

Дигитално процесирање на слика

Тема:

Детекција на регистарски таблички и препознавање на карактери

Опис

Во последните неколку години, светот на компјутерската визија бележи огромни напредоци кои се должат на нови технологии. Заради овие достигнувања, сега е возможно компјутерите со помош на компјутерската визија да ги надминат луѓето во решавање на проблемите од типот на класификација, препознавање на слики, детекција на објекти во слики, препознавање на лица, читање на текст од слика и така натаму.

Со помош на конволуциски невронски мрежи, ќе направиме апликација која ќе биде способна да исчита регистарска табличка од било која слика со возило. Потребно е прво со помош на еден модел да се лоцира регистарската табличка на сликата, па потоа со друг модел за препознавање на карактери на слика да се отчита што пишува на табличката.

Длабоки конволуциски невронски мрежи

Длабоки конволуциски невронски мрежи се алгоритми за обработка на слики. Тие се способни да класифираат слики, да препознаваат објекти во слики, да препознаваат облици, да препознаваат карактери или текст и слично. Овие невронски мрежи се составени од различни типови на слоеви, како конволуциски (convolutional), поврзани (dense), регуларизациски (regularization), акумулациски (pooling), сплескувачки (flatten) и многу други.

Модели за детекција на регистарски таблички

За целите на овој проект, креирани се 3 модели кои ќе предвидуваат каде се наоѓаат регистарските таблички во дадените слики со помош на библиотеката Keras и програмскиот јазик Python.

Keras е бесплатна библиотека за креирање на невронски мрежи со API на високо ниво во Python и е интерфејс за библиотеката TensorFlow. TensorFlow е библиотека на ниско ниво која нуди интеракција со графички карти преку CUDA и е наменета за машинско учење и вештачка интелигенција.

Податочното множество (dataset) е достапно [тука](#), и се состои 433 слики од автомобили со регистарски таблички и ознаки за тоа каде се наоѓаат регистарските таблички во сликите во описаните со четири координати на краевите на правоаголникот кој ја опишува

регистарската табличка (bounding box). Координатите се во XML формат. Истото податочно множество е користено за сите модели за да може да бидат тестирали со него.

Моделите се тренирани на графичката карта NVIDIA RTX 2080 Ti со 11 GB VRAM (графичка меморија) на локален компјутер.

Подолу се описаны архитектурите на користените модели и параметрите со кои се тренирани.

Модел 1

ResNet, или Residual Network, е популарна конволуциска невронска мрежа често пати користена во светот на компјутерската визија. Оваа мрежа била многу успешна со тоа што добила прво место на натпреварот за класификација ILSVRC 2015 со грешка од само 3.57%.

Невронските мрежи кои работат во полето на компјутерската визија се многу комплексни, односно многу длабоки, и притоа тие дополнителни слоеви дозволуваат решавање на покомплексни проблеми затоа што различни слоеви може да бидат тренирани за различни задачи. Но, како што расте длабочината на невронската мрежа, се појавува проблемот на деградација, што значи дека, колку се зголемува длабочината на невронската мрежа, нивото на прецизност се намалува. Невронските мрежи ResNet се креирани за да го решат точно овој проблем, со тоа што „скокаат“ слоеви за да не се деградира прецизноста. Се покажува дека, како резултат на ова, ResNet невронските мрежи дозволуваат да се тренираат многу подлабоки невронски мрежи од што било претходно можно.

Постојат различни варијатни од ResNet кои работат на истиот концепт, но со различен број на слоеви. Ние ја користиме конволуциската невронска мрежа ResNet50, која е 50 слоеви длабока. Со помош на библиотеката Keras, ние можеме да вшитаме претренирана верзија од моделот на повеќе од еден милион слики од базата на слики ImageNet. Овој модел е способен да класифицира слики во 1000 разни категории на објекти, меѓу кои различни животни, тастатури, глувчиња, регистарски таблици и многу други.

