

Использование свёрточных сетей для решения задачи классификации изображений

Роман Первутинский

Аннотация

Свёрточные сети являются эффективным методом классификации изображений. Данная работа исследует возможности разных по сложности сетей для решения задачи классификации дорожных знаков.

Abstract

Convolutional neural networks provide an effective method for image classification. This paper studies the capabilities of networks of different complexity for traffic signs classification.

Введение

В решениях многих простых задач классификации используются простые модели, такие, как логистическая регрессия и решающие деревья. Однако они плохо подходят для задач с большим признаковым пространством (изображения, видео, большие тексты). Например, даже изображение с разрешением 30×30 пикселей и тремя каналами имеет 2700 признаков.

В классификации изображений распространение получили свёрточные сети, однако они сложны (как для использования, так и вычислительно) и требуют тонкой подстройки под задачу. Для этого параллельно со свёрточными слоями используются слои с макс-пулингом и нормализацией, Dropout-слои.

Рассмотрим их применение на задаче классификации дорожных знаков. На вход подаются изображения низкого разрешения (для простоты масштабируем их до 30×30), которые нужно распределить по 43 группам.



Рисунок 1: Примеры изображений

Описание экспериментов

Нам дано два набора данных:

- Тренировочные (~40000 вхождений)
- Тестовые (~12000 вхождений)

Разделим их на датасеты для тренировки, валидации и тестирования и будем использовать их далее.

Логистическая регрессия

Для проверки корректности данных используем самую простую модель из доступных: логистическую регрессию (`sklearn.linear_model.LogisticRegression` с параметрами по умолчанию).

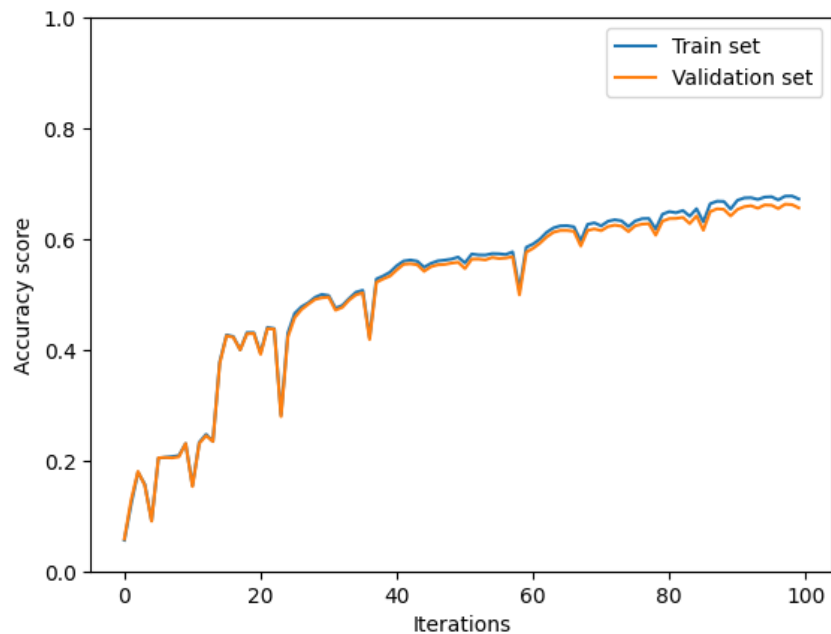


Рисунок 2: Изменение точности во время обучения модели 1

Как видно на рисунке, точность увеличивается очень медленно. После большого количества итераций, получаем итоговый результат:

- 92% на датасете для валидации
- 83% на датасете для тестирования

Простейшая нейронная сеть

В качестве нижней границы и чтобы проверить работоспособность функций для нейронных сетей, рассмотрим сеть с всего одним полносвязным слоем (`torch.nn.Linear`).

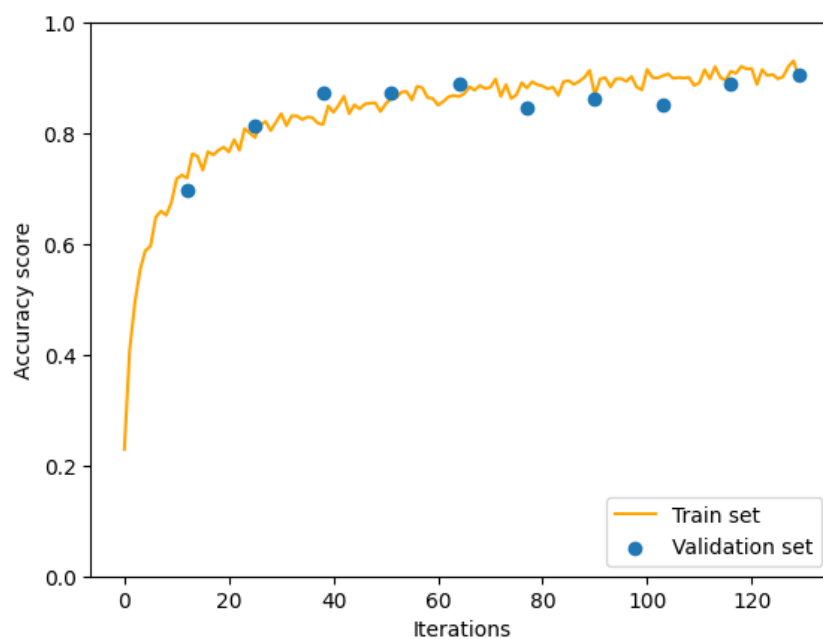


Рисунок 3: Изменение точности во время обучения модели 2

Можно заметить, что точность увеличивается гораздо быстрее, однако через какое-то время упирается в потолок.

