Mapping Visual Themes among Authentic and Coordinated Memes

Keng-Chi Chang

UC San Diego

July 6, 2022

Political Networks Conference (PolNet)

Background & research question

- Huge parts of political social media
  - 490K Twitter users, 19% tweets are memes, 30% political Du, Masood, and Joseph 2020

Background & research question

- Huge parts of political social media
  - 490K Twitter users, 19% tweets are memes, 30% political — Du, Masood, and Joseph 2020
- Information campaigns share a lot of images and memes

Background & research question

- Huge parts of political social media
    - 490K Twitter users, 19% tweets are memes, 30% political Du, Masood, and Joseph 2020
- Information campaigns share a lot of images and memes
- Russian Internet Research Agency (IRA) shared 1.8M images during 2016 election

Background & research question

- Huge parts of political social media
  - 490K Twitter users, 19% tweets are memes, 30% political Du, Masood, and Joseph 2020
- Information campaigns share a lot of images and memes
- Russian Internet Research Agency (IRA) shared 1.8M images during 2016 election
  - IRA employees are assessed for meme-making abilities DiResta, Grossman, and Siegel 2021

Background & research question

- Huge parts of political social media
    - 490K Twitter users, 19% tweets are memes, 30% political — Du, Masood, and Joseph 2020
- Information campaigns share a lot of images and memes
- Russian Internet Research Agency (IRA) shared 1.8M images during 2016 election
    - IRA employees are assessed for meme-making abilities — DiResta, Grossman, and Siegel 2021
- Observationally, what distinguishes state-linked memes from authentic memes?

Methods overview

1. Classify IRA images into memes vs. non-memes
   - Focus on IRA images predicted as memes (coordinated memes)

Methods overview

1. Classify IRA images into memes vs. non-memes
   - Focus on IRA images predicted as memes (coordinated memes)
2. Extract embeddings of visual feature *jointly* for both authentic memes (from Reddit) and coordinated memes (from IRA) using DeepCluster Caron et al. 2019

Methods overview

1. Classify IRA images into memes vs. non-memes
   - Focus on IRA images predicted as memes (coordinated memes)
2. Extract embeddings of visual feature *jointly* for both authentic memes (from Reddit) and coordinated memes (from IRA) using DeepCluster Caron et al. 2019
3. Cluster memes based on the vectors of embeddings using K-means

Methods overview

1. Classify IRA images into memes vs. non-memes
   - Focus on IRA images predicted as memes (coordinated memes)
2. Extract embeddings of visual feature *jointly* for both authentic memes (from Reddit) and coordinated memes (from IRA) using DeepCluster Caron et al. 2019
3. Cluster memes based on the vectors of embeddings using K-means
4. Label the clusters based on representative images and compare the difference in proportions between coordinated and authentic memes

Data collection

- A random sample of 26K images shared by IRA on Twitter

Data collection

- A random sample of 26K images shared by IRA on Twitter
- 26K authentic memes collected from the `r/meme` subreddit
  - Some can still be coordinated, but percentage should be low

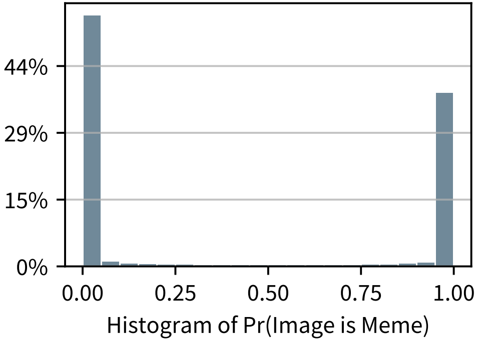
Data collection

- A random sample of 26K images shared by IRA on Twitter
- 26K authentic memes collected from the `r/meme` subreddit
  - Some can still be coordinated, but percentage should be low
- 15K non-meme image-with-text data as negative samples (step 1)
  - COCO-Text: large scale dataset for text recognition in natural scenes (not superimposed)
  - Don't want models simply picking up textual features in images and classifying images into with vs. without text

