Beyond Prediction: Identifying Latent Treatments in Images

Alex Pugh
Lecturer
Rice University

Michelle Torres
Assistant Professor
UCLA

February 11, 2025

AMERICAS-TEST-2 MARCH 3, 2017 / 2:13 PM / UPDATED 6 YEARS AGO

Exclusive: Trump administration considering separating women, children at Mexico border

AMERICAS-TEST-2 MARCH 3, 2017 / 2:13 PM / UPDATED 6 YEARS AGO

Exclusive: Trump administration considering separating women, children at Mexico border

Trump moves to end 'catch and release', prosecuting parents and removing children who cross border

Questionable court policy separates parents, children

AMERICAS-TEST-2 MARCH 3, 2017 / 2:13 PM / UPDATED 6 YEARS AGO

Exclusive: Trump administration considering separating women, children at Mexico border

Trump moves to end 'catch and release', prosecuting parents and removing children who cross border

Questionable court policy separates parents, children

Attorney General Sessions Delivers Remarks Discussing the Immigration Enforcement Actions of the Trump Administration

San Diego, CA ~ Monday, May 7, 2018

If you are smuggling a child, then we will prosecute you and that child will be separated from you as required by law.

AMERICAS-TEST-2 MARCH 3, 2017 / 2:13 PM / UPDATED 6 YEARS AGO

Exclusive: Trump administration considering separating women, children at Mexico border

Trump moves to end 'catch and release', prosecuting parents and removing children who cross border

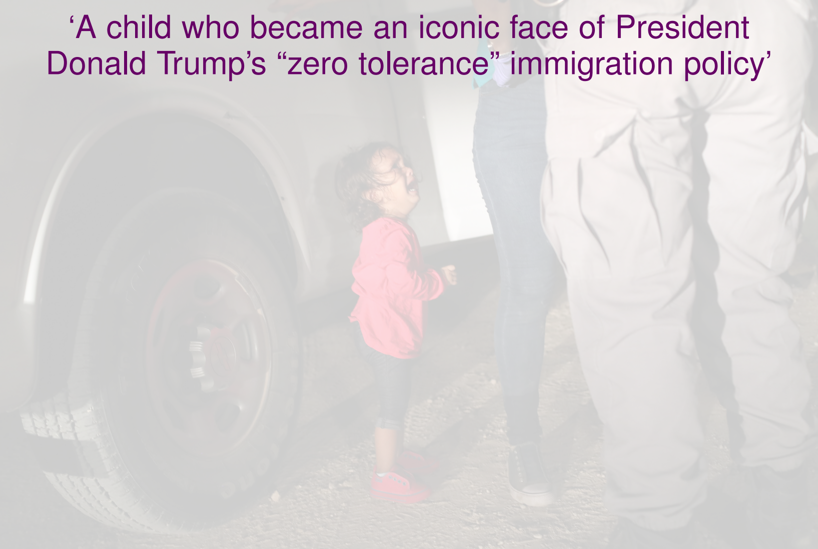
Questionable court policy separates parents, children

Attorney General Sessions Delivers Remarks Discussing the Immigration Enforcement Actions of the Trump Administration

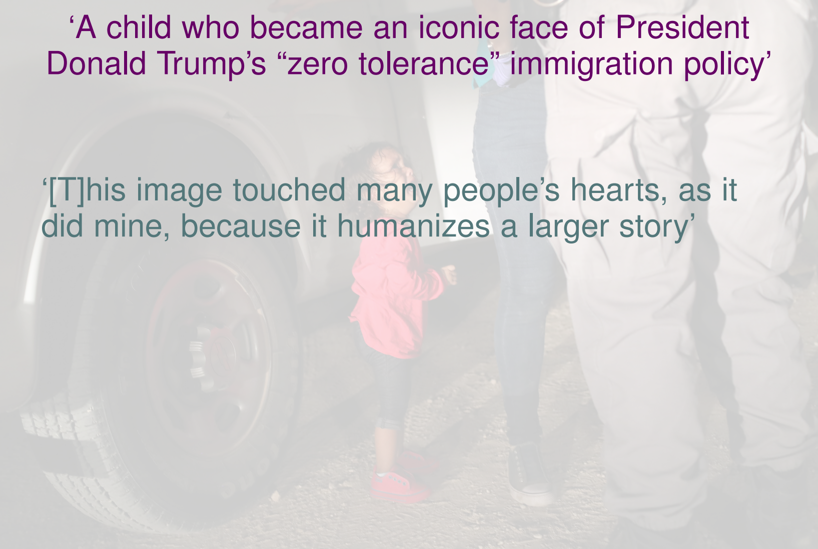
San Diego, CA ~ Monday, May 7, 2018

If you are smuggling a child, then we will prosecute you and that child will be separated from you as required by law.

A family was separated at the border, and this distraught father took his own life

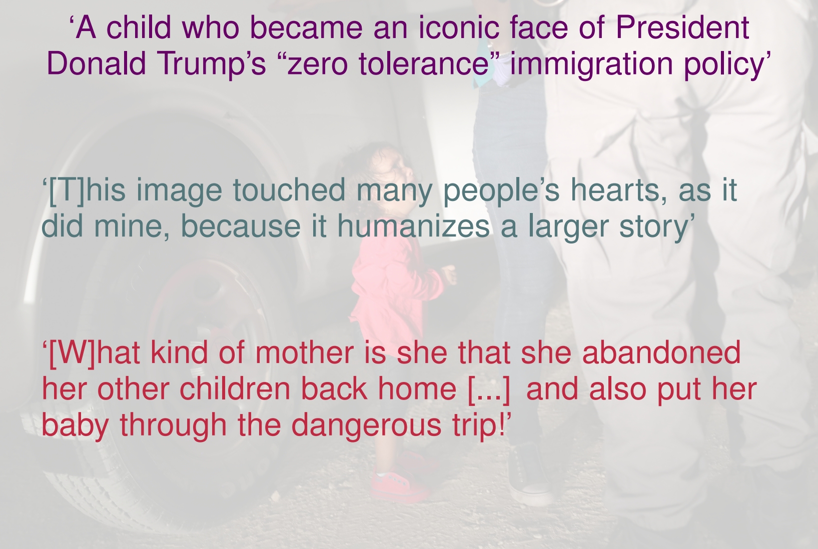


‘A child who became an iconic face of President Donald Trump’s “zero tolerance” immigration policy’



‘A child who became an iconic face of President Donald Trump’s “zero tolerance” immigration policy’

‘[T]his image touched many people’s hearts, as it did mine, because it humanizes a larger story’



‘A child who became an iconic face of President Donald Trump’s “zero tolerance” immigration policy’

‘[T]his image touched many people’s hearts, as it did mine, because it humanizes a larger story’

‘[W]hat kind of mother is she that she abandoned her other children back home [...] and also put her baby through the dangerous trip!’

