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.

Sample photos from our dataset for the topic of immigration (top) and abortion (bottom). We observe that sources from different sides of the political spectrum portray the same subject differently.

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

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.

many black women are more liberal than conservative
Guess: L

Most african american women lean left
Guess: L

Guessed incorrectly

Lots and lots of guns.
Guess: R

Guns and flag
Guess: R

The right would like a more militant George Washington
Guess: R

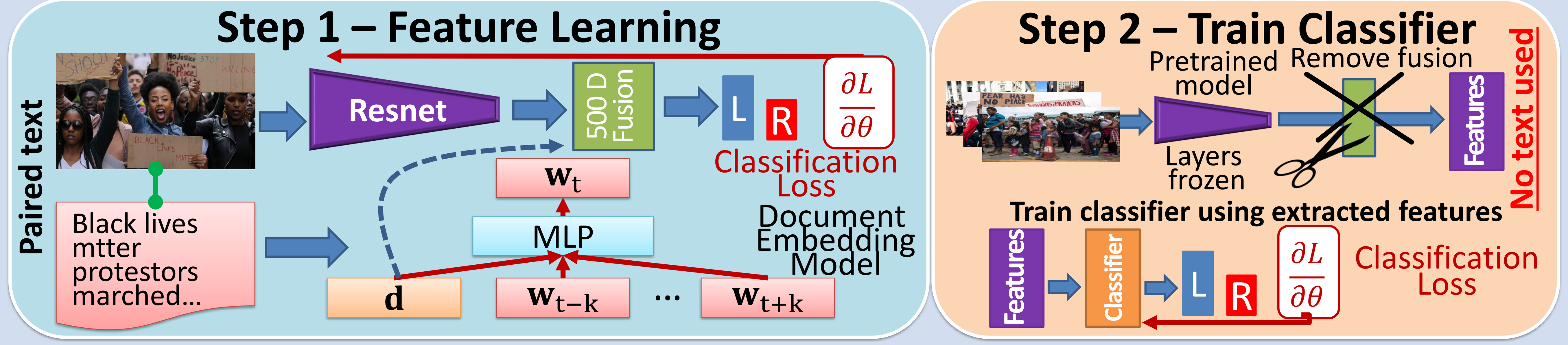
Due to the LGBT flag in the street.
Guess: L

Definitely left because this is the colors for the gay flag…
Guess: L

Colors on the road
Guess: L

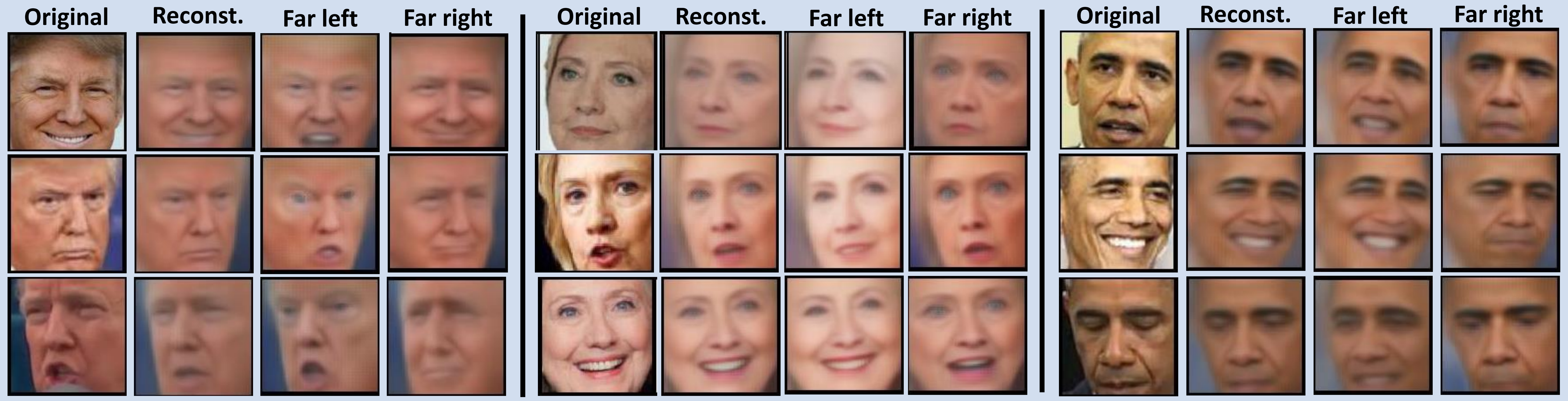
Approach for Predicting Visual Political Bias

We propose a two-stage approach. In stage 1, we learn visual features jointly with paired 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.



Generating Politically Biased Faces

We train an autoencoder on faces of politicians. We condition on automatically predicted facial attributes / expressions and latent features. We learn the avg. portrayal of each politician in the L/R and modify imgs. to be more L/R leaning.



Quantitative Evaluation

	Resnet	Joo (CVPR 2014)	Human Concepts	OCR	Ours	*Ours (U.B.)*
Weakly Supervised	0.678	0.670	0.675	0.686	**0.712**	0.803
Human Labels	0.590	0.593	0.587	0.613	**0.620**	0.626

We test our approach at politics classification vs. several baselines. Joo uses hand crafted features for visual persuasion, human concepts uses concepts mentioned by MTurkers, and OCR uses optical character recognition. We show our approach outperforms these by large margins on both weakly supervised and human labels.

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 protestors, "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.