Простая свёрточная нейронная сеть

Рассмотрим сеть с одним свёрточным слоем и двумя полносвязными слоями:

- Свёрточный слой
 - `nn.Conv2d` на 16 фильтров (с `nn.ReLU`)
 - `nn.MaxPool2d`
 - `nn.BatchNorm2d`
- `nn.Flatten`
- Полносвязные слои
 - `nn.Linear` с 256 нейронами (с `nn.ReLU`)
 - `nn.Linear` с логсофтмаксом

Такая сеть, несмотря на свою простоту, уже должна показывать хорошие результаты.

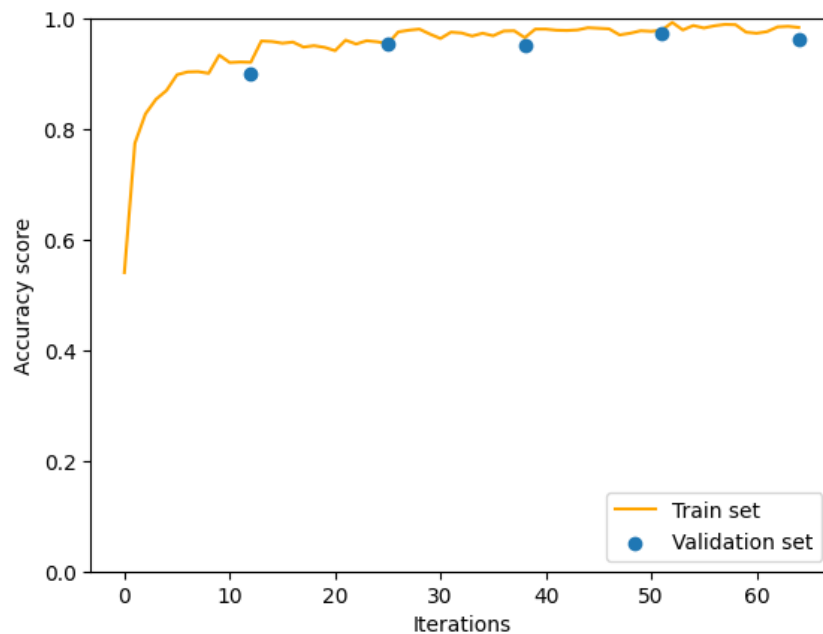


Рисунок 4: Изменение точности во время обучения модели 3

По сравнению с моделью 2, мы вдвое уменьшили количество эпох, однако точность модели уже к третьей эпохе превышает 90%.

Более сложная свёрточная нейронная сеть

Рассмотрим сеть, аналогичную предыдущей, но с двумя слоями `nn.Conv2d` (один на 16, второй на 32 фильтра). Основная цель — сравнить их.

