

# **AI기법과 활용**

**Week-07. Optical character recognition (OCR)**

**2022-Summer 서중원**

# OCR

Try the API

Faces

Objects

Labels

Logos

Text

Properties

Safe Search



스크린샷 2020-10-17 오전 1.26.50.png

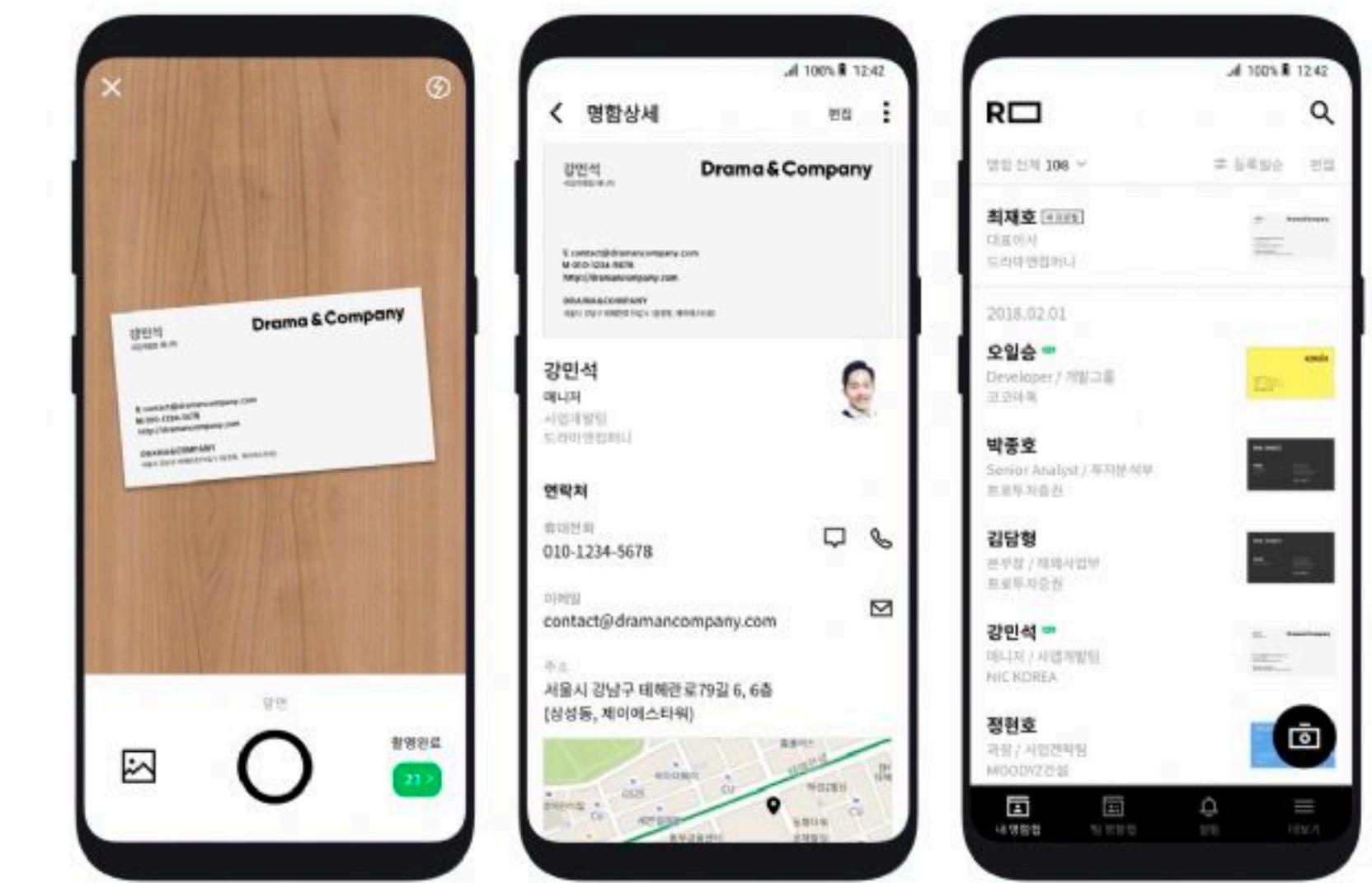
+Page 1

+Block 1

+Paragraph 1

IRONMAN

