

# Design and Evaluation of License Plate Recognition Model with Pseudo-labeling and BiLSTM

## ABSTRACT

Deep learning algorithms are being used to detect and recognize car license plate for intelligent vehicle control systems. License plate recognition goes through the stages of object detection and character recognition. Character recognition is mainly trained using supervised learning method while it consumes a lot of cost in labeling the collected dataset. In this paper, we propose a license plate detection and recognition model using a semi-supervised learning method. The proposed model consists of a VGG-16-based detector and a CNN- and BiLSTM-based recognizer. The proposed model is an end-to-end model that does not require third-party modules such as OCR. The proposed model utilizes unlabeled data and shows higher performance than when only labeled data is used. The proposed model can also be used for other semi-structured documents and function as a labler for large-scale data.

**Keywords :** license plate, semi-supervised learning, object detection, character recognition

## Pseudo-labeling 과 BiLSTM 을 이용한 자동차 번호판 인식 모델의 설계와 평가

### 요 약

지능형 차량 관제 시스템을 위한 자동차 탐지 및 인식 기술에 딥러닝 알고리즘이 활용되고 있다. 번호판 인식 기술은 객체 탐지와 문자 인식의 단계를 거치며 문자 인식의 경우 지도학습 방식으로 주로 훈련된다. 지도학습 방식의 훈련은 수집된 데이터에 대한 라벨링 작업에 많은 비용을 소모한다. 본 논문에서는 준지도학습 방식을 통한 번호판 탐지 및 인식 모델을 제안한다. 제안 모델은 VGG-16 기반의 탐지기와 CNN 및 BiLSTM 기반의 인식기로 구성되며 OCR 과 같은 별도의 모듈을 필요로 하지 않는 종단간 모델이다. 제안 모델은 레이블 정보가 없는 데이터를 활용하여 레이블 정보를 가진 데이터만을 사용할 때보다 높은 성능을 보인다. 제안 모델은 다른 반정형 문서에도 활용될 수 있으며 대규모 데이터에 대한 레이블러로서의 기능을 할 수 있다.

**키워드 :** 자동차 번호판, 준지도학습, 객체 탐지, 문자 인식

### 1. 서 론

자동차 번호판의 탐지 및 인식은 지능형 차량 관제 시스템에서 중요한 연구분야로 주차 제어, 교통 관리 등 다양한 분야에서 활용된다[1]. 영상으로부터 번호판을 인식 시, 왜곡된 번호판으로 인해 문자의 세그멘테이션과 인식의 정확도가 하락한다. 이를 해결하기 위해 딥러닝 알고리즘을 활용한 연구가 도입되고 있다[2].

딥러닝 기반 번호판 인식 기술은 객체 탐지와 문자 인식을 수행하는 2 단계 방식으로 진행된다. 객체 탐지

모델과 문자 인식 모델은 학습 데이터의 레이블 정보를 이용해 신경망의 추론과 실제 정답 간 오차가 작아지도록 학습하는 지도학습 방식으로 훈련된다. 지도학습 방식의 훈련은 학습 데이터셋 준비 과정에서 데이터 수집과 더불어 수집된 만큼의 추가적인 레이블링 과정이 수반된다.

본 논문에서는 수집된 이미지 보다 적은 양의 레이블 정보로 번호판 탐지 및 인식을 진행하는 Pseudo label[3] 기반의 준지도학습 모델을 설계하고 평가한다. VGG-16(Visual Geometry Group 16)[4]에 기반한