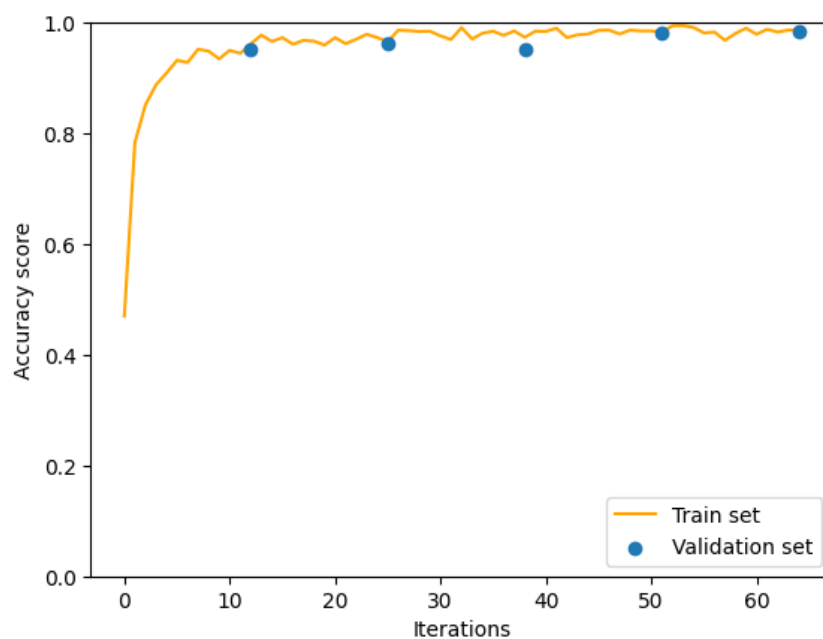


Рисунок 5: Изменение точности во время обучения модели 4

Финальные результаты

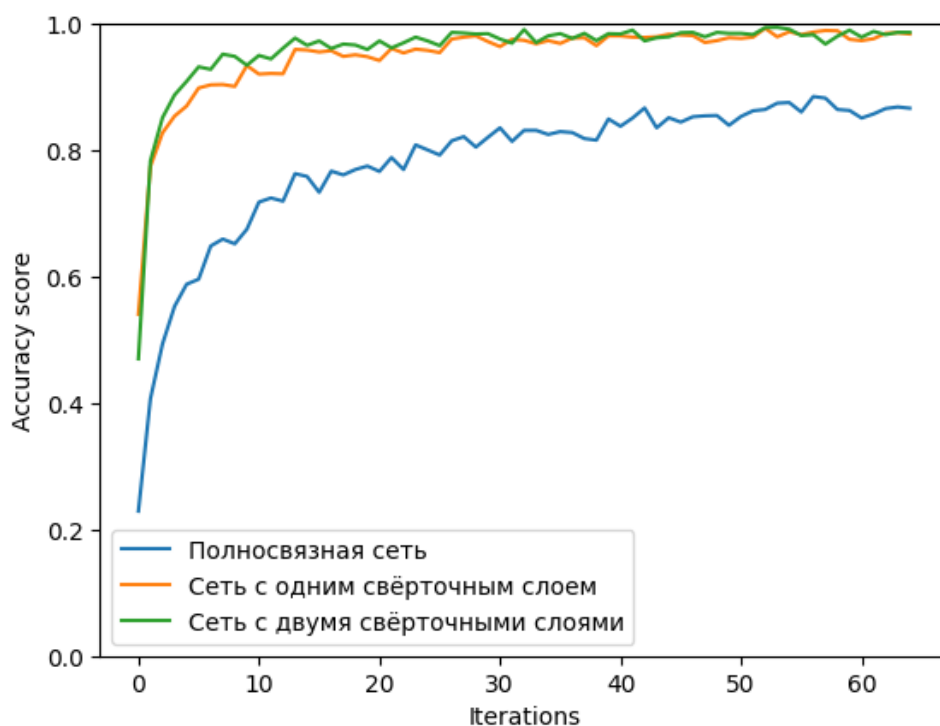


Рисунок 6: Точность на датасете для валидации

На этом графике можно видеть различие между моделями 3 и 4: вторая обучается быстрее, несмотря на то, что отличия между ними небольшие.

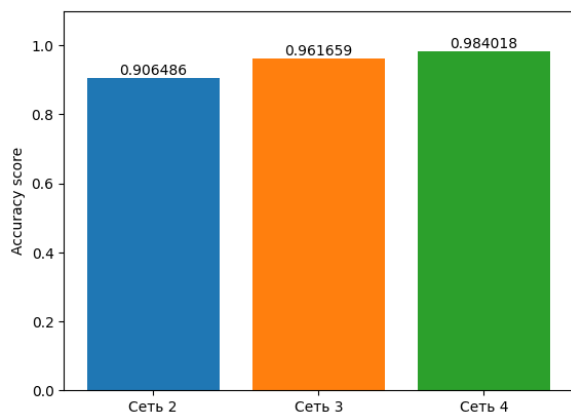


Рисунок 7: Точность на датасете для валидации

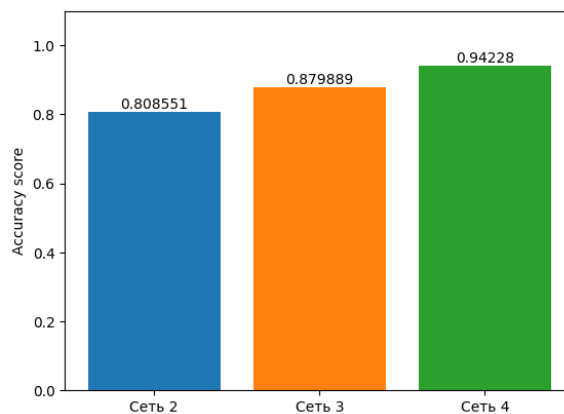


Рисунок 8: Точность на датасете для тестирования

Итоговая полученная точность на тестовых данных: 94%.

Заключение

В данной работе представлены модели хорошо работающих свёрточных сетей для решения представленной задачи. Точность можно увеличивать за счёт двух факторов:

- Количество эпох
- Количество и композиция слоёв

При увеличении либо того, либо другого обучение занимает гораздо больше времени. Поэтому для достижения наших целей это не целесообразно. Однако данная работа показывает, что свёрточные сети позволяют получить хорошие результаты даже при очень небольшом количестве слоёв, и есть потенциал для увеличения точности.