# OCR 어디에 활용될까?



# OCR

## 어디에 활용될까?

### Millions of historic newspaper images get the machine learning treatment at the Library of Congress

Devin Coldewey @techcrunch / 3:38 AM GMT+9 • May 8, 2020

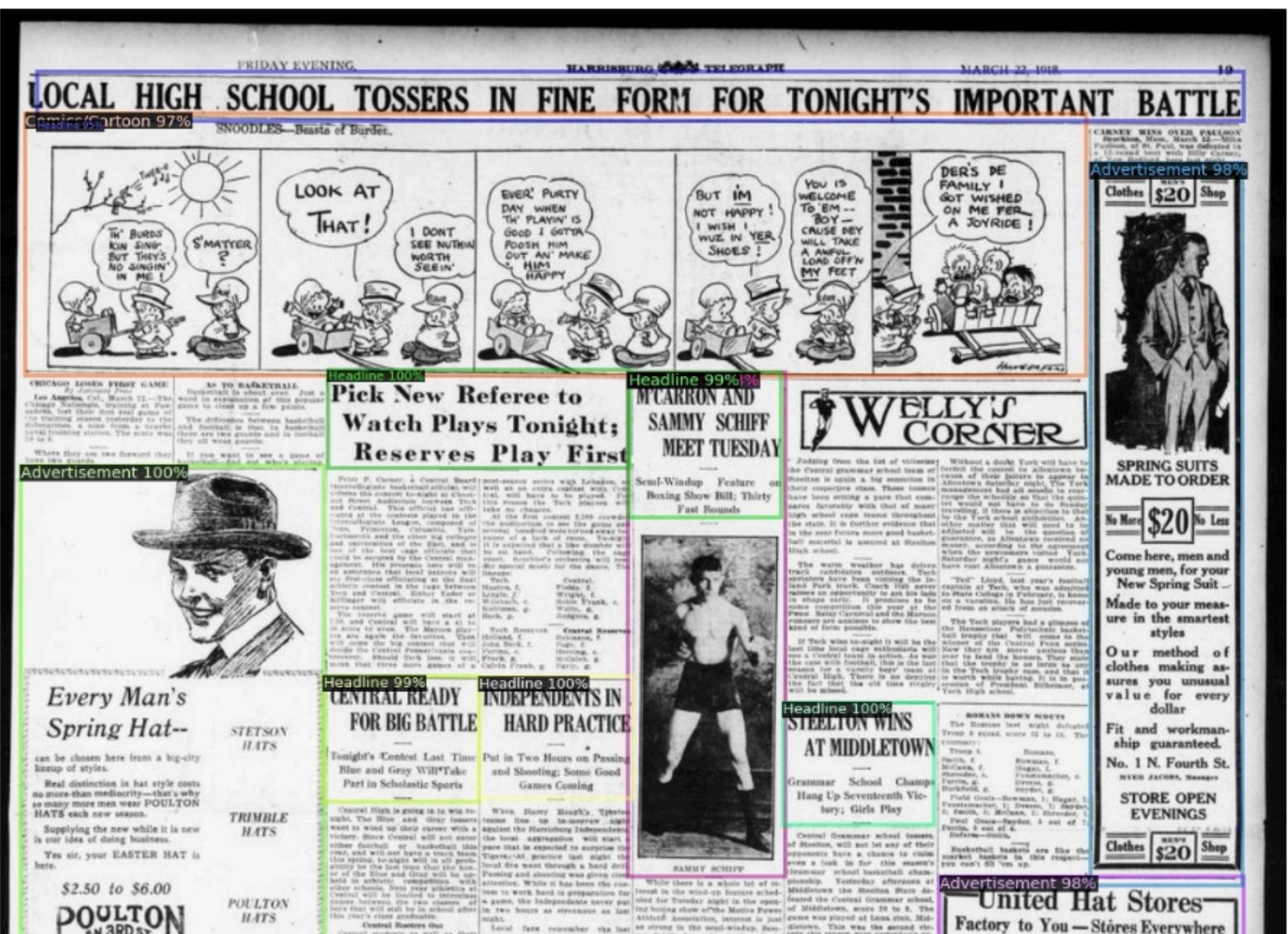
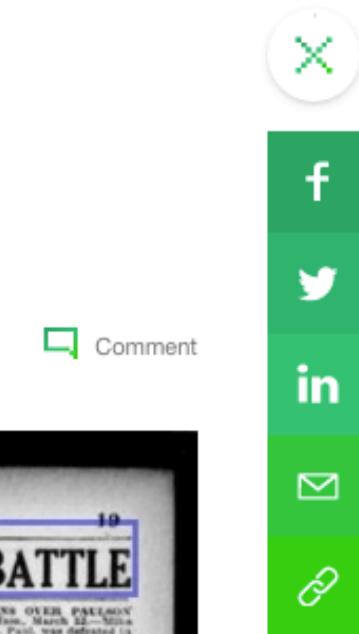