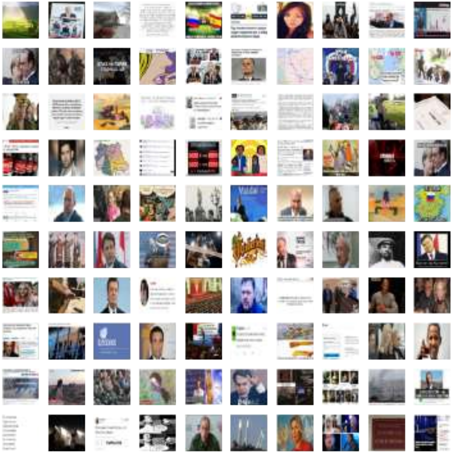
Example from COCO-Text Dataset

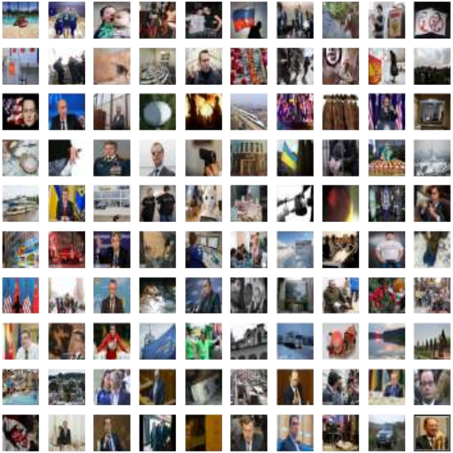
Classify IRA images into memes

- Data: memes from Reddit vs. non-meme images-with-text (COCO-text)
- Model: ResNet-50
- Performance: Accuracy > 0.97 on test set
- Apply the trained model on IRA images



Predicted IRA memes vs. non-memes





Identify themes in memes

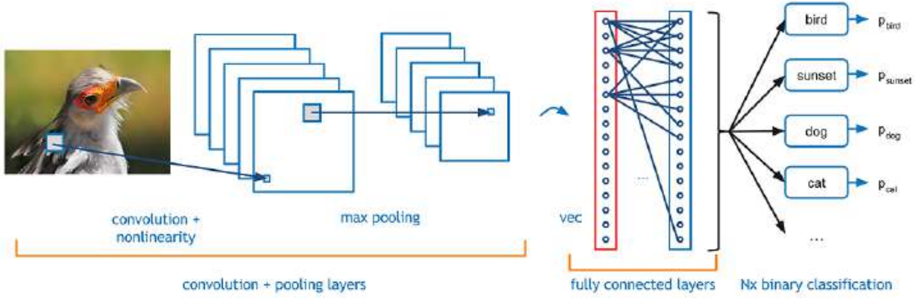
- **Bag-of-Visual-Words**: useful for local features — Sivic and Zisserman 2003; Torres 2018
  - Identifies the keypoints using algorithms like SIFT
  - Builds patches around the neighbor of keypoints
  - Cluster the patches and count the occurrence of clusters

Identify themes in memes

- **Bag-of-Visual-Words**: useful for local features — Sivic and Zisserman 2003; Torres 2018
  - Identifies the keypoints using algorithms like SIFT
  - Builds patches around the neighbor of keypoints
  - Cluster the patches and count the occurrence of clusters

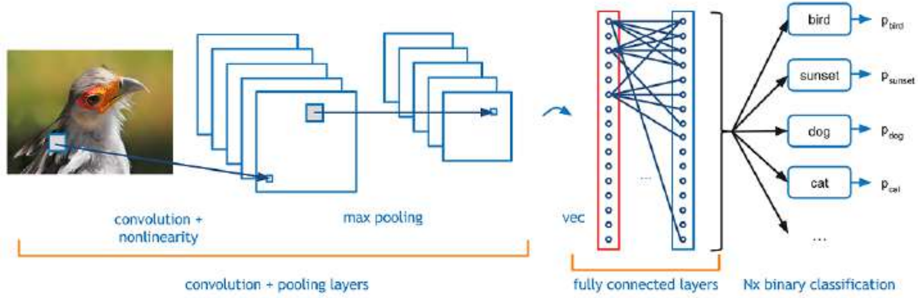
- Modern approaches use **transfer learning**
  - Pretrain on some large-scale data and apply it on another data
  - But many vision tasks are NOT useful for social scientists…

Many vision tasks are NOT useful



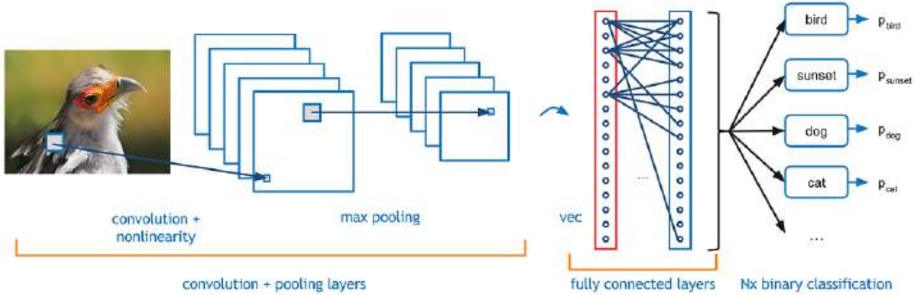
- Most pretrained classification labels are not of interest (dog breeds!)

