

[Detection of Respiratory diseases using respiratory sounds]

정새하, 주지윤, 최은진

인하대학교 컴퓨터공학과

요 약

기술의 진보로 인해 e-health 에 대한 관심이 늘어남에 따라 있으며 이로 인한 e-health device 소프트웨어에 대한 수요가 증가하고 있다. 청진의 중요성과 간단한방식에 착안하여 호흡기 소리로 청진을 할 때 판단의 요소가 되는 부잡음인 천명, 수포음 등이 검출하고 분류가 가능한 점을 이용하여 머신 러닝, 노이즈 제거, 주파수 분리 및 추출기술 등을 통해 폐렴, 기관지염과 같은 몇 가지 호흡기 질환을 자동으로 진단하는 프로그램을 개발하고자 한다.

1. 서론

[개발의 동기]

의학기술의 발달 및 정보처리의 기술이 빠르게 발전함과 함께 portable healthy device 기술의 진보로 e - health 에 관한 관심과 수요가 높아지고 있다. 이러한 배경속에서, 청진이 실시가 용이하며 오디오 기술과 밀접히 연관되어 있다는 점과 내과진료시 항상 실시될 만큼 중요한 지표라는 점에 주목하였다. [1]에 따르면 현재의 청진은 청진기를 사용하는 관찰자와 환자사이의 소통으로 국한되어 있어 관찰자의 주관적인 판단으로만 이루어지고 있다. 이러한 폐음을 녹음하여 정보화된 방식으로 접근한다면 상황에 따른 소리의 양상을 자동적으로 분류하여 조금 더 구체화된 분석을 할 수 있을 것이다. 또한 [2,3]에 따르면 노이즈와 청진기의 소리전달기술의 제한으로 어려움을 겪고 있는 청진에서 소리의 증폭, 노이즈 제거 기술을 사용하여 보다 깨끗한 청진음을 바탕으로 폐음의 특징적인 요소들을 분류하여 정확한 분석을 이끌어 내는데 효과를 볼 수 있을 것이다. Neural network 와 machine learning, 푸리에 변환, frequency decomposition 을 통한 computerized 된 폐음 분석 software 는 e-health 의 강력한 장치가 될 것이다.

[개발의 필요성 및 기대효과]

호흡기 질환의 진단에서 폐음 청진은 가장 기본적인 검사이다. 청진도구만 있다면 언제든지 실시될 수 있으며 연속적인 검사도 용이하기 때문에 환자의 경과를 관찰하기에 적합하다. 간단한 검사로 편리성을 제공하면서도 질환의 판단의 중요한 지표가 되는 중요한 검사이다. ^{3,4)}국내에서 시행된 개원 가정의의 직무분석과 관련된 연구에서도 수행 작업별 중요도에서 폐음 청진을 포함하는 신체 진찰은 상당한 중요도를 가진 것으로 나타났다.

호흡의 강도는 폐용량, 혈류율, 검출부위에 따라 변화하며, 가슴에서 청진 된 폐음은 폐의 환기의 분포상태, 가슴 조직의 두께 등의 차이에 의해 소포성의 흡기음이 일정치 않은 변화를 보인다. 이러한 청진은 유용한 정보를 제공하나 청진음과 진단에 대한 정확한 이론이 정립 되어있지 않은 상태이다. 이러한 시점에서 호흡음의 정량적인 분석과 진단관계에 대한 연구의 필요성이 대두되고 있다. 본 연구는 녹음된 폐음의 데이터셋을 기반으로 머신 러닝을 이용하여 미리 정의된 7 가지의 폐질환을 추측하는 소프트웨어를 개발하고자 한다

[관련연구]

Hai Chen(2019)의 Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks 연구는 OST 와 ResNet 네트워크를 사용하여 부잡음인 천명(wheeze)과 수포음(crackle) 그리고 호흡음을 분류(classification)하는 법에 대해 다루고 있다. 정확도가 높은 결과를 냈으나 본 연구에서 진행하려는 것과는 달리 분류해낸 소리를 통해 병을 찾아내지는 않는다.

E. Andrès (2018)의 Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0 연구는 현대의 인공지능과 인공 신경망, 퍼지(fuzzy) 시스템 및 유전자 알고리즘을 기반으로 청진 소리를 수집하고 알려진 병리학적 소리를 제공하는 데 사용되는 다양한 기술들에 대해 다루고 있다.

