

Snakes on a train

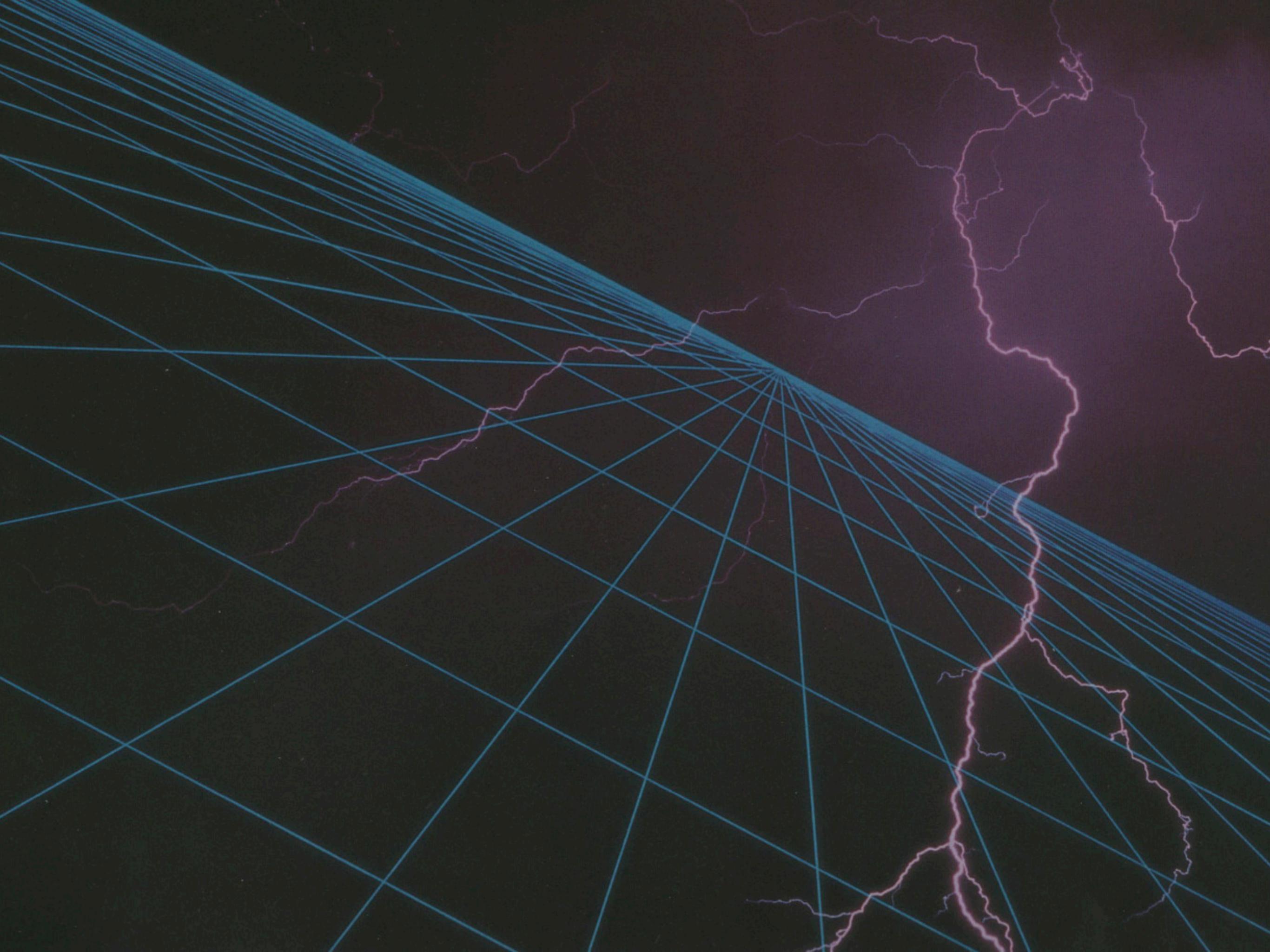
Extracting text from images



Evan Richards

<https://github.com/Eunoia/snakes on a train>

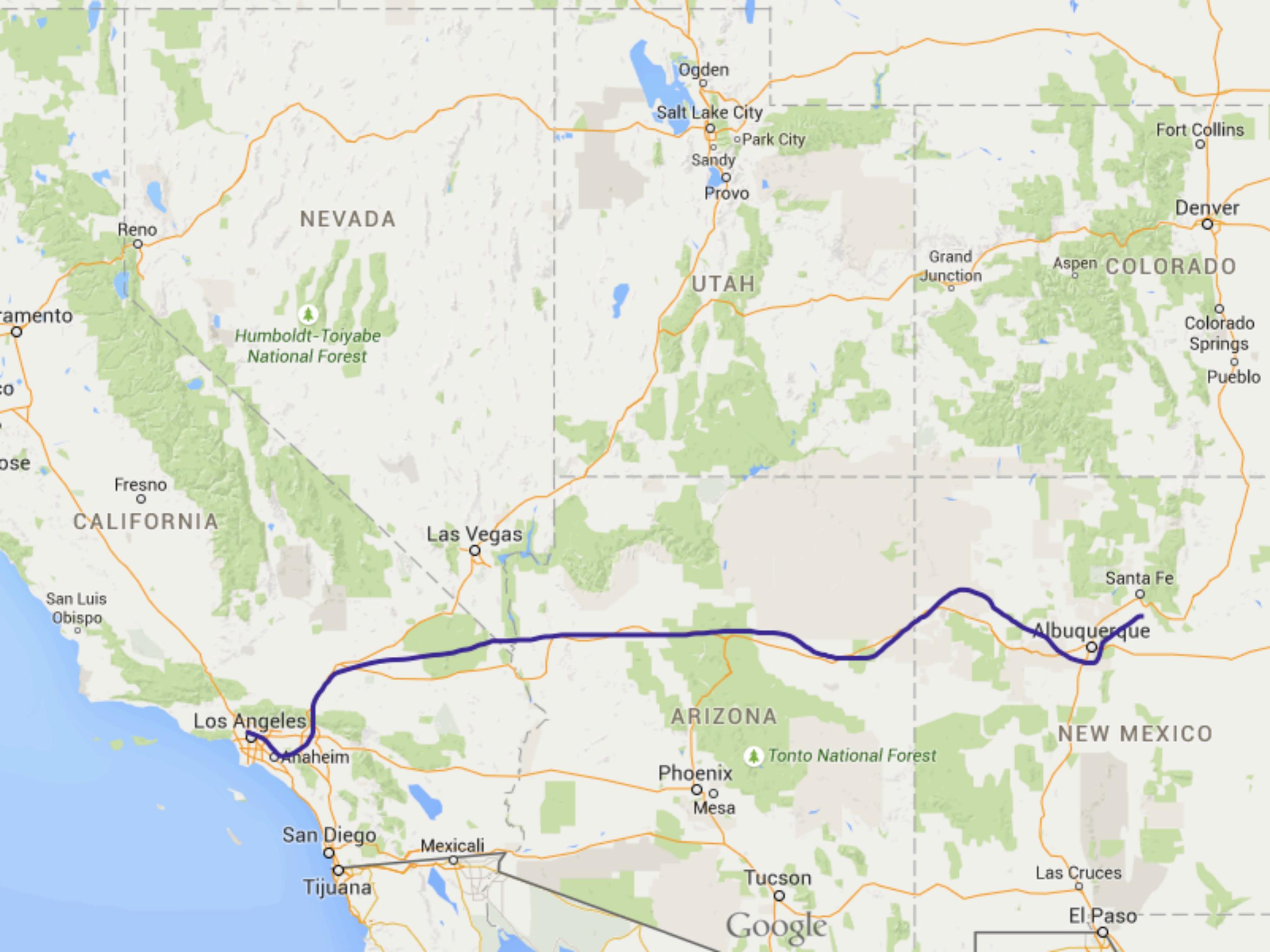




Python

Tesseract







eTicket

PRESENT THIS DOCUMENT FOR BOARDING

RESERVATION NUMBER 366B20

RES# 366B20-19JUL15

LAX ➤ LMY
LOS ANGELES, CA LAMY, NM

One-Way

JULY 26, 2015

TRAIN	SOUTHWEST CHIEF	LOS ANGELES - LAMY	DEPARTS	ARRIVES (Mon Jul 27)
4	Jul 26, 2015	1 Reserved Coach Seat	6:15 PM	1:17 PM



eTicket

PRESENT THIS DOCUMENT FOR BOARDING

RESERVATION NUMBER 366B20

RES# 366B20-19JUL15

LAX



LAMY

LOS ANGELES, CA

LAMY, NM

One-Way

JULY 26, 2015