Невронската мрежа ResNet50 прима слики кои се 224x224 пиксели големи во 3 канали (RGB, BGR, итн).

Нашиот модел се состои од следните секвенцијално поврзани слоеви:

- ResNet50
- GlobalAveragePooling2D
- Flatten

- Dense(1024 неврони, ReLU активацијска функција)
- Dense(4 неврони, ReLU активацијска функција)

Невронската мрежа е 54 слоеви длабока со вкупно 25,689,988 параметри, од кои 2,102,276 се тренирачки, а 23,587,712 не се тренирачки. За оптимизатор користиме Adam со стапка на учење 0.001 и за функција на загуба користиме средноквадратно растојание помеѓу предвидените и вистинските вредности.

Првиот слој е ResNet50, односно архитектурата на невронската мрежа ResNet50. Слојот GlobalAveragePooling2D ја враќа средната вредност на секоја од матриците на каналите на сликата. Слојот Flatten ги претвара мултидимензионалните низи во еднодимензионални. Слоевите Dense се целосно поврзаните слоеви со неврони и соодветна активацијска функција. На првиот слој оневозможуваме тренирање, затоа што е веќе претрениран на сликите од ImageNet и нема потреба уште еднаш.

Невронската мрежа се тренира на податочното множество во сооднос 80% тренирачко, 10% валидациско, 10% тестирачко множество. Сликите се подготвуваат за тренирање со тоа што ќе бидат претворени во 224x224 големина и вредностите на каналите ќе бидат скалирани од 0 - 255 во 0 - 1. На сличен начин се скалираат и координатите на регистарските таблички во ранг од 0 до 1.

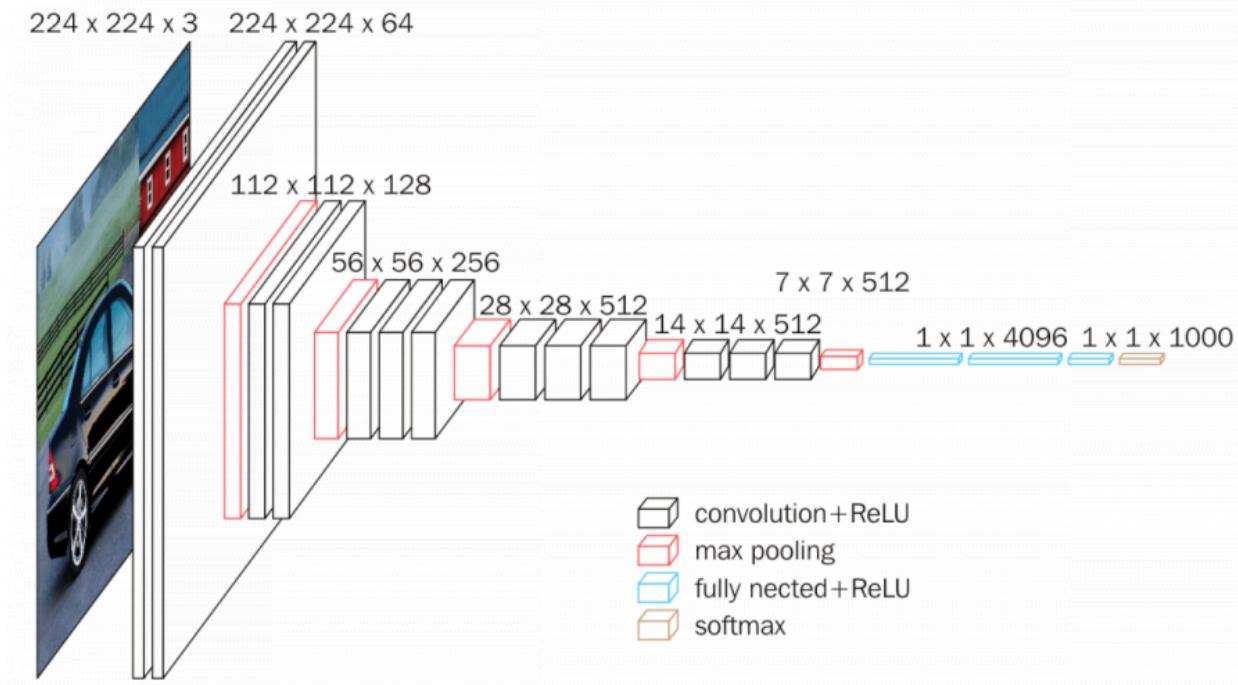
Моделот е трениран во 100 и 500 епохи за споредба на перформансите.

