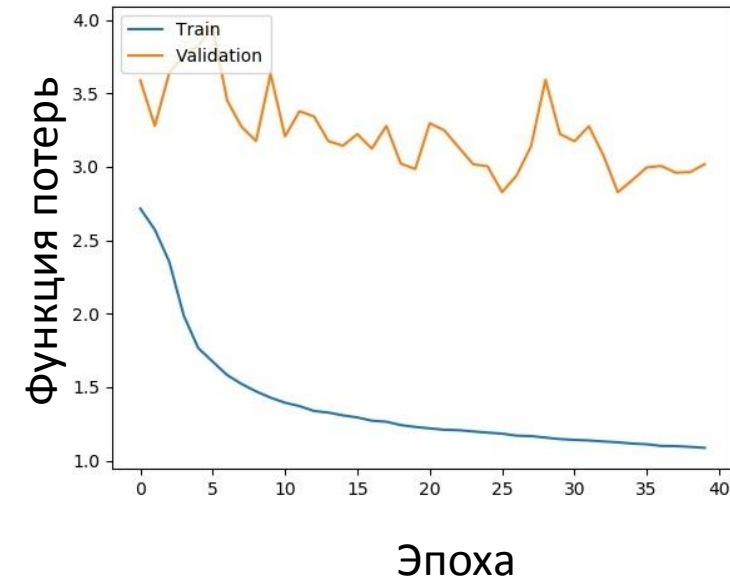
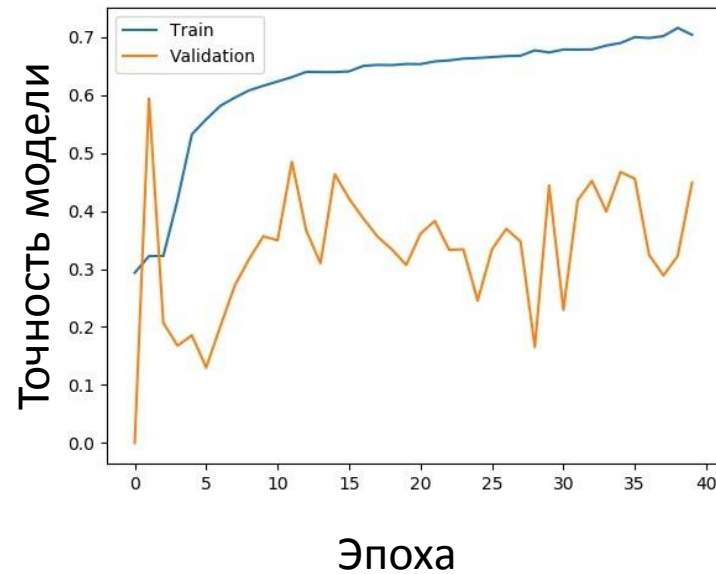


Рис 7: Точность и функция потерь при обучении с комбинациями линий

Заключение

- Для классификации спектральных линий блазаров были применены методы машинного обучения.
- Показано, что нейронные сети могут использоваться для классификации спектральных линий даже при низких интенсивностях.
- Модель успешно идентифицирует одиночные серии линий, но не может быть применена для спектров с комбинациями разных серий.
- В дальнейшем предлагается использование других методов машинного обучения, например, сегментации изображений.

Спасибо за внимание!