Image Credits: Library of Congress

# OCR

## OCR Pipeline

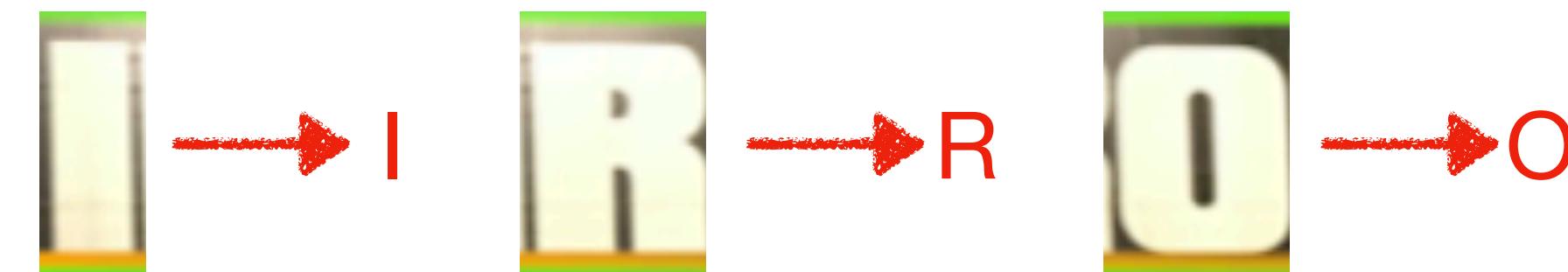
- Text Detection



- Character Segmentation



- Character Classification

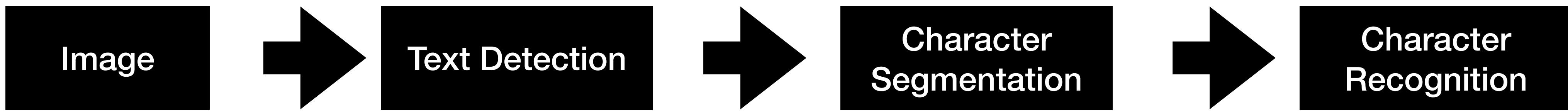


- Cleaning
- Use language model

1R0 -> IRO  
Helo -> Hello