Nabila Husna Mohd Johari(2018)의 Distinctive Features for Classification of Respiratory Sounds Between Normal and Crackles Using Cepstral Coefficients 연구에서는 Mel-frequency Cepstral Coefficient(MFC)를 사용하여 정상적인 호흡소리와 수포음이 섞인 호흡소리에서 feature 을 추출해낸다. 또한 coefficient 기반의 cepstral 의 평균과 표준 편차(SD)와 같은 통계적 계산을 통해 정상적인 소리와 수포음 사이의 차이를

구별한다.

Dalal Bardou(2018)의 Lung sounds classification using convolutional neural networks 연구는 폐의 소리를 분류하는 법에 대해 다루었으며, 손으로 제작된 feature 기반의 분류기보다 CNN(Convolutional Neural Network)을 이용한 방법이 효과적이라고 설명한다. 하지만 이 연구도 마찬가지로 소리를 분류해 낼 뿐 그를 통해 관련 병을 찾아내지는 않는다.

[개발하고자 하는 기술]

우리는 디지털 청진기 및 기타 녹음 기술을 사용하여 녹음한 호흡기 소리를 기계 학습을 사용하여 천식, 폐렴 및 세기관지염과 같은 호흡기 질환을 자동으로 진단하는 프로그램을 만들고자 한다.

호흡기 소리로 청진을 할 때에는 크게 폐음과 부잡음을 나눠서 듣는다. 기존의 연구들은 대부분이 부잡음 중에 crackle 과 wheeze 의 검출에 집중하였다. 우리는 부잡음의 검출을 넘어서 병의 진단에 직접적인 도움을 줄 수 있도록 병에 걸렸을 확률 도출을 목적으로 한다.

이 주제는 [5]에서 소개하는 데이터셋을 사용하여 ICBHI 2017 Challenge 의 목적이기도 했던 주제이기도 하다. 하지만 기존의 연구들은 위처럼 병 자체의 진단에 집중하지 않았고 딥러닝 기술을 접목시키기보다 feature 선정에 집중하였다는 부분에서 우리와 차별점이 있다.

우리의 검출 질병은 총 7 가지로 URTI(상부 호흡기 감염), Asthma(천식), COPD(만성 폐쇄성 폐질환), Pneumonia(폐렴), Bronchiectasis(기관지확장증), Bronchiolitis(모세기관지염), LRTI(하부 호흡기 감염)이 있다.

여기서 COPD 의 경우 crackle 과 wheeze 모두 나타나기도 하고 둘 중 하나만 나타나기도 한다. 따라서 부잡음과 같은 호흡기 소리에서 특유의 패턴만 검출해내는 것이 아니라 전체 파형을 봐야한다.

Crackle 은 수포음이라고 하며 정상적인 폐음이 들리지 않으면서 나는 짧고 폭발적인 Non Musical Sound 이고, wheeze 는 천명음이라고 하며 쉼쉼 거리는 것을 말한다.

기존의 연구들은 이런 특징 검출에만 집중을 하였기 때문에 LPCC, MFCC, 새로 제안하는 transformer 등으로 feature 를 뽑은 뒤에 CNN 이나 perceptron 등의 머신 러닝 모델에 넣는 식으로 모델을 설계하였다.

우리의 모델은 생활 노이즈를 제거하기 위한 전처리를 거친 후 RNN 계열 딥러닝 네트워크에 넣는 방향으로 설계하여 전체 파형을 보다 잘 볼 수 있도록 할 예정이다.

네트워크는 전체적으로는 LSTM 을 쌓은 구조를 사용할 예정이며 실험환경에 여유가 된다면 부잡음이 병의 진단에 중요한 역할을 하는 만큼 부잡음을 검출하여 해당하는 부분들에 attention 을 주는 방향도 고려하고 있다.

최종 결과값은 위의 질병 7 가지와 건강한 상태를 더한 8 가지 상태 중 가장 가능성이 높은 상태와 그 상태로 진단할 확률 값을 제공할 예정이다.