Model 2

VGG16 е друга популарна конволуциска невронска мрежа предложена од К. Симонјан и А. Зисерман од универзитетот Оксфорд. Овој модел е еден од најдобрите на натпреварот ILSVRC 2014 тестиран врз податочното множество на ImageNet со 92.7% прецизност при класификација на слики. Овој модел е многу добар за feature extraction (дескриптори на слики).

Невронската мрежа VGG16 е составена од 16 слоеви кои имаат тежини и 5 кои немаат. Филтрите на сите конволуциски слоеви се од големина 3x3, додека pooling (акумулацијски) слоеви се врз база на максимум со димензии 2x2 и чекор 2. Со помош на библиотеката Keras, можеме да вчитаме веќе претренирана верзија од моделот на сликите на податочното множество ImageNet.

Невронската мрежа VGG16 прима слики со големина 224x224 во 3 канали.



Сл. 1: приказ на слоевите на VGG16 невронската мрежа

Составот на слоевите на невронската мрежа VGG16 е:

- влез од големина ($224, 224, 3$)
- два конволуциски слоеви со 64 филтри
- maxpool слој
- два конволуциски слоеви со 128 филтри
- maxpool слој
- три конволуциски слоеви со 256 филтри
- maxpool слој
- три конволуциски слоеви со 512 филтри
- maxpool слој
- Dense(4096 неврони)
- Dense(1000 неврони, softmax активацијска функција)

Maxpool слојот е акумулацијски слој со максимално акумулирање, додека Dense слојот е целосно поврзан слој од неврони.

Нашиот модел се состои од следните секвенцијално поврзани слоеви:

- VGG16

- Flatten
- Dense(128 неврони, ReLU активацијска функција)
- Dense(64 неврони, ReLU активацијска функција)
- Dense(64 неврони, ReLU активацијска функција)
- Dense(4 неврони, sigmoid активацијска функција)

VGG16 слојот го оневозможуваме да биде трениран, затоа што доаѓа претрениран. Flatten слојот ги претвара мултидимензионалните низи во еднодимензионални.

Нашата невронска мрежа има 17,938,756 вкупно параметри, од кои 3,224,068 се тренирачки, а останатите 14,714,688 се нетренирачки. За оптимизатор користиме Adam со стапка на учење 0.001 и за функција на загуба користиме средноквадратно растојание помеѓу предвидените и вистинските вредности.

Невронската мрежа се тренира на податочното множество во сооднос 80% тренирачко, 10% валидациско, 10% тестирачко множество. Сликите се подготвуваат за тренирање со тоа што ќе бидат претворени во 224x224 големина и вредностите на каналите ќе бидат скалирани од 0 - 255 во 0 - 1. На сличен начин се скалираат и координатите на регистарските таблички во ранг од 0 до 1.

Моделот е трениран во 50 и 500 епохи за споредба на перформансите.

Модел 3

InceptionResNetV2 е популарна конволуцијска невронска мрежа која е најдлабока од сите што ги пробуваме, односно е 164 слоеви длабока. На првата слика се описаны слоевите на невронската мрежа. На втората слика се деловите Inception-A, Inception-B, Inception-C. На третата слика се описаны деловите Reduction-A и Reduction-B. На последната слика е даден Stem делот. Оваа невронска мрежа бара поголеми компјутерски ресурси за тренирање.