# OCR

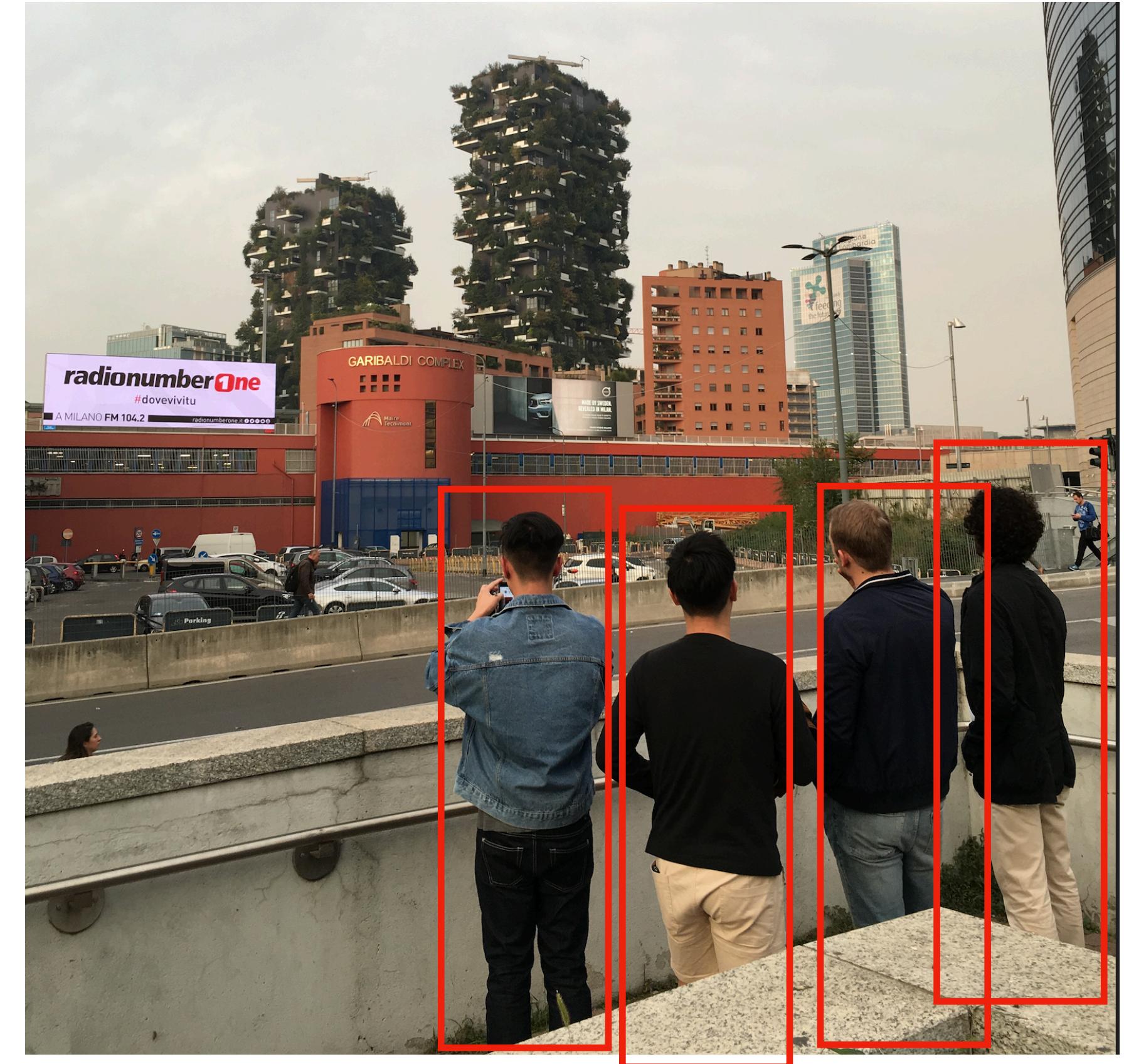
## OCR Pipeline: How it works



각 단계별 성능이 독립적으로 중요

# OCR

## Text Detection vs Object Detection



# 보행자 감지 (Pedestrian Detection)

$x = 82 \times 36$  의 이미지

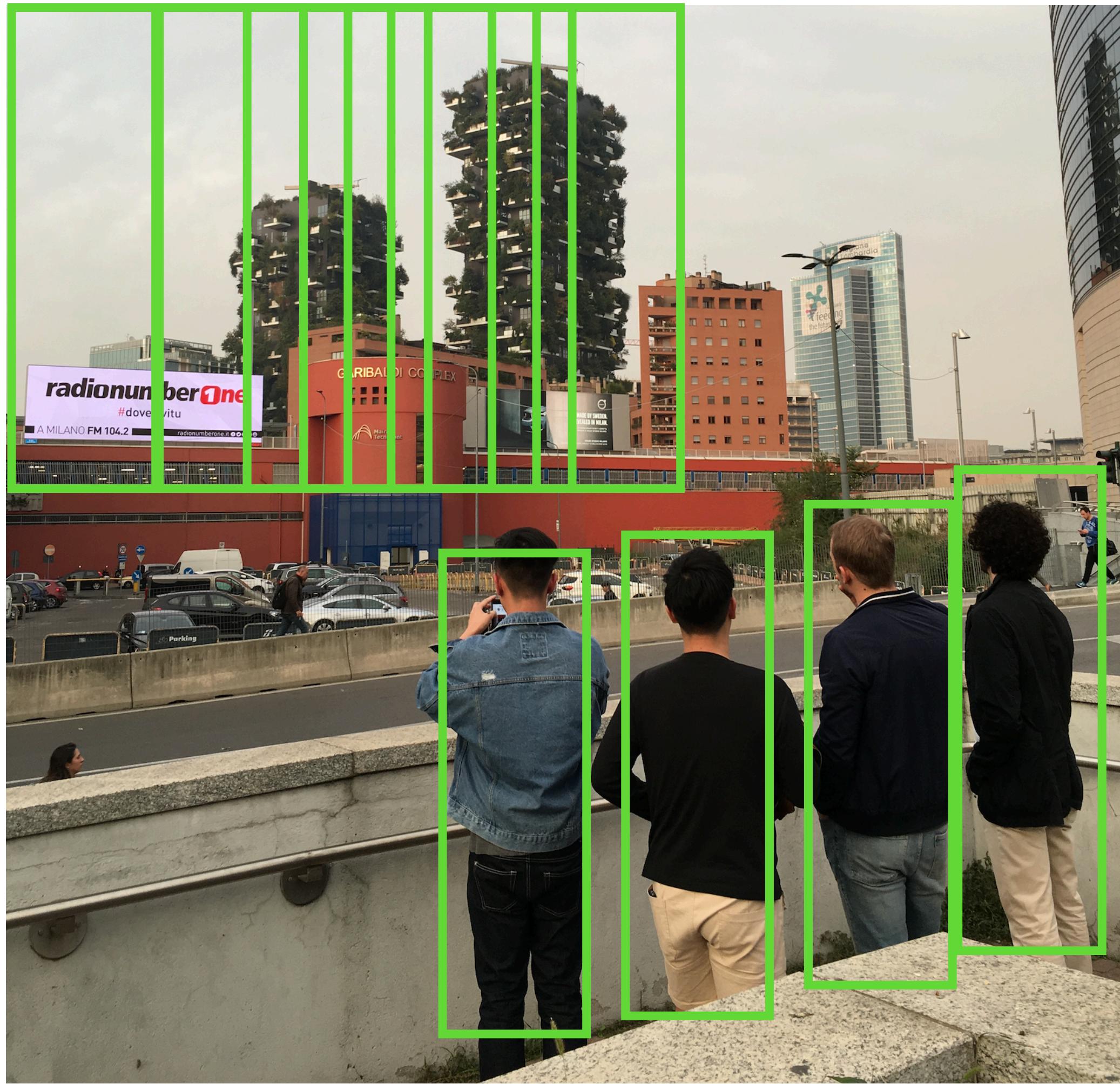


Positive examples ( $y=1$ )