탐지기, CNN(Convolutional Neural Network) [5] 과 BiLSTM(Bidirectional Long Short-Term Memory) [6] 에 기반한 인식기로 구성된 제안 모델은 레이블 정보가 있는 데이터와 레이블 정보가 없는 데이터를 동시에 활용하여 학습할 수 있으며 OCR 과 같은 제 3의 모듈에 의존하지 않는 종단간(end-to-end) 방식의 모델이다. 제안된 준지도학습 기반의 번호판 인식 모델은 동일한 크기의 레이블 정보를 가진 학습 데이터에 대해 지도학습 방식을 이용한 모델과 비교하여 개선된 성능을 보인다. 추가적으로 레이블 정보가 있는 데이터와 레이블 정보가 없는 데이터의 비율에 따른 성능 비교를 통해 레이블 정보가 없는 데이터가 레이블 정보가 있는 데이터에 비해 과하게 많을 경우 준지도학습의 성능에 악영향을 줄 수 있음을 보인다.

## 2. 관련 연구

기존에 제안된 번호판 인식 모델의 자동차 번호판 탐지는 기계학습, 딥러닝, 영상처리 기법 등이 적용되었다. 탐지된 번호판으로부터의 문자 또는 숫자에 대한 인식 모델은 딥러닝 모델이 선택되었다. SVM(support vector machine) [7]을 적용하여 번호판을 탐지하고 탐지된 모델에 대하여 CNN 알고리즘을 사용하는 모델이 제안되었다[8]. 제안 모델은 Canny 윤곽선 검출 알고리즘을 이용하여 번호판 후보 이미지들을 탐지하고 사전에 지도학습된 SVM을 적용하여 번호판 객체를 선출한다. 번호판 객체의 가로세로 영역 히스토그램을 계산하여 문자와 숫자들을 분할한다. 중형비 기반이 아닌 탐지 위치를 기준으로 인식하여 번호판 규격이 바뀌어도 탐지가 가능하였다. 분할된 문자와 숫자는 CNN 모델을 훈련하여 인식하였다.

탐지 단계에서 YOLOv4[9]를, 인식 단계에서 CNN을 이용한 모델이 제안되었다. 제안 모델은 번호판 탐지, 번호판 영상 처리, 번호판 인식의 단계를 수행한다[10]. 자음과 모음이 분리되어 있어 오인식이 잦은 한글의 인식률을 향상시키기 위해 영상 침식(erosion), 분리된 자음과 모음 영역 탐지, 음절 영역 추출과 같은 3 가지 영상 처리 알고리즘을 사용한다.

검출과 인식이 별개의 모델이 아닌 하나의 모델 내에서 이루어지는 SSD(Single Shot Detection) [11] 기반의 모델이 제안되었다. 제안 모델은 기존의 SSD에서 모델 파라미터의 개수를 줄이기 위해 병목 블록을 사용

하여 특징맵의 차원을 축소하는 과정을 거친다. 파라미터의 감소로부터 모델의 학습 및 추론 속도 향상의 효과를 기대한다.

오픈소스 라이브러리 easyOCR 은 탐지 단계에서 CRAFT (Character Region Awareness for Text Detection) 모델을 사용한다[12]. 글자의 지역 임계값과 글자 간 유연 임계값을 설정하여 문자를 단어 단위의 경계 박스로 구분한다. 인식 단계에서는 CRNN(Convolutional Recurrent Neural Network) [13]을 사용한다.

## 3. 제안 방법

### 3.1 번호판 객체 탐지

$256 \times 256$  크기의 단일채널 흑백 이미지에 존재하는 차량에 대하여 번호판 탐지를 위한 지역화(localization)를 수행한다. VGG-16 으로부터 이미지의 특징맵을 추출하고 합성곱 연산을 수행하여 자동차 번호판 영역을 추출한다. 그림 1 은 입력 영상으로부터 번호판 객체를 추출하는 모델 구조이다. 입력 영상에 대한 VGG-16의 출력은  $512 \times 8 \times 8$ 의 크기이며 출력층의 완전연결계층은 512 채널의 특징맵을 4 채널의 벡터로 축소한다. 계산 복잡도를 줄이기 위해 합성곱 연산을 3 번에 나누어 거쳐 차원 축소를 수행한다. 모델의 출력층으로부터 경계 박스 좌표를 나타내는 4 개 채널의  $1 \times 1$  벡터가 출력되며 각 벡터는 경계 박스의 좌상단 x 좌표, 좌상단 y 좌표, 너비, 그리고 높이를 나타낸다. 탐지 모델로부터 출력된 좌표 정보를 통해 번호판 영역을 추출한다.