USING IMAGES IN SOCIAL SCIENCES

USING IMAGES IN SOCIAL SCIENCES

- Studying the effect of presenting information through images
  - Labels vs. Images to signal race/ethnicity (Abrajano, Elmendorf, & Quinn 2018)
  - Visual cues and political knowledge (Prior 2014)

USING IMAGES IN SOCIAL SCIENCES

- Studying the effect of presenting information through images
  - Labels vs. Images to signal race/ethnicity (Abrajano, Elmendorf, & Quinn 2018)
  - Visual cues and political knowledge (Prior 2014)
- Measurement
  - Electoral fraud (Cantú 2019)
  - Displays of emotion (Boussalis et al. 2021)
  - Rural electrification and service provision (Min 2015)

USING IMAGES IN SOCIAL SCIENCES

- Studying the effect of presenting information through images
  - Labels vs. Images to signal race/ethnicity (Abrajano, Elmendorf, & Quinn 2018)
  - Visual cues and political knowledge (Prior 2014)
- Measurement
  - Electoral fraud (Cantú 2019)
  - Displays of emotion (Boussalis et al. 2021)
  - Rural electrification and service provision (Min 2015)
- Use images as a vehicle for a complex treatment
  - Masculinity/femininity (Bauer & Carpinella 2018)
  - Police militarization (Mummolo 2018)
  - Level of conflict on attitudes towards protesters (Torres 2022)

EVALUATING THE IMPACT OF CONFLICT

EVALUATING THE IMPACT OF CONFLICT

EVALUATING THE IMPACT OF CONFLICT

EVALUATING THE IMPACT OF CONFLICT

EVALUATING THE IMPACT OF CONFLICT

How can we estimate causal effects of latent treatments in multi-dimensional interventions like images?

How can we estimate causal effects of latent treatments in multi-dimensional interventions like images?

How can we learn more about the relationship between visual features and outcomes?

How can we estimate causal effects of latent treatments in multi-dimensional interventions like images?

How can we learn more about the relationship between visual features and outcomes?

Use computer vision to go beyond prediction

Multi-dimensional Interventions and Latent Treatments

Multi-dimensional Interventions and Latent Treatments

- Text, speech, or visual content are vehicles to deliver treatment of interest

Multi-dimensional Interventions and Latent Treatments

- Text, speech, or visual content are vehicles to deliver treatment of interest
- They may also contain many other important features that impact outcome

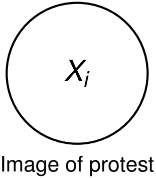
Multi-dimensional Interventions and Latent Treatments

- Text, speech, or visual content are vehicles to deliver treatment of interest
- They may also contain many other important features that impact outcome
  - Latent treatment of interest: treatment of interest that cannot be *directly* manipulated (or *independently* from other features →) *level of conflict*

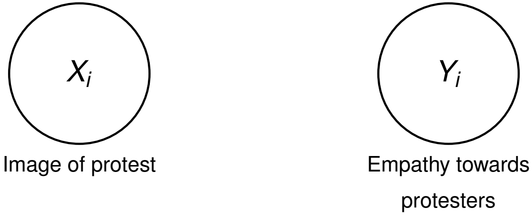
Multi-dimensional Interventions and Latent Treatments

- Text, speech, or visual content are vehicles to deliver treatment of interest
- They may also contain many other important features that impact outcome
  - Latent treatment of interest: treatment of interest that cannot be *directly* manipulated (or *independently* from other features →) *level of conflict*
  - Unmeasured latent treatments: other dimensions potentially confounding or interacting with measured treatment → *magnitude*

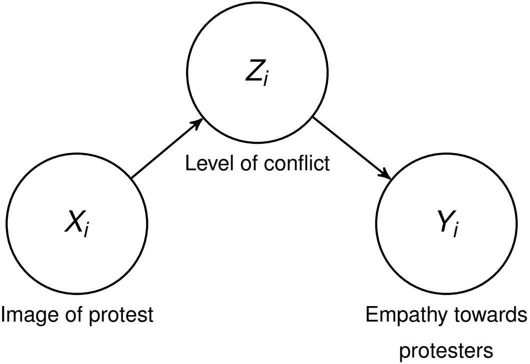
UNDERSTANDING THE PROBLEM: DAG



UNDERSTANDING THE PROBLEM: DAG

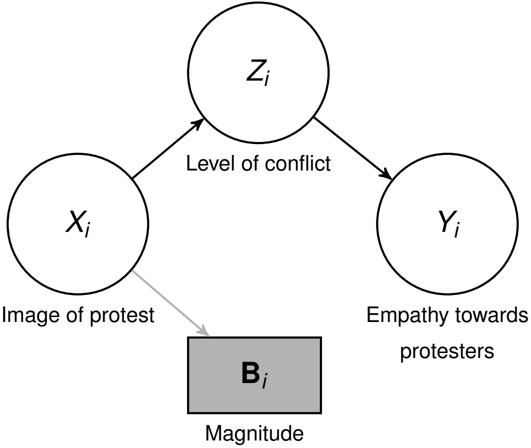


UNDERSTANDING THE PROBLEM: DAG



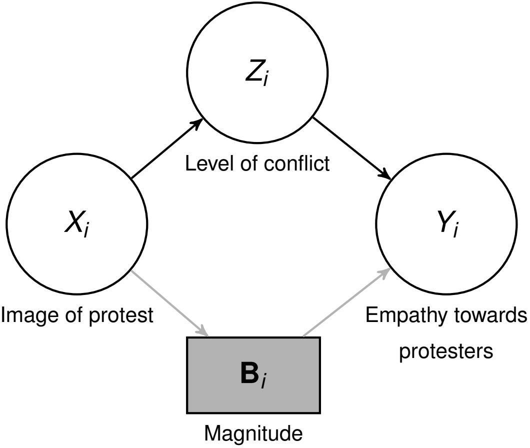
$g()$ maps features of X to Z: $Z_i \equiv g(\mathbf{X}_i) \rightarrow$ It is known!

UNDERSTANDING THE PROBLEM: DAG

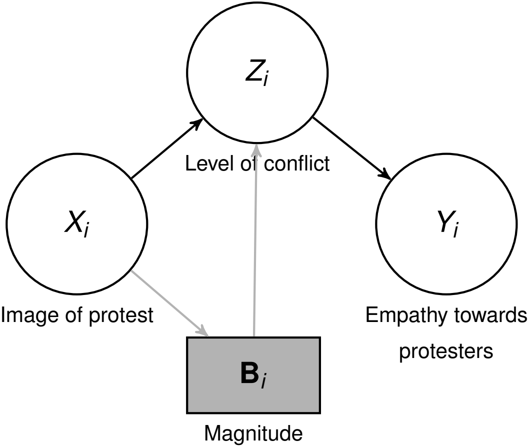


$h()$ maps features of X to $\mathbf{B}$: $\mathbf{B}_i \equiv h(\mathbf{X}_i) \rightarrow$ It is NOT known

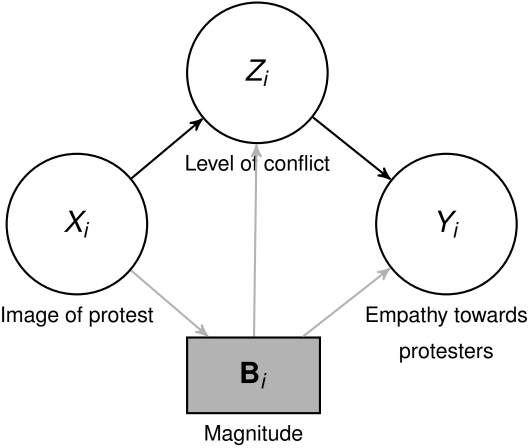
UNDERSTANDING THE PROBLEM: DAG



UNDERSTANDING THE PROBLEM: DAG



UNDERSTANDING THE PROBLEM: DAG



Omitted variable bias scenario

UNDERSTANDING THE PROBLEM IN A NUTSHELL

Understanding the problem in a nutshell

- Randomization of intervention is insufficient to identify causal effects
  - Latent treatment of interest not **directly** randomized
  - Unmeasured latent features might confound estimate of causal effect

UNDERSTANDING THE PROBLEM IN A NUTSHELL

- Randomization of intervention is insufficient to identify causal effects
  - Latent treatment of interest not **directly** randomized
  - Unmeasured latent features might confound estimate of causal effect

What should we do?

