Structuring Quantitative Image Analysis with Object Prominence

120th APSA Annual Meeting
Philadelphia—5-8 September 2024

Christian Arnold (University of Birmingham)
Andreas Küpfer (TU Darmstadt)

06 September 2024

What Do We See in Images?

The Challenge of Understanding the Message of an Image

Three Paradigms of Analyzing the Message of an Image

- Annotating semantic units, i.e., objects and people (e.g., *Loken 2021, Torres 2024*)
- Operationalizing latent characteristics of an image (*Peng 2022*)
- Raw images and deep learning to classify directly (*e.g., Gasparyan and Sirotkina 2024 (APSA Working Paper), Joo and Steinert-Threlkeld 2022, Torres and Cantú 2022*)

Three Paradigms of Analyzing the Message of an Image

- **Annotating semantic units, i.e., objects and people (e.g., *Loken 2021, Torres 2024*)**
- Operationalizing latent characteristics of an image (*Peng 2022*)
- Raw images and deep learning to classify directly (*e.g., Gasparyan and Sirotkina 2024 (APSA Working Paper), Joo and Steinert-Threlkeld 2022, Torres and Cantú 2022*)

Why 'Just' Annotating Images is Not Enough

Semantic Structure of Modalities

- ☺ Audio can be transcribed
- ☺ Transcript consists of words
- 😐 Image is built from pixels in a rectangular shape

Why 'Just' Annotating Images is Not Enough

Semantic Structure of Modalities

- ☺ Audio can be transcribed
- ☺ Transcript consists of words
- 😐 Image is built from pixels in a rectangular shape

We can use **existing models**, **supervised learning**, or **unsupervised classification** to annotate semantic units on images.

Why ’Just’ Annotating Images is Not Enough

Semantic Structure of Modalities

- ☺ Audio can be transcribed
- ☺ Transcript consists of words
- 😐 Image is built from pixels in a rectangular shape

We can use **existing models**, **supervised learning**, or **unsupervised classification** to annotate semantic units on images. **But what parts of an image matter?**

Object Prominence as Structuring Principle

But what parts of an image matter?

Object Prominence as Structuring Principle

But what parts of an image matter?

Salience maps highlight **high-attentive** areas where the human eye focuses on perceiving **recognized objects** *(Koch and Ullman 1985)*:

$$\text{Object Perception} = \text{Object Recognition} \times \text{Object Attention} \tag{1}$$

Object Prominence as Structuring Principle

But what parts of an image matter?

Salience maps highlight **high-attentive** areas where the human eye focuses on perceiving **recognized objects** *(Koch and Ullman 1985)*:

$$\text{Object Perception} = \text{Object Recognition} \times \text{Object Attention} \tag{1}$$

Our proposed modeling approach:

$$\text{Object Prominence} = \text{Object Detection} \times \text{Object Salience} \tag{2}$$

Measuring Object Prominence

Having **detected objects**, how can we measure their **salience** in practice?

Measuring Object Prominence

Having **detected objects**, how can we measure their **salience** in practice?

Size and Centeredness

Larger objects close to the center correlate with higher prominence.

Depth

Estimate the distance of each pixel to the camera to indicate higher prominence for closer objects.

Salient Object Detection

Combining various image features, SOD models estimate the salience of objects directly.

Measuring Object Prominence

Having **detected objects**, how can we measure their **salience** in practice?

Size and Centeredness

Larger objects close to the center correlate with higher prominence.

Depth

Estimate the distance of each pixel to the camera to indicate higher prominence for closer objects.

Salient Object Detection

Combining various image features, SOD models estimate the salience of objects directly.

Implementation Effort / Interpretability

Low / High

Medium / Medium

Model-Dependent

Application 1: Improving Idealpoint Estimation from Visual Bag of Words (VBoW)

Scaling Images of US Newspapers

Multimodal News Article Dataset *(Thomas and Kovashka 2019)*

- 59,361 news articles from eight US newspapers spanning across five political issues
- Text-image pairs

Scaling Images of US Newspapers

Multimodal News Article Dataset *(Thomas and Kovashka 2019)*

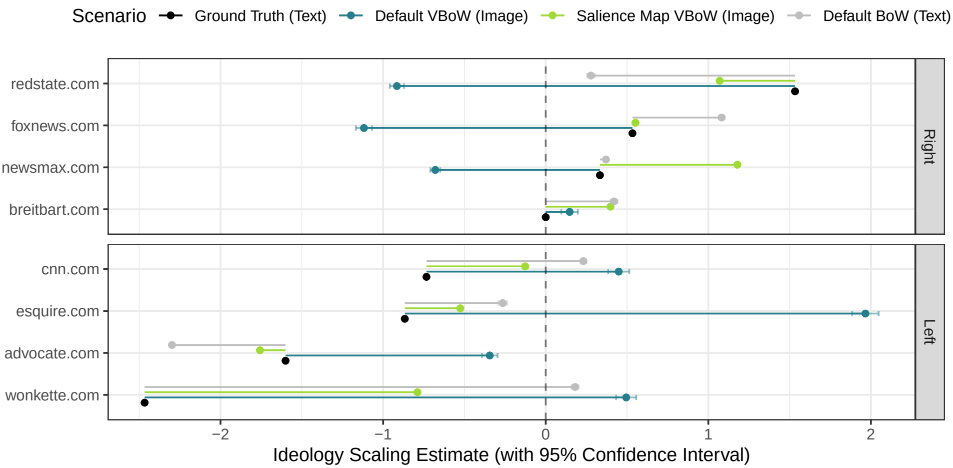
- 59,361 news articles from eight US newspapers spanning across five political issues
- Text-image pairs

Salience Maps and Visual Bag of Words in Practice

- Object Detection: Unsupervised Visual Bag of Words approach (*Torres 2024*)
- Object Salience: Structuring (non-)attentive areas with salient object detection (*Zhang et al. 2015*)

What Does the Model See in Images?