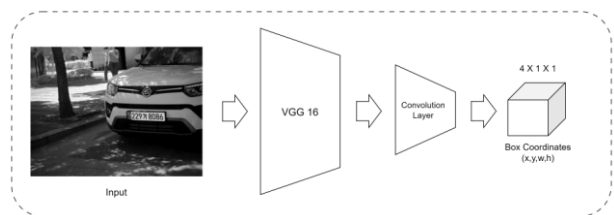


Fig 1. VGG-16 based license plate detection model

### 3.2 문자 인식

그림 2 는 탐지 모델로부터 출력된 번호판 영역 내의 문자를 인식하는 문자 인식 모델의 구조이다. 문자 인식 모델은 CNN 계층과 BiLSTM 계층으로 구성된다.

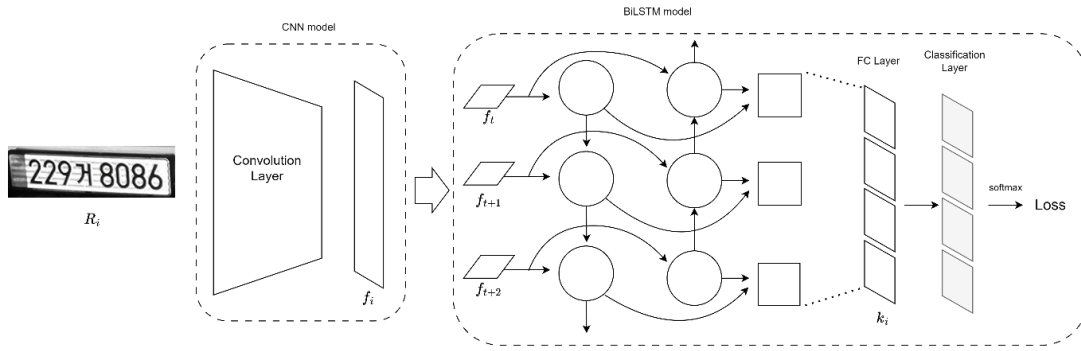


Fig 2. BiLSTM based character recognition model

CNN 계층은 번호판 영역  $R_i$ 를 자동차 번호판에 대한 특징 정보를 담고 있는 행렬  $f_i \in \mathbb{R}^{n \times 512}$ 로 변환한다. CNN 계층은 커널을 통해 입력 값을 추상화하고 평균 풀링 계층을 통해 특정 윈도우 내 출력의 평균값을 구해 압축한다. 계층이 깊어질수록 커널은 높은 수준의 특징을 추출하기 때문에 자동차 번호판 이미지에 존재하는 얼룩, 비, 눈과 같은 노이즈의 영향을 덜 받는다 [14].

BiLSTM 모델은 한 시점에서 양방향 순서에 대한 공간 정보를 추출할 목적으로 학습하는 모델이다. 시계열 데이터로 양방향으로 학습을 하면서 특정 시점에 대해 이전 시점 상태와 이후 시점 상태 간의 관계를 기억할 수 있다. 이로 인해 모델은 문자와 문맥 특징에 강건하여 특징 부족에 대한 영향을 덜 받는다 [15]. 제안 모델의 BiLSTM 계층은 단어를 처리하는 Embedding 계층으로서 특징 벡터  $f_i$ 를 시점  $t$ 에 대한 벡터  $\mathbb{R}^{1 \times 256}$ 로 처리하였다. 그 결과, 행렬  $f_i$ 를  $n$  개의 시점을 가진 시계열 데이터로 변환하여 행렬  $k_i \in \mathbb{R}^{n \times 256}$ 를 출력한다.