Архитектура модели

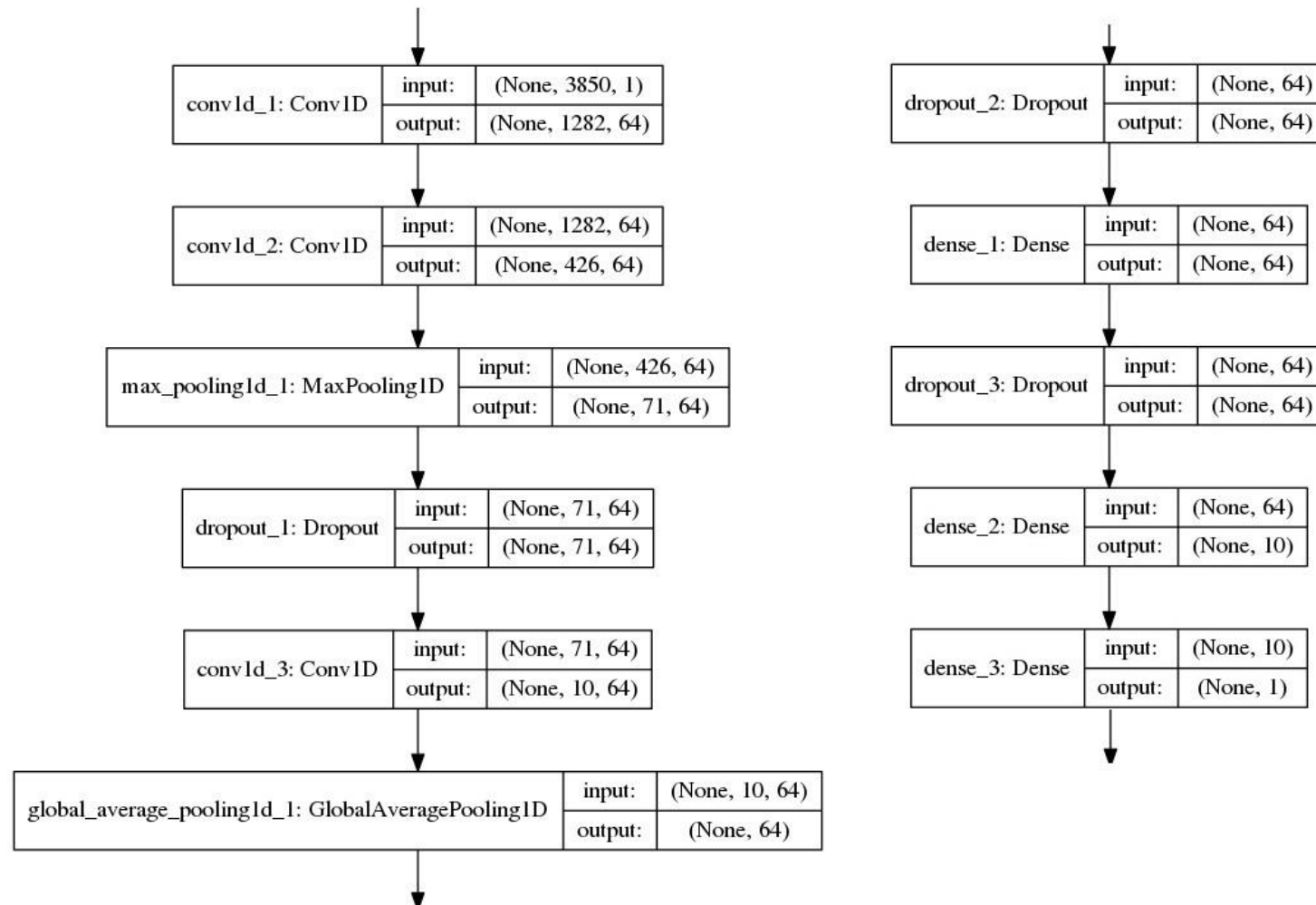


Рис. 8: Графическое изображение архитектуры

Наборы данных и параметры модели

- **Результаты-2:** На рис. 6 сеть обучалась на датасете с пятью разными сериями линий, по 2000 спектров на каждую серию. Красное смещение подобрано так, чтобы от каждой серии оставалась хотя бы одна линия.
- **Результаты-3:** На рис. 7 датасет такой же, как и в случае рис. 6, но с добавленными попарными комбинациями серий.
- В обоих случаях проверочная выборка составляет 10% от тренировочной.
- В качестве функции потерь использована кросс-энтропия
$$H(p, q) = - \sum_x \sum_{i=1}^n p_i(x) * \log(q_i(x))$$