[개발계획]

연구에 사용될 데이터셋은 ICBHI 2017 Challenge 의 'Respiratory Sound Database'로 kaggle 에도 올라와 있다. 해당 데이터셋에는 126 명의 환자에서 수집한 1864 회의 호흡 주기가 포함된 총 5.5 시간의 호흡 녹음본이 존재한다. 이 중 1864 개는 수포음(crackle), 886 개는 천명(wheeze), 그리고 506 개는 수포음과 천명 모두 포함되어 있다. 또한 깨끗한 호흡 소리와 실제 생활 조건과 같은 잡음들이 포함된 소리가 함께 존재한다. 그렇기 때문에 잡음이 섞인 소리일 경우 노이즈 캔슬링(Noise Cancelling, NC)을 통해 되도록 깨끗한 호흡 소리를 추출할 수 있도록 할 계획이다.

호흡음은 크게 정상 호흡음(Breathing sound), 이상 호흡음(Abnormal breath sound) 그리고 부잡음으로 나눌 수 있다. 본 연구는 앞서 얻은 깨끗한 호흡 소리를 정상 호흡음과 이상 호흡음을 포괄하는 호흡음과 그 이외에 들리는 소리인 부잡음으로 나눈다. 그 후, 부잡음 또한 크게 수포음(crackle)과 천명(wheeze) 2 가지로 분류하여 각 소리의 패턴을 찾아 정립한 다음 딥러닝(Deep Learning)을 사용하여 분석할 계획이다.

또한 본 연구에서 찾아내려고 하는 호흡기 질환을 가진 환자의 호흡음이 어떤 특징을 나타내는지, 어느 음에서 이상을 보이는지 분석하고 학습할 것이다.

그리하여 최종적으로 학습된 결과를 바탕으로 해당 데이터가 COPD(만성 폐쇄성 폐질환), LRTI(하부 호흡기 감염), URTI(상부 호흡기 감염), Asthma(천식), Pneumonia(폐렴), Bronchiectasis(기관지 확장증), Bronchiolitis(모세기관지염)과 같은 호흡기 질환을 가진 환자의 호흡음인지 건강한 사람의 호흡음인지 구별하도록 할 계획이다.

2. 개발내용

개발하고자 하는 모델은 녹음된 호흡기 소리가 데이터로 주어졌을 때, 이를 분석하여 질병가능성을 예측하고 tracking 하는 모델이다.

이때 데이터는 아래의 전처리 과정들 중에 하나를 통해 최적화하여 아래의 신경망 구조의 input으로 주어진다. 이후 신경망을 거쳐 앞서 언급하였던 질병들 중 예측되는 질병과 그에 대한 확률을 결과로 출력한다.

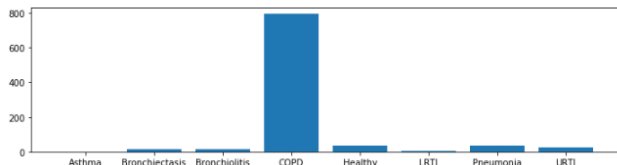
2.1 전처리

전처리를 통해 데이터의 노이즈를 줄이고 신경망 구조에 알맞은 형태로 가공한다면 더 나은 결과를 얻을 수 있다. 이때, 적합할 것이라 예상되는 전처리 과정을 다음과 같이 명시하고, 실험을 통해 각각의 방법을 채택하였을 때, 제일 높은 정확도를 보이는 방법을 최종적인 전처리 과정으로 한다.

2.1.1. Data Crop

연구에 사용된 데이터셋은 ICBHI 2017 Challenge 의 ‘Respiratory Sound Database’ 로 데이터의 분포가 다음과 같다.

```
Asthma : 1
Bronchiectasis : 16
Bronchiolitis : 13
COPD : 793
Healthy : 35
LRTI : 2
Pneumonia : 37
LRTI : 23
<BarContainer object of 8 artists>
```



그림과 같이 데이터셋이 COPD 에 집중되어 있어 편향된 학습을 야기하므로 이를 보완하기 위해 부트스트래핑 방법을 사용한다. 부트스트래핑은 데이터셋의 분포가 고르지 않은 경우 사용되며 데이터셋에서 중복을 허용하는 random sampling 을 반복하여 병들이 모집단의 데이터 분포와 비슷한 확률로 실험 될 수 있도록 한다.