Negative examples ( $y=0$ )

# Sliding Window Detection



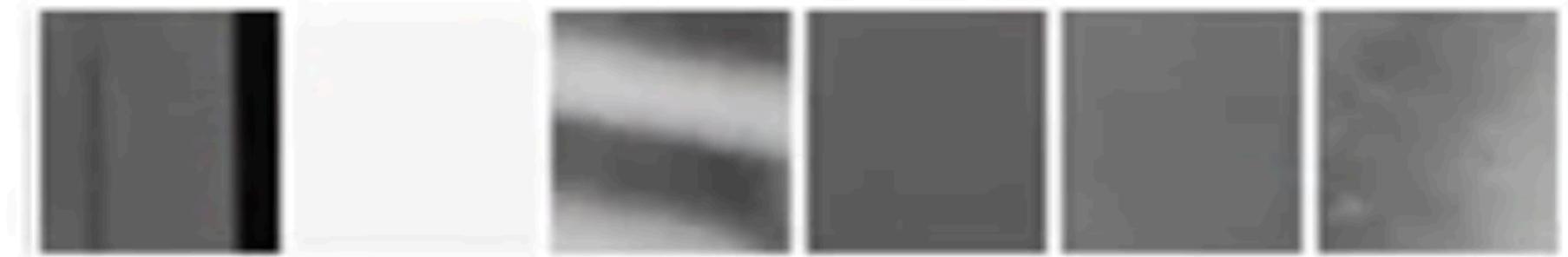
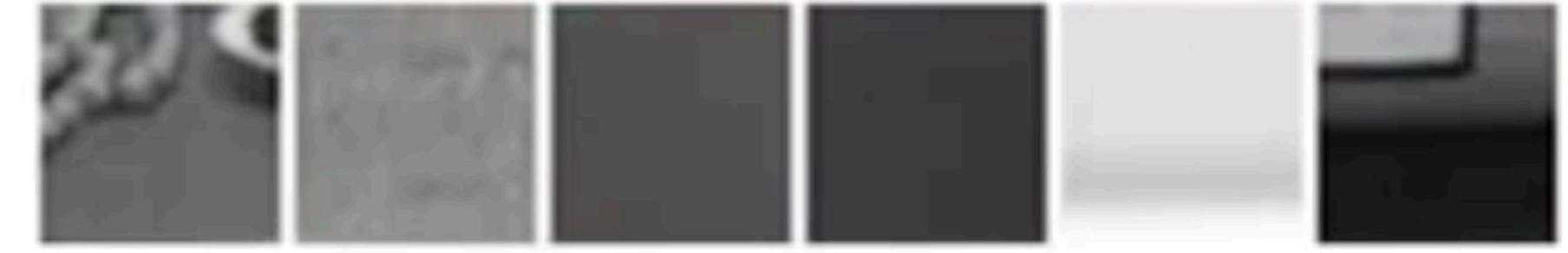
# Text Detection