시계열 데이터에서 최종적인 문자열을 구하기 위해서는 모델 출력의 차원 크기를 문자의 클래스 개수로 변형해야 한다. 출력 값에 대해서 완전연결 계층을 통해 클래스 개수로 변환한 뒤 소프트맥스 함수를 적용하여 확률 분포를 구하며 확률 분포 값과 Loss 함수를 통해 모델을 학습한다.

OCR(Optical Character Recognition) 기반의 문자 인식은 이미지가 깨끗하지 못하거나 정렬되어있지 않으면 문자가 오인식, 혹은 미탐되는 문제가 발생할 수 있어 이미지의 전처리 과정이 중요하게 다뤄진다. 실제 환경에서의 입력 이미지에 대해 보편적인 전처리 방식을 찾는 것은 많은 비용을 요구하고 전처리 및 OCR에 의한 오류가 후속 처리에까지 부정적인 영향을 미칠 수

있다 [16]. 반면 제안된 문자 인식 모델은 인코더와 디코더로 구성된 종단간(end-to-end) 방식의 모델로서 입력 이미지에서 문맥 정보를 추출하고 원하는 출력으로 직접 매핑하기 때문에 다른 모델에 의존하지 않으며 기존 OCR 방식의 문제로부터 자유롭다.

### 3.3 Loss 함수

제안 모델은 번호판 탐지 모델과 문자 인식 모델의 학습을 위한 Loss 함수로 각각 MSE(Mean Squared Error) Loss 와 CTC(Connectionist Temporal Classification) Loss 를 사용한다. MSE Loss 는 식 (1) 과 같이 입력 데이터에 대하여 실제값과 예측값의 차이를 제공한 후 평균을 계산한 값이다. 번호판 탐지 모델은 예측하고자 하는 값이 번호판 경계 박스의 좌표값, 연속적인 실수이기에 예측값과 실제값의 오차를 계산하여 이를 줄이는 방향으로 학습을 진행한다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (1)$$

문자 인식 모델의 학습에 사용되는 CTC 는 입력 음성 프레임 시퀀스 또는 단어 시퀀스 간에 명시적인 정렬(alignment) 없이도 학습할 수 있는 기법이다. 입력 프레임에 레이블을 각각 명시적으로 부여하지 않아도 모델을 학습할 수 있다.

$$p(\pi|X) = \prod_{t=1}^T y_{\pi_t}^t \quad (2)$$

$$p(Y|X) = \sum_{\pi \in B^{-1}(Y)} p(\pi|X) \quad (3)$$

식 (2)는 입력 데이터가 주어졌을 때 모든 가능한 경로의 확률을 나타낸 것이다.  $t$  는 시계열 데이터에서

시간을,  $\pi_t$ 는 t일 때, 선택된 상태 또는 출력 값을 의미한다. 이때 CTC 는 각 상태를 조건부 독립으로 가정하기 때문에 입력 데이터에 의해서만 상태의 값이 변하며, 상태들 간 연관성을 가지지 않는다. 식(3)은 시계열 데이터 X 가 주어졌을 때 정답 레이블 Y 가 나타날 확률을 나타낸 것으로 상정 가능한 모든 경로( $\pi$ )들의 확률 합이다.  $B^{-1}(Y)$ 는 공백과 중복된 레이블을 제거하여 정답 레이블 Y 가 될 수 있는 모든 가능한 경로들의 집합이다[17].