또한 이미지데이터에 대해 동일한 이미지를 회전 이동, 평행이동, 밝기조절, 확대/축소 등을 통해 이미지를 얻어내고 그 데이터를 train set 에 추가하여 학습시키는 data augmentation 에 대해서도 고려되었으나 소리 signal 은 시간과 주파수 domain 으로 정의되어 표현되어 있으므로 이러한 변경이 불가능 하다. 임의의 노이즈를 추가하거나 시간 축으로 늘리는 것은 데이터가 다른 병의 데이터로 바뀔 수 있는 위험이 있으므로 길이가 제각각인 sound 파일을 같은 시간 간격으로 자르는 것으로 data 에 변화를 준다.

data 의 crop 은 다음의 두 가지 방식이 사용된다. 첫번째는, 가장 짧은 데이터의 길이, 즉 7 초를 기준으로 오버랩 구간 없이 audio 파일을 자른다. 두번째는 4 초를 기준으로 audio 파일을 자르고 1 초씩 오버랩하여 audio 데이터셋을 구성한다. 앞의 방법 중 가장 효율적인 방식을 사용하기로 한다.

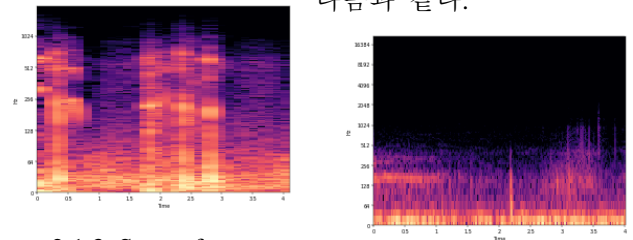
2.1.2. STFT

시간영역의 신호를 작은 세그먼트로 나누어 그 동안은 신호가 stationary 하다고 가정하여 푸리에 변환을 한다. 가우시안(Gaussian) 윈도우 함수를 사용하며 윈도우길이가 고정되기 때문에 일정한 해상도

를 가진다. 또한 시간영역의 해상도와 주파수 영역의 해상도는 반비례하게 된다. STFT 의 식은 다음과 같다.

$$STFT(\tau, \Omega) = \int_{-\infty}^{\infty} [f(t)w^*(t-\tau)]\exp(-j\Omega t)dt \quad (1)$$

여기서 $f(t)$ 는 시간 영역의 신호이며, $w(t)$ 는 윈도우 함수이다. 결과로 주파수와 시간 영역의 신호가 $M*N$ 인 STFT 행렬이 얻어진다. 하지만 이는 너무 차원이 크기 때문에 $K*J$ 차원으로 압축시켜 준다. 이때 $K*J$ 의 하나의 셀은 $M/K * N/J$ 크기의 행렬의 모든 셀의 값을 합하여 하나의 셀로 압축한 행렬이다. 이때 이 행렬은 신경망의 입력으로 사용될 특성 벡터가 된다. 따라서 STFT 는 특정과형에서 주파수 성분의 분포를 파악하는데 용이하므로 Local 의 특징을 추출하기에 적합하다. STFT 처리된 이미지는 다음과 같다.



2.1.3. S-transform

S-transform 은 시간주파수 분포로, 단기 푸리에 변환 (STFT)을 가우시안 함수를 사용하여 일반화한 것으로, 웨이블릿 변환(wavelet transform)을 연장하고 단점을 일부 극복한 방법이다. STFT 의 확장된 방법으로 가장 큰 차이점은 transform 에 사용되는 윈도우이다. STFT 는 fix 된 사이즈의 윈도우를 통해 transform 되기 때문에 원하지 않는 정보의 손실이나 노이즈가 생성될 수 있다. 특히 고주파에서 distortion 이 발생하게 된다. 하지만 S-transform 은 가변하는 크기의 윈도우를 채택하여 위와 같은 단점을 극복한다. 즉, cross-term problem 이 없으므로 저주파와 고주파 모두 적절한 transform 이 일어날 수 있으며 따라서, signal 의 특징을 파악하는데 좀 더 용이하다. 윈도우의 사이즈는 sin 파의 modulation 을 시간 축에 고정하고 이것으로 s 변환에서 가변하는 가우스 윈도우 확장 및 번역을 국소화 함으로써 정해진다. gabor 변환보다 더 나은 명료성을 산출한다. window 가 좁으며 고주파에 대해 더 나은 시간 해상도를 산출하며, 저주파에서는 더 나은 주파수 해상도를 가진다.