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts
3. Discover latent treatments in intervention
   - Generate training and test sets

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts
3. Discover latent treatments in intervention
   - Generate training and test sets
   - Use supervised Indian Buffet Process (sIBP) on training set matrix to infer k latent treatments

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts
3. Discover latent treatments in intervention
   - Generate training and test sets
   - Use supervised Indian Buffet Process (sIBP) on training set matrix to infer k latent treatments
   - Select best model configuration based on quantitative measures and qualitative inspection of identified treatments

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts
3. Discover latent treatments in intervention
   - Generate training and test sets
   - Use supervised Indian Buffet Process (sIBP) on training set matrix to infer k latent treatments
   - Select best model configuration based on quantitative measures and qualitative inspection of identified treatments
4. Estimate the effect of latent treatments

F&G TO THE RESCUE

Develop a process for identifying latent treatments and adjusting for them to estimate the casual effect of the latent treatment of interest

1. Get responses, Y_i
   - Randomly assign intervention to respondents and obtain Y_i
2. Build vocabulary and extract features to construct intervention-feature matrix
   - Document-Term Matrix with texts
3. Discover latent treatments in intervention
   - Generate training and test sets
   - Use supervised Indian Buffet Process (sIBP) on training set matrix to infer k latent treatments
   - Select best model configuration based on quantitative measures and qualitative inspection of identified treatments
4. Estimate the effect of latent treatments
   - Infer latent treatments in the test set
   - Estimate effect of latent treatments using regression in test set

Our Contribution: Adapting and Extending the Framework

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images
  1. Taking multidimensionality serious

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images
  1. Taking multidimensionality serious
  2. “Mechanical” challenge of tokenizing images

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images
  1. Taking multidimensionality serious
  2. “Mechanical” challenge of tokenizing images
     - New framework for building a “DTM” for images → Image-Visual Word Matrix

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images
  1. Taking multidimensionality serious
  2. “Mechanical” challenge of tokenizing images
     - New framework for building a “DTM” for images → Image-Visual Word Matrix
     - New proposal to create and visualize tokens for validation and interpretation purposes

Our Contribution: Adapting and Extending the Framework

- CHALLENGE: How to use this with images
  1. Taking multidimensionality serious
  2. “Mechanical” challenge of tokenizing images
     - New framework for building a “DTM” for images → Image-Visual Word Matrix
     - New proposal to create and visualize tokens for validation and interpretation purposes
  3. Theoretical concerns regarding assumptions

TAKING MULTIDIMENSIONALITY SERIOUS

Taking multidimensionality serious

- Latent treatments are difficult to convey

TAKING MULTIDIMENSIONALITY SERIOUS

- Latent treatments are difficult to convey
- Images might help BUT consider information equivalence
  (Dafoe, Zhang, & Caughey 2018)

Taking multidimensionality serious

- Latent treatments are difficult to convey
- Images might help BUT consider information equivalence (Dafoe, Zhang, & Caughey 2018)
  1. Keep everything constant (!)

TAKING MULTIDIMENSIONALITY SERIOUS

- Latent treatments are difficult to convey
- Images might help BUT consider information equivalence
  (Dafoe, Zhang, & Caughey 2018)
  1. Keep everything constant (!)
  2. Average other factors

Taking multidimensionality serious

- Latent treatments are difficult to convey
- Images might help BUT consider information equivalence (Dafoe, Zhang, & Caughey 2018)
  1. Keep everything constant (!)
  2. Average other factors

Our suggestion

Use a large pool of images representing your treatment of interest

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): ✓

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): $\checkmark$
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): ✓
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:
  1. An outcome vector, Y_i containing the responses associated with each document

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): ✓
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:
  1. An outcome vector, Y_i containing the responses associated with each document
     - ✓ Responses from experimental vignette including images, or

REQUIREMENTS FOR ADAPTING F&G FRAMEWORK

- Relies on having a large pool of texts (*): ✓
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:
  1. An outcome vector, Y_i containing the responses associated with each document
     - ✓ Responses from experimental vignette including images, or
     - ✓ Labels generated by coders (non-experimental set-up)

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): ✓
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:
  1. An outcome vector, Y_i containing the responses associated with each document
     - ✓ Responses from experimental vignette including images, or
     - ✓ Labels generated by coders (non-experimental set-up)
  2. A matrix of word counts per document, $\mathbf{X}_i$

Requirements for Adapting F&G Framework

- Relies on having a large pool of texts (*): ✓
- Mechanical requirements to use sIBP for discovery of latent treatments in texts:
  1. An outcome vector, Y_i containing the responses associated with each document
     - ✓ Responses from experimental vignette including images, or
     - ✓ Labels generated by coders (non-experimental set-up)
  2. A matrix of word counts per document, $\mathbf{X}_i$
     - **✗ Challenge is creating $\mathbf{X}_i$ for images: defining and measuring features**

Challenges of Tokenizing Images

- Texts can be decomposed into words, *n*-grams, sentences = meaningful tokens(*)

Challenges of Tokenizing Images

- Texts can be decomposed into words, *n*-grams, sentences = meaningful tokens(*)
- Images can be decomposed into pixels $\neq$ meaningful tokens

Challenges of Tokenizing Images

- Texts can be decomposed into words, *n*-grams, sentences $=$ meaningful tokens(*)
- Images can be decomposed into pixels $\neq$ meaningful tokens
- Edges, colors, shapes and other features give meaning to images

CHALLENGES OF TOKENIZING IMAGES

- Texts can be decomposed into words, *n*-grams, sentences = meaningful tokens(*)
- Images can be decomposed into pixels $\neq$ meaningful tokens
- Edges, colors, shapes and other features give meaning to images
- Construct “visual words” based on those elements

Constructing Visual Words

Constructing Visual Words

1. Identification of blocks in images

Constructing Visual Words

1. Identification of blocks in images
2. Extraction of features using a CNN

Constructing Visual Words

1. Identification of blocks in images
2. Extraction of features using a CNN
3. Construction of visual vocabulary based on clustering features

Constructing Visual Words

1. Identification of blocks in images
2. Extraction of features using a CNN
3. Construction of visual vocabulary based on clustering features
4. Construction of Image-Visual Word matrix

DIVISION OF IMAGES INTO BLOCKS

(a) Original image (resized)

Division of Images into Blocks

(a) Original image (resized)

(b) Image divided into 32×32 pixels blocks

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Brief crash course on Convolutional Neural Networks

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Brief crash course on Convolutional Neural Networks

- CNNs composed of many connected layers that represent content of an image

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Brief crash course on Convolutional Neural Networks

- CNNs composed of many connected layers that represent content of an image
- They scan images and create “feature maps” that convey information about the presence of complex features and their interactions

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Brief crash course on Convolutional Neural Networks

- CNNs composed of many connected layers that represent content of an image
- They scan images and create “feature maps” that convey information about the presence of complex features and their interactions
- Final *dense* layer outputs probability of image belonging to each potential label

Feature Extraction with CNNs

- Image blocks serve as the inputs of a CNN: an image composed of mini images

Brief crash course on Convolutional Neural Networks

- CNNs composed of many connected layers that represent content of an image
- They scan images and create "feature maps" that convey information about the presence of complex features and their interactions
- Final *dense* layer outputs probability of image belonging to each potential label
- These probabilities are obtained through back-propagation and the minimization of prediction error in a training set (based on labeled data)

FEATURE EXTRACTION USING CNNS

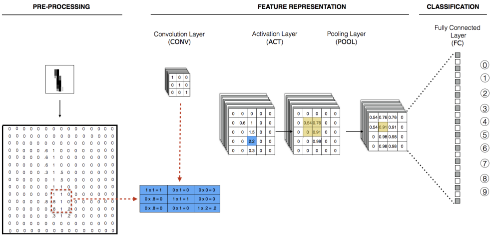


Figure 1. Example of a convolutional neural network structure.