CTC 는 조건부 독립시행으로, 모든 결과가 다른 결과와 조건부적으로 독립된다는 특징을 가지고 있기에 이전 시점의 출력에 영향을 받는 조건부 종속(conditionally dependent)에 해당되는 데이터와는 적합하지 않다. 자동차 번호판의 경우, 배열된 문자들이 서로 독립적이므로 CTC Loss 를 적용하기에 적합하다. 또한 CTC 는 입력 받는 확률 벡터의 차원에 공백을 포함하여 입력과 출력 간의 지정된 정렬을 요구하지 않기에 입력과 출력 간의 시퀀스 길이가 동일하지 않아도 된다. 이를 통해 모든 가능한 정렬의 확률 값을 더하여 최종 출력 값을 결정하며, 다중 입력 값을 하나의 결과값으로 매칭하는 다대일 방식을 보장한다.

한국의 자동차 번호판은 등록 시기, 차량 종류 등에 따라 문자 길이가 다르다. 일반 차량은 하나의 한글을 포함하는 7,8 개 문자로 구성되며 특수 차량(택시, 버스, 이륜차)의 경우, 지역 이름을 반영하여 3 개의 한글을 포함한 9 개의 문자로 구성된다. 이러한 출력 값의 시퀀스 길이 차이를 고려하여 CTC 를 적용한다. CTC 는 시간 순으로 단조 정렬(monotonic alignment)을 하기에 출력 값에 대해서 순서를 고려하며, 반복된 값을 허용하지 않는다. 따라서, 본래 문자열과 같은 순서로 context 를 만들고 반복되는 값을 제거하여 최종 값을 출력할 수 있다.

### 3.4 Pseudo label 기법

제안 모델은 문자 인식 모델의 학습 단계에서 Pseudo label 기법을 활용한다. Pseudo label 은 레이블 정보가 없는 데이터를 활용하여 지도학습 기반의 분류 모델을 학습하는 준지도학습 기법이다. 학습 과정 중 레이블 정보가 없는 데이터의 클래스를 분류하고 예측된 클래스를 다음 학습 단계 때 레이블 정보가 없는 데이터의 레이블  $y_i$  로서 활용한다. 레이블 정보가 없는 데이터에 대해서 최대 예측 확률값  $f_i(x)$ 을 가진 클래스

를 레이블로 매핑한다.

$$y'_i = \begin{cases} 1, & \text{if } i = \operatorname{argmax}_{i'} f_{i'}(x) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Pseudo label 을 통해 학습 시, 학습 초기에는 레이블 정보를 가진 데이터만을 사용해 지도학습을 수행한다. 레이블 정보를 가진 데이터로 일정 횟수 동안 사전 학습을 진행한 다음 레이블 정보가 없는 데이터를 추가로 활용하여 준지도학습을 수행한다.

준지도학습 단계에서의 손실함수는 레이블이 있는 데이터의 CTC 오차  $L(y_i^m, f_i^m)$ 와 레이블이 없는 데이터의 CTC 오차  $L(y_i'^m, f_i'^m)$ 에  $\alpha$ 를 곱한 후의 평균합이다.

$$L(Y, X) = \sum_{(X, Y) \in D} -\log p(Y|X) \quad (5)$$

$$L = \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^C L(y_i^m, f_i^m) + \alpha \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^C L(y_i'^m, f_i'^m) \quad (6)$$

$\alpha$ 는 학습 성능의 향상을 위해 레이블이 있는 데이터의 오차와 레이블이 없는 데이터의 오차 간 균형을 맞추는 계수이다.  $\alpha$ 값이 과하게 클 경우, 레이블이 있는 데이터에 의한 학습에까지 영향을 미치고 너무 작을 경우, 레이블이 없는 데이터를 사용하여 얻는 효과가 미비하다. 학습이 진행됨에 따라  $\alpha$ 값을 서서히 증가시켜 과적합을 방지하고 최적화 과정 중 극소점(local minima)문제를 예방한다[3].

$$\alpha = \begin{cases} 0, & \text{iteration} < h_1 \\ \frac{3(\text{iteration} - h_1)}{h_2 - h_1}, & h_1 \leq \text{iteration} < h_2 \\ 3, & \text{iteration} \geq h_2 \end{cases} \quad (7)$$