TRAIN

4

SOUTHWEST CHIEF

Jul 26, 2015

LOS ANGELES - LAMY

1 Reserved Coach Seat

DEPARTS

6:15 PM

ARRIVES (Mon Jul 27)

1:17 PM



Convert PDF to JPG

• 0,0

Crop the Image

section = self.image.crop((x, y, i, j))



LOS ANGELES, CA



UMY

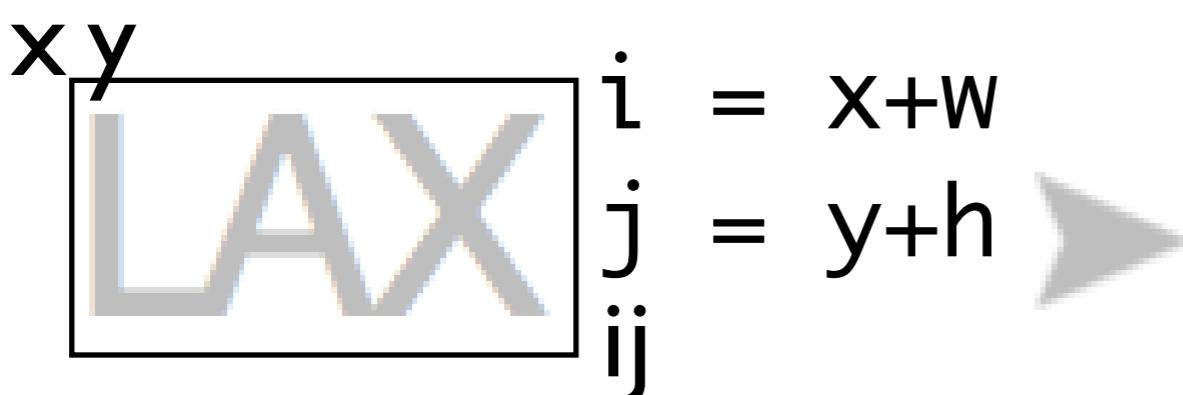
LAMY, NM

image.size()

•0,0

Crop the Image

section = self.image.crop((x, y, i, j))



RES# 366B20-19JUL15

LOS ANGELES, CA



LAMY, NM

image.size()

debugging

- `section.save("testing.jpg")`

SystemError:
tile cannot extend
outside image

image_to_string



Get in touch

- @evan on nmdevs.com
- <http://eunoia.github.io/hello/>
- Let's talk about your business.