Many vision tasks are NOT useful



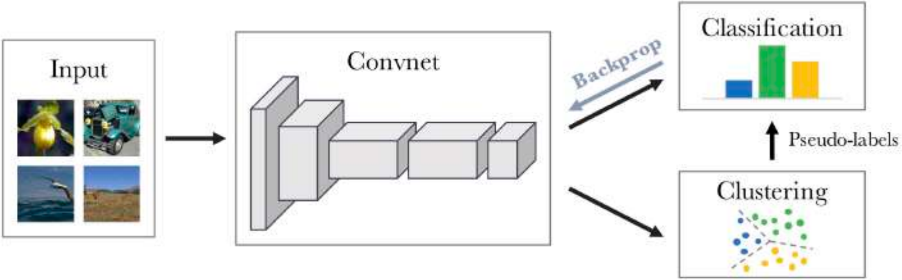
- Most pretrained classification labels are not of interest (dog breeds!)
- Most models are optimized for linear classification

Many vision tasks are NOT useful



- Most pretrained classification labels are not of interest (dog breeds!)
- Most models are optimized for linear classification
- We often want to know about the relationship between images

DeepCluster (Facebook AI Reserach) Caron et al. 2019

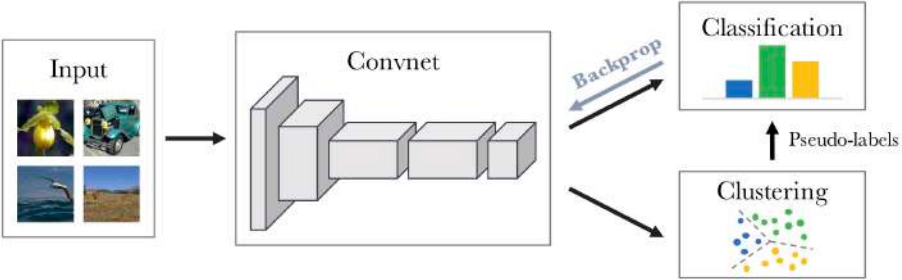


Caron et al. 2019

- Cluster assignments as pseudo-labels to train the CNN iteratively

DeepCluster (Facebook AI Reserach)

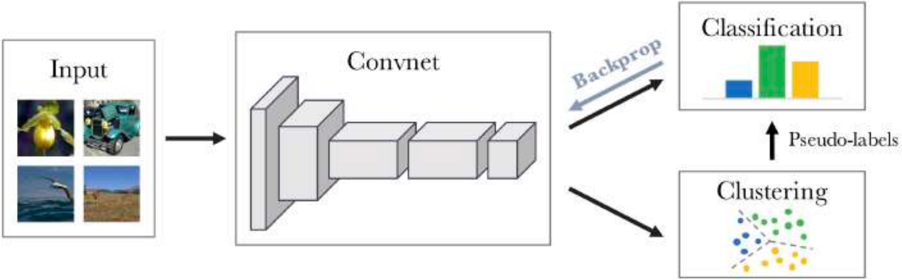
Caron et al. 2019



Caron et al. 2019

- Cluster assignments as pseudo-labels to train the CNN iteratively
- Map each meme into 4096-dimensional visual embedding space

DeepCluster (Facebook AI Reserach) Caron et al. 2019

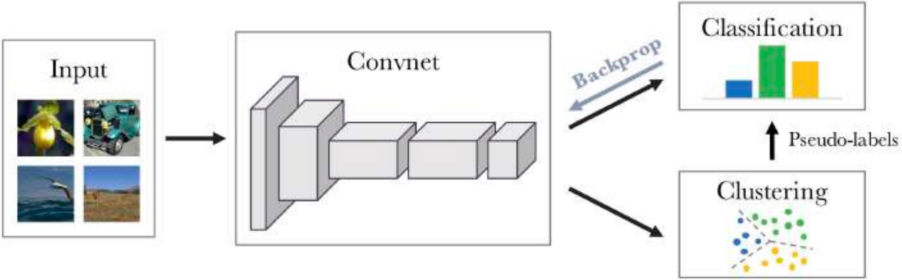


Caron et al. 2019

- Cluster assignments as pseudo-labels to train the CNN iteratively
- Map each meme into 4096-dimensional visual embedding space
- Open source, provide weights pre-trained on ImageNet as initialization