Salience Maps Improve Scaling Images of US Newspapers



Application 2: When Importance Itself is a Relevant Object of Study

Gender Bias in US Presidential Campaign Videos

Wesleyan Media Project *(Fowler, Franz, Ridout, and Baum, 2020/2023)*

- 1,934 video advertisements of candidates during the 2016 and 2020 US presidential races
- Hand-coded meta information on the candidates name, gender, party, or visibility in the video

Gender Bias in US Presidential Campaign Videos

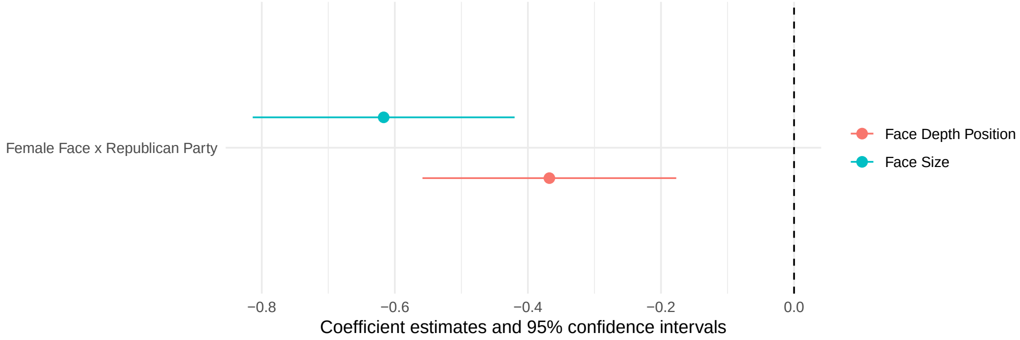
Wesleyan Media Project *(Fowler, Franz, Ridout, and Baum, 2020/2023)*

- 1,934 video advertisements of candidates during the 2016 and 2020 US presidential races
- Hand-coded meta information on the candidates name, gender, party, or visibility in the video

Are Women Less Prominent on Republicans' US Presidential Campaign Videos?

- Object Detection: Face detection and gender classification *(Serengil and Ozpinar 2024)*
- Object Salience: Normalized face depth *(Bhat et al. 2023)* and size

Women are Less Prominent on Republicans' US Presidential Campaign Videos!



Conclusion

A Framework to Analyze the Complex Semantic Structure of Images

- Identifying semantic units and their contextual meaning from text and audio is straightforward nowadays
- The message of an image is more than 'just' the objects in it
- To analyze images more thoroughly, we propose to model

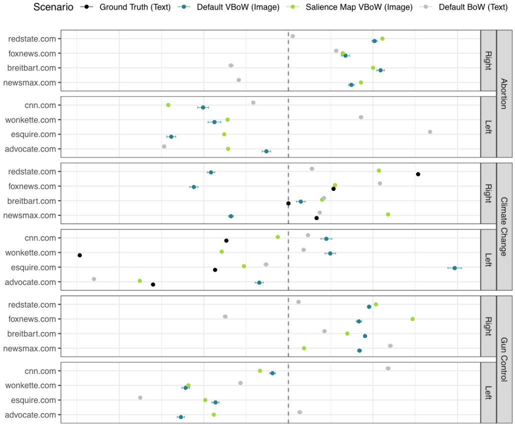
$$\text{Object Prominence} = \text{Object Detection} \times \text{Object Salience}. \tag{2}$$

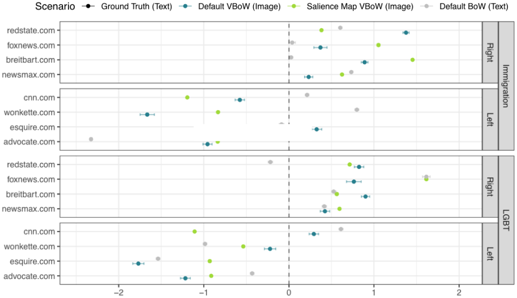
Structuring Quantitative Image Analysis with Object Prominence

Christian Arnold (University of Birmingham)
Andreas Küpfer (TU Darmstadt)

Working Paper:

Appendix: Scaling Images of US Newspapers by Issue





Appendix: Gender Bias in US Presidential Campaign Videos

	Depth Model	Face Size Model
Gender: Female	-0.01	0.48^{***}
	(0.04)	(0.05)
Party: Republican	-0.59^{***}	0.43^{*}
	(0.09)	(0.19)
Gender: Female x Party: Republican	-0.37^{***}	-0.62^{***}
	(0.10)	(0.10)
Num. obs.	67575	67616
Num. groups: Candidate ID	52	52
Num. groups: Candidate Visible	2	2
Num. groups: Election Year	2	2
Pseudo R^2	0.06	0.02

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$