$h_1 = 2000, h_2 = 8000$

레이블 정보가 없는 문자 이미지의 모델 출력은 학습이 진행됨에 따라 하나의 문자를 나타내는 레이블로 수렴하는 원-핫 벡터(one-hot vector)에 근접한다. 이는 레이블 정보가 없는 데이터의 입력에 대해 모델 출력의 엔트로피가 낮아지는 엔트로피 정규화(entropy regularization) 효과를 보이며 레이블 정보가 없는 데이터들에 대한 결정 경계(decision boundary)는 데이터들의 확률밀도가 적은 곳에 위치하게 되어 데이터들이 잘 구분될 수 있다.

Table 1. Comparison of recognition model accuracy

Number of Labeled Data	Number of Unlabeled Data					
	0	5,000	10,000	15,000	20,000	30,000
2,000	94.8%	98.1%	98.4%	98.0%	97.9%	96.6%
4,000	98.0%	98.2%	98.8%	99.0%	98.9%	98.6%
6,000	98.8%	98.9%	99.1%	99.4%	99.4%	99.4%

#### 4. 실험 결과

번호판 탐지 실험에 사용된 차량 이미지 데이터셋은 학습 데이터셋 2,000 장과 평가 데이터셋 3,746 장으로 구성하며 모델 평가지표는 정확도(%)이다. 번호판에 사용되는 문자 개수가 6 개 이상인 점과 객체로 탐지될 수 있는 잡음 등을 고려하여 탐지 모델의 출력인 경계 박스 내의 문자 객체가 6 개 이상일 시 정답으로 인식한다. 평가 데이터셋에 대해 탐지 성능 실험한 결과 총 3,746 장의 이미지에 대해 108 건의 오탐을 제외한 97.11%의 정확도를 보였다. 학습 데이터셋에서 신형 번호판과 구형 번호판 비율의 불균형이 존재해 상대적으로 학습 데이터가 적은 구형 번호판에서 대부분의 오탐이 발생하였다.

제안된 문자 인식 모델의 경우, 지도학습과 준지도학습 간 성능, 그리고 레이블과 레이블이 없는 데이터의 비율에 따른 성능 비교를 수행한다. 학습 데이터셋에서 각 실험에 사용된 레이블 정보를 가진 데이터의 수는 각각 2,000 장, 4,000 장, 6,000 장으로, 레이블 정보가 없는 데이터는 0 장, 5,000 장, 10,000 장, 15,000 장, 20,000 장, 30,000 장으로 분류하였다. 평가 데이터셋은 차량 번호판 사진 3,000 장이며 평가 지표는 문자 인식에 대한 정확도(%)이다.

지도학습의 경우 레이블 데이터만을 사용하여 2,000 번 학습하였다. 반면, 준지도학습의 경우, 레이블 데이터만으로 2,000 번 학습을 한 뒤, 레이블 데이터와 레이블이 없는 데이터 미니 배치를 동시에 6,000 번 학습을 하였다. 레이블 데이터와 레이블이 없는 미니 배치 크기는 각각 192, 512 개이다. 표 1 은 지도학습과 준지도학습에 따른 실험 결과를 비교한 것이다. 동일한 크기의 레이블 정보를 가진 데이터셋에 대해 준지도학습 방식이 지도학습 방식보다 높은 정확도를 보인다.

학습 데이터셋 내에서 레이블 정보를 가진 데이터보다 레이블 정보가 없는 데이터가 차지하는 비율이 과하

게 높을 경우, 이전보다 정확도가 낮아지는 모습을 보인다. 레이블 정보가 없는 데이터에 대한 예측을 교정할 수 없는 Pseudo label 기법의 특성 상 레이블 정보가 있는 데이터와 없는 데이터의 비율이 크게 차이날수록 잘못된 예측으로 인한 이상치(outlier)가 전체 모델을 오염시키는 영향이 커지는 것으로 분석된다.