FEATURE EXTRACTION USING CNNS

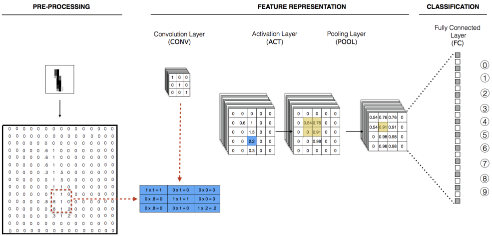


Figure 1. Example of a convolutional neural network structure.

- Use pre-trained or tuned model on each block of an image

FEATURE EXTRACTION USING CNNs

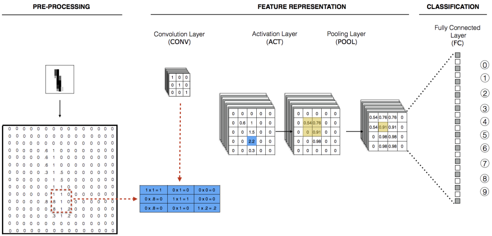


Figure 1. Example of a convolutional neural network structure.

- Use pre-trained or tuned model on each block of an image
- The CNN creates feature maps of “elements/descriptors” that can be found in an image

Feature extraction using CNNs

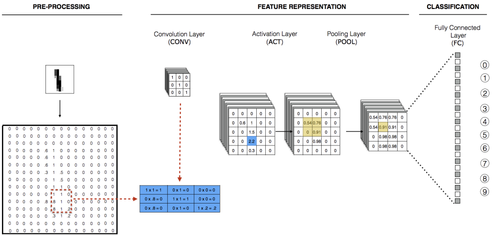


Figure 1. Example of a convolutional neural network structure.

- Use pre-trained or tuned model on each block of an image
- The CNN creates feature maps of “elements/descriptors” that can be found in an image
- Remove the dense layer (the final one) and keep an appropriate feature map vector → “Predictors” → Strongly associated to the treatment of interest.

Feature extraction using CNNs

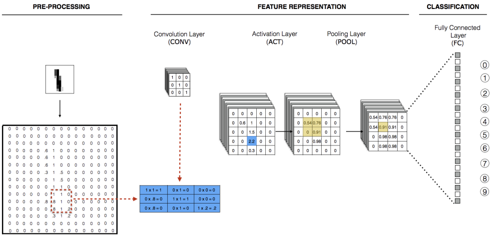


Figure 1. Example of a convolutional neural network structure.

- Use pre-trained or tuned model on each block of an image
- The CNN creates feature maps of “elements/descriptors” that can be found in an image
- Remove the dense layer (the final one) and keep an appropriate feature map vector → “Predictors” → Strongly associated to the treatment of interest.
- = Each image is described by vector of size *number of blocks* × *number of features from CNN*

FEATURE EXTRACTION USING CNNS

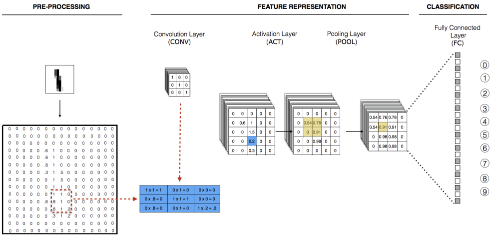


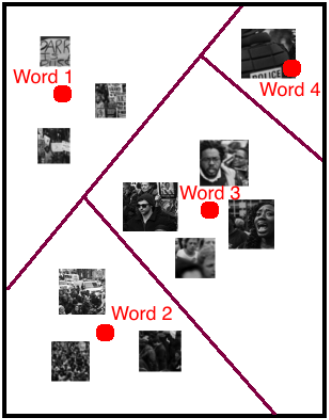
Figure 1. Example of a convolutional neural network structure.

- Use pre-trained or tuned model on each block of an image
- The CNN creates feature maps of “elements/descriptors” that can be found in an image
- Remove the dense layer (the final one) and keep an appropriate feature map vector → “Predictors” → Strongly associated to the treatment of interest.
- = Each image is described by vector of size *number of blocks* × *number of features from CNN*
- In our applications, this is 70×2,048

Clustering Features to Build Visual Vocabulary

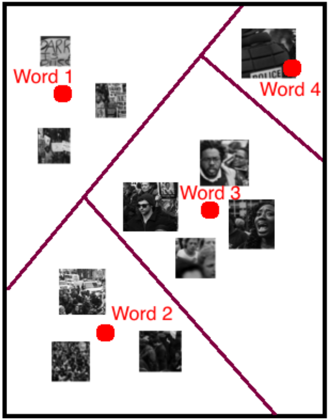
CLUSTERING FEATURES TO BUILD VISUAL VOCABULARY

- Need for tokens → Words in columns of a DTM



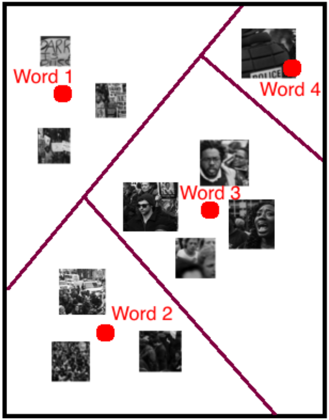
CLUSTERING FEATURES TO BUILD VISUAL VOCABULARY

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)



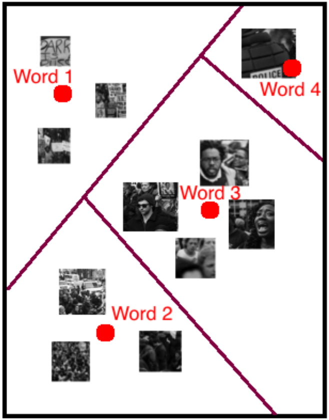
Clustering Features to Build Visual Vocabulary

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)
- Cluster randomly selected sample of feature vectors



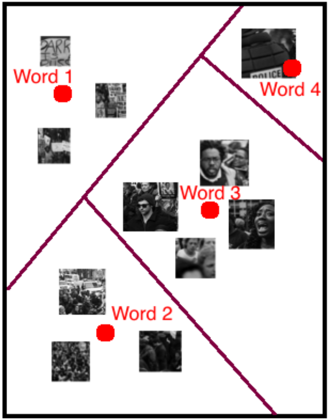
Clustering Features to Build Visual Vocabulary

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)
- Cluster randomly selected sample of feature vectors
- Centroid of cluster is the "visual word"



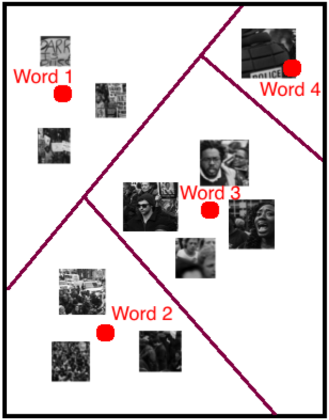
Clustering Features to Build Visual Vocabulary

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)
- Cluster randomly selected sample of feature vectors
- Centroid of cluster is the "visual word"
- Why do we do this?