DeepCluster (Facebook AI Reserach)

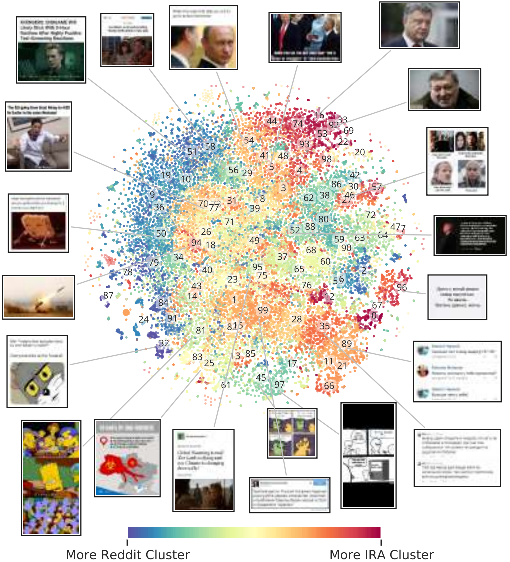
Caron et al. 2019



Caron et al. 2019

- Cluster assignments as pseudo-labels to train the CNN iteratively
- Map each meme into 4096-dimensional visual embedding space
- Open source, provide weights pre-trained on ImageNet as initialization
- Cluster on the embedding space (set $K = 100$) and label the clusters

t-SNE projection of the visual embedding



16: Male politician face

16

93: Gendered female image

93

94: Missles/bomb/war zone

94




















14: Russian symbols

14


















31: Military/police

31













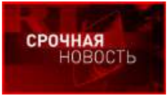


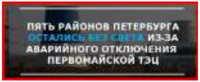


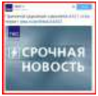



27: Slogans

27

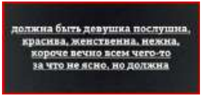






52: Quotes/face

52



48: Politician speaking

48




















45: 4 panel comics

45




















91: SpongeBob/Rick and Morty

91

51: Movie scene memes

51














87: Black Hackerman meme

87













32: Unsettled Tom meme

32









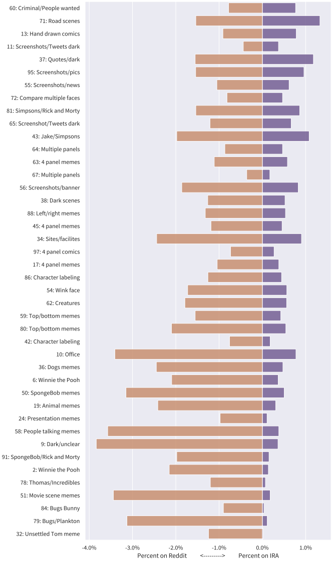


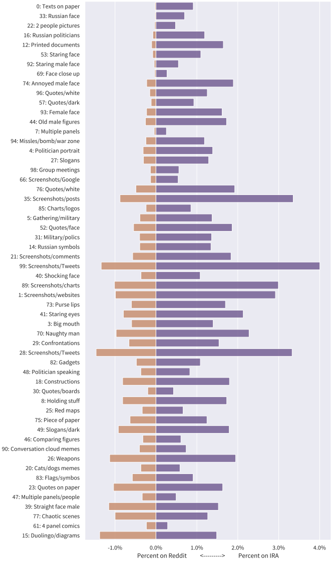






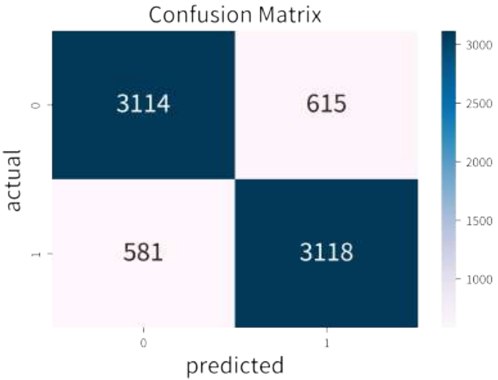
Theme distribution





Can machine discern coordinated memes?

