

Paddle OCR



Next Generation
OCR Product Makes
Data Extraction Easier

Zecong Zhao

zhao1560@umn.edu



Yufan Li

yufan001@umn.edu



Runan Cao

cao00235@umn.edu



Yujun Wang

wang9936@umn.edu



Team 9

Have you ever met
those problems
during work?



“The client sent me some **table screenshots**, Some of them are even **curved pictures**.
I need to convert them
into Excel tables.”

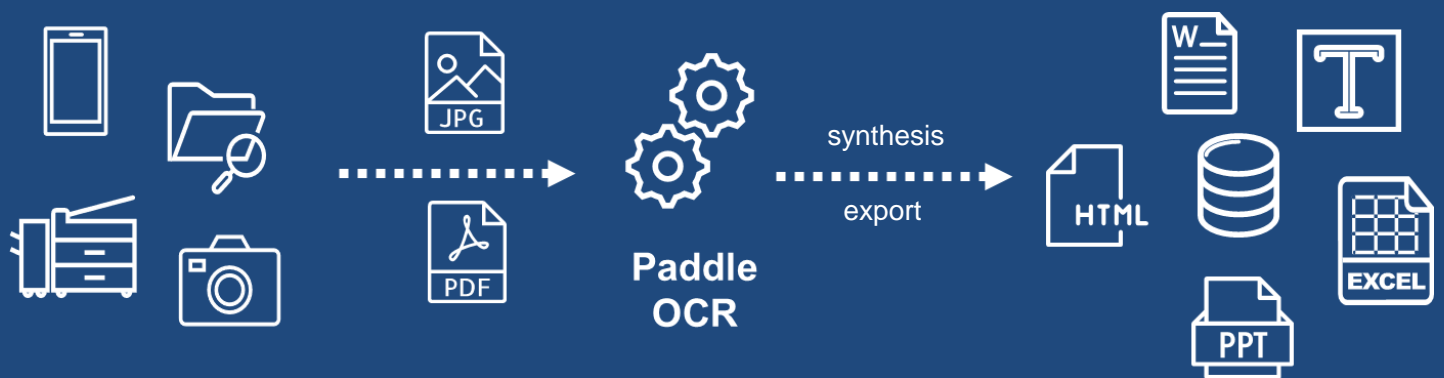
“The manager asked
me to organize **tons
of antique physical
documents** into
electronic data.”

What can PaddleOCR do for you?

PaddleOCR helps you extract text
and tables from pictures or documents
using just a line of code in Python.

It can also reduce human error, save
time and storage when it comes to
heavy loads and limited time.

How does PaddleOCR work?



Why PaddleOCR is your first choice?



7:1

0.33s : 0.048s

Compared with **EasyOCR**,
PaddleOCR is much more **faster**
and **ultra lightweight**.



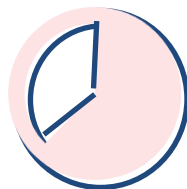
1:2

0.22 : 0.52

More **accurate** recognition result
with a higher F1 score.
PaddleOCR is **popular** in Github!

218M : 3.5M

62:1



13k : 18k

1:1.5



Promising potential cross industries!

Banking



Digitize and store
physical documents;
Scan-to-Pay ;
Prevent Fraud.

Tourism



Scan id when book and
check in a flight or a
hotel.

Retail



Scan and extract data
from bill of Lading,
Packing List, Invoices,
Purchase Orders and
more.

Health Care



Digitize data from
reports containing X-
rays, patient's history,
treatments, or
diagnostics, tests.