Невронската мрежа InceptionResNetV2 прима слики со големина 224x224 во 3 канали.

Моделот е составен од следните секвенцијално поврзани слоеви:

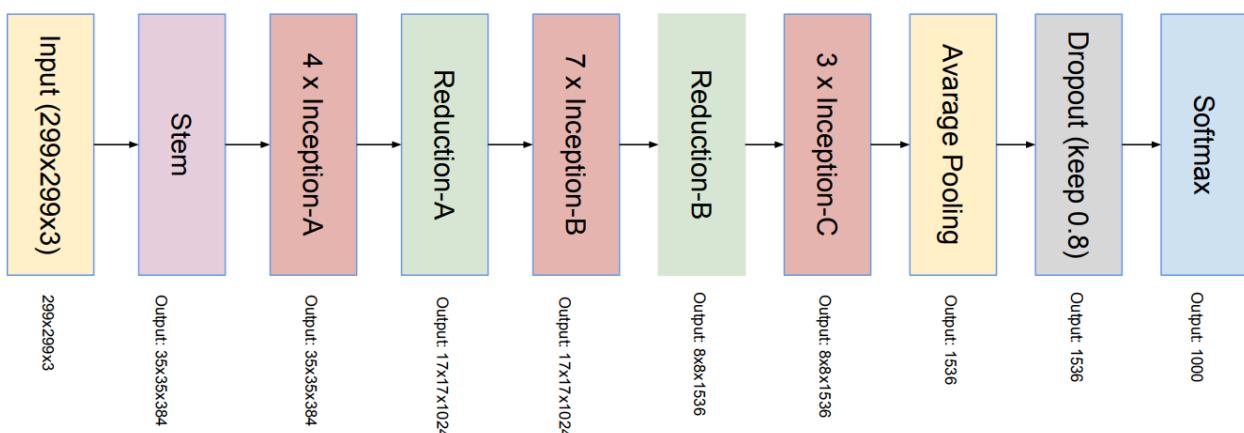
- InceptionResNetV2
- Flatten
- Dense(500 неврони, ReLU активацијска функција)
- Dense(250 неврони, ReLU активацијска функција)
- Dense(4 неврони, sigmoid активацијска функција)

