Tomer Biton

HF model of the weekend

naver-clova-ix/donut-base

*naver-clova-ix/donut-base

Tomer Biton

What Donut Actually Is

Document Understanding Transformer

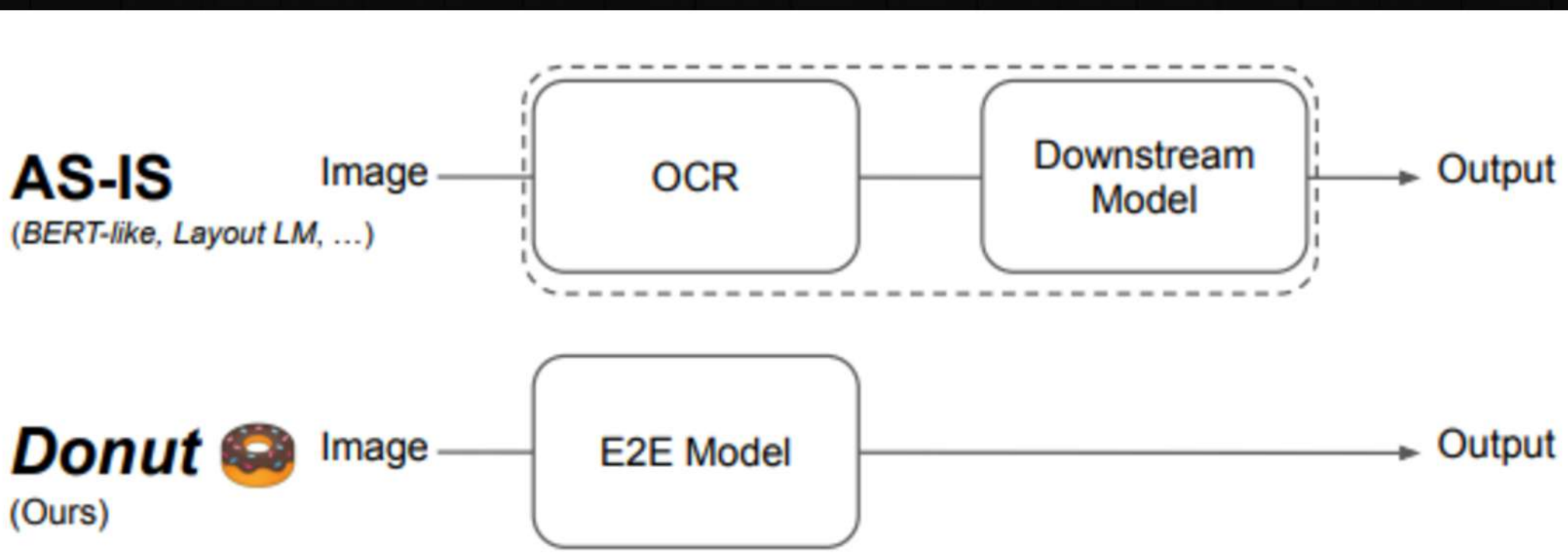
If you think OCR is just “extract text from images,” Donut will change your mind.
Unlike traditional OCR pipelines that need

→ text detector
→ text recognizer
→ layout model
→ post-processing scripts

Donut does everything end-to-end.
No bounding boxes.
No Tesseract.
No image preprocessing.
Just image in → JSON out.

Tomer Biton

How It Works Internally



Donut is not an OCR engine.
It's a Vision-Language Transformer that understands documents.

Tomer Biton

What I built

naver-clova-ix/donut-base-finetuned-cord-v2

Automated Document Processing System
This system is designed for the Optical Character Recognition (OCR) and data extraction from receipts and invoices.

Core Functionality
The system ingests the visual data from an image or PDF document and employs intelligent parsing to identify and extract key transactional and financial information (e.g., vendor, date, total amount, line items).

```
{
  "invoice_number": "INV-3337",
  "date": "Jan 25, 2016",
  "items": [
    {"description": "Web design", "qty": 1, "price": 85}
  ],
  "total": 93.50
}
```

*expected output

Tomer Biton

What I Learned This Weekend

Small model, big takeaways.

What makes Donut powerful

- OCR without OCR - it doesn't detect text; it reads documents holistically
- Layout-aware - understands fields, tables, relationships
- Multilingual - supports dozens of languages
- Trained for finance - invoices, receipts, forms
- Small enough for CPU - works on Macs without GPU