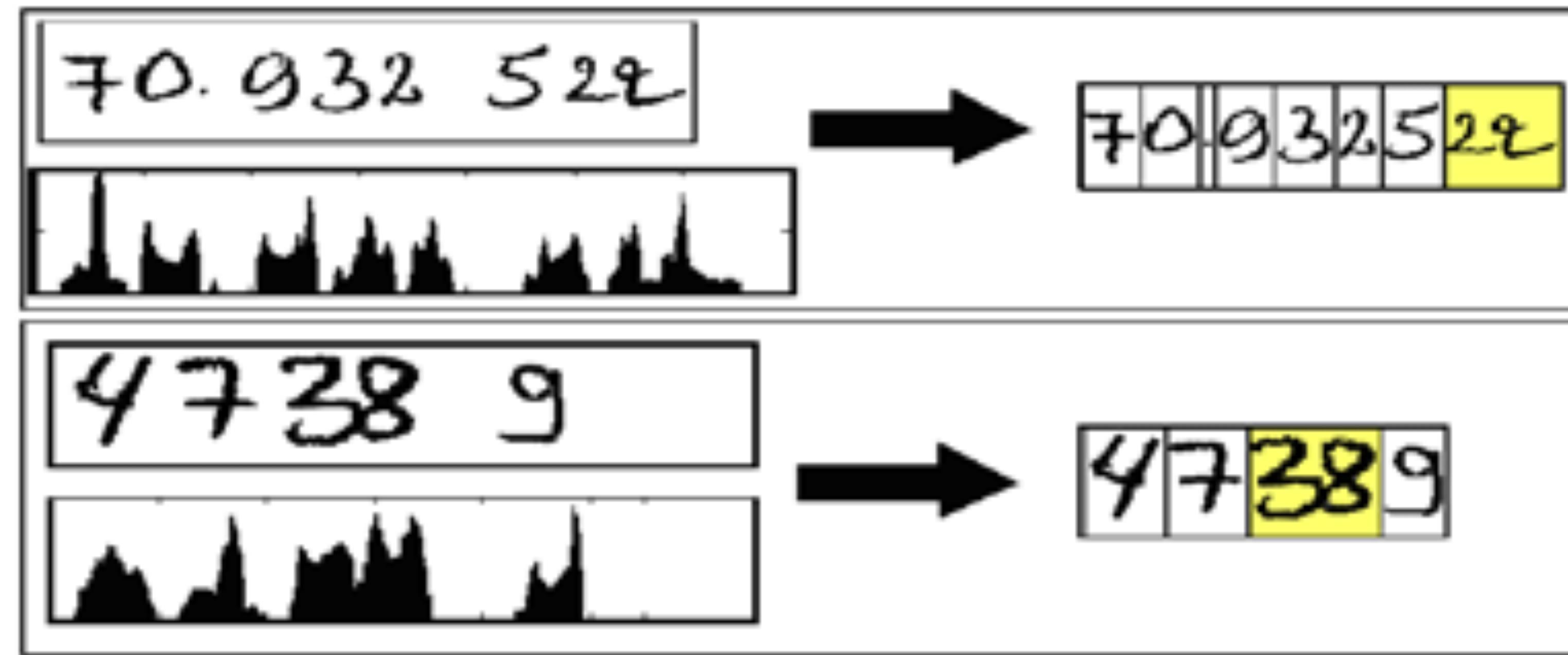
Positive examples ( $y=1$ )



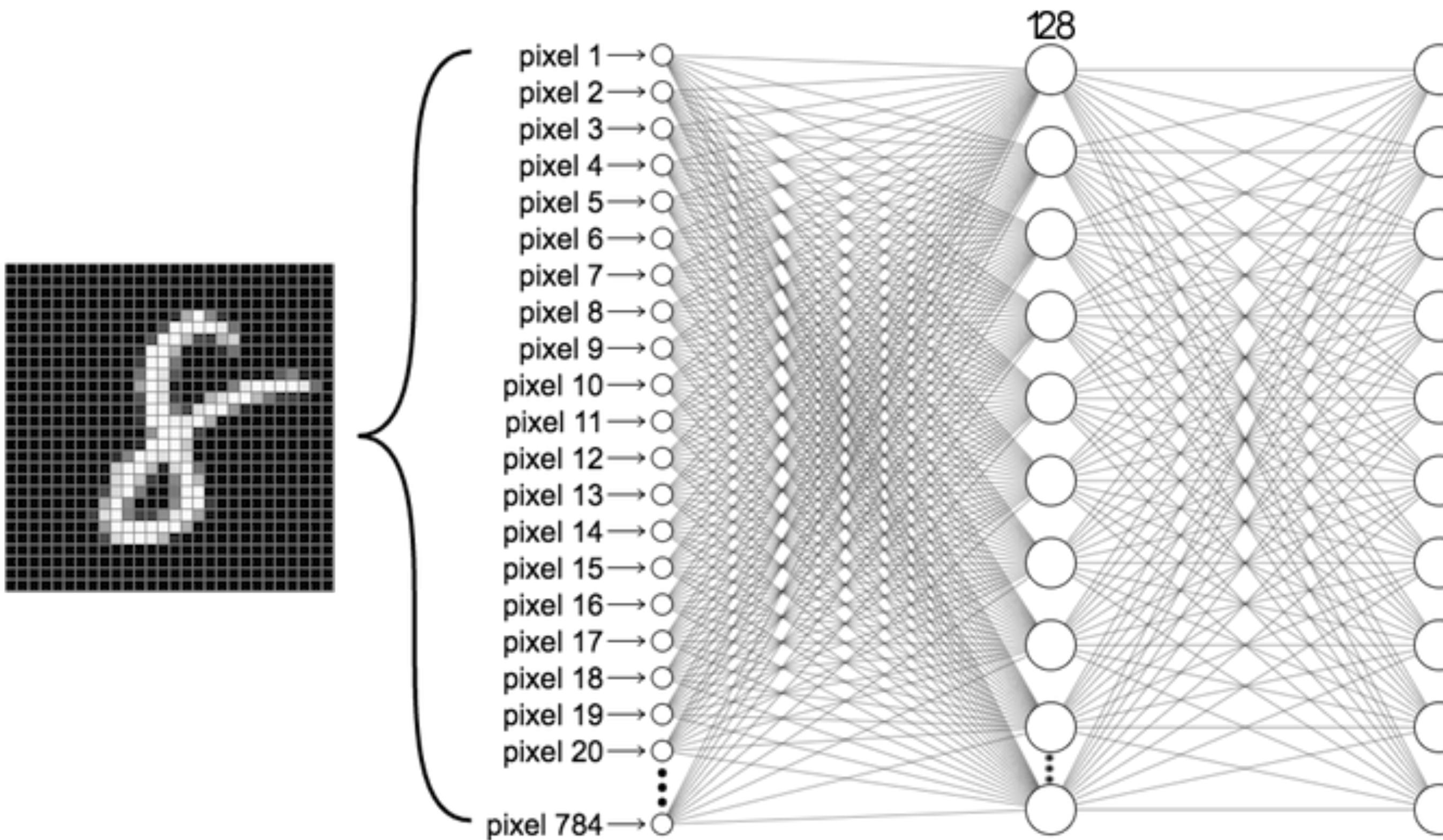
Negative examples ( $y=0$ )