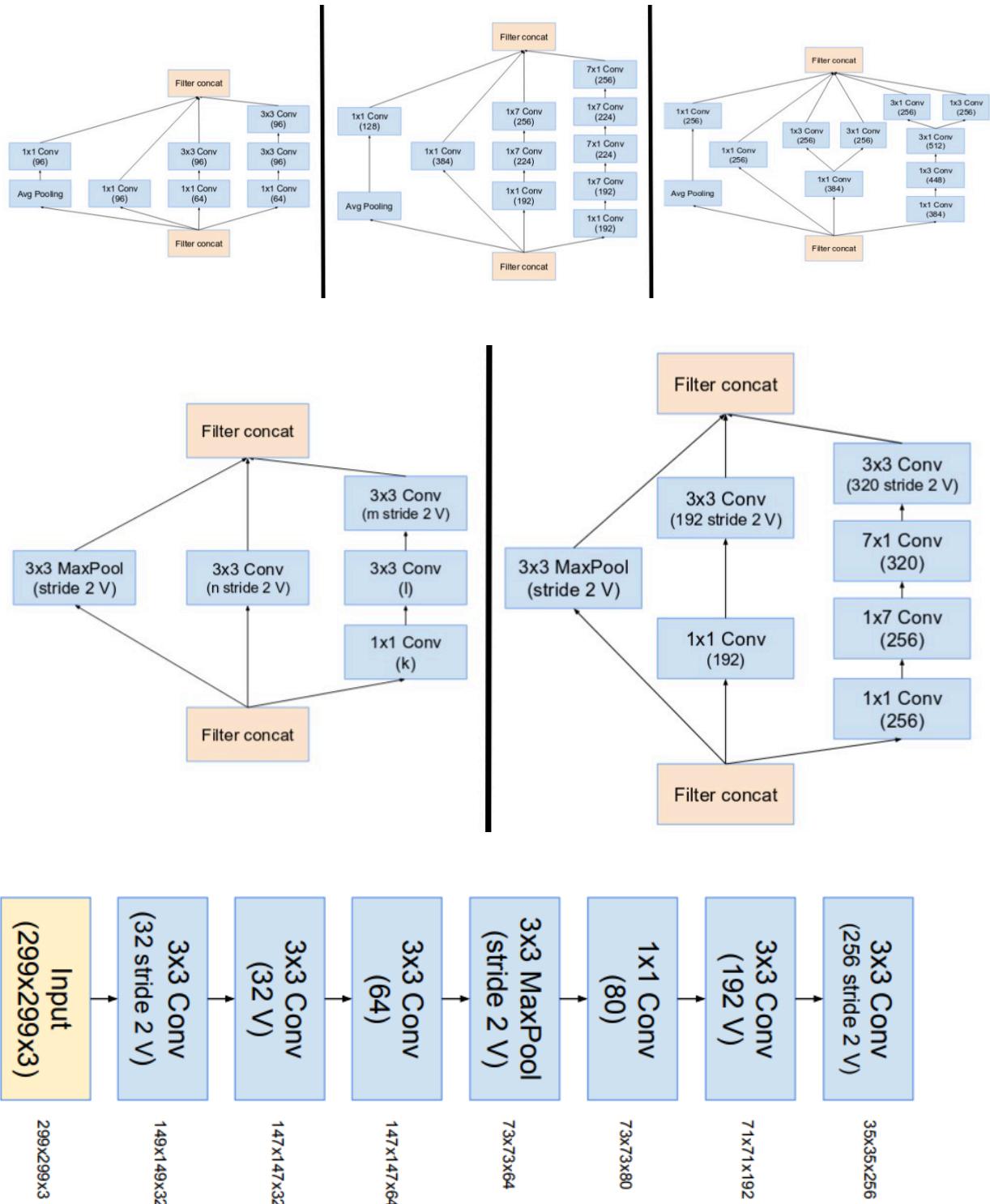
Поради фактот што оваа невронска мрежа е многу комплексна и многу длабока, не можеме да го вчитаме целото податочно множество (dataset) во меморија наеднаш. Затоа, креираме генератор за податочното множество со кој можеме да тренираме без да го имаме целото вчитано во меморија. Со делови од големина 64 се тренира моделот, односно 64 слики наеднаш вчитува во меморија.

Како и кај другите модели, ги користиме претренираните тежини на моделот од ImageNet податочното множество. Моделот има 73,663,490 параметри, од кои 73,602,946 се тренирачки, а 60,544 не се тренирачки. Го користиме оптимизаторот Adam со рата на учење од 0.0001. За функција на загуба го користиме средноквадратното растојание меѓу вистинската и предвидената вредност.

Невронската мрежа се тренира на податочното множество во сооднос 80% тренирачко, 10% валидациско, 10% тестирачко множество. Сликите се подготвуваат за тренирање со тоа што ќе бидат претворени во 224x224 големина и вредностите на каналите ќе бидат скалирани од 0 - 255 во 0 - 1. На сличен начин се скалираат и координатите на регистарските таблички во ранг од 0 до 1.

Моделот го тренираме во 15 и 50 епохи.





Сл. 2, 3, 4, 5: приказ на слоевите на InceptionResNetV2 невронската мрежа

Модел за препознавање на карактери

Користената библиотека за препознавање на карактери е EasyOCR, која е имплементирана во библиотеката PyTorch. Моделот на EasyOCR ја користи техниката CRAFT (Character-Region Awareness For Text detection) која служи за препознавање на текст. Овој алгоритам ги истражува „текстуалните области“ во секој регион со знаци и афинитетот меѓу знаците.