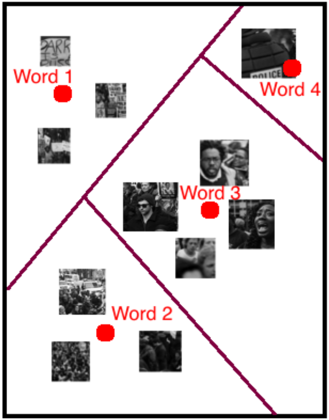
Clustering Features to Build Visual Vocabulary

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)
- Cluster randomly selected sample of feature vectors
- Centroid of cluster is the "visual word"
- Why do we do this?
  - Similar features = Same concept



Clustering Features to Build Visual Vocabulary

- Need for tokens → Words in columns of a DTM
- Define *v* clusters (= # of desired visual words)
- Cluster randomly selected sample of feature vectors
- Centroid of cluster is the "visual word"
- Why do we do this?
    - Similar features = Same concept
    - Reduce potential sparsity in IVWM



Visualizing Visual Words

- Blocks that belong to a given cluster are similar in terms of feature vectors

Visualizing Visual Words

- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar

Visualizing Visual Words

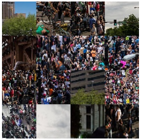
- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster

Visualizing Visual Words

- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the “average” block representing the cluster

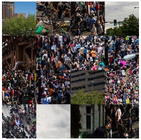
Visualizing Visual Words

- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct "visual words" using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the "average" block representing the cluster



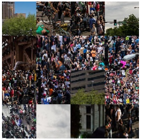
VISUALIZING VISUAL WORDS

- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the “average” block representing the cluster



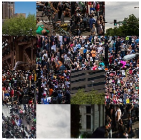
VISUALIZING VISUAL WORDS

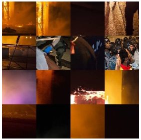
- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the “average” block representing the cluster



Visualizing Visual Words

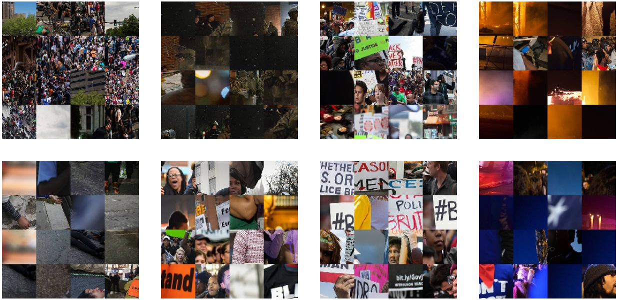
- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the “average” block representing the cluster





VISUALIZING VISUAL WORDS

- Blocks that belong to a given cluster are similar in terms of feature vectors
- Should look visually similar
- Construct “visual words” using the 16 feature vectors closest to each of the centroid of the cluster
- E.g. the most similar blocks to the “average” block representing the cluster



Building the IVWM to emulate DTM

BUILDING THE IVWM TO EMULATE DTM

Count the number of times each visual word appears in an image

Building the IVWM to emulate DTM

Count the number of times each visual word appears in an image

- Also not trivial...

Building the IVWM to emulate DTM

Count the number of times each visual word appears in an image

- Also not trivial...
- Assign each feature vector to the most similar visual word in the vocabulary

BUILDING THE IVWM TO EMULATE DTM

Count the number of times each visual word appears in an image

- Also not trivial...
- Assign each feature vector to the most similar visual word in the vocabulary
  - Compute the Euclidean distance between each feature vector and the centroids of the clusters

Building the IVWM to emulate DTM

Count the number of times each visual word appears in an image

- Also not trivial...
- Assign each feature vector to the most similar visual word in the vocabulary
  - Compute the Euclidean distance between each feature vector and the centroids of the clusters
  - Assign feature vector to visual word with shortest distance to centroid

QUICK STOP/WARNING

QUICK STOP/WARNING

- *Technically* we can run the sIBP already

QUICK STOP/WARNING

- *Technically* we can run the sIBP already
- ...and continue with the estimation of causal effects

Quick stop/warning

- *Technically* we can run the sIBP already
- ...and continue with the estimation of causal effects
- But what about the translation of assumptions to the world of images?

Quick stop/warning

- *Technically* we can run the sIBP already
- ...and continue with the estimation of causal effects
- But what about the translation of assumptions to the world of images?
- Our paper discusses translation, violations, and potential alternatives to maximize fulfillment

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to "alert" people about climate change and trigger pro-environment attitudes and behavior.

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to "alert" people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the "main subject" in pictures of climate change affect the perceptions of and reactions to it?

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to "alert" people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the "main subject" in pictures of climate change affect the perceptions of and reactions to it?
  - Latent treatment of interest: "victim" of climate change (Animal, Human, Object/Scene)

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to “alert” people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the “main subject” in pictures of climate change affect the perceptions of and reactions to it?
    - Latent treatment of interest: “victim” of climate change (Animal, Human, Object/Scene)
    - Outcome: Evaluations of whether climate changes affects own’s family and society

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to “alert” people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the “main subject” in pictures of climate change affect the perceptions of and reactions to it?
  - Latent treatment of interest: “victim” of climate change (Animal, Human, Object/Scene)
  - Outcome: Evaluations of whether climate changes affects own’s family and society
- Expectations:

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to "alert" people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the "main subject" in pictures of climate change affect the perceptions of and reactions to it?
  - Latent treatment of interest: "victim" of climate change (Animal, Human, Object/Scene)
  - Outcome: Evaluations of whether climate changes affects own's family and society
- Expectations:
  - Images with humans generate **stronger** perceptions of family/society being affected by climate change in comparison to images with animals

Application: Framing Climate Change

- Motivation: Campaigns and ads with messages with varying content trying to “alert” people about climate change and trigger pro-environment attitudes and behavior.
- Question: Does presenting humans as the “main subject” in pictures of climate change affect the perceptions of and reactions to it?
  - Latent treatment of interest: “victim” of climate change (Animal, Human, Object/Scene)
  - Outcome: Evaluations of whether climate changes affects own’s family and society
- Expectations:
  - Images with humans generate **stronger** perceptions of family/society being affected by climate change in comparison to images with animals
  - Images with objects generate **weaker** perceptions of family/society being affected by climate change in comparison to images with animals

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473
  - Testing: 714

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473
  - Testing: 714
- Corpus of 722 images from the *Climate Visuals* library: a project of *Climate Outreach* that compiles images that visualize the causes, effects, and solutions to climate change.

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473
  - Testing: 714
- Corpus of 722 images from the *Climate Visuals* library: a project of *Climate Outreach* that compiles images that visualize the causes, effects, and solutions to climate change.
- $\Rightarrow$ Extracted 70 feature vectors of size 2,048 per image, and clustered 30% of them into 750 visual words

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473
  - Testing: 714
- Corpus of 722 images from the *Climate Visuals* library: a project of *Climate Outreach* that compiles images that visualize the causes, effects, and solutions to climate change.
- ⇒ Extracted 70 feature vectors of size 2,048 per image, and clustered 30% of them into 750 visual words
  - Extraction using a tuned `ResNet50`: the last 2 (of 50) layers were re-trained using treatment labels

Application: Framing Climate Change, cont.

