

# Text detection in screen images

CSE 576, Computer Vision

Dominik Moritz, @domoritz

in collaboration with Jevin West, Poshen Lee, Joe Redmon

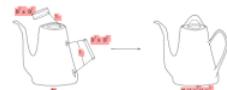


Figure 4: The attachment of a 1-handle and a 2-handle to a 3-manifold  $N$  creates a new 3-manifold.

## Goal

Researchers (AI2, Stanford HCI, eScience Institute) want to **extract data from figures** from research papers. However, these projects rely on embedded text or annotations. The goal of my class project is to implement **robust text extraction** (text detection and text recognition) in screen images, more specifically text in charts in research papers. Text detection in screen images and research papers is different from text detection in natural images because the features of text and other elements are similar in screen images.

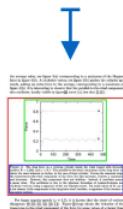
## Approach

Extracting text from images consists of two steps: **text detection** and **text recognition**. For the second step, we rely on **existing OCR software**.

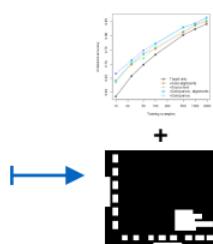
We train a **Convolutional Neural Network (CNN)** to decide for each pixel, whether it is text or not. We decided against generating training data and instead use extracted figures from a large corpus of papers from various sources.

Ton of papers from  
• ACL Anthology  
• arXiv  
• Pubmed  
• CiteSeer, JSTOR,  
PLoS

Preparation  
Usage

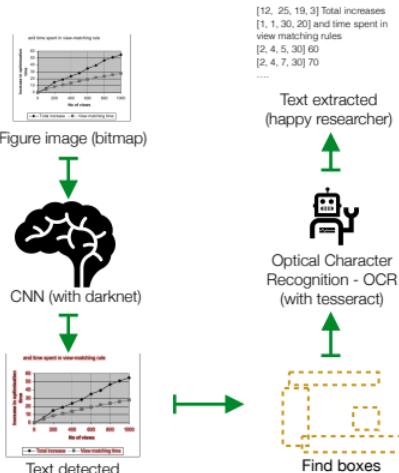


Extract figures with  
embedded text  
(with pdffigures)



Label + Image  
(with openCV, Python,  
imagemagick, AWS, Parallel)

Screenshots from Jeff Dean's talk at UW, which shows their use of Google Brain for text detection and recognition. We use a similar approach for text in screen images.



Text detected