

Соревнование по распознаванию регистрационных знаков

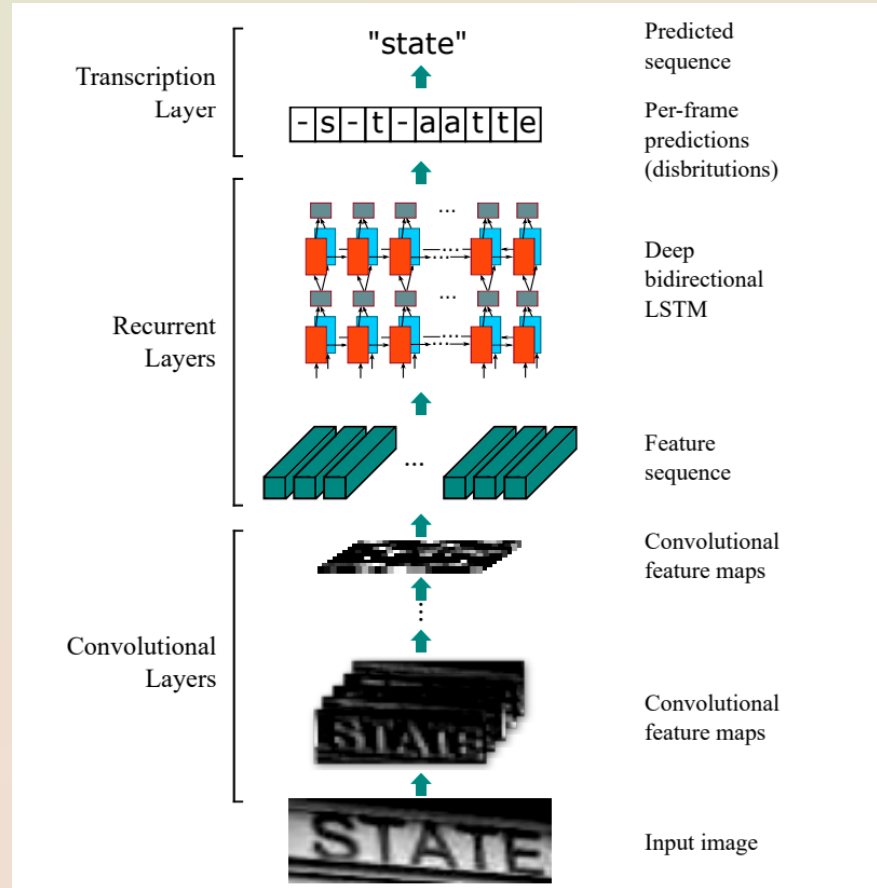


Public score: 0.04516

Private score: 0.05597

AI Masters
Сергей Руднев

Пайплайн решения



Извлечение признаков

Входное изображение пропускается через последовательность свёрточных слоев, в результате чего получается тензор признаков размера $C \times H \times W$ или последовательность длины W , каждый элемент которой соответствует определенной области на исходном изображении.

В качестве модели на этом этапе использовалась предобученная полносвёрточная часть ResNet18 (без двух последних слоёв AdaptiveAvgPool2d и Linear)

Предсказание вероятностей символов алфавита для каждого элемента последовательности

Полученная на первом шаге последовательность признаков пропускается через рекуррентную сеть, в результате чего на выходе образуется последовательность той же длины. Размер каждого элемента равен мощности алфавита + 1 (этот "+1" - символ пробел для использования CTC-Loss)

В качестве рекуррентной сети будем использовать GRU. Результат предсказаний после GRU дополнительно проведем через линейный слой для формирования итоговой матрицы с logits.

Декодирование и вычисление CTC-Loss

Декодируем полученные на предыдущем шаге распределения в итоговый текст и вычислим CTC-Loss между полученными распределениями и ground-truth-последовательностями.

Предсказанное распределение для картинки из валидации:



Обучение

Обучение построенной модели происходило на двух датасетах(с простыми номерами и со сложными). Было выбрано довольно большое число эпох(50), что и позволило добиться довольно хорошего результата