Augraphy: Data Augmentation for Document Images

Alexander Groleau, Stefan Larson, Kok Wei Chee, Jonathan Boarman

X⇒ Sparkfish

<u>github.com/sparkfish/augraphy</u>

LightingGradient Original

Jpeg

This work introduces *Augraphy*, a

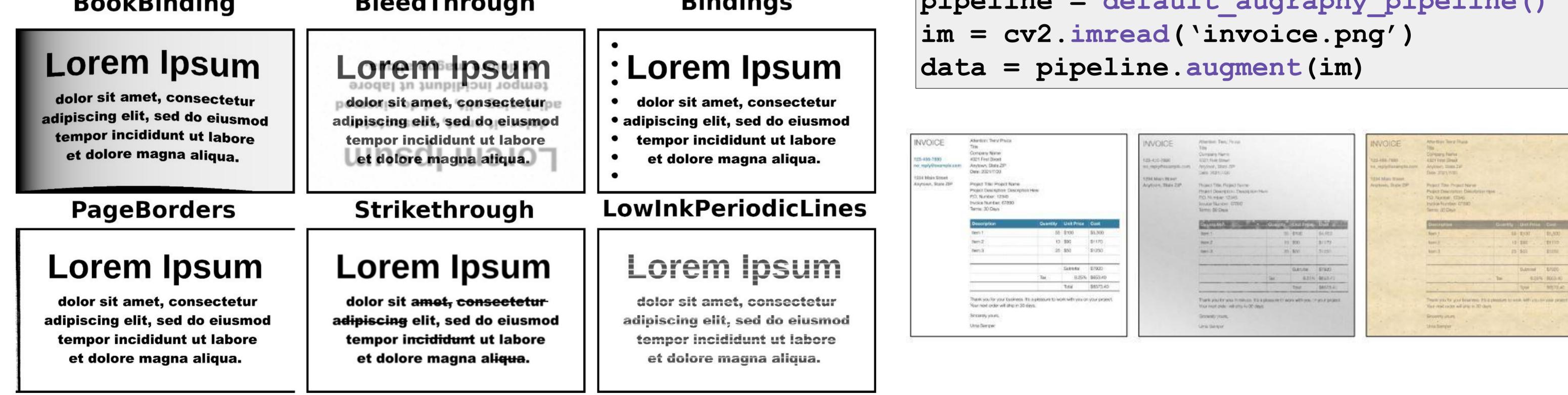
Lorem Ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	Loren lpsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	Lorem Ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	Python library for data augmentation pipelines for document image analysis. Augraphy uses many different augmentation strategies to produce augmented versions of clean document images that appear as if they have been
Folding	Faxify	DirtyDrum	
Loren Ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	Lorem Ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	LOCEM IDSUM dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	
PencilScribbles	BadPhotocopy	LetterPress	distorted due to noisy paper
dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	Locan Ipsum dolor sit amet, consectetur dipiscing elit, sed do eiusmod compor inciditdunit ut labore et dolore magna allqua:	Lorem Ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.	<pre>printing, faxing, scanning, or copy machine processes. from augraphy import *; import cv2</pre>

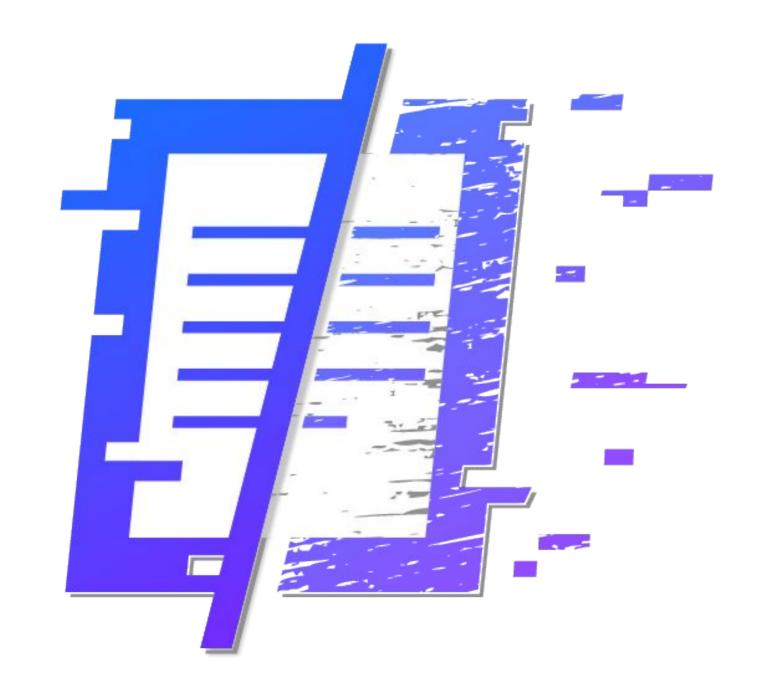
BookBinding

BleedThrough

Bindings

pipeline = default augraphy pipeline()







augraphy