Карактерите се препознаваат со помош на ResNet моделот кој исто така и ние го пробавме за детекција на регистарските таблички.

Тестирање на моделите за детекција на регистарски таблици

На сите модели е користено истото податочно множество, и притоа е направена истата поделба на податочното множество (80% тренирање, 10% валидација, 10% тестирање). Затоа, моделите можеме да ги тестираме на истите слики кои тие не ги виделе при тренирање.

Модел 1

Првите 3 слики се од моделот трениран во 100 епохи, а вторите 3 се од моделот трениран во 500 епохи.

На првата слика и двата модели промашуваат целосно, но на другите слики моделот со 100 епохи подобро предвидува од моделот во 500 епохи, иако тој во 500 епохи дава помали bounding boxes, но не целата регистарска табличка се наоѓа во правоаголникот (bounding box).



Сл. 6: модел 1, 100 епохи, тест слика 1



Сл. 7: модел 1, 100 епохи, тест слика 2



Сл. 8: модел 1, 100 епохи, тест слика 3



Сл. 9: модел 1, 500 епохи, тест слика 1



Сл. 9: модел 1, 500 епохи, тест слика 2



Сл. 9: модел 1, 500 епохи, тест слика 3

Модел 2

Првите 3 слики се од моделот трениран во 50 епохи, а вторите 3 се од моделот трениран во 500 епохи.

И двета модели доста добро предвидуваат каде се наоѓа регистарската табличка на повеќето слики, но сепак и двете промашуваат некои, како на пример првата слика.



Сл. 10: модел 2, 50 епохи, тест слика 1



Сл. 11: модел 2, 50 епохи, тест слика 2



Сл. 11: модел 2, 50 епохи, тест слика 3



Сл. 12: модел 2, 500 епохи, тест слика 1



Сл. 13: модел 2, 500 епохи, тест слика 2



Сл. 14: модел 2, 500 епохи, тест слика 3

Модел 3

Првите 3 слики се од моделот трениран во 15 епохи, а вторите 3 се од моделот трениран во 50 епохи.

Моделот во 50 епохи предвидува малце подобро од моделот во 15 епохи, но не е значително подобро. На некои слики и моделот во 15 епохи предвидува поголем дел од регистарската табличка.



Сл. 15: модел 3, 15 епохи, тест слика 2



Сл. 16: модел 3, 15 епохи, тест слика 1



Сл. 17: модел 3, 15 епохи, тест слика 3



Сл. 18: модел 3, 50 епохи, тест слика 2



Сл. 19: модел 3, 50 епохи, тест слика 1



Сл. 19: модел 3, 50 епохи, тест слика 3

Тестирање на моделот за препознавање на карактери

Користејќи го моделот 2 за детекција на регистарските таблички и моделот за препознавање на карактери, еве неколку слики и исчитан текст од тестирачкото множество:



Сл. 20: пример тест сликата

Излез:

Lo

83L4AF

Dautojctyid #ERToro

Fnsentee



Сл. 21: тест слика 4

Излез:

LR33 TEE



Сл. 22: тест слика 5

Излез:

@LZC N 5617



Сл. 23: тест слика 1

Излез:

Не беше препознаен текст на оваа слика



Сл. 24: тест слика 3

Излез:

HH15BD8877

Користена литература

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arXiv: 1512.03385, 2015.

Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv: 1602.07261v2, 2016.

Karen Simonyan, Andrew Zisserman. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. arXiv: 1409.1556, 2015.

Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee. Character Region Awareness for Text Detection. arXiv: 1904.01941, 2019.

Baoguang Shi, Xiang Bai, Cong Yao. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. arXiv: 1507.05717, 2015.