- Qualtrics survey with 2,187 respondents
  - Training: 1,473
  - Testing: 714
- Corpus of 722 images from the *Climate Visuals* library: a project of *Climate Outreach* that compiles images that visualize the causes, effects, and solutions to climate change.
- $\Rightarrow$ Extracted 70 feature vectors of size 2,048 per image, and clustered 30% of them into 750 visual words
  - Extraction using a tuned `ResNet50`: the last 2 (of 50) layers were re-trained using treatment labels
- sIBP with $k = 5$ latent treatments

EXAMPLES OF IMAGES IN EACH TREATMENT GROUP

(a) Animal

(b) Human

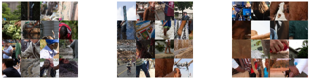
(c) Object/Scene

Results I: Top Words

Results I: Top Words

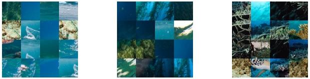
Z1: Snow, Ice & Sky

Z2: Body parts & Textures



Z3: Fire, warm, & reds

Z4: Water, Blue, & Waves



Z5: Sand, Dry, & Heat

Results II: Images featuring each latent component prominently

Results II: Images featuring each latent component prominently

Z1: Ice, Snow & Sky



Z2: Body parts & Textures



Z4: Fire, warm, & reds

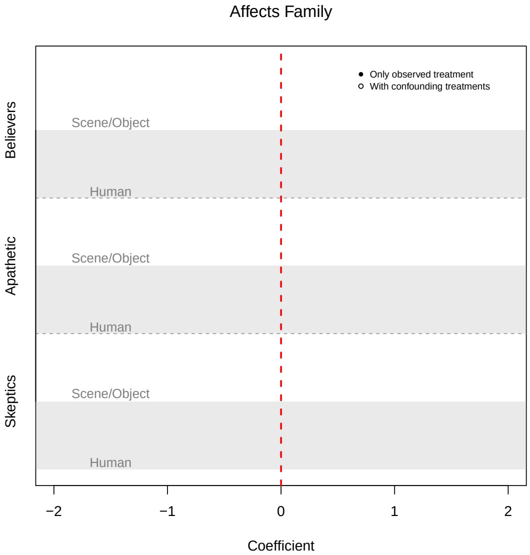


Z5: Water, Blue, & Waves

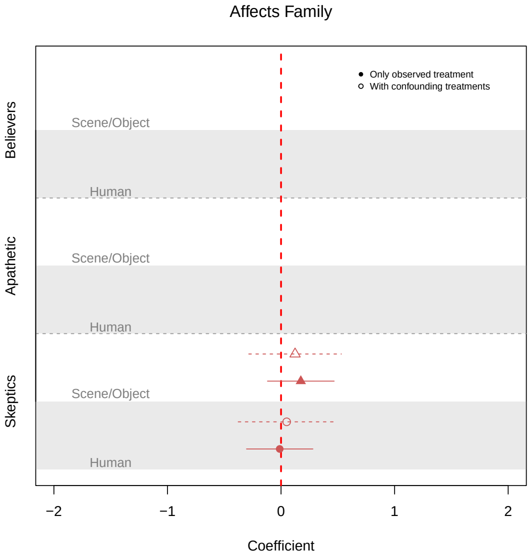


Results III: Estimation of main treatment effects, Affects family

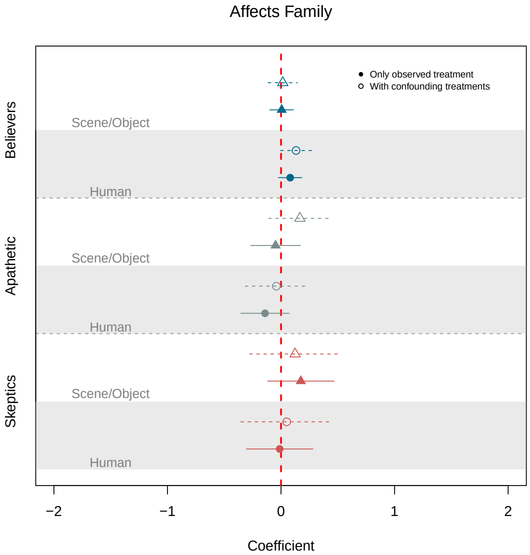
Results III: Estimation of main treatment effects, Affects family



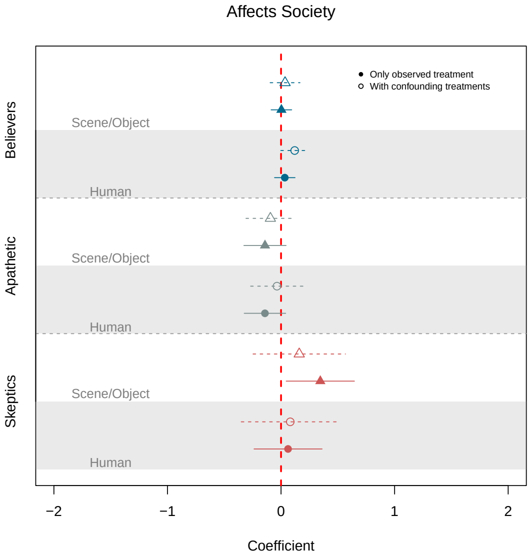
Results III: Estimation of main treatment effects, Affects family



Results III: Estimation of main treatment effects, Affects family



Results III: Estimation of main treatment effects, Affects society



Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements

Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict

Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict
- Identify the features in the images that influence respondents' classifications/answers

Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict
- Identify the features in the images that influence respondents' classifications/answers
- Outcome: low, medium or high conflict

Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict
- Identify the features in the images that influence respondents' classifications/answers
- Outcome: low, medium or high conflict
- 1,487 images from the BLM protests sampled from a collection compiled by *Getty Images* during the 2014 BLM protests

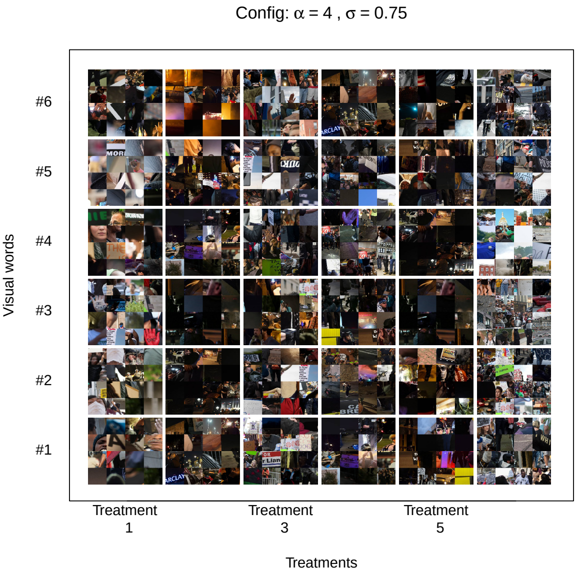
Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict
- Identify the features in the images that influence respondents' classifications/answers
- Outcome: low, medium or high conflict
- 1,487 images from the BLM protests sampled from a collection compiled by *Getty Images* during the 2014 BLM protests
- Survey on Lucid with 1,478 respondents

Application 2: BLM Protests

- Exploring the effect of visual conflict on attitudes towards social movements
- First step: labeling images according to the level of conflict they depict
- Identify the features in the images that influence respondents' classifications/answers
- Outcome: low, medium or high conflict
- 1,487 images from the BLM protests sampled from a collection compiled by *Getty Images* during the 2014 BLM protests
- Survey on Lucid with 1,478 respondents
- 200 words with $k = 6$ treatments