$$S_x(t, f) = \int_{-\infty}^{\infty} x(\tau)|f|e^{-\pi(t-\tau)^2 f^2} e^{-j2\pi f\tau} d\tau$$

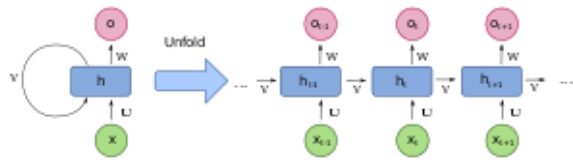
2.2 분류모델

우리의 분류 모델은 LSTM 구조, ResNet, CRNN 이 고려 대상이며 이 중 제일 적절한 모델을 채택하기로 한다. 다음은 이러한 Machine Learning 모델에 대

한 설명이다.

2.2.1 Recurrent Neural Network

RNN 은 은닉계층의 유닛들이 방향성 사이클을 포함하는 인공신경망의 일종이다. 이 구조는 은닉계층에 저장되어 있는 과거의 정보들을 현재의 입력값과 결합하여 사용할 수 있게 함으로써 입력 시퀀스에서 유닛 간의 상호적인 정보를 어느정도 유지할 수 있다. 인공신경망 학습에 적용되는 Back Propagation 알고리즘이 RNN 의 재귀적 구조에 의해 BPTT(Backward Propagation Through Time)으로 파생되었다. BPTT 알고리즘은 가장 먼저 신경망의 각 유닛들을 입력 시퀀스의 시간단계 만큼 펼쳐서 순환 구조를 순차적으로 나타낸다. 그 다음 back propagation 을 적용하여 각각의 weight 를 학습한다. 각 유닛의 가중치가 시간단계에 따라 변하지 않는다는 것이 중요하다. 그림 1 은 unfold 과정을 나타낸다. 그림과 같이 시퀀스의 길이가 늘어나게 되면 chain rule 에 의해 계산된 기울기를 전파해야 되는 레이어의 수가 증가하게 된다. 또 전파 시에 활성화 함수를 거쳐야 하므로 최종적인 기울기는 극도로 작아지는 vanishing gradient 문제를 마주하게 된다. 이에 대한 방안으로 최근 제안된 LSTM 구조는 가장 효과적인 방안을 제시하고 있다.



[그림 1] RNN

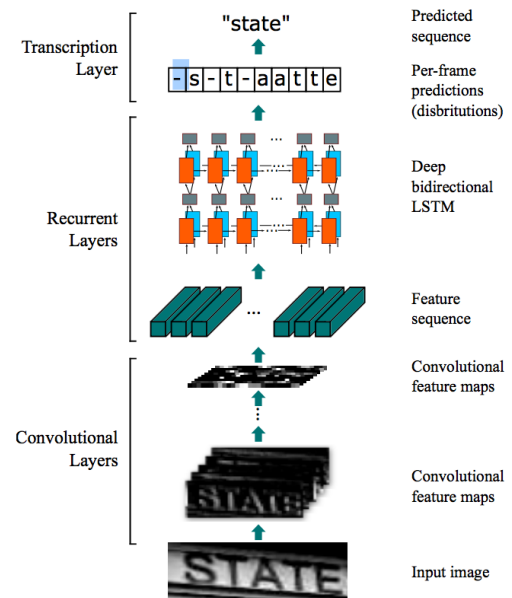
2.2.2 LSTM

LSTM 은 Long Short Term Memory network 의 약자로 Hochreiter & Schmidhuber 에 의해 제안된 RNN 구조이다. 전통적인 RNN 구조에서 은닉계층의 유닛들을 LSTM 블록으로 대체시킨 형태와 같다. LSTM 의 블록들은 재귀적 구조를 가진다. LSTM 블록 내부는 재귀적 구조를 가진 기억소자와 입력 게이트, forget 게이트, 출력 게이트 이렇게 총 3 가지의 게이트 유닛으로 이루어져 있다. LSTM 은 RNN 과 마찬가지로 은닉변수를 거쳐 최종 출력 값을 계산하지만, 이전에 게이트 유닛들을 이용하여 정보의 흐름을 조절한다. forget 게이트는 기억소자에 저장되어 있는 소자변수를 얼마나 잊어버릴지 결정한다. forget 게이트는 sigmoid 함수에 h 와 x 의 weighted sum 을 넣은 형태이다. forget 게이트 값이 0 이면 모두 잊어버린다는 의미를 가진다. 입력 게이트는 새로운 정보를 얼마나 소자변수로 가져갈지 결정하는 게이트이다. 새로운 정보는 tanh 함수를 이용한다. forget 게이트와 입력 게이트를 통해 조정된 과거와 새로운 정보는 식(4)에서와 같이 소자변수로 계산된다. 마지막으로 출력 게이트는 필터링을 거친 소자변수의 유용한