# Text Segmentation



# Character Recognition



# OCR을 직접 개발해서 사용하는게 좋을까?

- 특정 데이터에 특화되게 튜닝 하는 것은 OK
  - Transfer Learning, Fine-tuning
- 바닥부터 개발하는 것은..

For example, fully 80 percent of machine learning engineers with Ph.D.s. are scooped up by Google and Facebook, especially those that have patents. Yet, when we were conducting a search for one tech company, they refused to budge on their requirement for a Ph.D. As a result, a candidate with a master's degree that was well-qualified was hired by another company, and the job went unfilled for months.

실습 시작 전

## GCP 콘솔로 이동 후 OCR검색

The screenshot shows the Google Cloud Marketplace search results for "OCR". A red box highlights the search bar at the top and the "Cloud Vision API" entry in the list.

Google Cloud Marketplace search results for "OCR":

- Cloud Vision OCR On-Prem  
Google
- Cloud Vision API**  
Google Enterprise API
- Cullable PDF Toolkit  
Cullable API
- Flashpoint Intelligence Platform  
Flashpoint
- Healthcare ML  
Pluto7
- Intelligent Document Automation  
Virtusa Corporation
- Parashift Platform  
Parashift AG
- Patent PDF Samples with Extracted Structured Data  
Subsets of Patent Data
- Recognic  
Source Cospourcing Services Private Limited

Google Cloud Marketplace URL: <https://console.cloud.google.com/marketplace/product/google/vision.googleapis.com?q=search&referrer=se...>

## 사용 클릭

The screenshot shows a web browser window for the Google Cloud Marketplace. The URL in the address bar is `console.cloud.google.com/marketplace/product/google/vision.googleapis.com?q=search&referrer=search&authuser=2&project=glassy-life-350815&supportedpurview=project`. The page title is "Cloud Vision API – Marketplace". The main content area displays the "Cloud Vision API" product, which is described as "Image Content Analysis". Below the description are two buttons: a blue "사용" (Use) button with a red border, and a white "API 사용해 보기" (View API usage) button. A navigation bar at the bottom includes tabs for "개요" (Overview), "문서" (Documentation), and "지원" (Support). The "개요" tab is currently selected. The "개요" section contains a brief description of the API's features and its integration with Google Vision. The "추가 세부정보" (Additional Information) section provides details such as the service type (SaaS & APIs), last update date (22. 4. 30.), category (Machine learning, Big data, Google Enterprise APIs), and service name (vision.googleapis.com).

Cloud Vision API

Google Enterprise API

Image Content Analysis

사용

API 사용해 보기

개요

Integrates Google Vision features, including image labeling, face, logo, and landmark detection, optical character recognition (OCR), and detection of explicit content, into applications.

추가 세부정보

유형: SaaS & APIs

최종 업데이트: 22. 4. 30.

카테고리: Machine learning, Big data, Google Enterprise APIs

서비스 이름: vision.googleapis.com

## Text to speech 검색

The screenshot shows the Google Cloud API Marketplace search results for 'text to speech'. A red box highlights the search bar at the top. Another red box highlights the 'Cloud Text-to-Speech API' entry in the results.