- Logistic regression using only the 4096-d embeddings, 70/30 train/test split
- Training accuracy 0.90, AUC 0.91
- Testing accuracy 0.84, precision 0.84, recall 0.84, F_1-score 0.84



Limitations

- Reddit might not be the best group for comparison
  - Reddit is fairly comprehensive for memes in the US (entertainment & politics)

Limitations

- Reddit might not be the best group for comparison
  - Reddit is fairly comprehensive for memes in the US (entertainment & politics)
  - Russian IRA are outward-facing, targeted at US election

Limitations

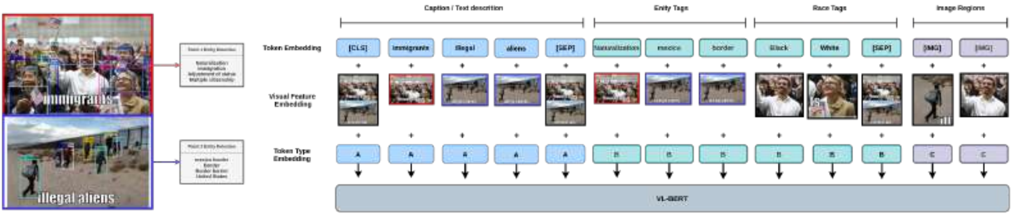
- Reddit might not be the best group for comparison
  - Reddit is fairly comprehensive for memes in the US (entertainment & politics)
  - Russian IRA are outward-facing, targeted at US election
  - Other data sources need additional human annotations for memes

Limitations

- Reddit might not be the best group for comparison
  - Reddit is fairly comprehensive for memes in the US (entertainment & politics)
  - Russian IRA are outward-facing, targeted at US election
  - Other data sources need additional human annotations for memes
- Not account for text, languages, panel structures, and tweet-level metadata

Next steps

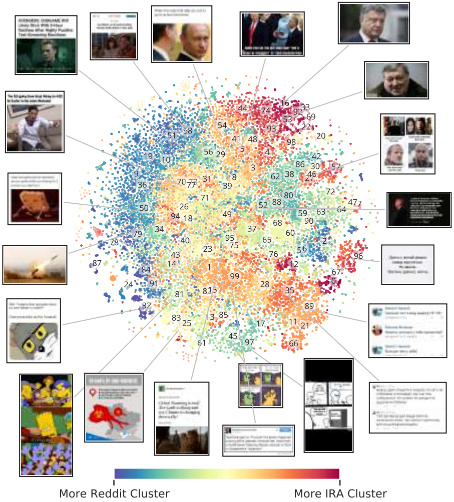
- Use multimodal transformer models (ex. VisualBERT) to extract feature embeddings that incorprate textual and text-scene information



Zhu 2020

- Preprocess memes to strip off "structures" unrelated to topic of memes
- More flexible clustering models to incorporate tweet-level covariates

Thank you!



Feedbacks, thoughts, comments welcome!!

`@kengchichang`

`https://arxiv.org/abs/2206.02306`

- Cluster on meaningful low-dim visual embeddings
- Authentic/coordinated memes share many themes
- Coordinated IRA memes military, gender, quotes
- Authentic Reddit memes comics, movie characters
- Logistic regression on visual embedding discern IRA memes with accuracy of 0.84

References I

DiResta, Renée, Shelby Grossman, and Alexandra Siegel. 2021. "In-House Vs. Outsourced Trolls: How Digital Mercenaries Shape State Influence Strategies." *Political Communication* 0 (0): 1–32. doi:10.1080/10584609.2021.1994065.

Du, Yuhao, Muhammad Aamir Masood, and Kenneth Joseph. 2020. "Understanding Visual Memes: An Empirical Analysis of Text Superimposed on Memes Shared on Twitter." *Proceedings of the International AAAI Conference on Web and Social Media* 14:153–164.

Sivic and Zisserman. 2003. "Video Google: A Text Retrieval Approach to Object Matching in Videos." In *Proceedings Ninth IEEE International Conference on Computer Vision,* 1470–1477 vol.2. October. doi:10.1109/ICCV.2003.1238663.

Torres, Michelle. 2018. "Give Me the Full Picture: Using Computer Vision to Understand Visual Frames and Political Communication." *Working Paper:* 30.

Zhu, Ron. 2020. "Enhance Multimodal Transformer With External Label And In-Domain Pretrain: Hateful Meme Challenge Winning Solution."