정보를 최종 은닉변수로 출력하게 된다. 이를 통해 gradient 가 소멸되어 학습되지 않는 문제를 해결하면서도 효과적인 RNN 구조를 사용할 수 있다.

2.2.3 CRNN

CRNN[8]은 DCNN(Deep Convolutional Recurrent Neural Network)와 RNN 의 조합으로 Convolutional Recurrent Neural Network 로 명명된다. CRNN 은 이미지에서 시계열 데이터를 인식하도록 특별히 설계된 새로운 신경망 모델이다. CRNN 은 다음의 그림과 같이 Transcription Layer, Recurrent Layer, Convolutional Layer 로 이뤄진다. Convolutional Layer 에서는 입력 이미지에서 convolution 과 max-pooling layer 을 통해 feature map 을 생성합니다. 여기서 추출된 feature sequence 는 Recurrent Layer 의 입력이 되며 sequence 내의 각 vector 는 수용 필드와 관련되며, 그 영역에 대한 이미지를 설명하는 것으로 이해될 수 있다. RNN 은 sequence 안에서 문맥 정보를 캡처하는 강력한 기능을 갖추는데, 이는 이미지 기반 sequence 인식을 위해 상황 별 신호를 사용하는 것이 각 symbol 을 독립적으로 처리하는 것보다 안정적이라는 장점을 제공한다. 또한 RNN 은 back-propagate 가 가능하므로 CRNN 에서 Recurrent layer 와 convolution layer 를 함께 학습할 수 있다. 이 때 Vanishing Gradient 문제를 방지하기 위해 RNN 모델로 양방향 LST 또는 GRU 를 사용한다. 양방향이므로 양방향의 컨텍스트가 유용하고 서로 보완적이다. 우리가 사용한 모델의 CNN 부분의 구조는 2.2.4 에서 설명할 ResNet 의 구조를 레이어 수 같은 부분들을 다르게 하여 사용하였다.



[그림] CRNN

2.2.4 ResNet

ResNet 은 Residual Block 의 개념을 CNN 에 도입함으로써 Gradient Vanishing 문제를 해결하려 한 인

참고문헌

- [1] Murphy RL, Vyshedskiy A, Power-Charnitsky VA, Bana DS, Marinelli PM, Wong-Tse A, Paciej R(2004). "Automated lung sound analysis in patients with pneumonia." *Resp Care*. 49(12):1490–1497.
- [2] Grenier MC, Gagnon K, Genest J, Durand J, Durand LG. (1998)"Clinical comparison of acoustic and electronic stethoscopes and design of a new electronic stethoscope." *Am J Cardiology*. 81:653–56.
- [3] Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, Malmurugan N. "Neural classification of lung sounds using wavelet coefficients."(2004) *Comput Biol Med*. 34:523–537.
- [4] Jauhar S(2006), "The Demise of the Physical Exam." *N Engl J Med*, 354(6):548-51
- [5] Mangione S, Duffy Fd(2003), "The teaching of chest auscultation during primary care training: has anything changed in the 1990s?", *chest*, 124(4):1430-
- [6] Rocha BM et al. (2019) "An open access database for the evaluation of respiratory sound classification algorithms" *Physiological Measurement* 40 035001
- [7] Hai Chen, Xiaochen Yuan, Zhiyuan Pei, Mianjie Li & Jianqing Li. (2019) "Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks" *IEEE Access*, vol7. 10.1109/ACCESS.2019.2903859
- [8] Baoguang Shi, Xiang Bai, Cong Yao(2015) "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition" *IEEE* 2015 10.1109/TPAMI.2016.2646371