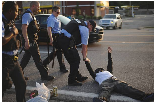
Results I: Top words



IMAGES FEATURING EACH OF THE LATENT TREATMENTS

Z1: Signs/ Hands

Z4: Pavement



Z5: Police/ Dark clothing

EFFECT OF LATENT TREATMENTS

	All	Democrats	Republicans	Independents
	(1)	(2)	(3)	(4)
Z1: Signs/Hands/Faces	−**0.080**	−0.040	−**0.102**	−**0.158**
	(0.024)	(0.037)	(0.040)	(0.061)
Z2: Night/Lights	0.018	0.020	0.011	0.035
	(0.025)	(0.036)	(0.043)	(0.064)
Z3: Close-up/Small group	−0.003	−0.012	0.028	−0.057
	(0.024)	(0.035)	(0.042)	(0.063)
Z4: Pavement	**0.086**	**0.099**	0.017	**0.189**
	(0.026)	(0.037)	(0.043)	(0.066)
Z5: Police/Dark clothing	**0.103**	0.070	**0.180**	0.049
	(0.027)	(0.037)	(0.043)	(0.069)
Z6: Daylight protester	−**0.138**	−**0.149**	−**0.099**	−**0.193**
	(0.025)	(0.035)	(0.042)	(0.064)
Constant	**1.997**	**1.971**	**2.006**	**2.057**
	(0.026)	(0.037)	(0.043)	(0.066)
N	4,580	2,245	1,642	693
Adjusted R^2	0.013	0.010	0.014	0.026

Bold coefficient: $p \leq 0.05$. Bootstrapped standard errors shown.

THINGS TO CONSIDER

- Coherence and symbolism of visual words

Things to consider

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept

THINGS TO CONSIDER

- Coherence and symbolism of visual words
    - Some tokens are similar in terms of features but not concept
    - Others are more synonyms than exclusive words

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction
  - Mapping of latent concepts to low-dimensional features

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction
  - Mapping of latent concepts to low-dimensional features
  - Sensitivity of results to feature definition/extraction

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction
  - Mapping of latent concepts to low-dimensional features
  - Sensitivity of results to feature definition/extraction
- Interpretation of treatments:

THINGS TO CONSIDER

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction
  - Mapping of latent concepts to low-dimensional features
  - Sensitivity of results to feature definition/extraction
- Interpretation of treatments:
  - “tea leaves reading”

Things to consider

- Coherence and symbolism of visual words
  - Some tokens are similar in terms of features but not concept
  - Others are more synonyms than exclusive words
  - Some visual words are plainly “bad”
- Dimension reduction
  - Mapping of latent concepts to low-dimensional features
  - Sensitivity of results to feature definition/extraction
- Interpretation of treatments:
  - “tea leaves reading”
  - Post-treatment treatments!

Further research

- Diagnosis and hyperparameter search:
  - Feature extraction: number and size of blocks, number of clusters, CNN model, basic pre-trained vs. transfer learning
  - "Curated" vocabulary
  - sIBP: number of treatments, model selection, qualitative assessment
- After latent treatment identification: power, refining experimental design
  - Photoshop...
  - ...or new techniques (currently experimenting with LVMs)
- More methods for disentangling the relationship between features and labels/outcomes: SPRAY, heatmaps, and more.

Thank you!

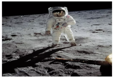
Alex Pugh, alexpugh@rice.edu
Michelle Torres, smtorres@ucla.edu

APPENDIX

Table of Contents

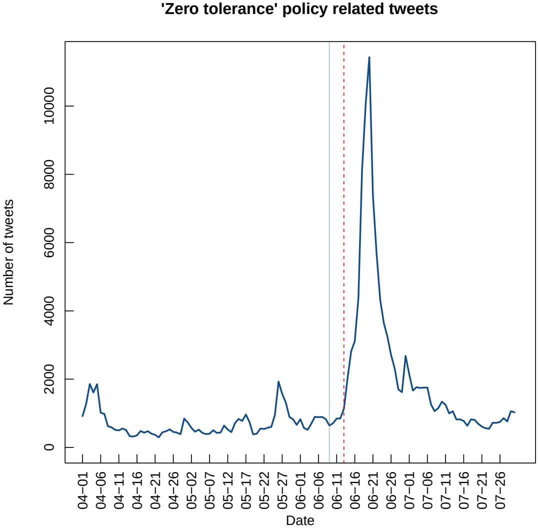
IMAGES ARE POWERFUL

10 Ph

- Activate unconscious cognitive processes (LeDoux 1986, Zajonc 1984)
- Affect attention and content processing (Smith et al. 2001)
- Increase the credibility of information: “see it to believe it” (Campbell 2004)
- Provide “easy to digest” rich information (Lang, Potter and Bolls 1999)
- Communicate and highlight particular messages (Barry 1997; Gamson 1989; Parry 2011)

PUBLIC DISCUSSION OF 'ZERO TOLERANCE' POLICY



Understanding the problem: Notation

- Objective: Understand how users respond to texts, $\mathcal{X}$
- Potential outcome: $Y_i(\mathbf{X}_i)$
- But, interest is in latent treatment's effect. Let $g : \mathcal{X} \rightarrow \{0, 1\}$ (presence or absence of treatment of interest)
- If $g(\mathbf{X}_i) = 1$, then latent treatment is present; if $g(\mathbf{X}_i) = 0$, then treatment is absent.
- Latent treatment in text assigned to respondent i: $Z_i \equiv g(\mathbf{X}_i)$
- (Very Likely) Assumption: There exists other set of unmeasured latent treatments, $\mathbf{B}_i \equiv h(\mathbf{X}_i)$
- Important: $h(\cdot)$ and $g(\cdot)$ capture all relevant features of the text.
- Thus, the potential outcome is $Y(\mathbf{X}_i) = Y(Z_i, \mathbf{B}_i)$
- g is known, but h isn't.

Constructing Visual Words: Overview

Original Steps (Grauman & Darrell, 2005)

1. Identification of key points
2. Description of key points based on pixel intensity
3. Construction of visual vocabulary based on clustering features
4. Construction of Image-Visual Word matrix

Proposed Process

1. **Identification of blocks in images**
2. **Extraction of features using a CNN**
3. Construction of visual vocabulary based on clustering features
4. Construction of Image-Visual Word matrix

Table index

Benefits of Blocking Compared to Key Point Detection

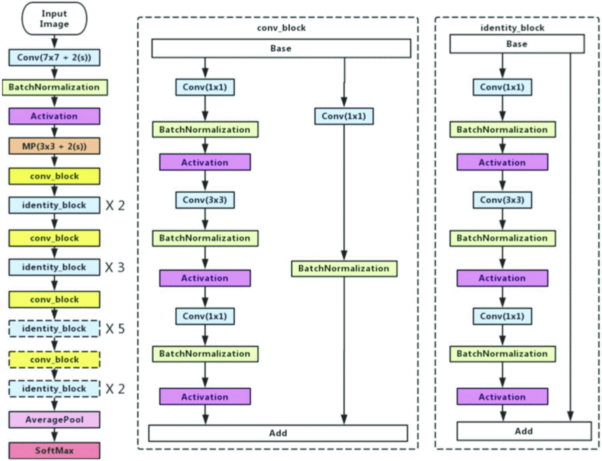
- Key Point Detection:
  - Identifies salient regions in images but...
  - ...can result in high variance in number of key points per image
  - Discussion of whether salient areas are the *only* ones providing information
  - Dependency on *another* hyperparameter
- Blocking:
  - Standardizes areas of the images that are worth focusing on
  - Size of blocks adaptable to image data complexity
    - Complex Images with multiple elements in smaller sizes = small blocks
    - Simple, parsimonious images with few elements = large blocks
    - Similar to transformer set-up

