Predicting the Politics of an Image Using Webly Supervised Data
Christopher Thomas and Adriana Kovashka
University of Pittsburgh

Introduction
The news media shape public opinion, and often, the visual bias they contain is evident for human observers. In this paper, we model visual political bias in contemporary media sources at scale, using webly supervised data. We release a dataset of over one million unique images and associated news articles from left- and right-leaning news sources and develop a method to predict the image’s political leaning. This problem is particularly challenging given the visual diversity of the data and the higher-level reasoning required to understand political bias.

Crowdsourced Annotations
Each image is associated with a weak bias label based on the bias of the source it was harvested from, but in addition, we wanted to test how well this assumed bias label correlated with humans’ understanding of bias. We collected 14,327 annotations, including image-text alignment, political bias labels, rationales, etc.

Approach for Predicting Visual Political Bias
We propose a two-stage approach. In stage 1, we learn visual features jointly paired with text for bias classification. In stage 2, we remove the need for text by training a classifier on top of stage 1. We show this sig. outperforms numerous baselines.

Politics Dataset
We harvested a dataset of over one million images and paired text articles from biased media sources. Our dataset contains images harvested on 20 different issues including abortion, climate change, gun control, immigration, welfare, etc. We harvest our data from over 500 media sources. We also performed a large scale human study using Amazon MTurk and collected annotations for a large number of images. Additional metadata such as image captions, page URLs, human annotations, etc. are available for many photos. We release our dataset for download at: www.cs.pitt.edu/~chris/politics

Quantitative Evaluation

Predicting Words from Images
We trained a model to predict words from images using our dataset. The model learns visual cues for each word, demonstrating the utility of exploiting text, even for purely visual classification. For example, “antifa” features black-clad protesters, “brutality” features police scenes and protests, “immigrant” features the border, and “LGBT” features pride flags.

Visual Explanations
We used Grad-CAM++ to compute attention maps for Ours vs. Resnet. We find our model most attends logos and faces of well-known public figures. Our method learned visual features (e.g. logos) that complement the text in stage 1, which later work even without the text.