Google Cloud - My First Project

API 및 서비스

사용 설정된 API 및 서비스

라이브러리

사용자 인증 정보

OAuth 동의 화면

도메인 확인

페이지 사용 동의

검색 text to speech

Cloud Speech-to-Text | Cloud Speech-to-Text  
문서

Cloud Text-to-Speech | Cloud Text-to-Speech  
문서

Speech-to-Text 빠른 시작(Node.js)  
대화형 튜토리얼

[문서 및 튜토리얼 결과 더보기](#)

MARKEPLACE

Cloud Speech-to-Text API  
Google Enterprise API

**Cloud Text-to-Speech API**  
Google Enterprise API

Cloud Video Intelligence API  
Google Enterprise API

Bespoken Automated Testing For Voice and Chat  
Bespoken

Mosaicx  
CX360

Multilingual Spoken Words Corpus - MLCommons Association  
BigQuery Public Data

Programmable Speech Analytics

4일 7일 14일  30일

https://console.cloud.google.com/marketplace/product/google/texttospeech.googleapis.com?q=search&referrer=search&authuser=2&project=glassy-life-350815&supportedpurview=project

## 사용 클릭

The screenshot shows a browser window with multiple tabs open at the top, including 'Cloud Text-to-Speech API - Met...', 'gcp text to speech - Google Se...', 'Text-to-Speech: 생동감 있는 음성...', '이미지의 텍스트 감지 | Cloud Visi...', 'Python Client for Google Cloud...', '빠른 시작: 클라이언트 라이브러리를...', 'Cloud Vision 문서 | Cloud Visi...', and others. The main content area is the Google Cloud Marketplace page for the 'Cloud Text-to-Speech API'. The page title is 'Cloud Text-to-Speech API' with a blue hexagonal icon containing a white speaker symbol. Below the title, it says 'Google Enterprise API'. A description states: 'Synthesizes natural-sounding speech by applying powerful neural network models.' A prominent blue button labeled '사용' (Use) is highlighted with a red rectangle. To its right is another button labeled 'API 사용해 보기' (View API usage). A tooltip below the buttons says '이 API를 사용 설정하려면 클릭합니다.' (Click to set up the API usage). At the bottom of the page, there are three tabs: '개요' (Overview), '가격 책정' (Pricing), and '문서' (Documentation). The '개요' tab is currently selected and active. The '개요' section contains a summary of the service: 'Synthesizes natural-sounding speech by applying powerful neural network models.' To the right of this summary is a '추가 세부정보' (Additional details) section with the following information: '유형: SaaS & APIs', '최종 업데이트: 22. 4. 30.', '카테고리: Google Enterprise APIs', and '서비스 이름: texttospeech.googleapis.com'.

Cloud Text-to-Speech API

Google Enterprise API

Synthesizes natural-sounding speech by applying powerful neural network models.

사용 API 사용해 보기

이 API를 사용 설정하려면 클릭합니다.

개요 가격 책정 문서

개요

Synthesizes natural-sounding speech by applying powerful neural network models.

추가 세부정보

유형: [SaaS & APIs](#)

최종 업데이트: 22. 4. 30.

카테고리: [Google Enterprise APIs](#)

서비스 이름: [texttospeech.googleapis.com](#)

## Speech to text click

The screenshot shows the Google Cloud Marketplace search results for the query "speech text". A red box highlights the search bar at the top of the modal window. Another red box highlights the "Cloud Speech-to-Text API" entry in the results list.

Google Cloud - My First Project

검색 speech text

API 및 서비스

사용 설정된 API 및 서비스

라이브러리

사용자 인증 정보

OAuth 동의 화면

도메인 확인

페이지 사용 동의

서비스 이름  
texttospeech.

설정항목

그래프 선택  
그래프 4개

필터: Version: v1 및 v1

응답 코드별

Cloud Speech-to-Text API  
Google Enterprise API

Cloud Text-to-Speech API  
Google Enterprise API

Cloud Video Intelligence API  
Google Enterprise API

Bespoken Automated Testing For Voice and Chat  
Bespoken

Mosaicx  
CX360

4일 7일 14일  30일

선택한 기간에는 데이터가 없습니다.

<https://console.cloud.google.com/apis/api/texttospeech.googleapis.com/metrics?project=glassy-life-350815&authuser=2&supportedpurview=project>

## 사용 클릭

The screenshot shows a browser window with the URL `console.cloud.google.com/marketplace/product/google/speech.googleapis.com?q=search&referrer=search&authuser=2&project=glassy-life-350815&supportedpurview=project`. The page is titled "Cloud Speech-to-Text API" and is part of the "Google Enterprise API" category. It features a blue icon representing speech recognition and the text "Speech recognition". Below the title, there are two buttons: a red-bordered "API 사용해 보기" button and a "API 사용해 보기" button. A navigation bar at the bottom includes tabs for "개요", "가격 책정", "문서", and "지원", with "개요" being the active tab.

**Cloud Speech-to-Text API**  
Google Enterprise API

Speech recognition

API 사용해 보기

개요 가격 책정 문서 지원

**개요**

Converts audio to text by applying powerful neural network models.

**추가 세부정보**

유형: [SaaS & APIs](#)  
최종 업데이트: 22. 4. 30.  
카테고리: [Machine learning](#), [Google Enterprise APIs](#)  
서비스 이름: [speech.googleapis.com](#)

**E.O.D**