Choosing a CNN

- Ideal: Select a CNN trained on images resembling concept of interest
- This is not trivial...
- Two Alternatives
  1. Use pre-trained model without the output layers
  2. Retrain some layers in existing model via transfer learning
- In our applications, we use ResNet50 trained on ImageNet (14 million+ images of 1,000 categories)
  - Not ideal in terms of "fit" but...
  - Offers a conservative test
  - It includes some interesting and relevant categories (e.g. police car, assault gun, aircraft, chainlink fence, etc.)
  - Interested in patterns not labels

Table index

ResNet50 architecture



OTHER ALTERNATIVES TO EXTRACT TOKENS

- Basic Histogram of Gradients (HoGs) → “Too simple” (*)
- Object detection in each image and feature count → A priori knowledge of what to find
- Layer-wise relevance propagation heatmaps → Information at the image level; isolated features are hard to track
- Transformers → better at prediction but less applicable to feature extraction. Similar issues with respect to interpretation(*)

BAD VISUAL WORDS

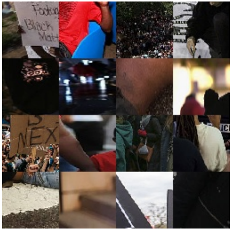
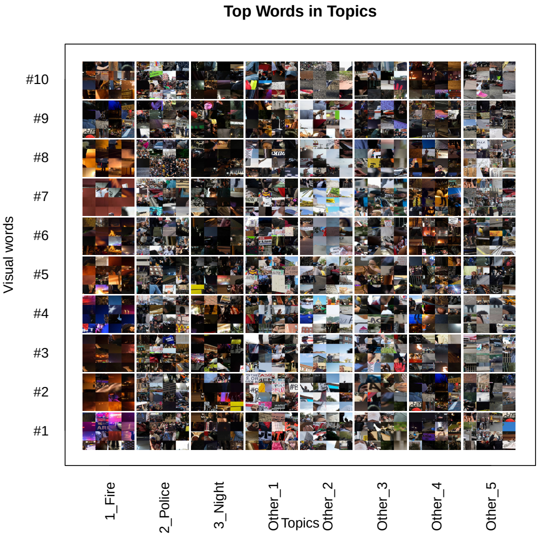


Table index

POTENTIAL SOLUTIONS: KEYATM



THEORETICAL ASSUMPTIONS I

SUTVA

For all individuals i and any X, X' such that $X_{j[i]} = X'_{j[i]}$, $Y_i(X) = Y(X')$

An individual's response to an image is only impacted by the assigned image

Potential violations:

- If coding multiple images, individuals' responses may be influenced by preceding images
- Analyst induced SUTVA violations if same images used for discovery of latent treatments and estimation of causal effects (Egami et al. 2018)

Theoretical Assumptions II

Ignorability and Positivity

For all individuals i, $Y_i(x) \perp\!\!\!\perp X_i$ *and* $Pr(X_i = x) > 0$ *for all* $x \in X$

- Independence of the treatment assignment from the potential outcomes
- Every treatment has a chance of being observed

Potential violations:

- Satisfied with proper randomization of images based on N individuals and n_t images per treatment
- Caution about the estimation of causal effects if missingness in coder labels, removal of low quality responses, attrition caused by the treatment

THEORETICAL ASSUMPTIONS III

Sufficiency

For all X and X' such that $g(x) = g(X')$, $E[Y_i(g(X))] = E[Y_i(g(X'))]$ and $Pr(Z_i = 1|\mathbf{B}_i = \mathbf{b})$

- Codebook function identifies all information in an image that is relevant to the response

Potential violations:

- The tokenization of images process may remove or reduce information about features that are relevant to latent treatment of interest and individuals' responses
- Might expect violations especially if tokenization process removes information regarding color

Theoretical Assumptions III, cont.

- Color of flags indicating ideological stand

Theoretical Assumptions IV

Common Support

For all X and X' such that $g(x) = g(X')$, $E[Y_i(g(X))] = E[Y_i(g(X'))]$ and $Pr(Z_i = 1|\mathbf{B}_i = \mathbf{b})$

- All combinations of latent treatments have a non-zero probability of being observed
- No aliasing between latent treatments

Potential violations:

- May be the assumption most likely violated with images
- Some treatment combinations may not be present in body of images because latent features naturally correlate
- Challenging to manipulate images to satisfy this assumption

THEORETICAL ASSUMPTIONS IV, CONT.

- Assume Z_1 is "children" and Z_3 is "fire." Unlikely to find both in one picture.
- However...
- Thus, be careful!

EXPERIMENTAL DESIGN: TREATMENT COMBINATIONS

Experimental design: Treatment combinations, cont.

Experimental design: Treatment Combinations, cont.

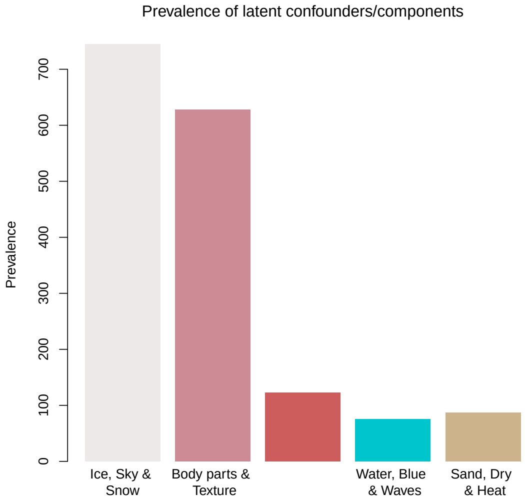
Climate Change: Dataset

- Climate Visuals library (link)
- Only Creative Commons images used
- Images visualize full concept of climate change
- Impacts, Causes, and Solutions
- Geographically and ethnically diverse in composition

Correlation between treatments

	Z1	Z2	Z3	Z4	Z5
Z1: Ice, sky & snow	1	0.461	0.015	−0.005	−0.041
Z2: Body parts & Texture	0.461	1	-0.014	−0.016	0.028
Z3: Fire, warm & reds	0.015	−0.014	1	0.057	0.059
Z4: Water, blue & waves	-0.005	−0.016	0.057	1	-0.021
Z5: Sand, dry & heat	-0.041	0.028	0.059	-0.021	1

Distribution of treatments in sample



Estimation of confounding effects, Affects family

	Climate change affects my family		
	Skeptic	Apathetic	Believers
Z1: Ice, sky & snow	−0.056	−0.139	−0.021
	(0.190)	(0.127)	(0.060)
Z2: Body parts & Texture	−0.091	−0.010	0.085
	(0.189)	(0.127)	(0.060)
Z3: Fire, warm & reds	−0.203	−0.027	−0.081
	(0.282)	(0.183)	(0.086)
Z4: Water, blue & waves	0.060	0.190	0.094
	(0.284)	(0.219)	(0.099)
Z5: Sand, dry & heat	0.123	0.362	0.034
	(0.346)	(0.224)	(0.087)
Constant	**2.249**	**2.762**	**3.341**
	(0.125)	(0.088)	(0.045)
N	182	252	760
R^2	0.007	0.020	0.005

Bold coefficient: $p \leq 0.05$

Effect of confounding treatments on treatment of interest

	Human Treatment
Z1: Ice, sky & snow	-0.222^{*}
	(0.030)
Z2: Body parts & Texture	0.243^{*}
	(0.030)
Z3: Fire, warm & reds	-0.032
	(0.043)
Z4: Water, blue & waves	-0.254^{*}
	(0.049)
Z5: Sand, dry & heat	0.060
	(0.046)
Constant	0.358^{*}
	(0.021)
N	1,194
Log Likelihood	-758.470
AIC	1,528.940

* $p < 0.05$

Table index