표 2 는 동일한 시험 데이터셋에 대해 제안 모델과 기존 지도학습 기반 문자 인식 모델의 정확도를 비교한 것이다. 비교 모델들은 사전 학습된 가중치를 이용하였으며 제안 모델이 비교 모델에 비해 17~40%개선된 성능을 보인다.

Table 2. Comparative results of character recognition models

Method	Accuracy (%)
donut	85.96
paddle	82.9
easyOCR	78.02
PororoOCR	73.46
SSD	66.50
SVM	53.52
Ours	99.40

#### 5. 결 론

본 논문에서 레이블 정보가 없는 데이터를 분류 모델 학습에 사용하는 준지도학습 기반의 번호판 탐지 및 인식 모델을 제안하였다. 제안 모델은 지도학습 기반 모델과 비교하여 향상된 성능을 보였으며 이러한 준지도학습 방식은 딥러닝 모델 구축에 있어 고비용이 소모되는 학습 데이터 수집과 레이블링 작업에 대한 소요를 줄일 수 있다. 제안 모델은 번호판뿐만 아닌 주민등록증, 영수증 등 문자 인식이 필요한 여러 반정형 문서에

활용할 수 있으며 연속된 비디오 스트림의 영역에서 또한 활용할 수 있다. 또한 제안모델은 준지도학습 방법을 통해 레이블 정보가 없는 데이터셋에 레이블을 부여하는 레이블러(labeler)로서 이용하여 데이터셋을 구축에 효율적으로 이용할 수 있다.

## References

- [1] Jin, Moon Yong, et al. "Real-Time Vehicle License Plate Recognition System Using Adaptive Heuristic Segmentation Algorithm." *KIPS Tr. Software and Data Eng* 3.9, 2014, 361–368.
- [2] Li, Hui, and Chunhua Shen. "Reading car license plates using deep convolutional neural networks and LSTMs." *arXiv preprint arXiv:1601.05610*, 2016.
- [3] Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." *Workshop on challenges in representation learning, ICML*. Vol. 3. No. 2. 2013.
- [4] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*, 2014.
- [5] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25, 2012.
- [6] Graves, Alex, and Jürgen Schmidhuber. "Frame-wise phoneme classification with bidirectional LSTM and other neural network architectures." *Neural networks* 18.5–6 (2005): 602–610.
- [7] Hearst, Marti A., et al. "Support vector machines." *IEEE Intelligent Systems and their applications* 13.4, 1998, 18–28.
- [8] Kim, Jung-Hwan, and Joonhong Lim. "License Plate Detection and Recognition Algorithm using Deep Learning." *Journal of IKEEE* 23.2, 2019, 642–651.
- [9] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934*, 2020.
- [10] Kim, Jae-Jung, and Chang-Bok Kim. "Implementation of Robust License Plate Recognition System using YOLO and CNN." *Journal of Korean Institute of Information Technology* 19.4, 2021, 1–9.
- [11] Liu, Wei, et al. "Ssd: Single shot multibox detector." *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer International Publishing, 2016.
- [12] Baek, Youngmin, et al. "Character region awareness for text detection." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.
- [13] Shi, Baoguang, Xiang Bai, and Cong Yao. "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition." *IEEE transactions on pattern analysis and machine intelligence* 39.11 (2016): 2298–2304.
- [14] Ilesanmi, Ademola E., and Taiwo O. Ilesanmi. "Methods for image denoising using convolutional neural network: a review." *Complex & Intelligent Systems* 7.5 (2021): 2179–2198.
- [15] Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991*, 2015.
- [16] Kim, Geewook, et al. "Donut: Document understanding transformer without ocr." *arXiv preprint arXiv:2111.15664* (2021): 15.
- [17] Hannun, Awni. "Sequence modeling with etc." *Distill* 2.11 (2017): e8.