Detecting Frames in News Headlines and Lead Images in U.S. Gun Violence Coverage

Isidora Chara Tourni   Taufiq Daryanto[*]   Fabian Zhafransyah[*]
Lei Guo   Edward Halim   Mona Jalal   Boqi Chen   Sha Lai
Hengchang Hu   Prakash Ishwar   Margrit Betke   Derry Tanti Wijaya[†]
Boston University
wijaya@bu.edu

Abstract

News media structure their reporting of events or issues using certain perspectives. When describing an incident involving gun violence, for example, some journalists may focus on mental health or gun regulation, while others may emphasize the discussion of gun rights. Such perspectives are called "frames" in communication research. We study, for the first time, the value of combining lead images and their contextual information with text to identify the frame of a given news article. We observe that using multiple modes of information(article- and image-derived features) improves prediction of news frames over any single mode of information when the images are relevant to the frames of the headlines. We also observe that frame image relevance is related to the ease of conveying frames via images, which we call *frame concreteness*. Additionally, we release the first *multimodal* news framing dataset related to gun violence in the U.S., curated and annotated by communication researchers. The dataset will allow researchers to further examine the use of multiple information modalities for studying media framing.

1 Introduction

Media framing refers to the journalistic practice of selecting aspects of a perceived reality and making them more salient in news coverage (Entman, 1993; Reese et al., 2001). In political communication, for example, news framing is helpful as it reveals how the news article is structured to promote a certain side of the political spectrum, thus influencing the public opinion in a particular way.

Journalists have been using both text and images to frame news stories. Images in news stories can help convey controversial or provocative meanings that would otherwise be unpalatable to the news audience, if it were spelled out in text (Messaris and Abraham, 2001). While text is more influential in changing opinions, visuals elicit more attention and emotional reactions, resulting in behavioral change (Coleman and Wu, 2015; Dan, 2017; Powell et al., 2015). Lead images may carry additional background knowledge about the event (e.g., showing well-known people and locations). An image showing a school, for example, might suggest an article about gun violence focuses on the "School/Public Safety" frame. Text and images thus work in tandem to create a holistic perception of news and must be considered together when analyzing news frames (Wessler et al., 2016).

Given the importance of visuals in media framing and the rising gun violence in the U.S. (Guo et al., 2021), we extend the Gun Violence Frame Corpus (GVFC) (Liu et al., 2019), which contains news headlines related to U.S. gun violence and their domain-expert frame annotations, by retrieving the lead images of the articles and obtaining their relevance annotations from communication domain experts (i.e., an image is annotated as *relevant* if it expresses the annotated headline frames). Notably, about half of the time, the images presented do not express the annotated headline frames (Table 1). This might be explained from the journalism research perspective, as reporters and photographers do not necessarily work together seamlessly in the newsroom as they occupy distinct occupational roles and often compete for control over how a story may be packaged and presented as a final product (Lowrey, 2002). Hence, in addition to communication scholars benefiting from tools that can analyze, on large scale, images and headlines in tandem for frames, newsroom editors would benefit from tools that can identify images that help depict the main thrust of the story's focus (Caple, 2010). Such tools do not yet exist, and our work addresses this need.

In this work, we comprehensively explore the use of multimodal information from news articles

[*] Institut Teknologi Bandung
[†] Corresponding Author

Findings of the Association for Computational Linguistics: EMNLP 2021, pages 4037–4050
November 7–11, 2021. ©2021 Association for Computational Linguistics

i.e., headlines or summaries, and their lead images i.e., the images, categories of objects in the images, or the background/real-world knowledge contained in them, for predicting frames. Our results show that for news articles with relevant images, using only image-derived features or only article-derived features (i.e., headlines and/or extractive summaries) yields less accurate frame predictions than our multiple modalities approach. When considering articles with irrelevant images, the accuracy of the multimodal approach is comparable to that based on only article-derived features. We also observe that adding image contextual information using the Google Web Entity Tagger API[1] or an entity-aware news image caption generation model (Tran et al., 2020) or by asking humans to annotate the central subject of the image in terms of pre-defined categories that cue gun violence frames, such as politician (*politics* frame), legislative buildings (*gun control* frame), school/campus (*school/public safety* frame), etc., improves the performance of frame prediction, compared to using raw images alone. The API tags capture background information associated with an image from the Web, such as the list of named entities in the image, by finding similar images in the Web and parsing the associated web page contents. News image captions capture real-world information in the image, e.g., the names of people and objects, by learning to associate words in the article text with faces and objects. Human annotations of the central subject of the image in terms of categories such school or legislative buildings, capture the annotators' background knowledge of the identities of entities in the image.

Overall, our contributions are the following: (1) A well-curated *multimodal text-image framing dataset with expert annotations*[2]: With the goal of predicting frames based on multiple information modalities, we augment GVFC by using the article URLs to retrieve lead images of articles and annotate the images for their visual framing labels, which include (a) the Subject/Race/Ethnicity (SRE) annotations of the central *subject* of the image (i.e., suspect vs. victim vs. politician, etc.) and whether the image contains anything related to *race/ethnicity*, and (b) the image relevance annotations i.e., whether the images are relevant to the annotated frames of their headlines. In addition to frame annotations of news article headlines and lead images, for each image we provide its URL, Web Entity tag (API tag), caption generated using a state-of-the-art news image captioning system (Tran et al., 2020), and the article summary generated by a state-of-the-art extractive summarization system (Liu and Lapata, 2019). (2) *Comprehensive study and development of methods to combine multimodal information to predict article frames and image relevance*: We explore various approaches to predict image relevance and article frames using information from both article and lead images, using BERT (Devlin et al., 2018) to represent text and a deep convolutional neural network ResNet-50 (He et al., 2016) to represent raw images. (3) *Frame concreteness*: We propose a novel method for measuring the ease of conveying frames through images via the concreteness of words in its headlines, i.e., the ease of identifying tangible concepts and mental images that arise in correspondence to words (Paivio et al., 1968), and relate frame image relevance to frame concreteness.

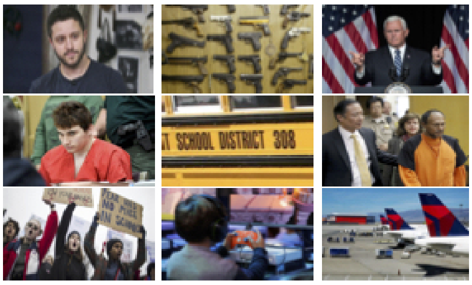


Figure 1: Sample images for each frame (from left to right and top to down): 2nd Amendment, Gun Control, Politics, Mental Health, School/Public Space Safety, Race/Ethnicity, Public Opinion, Society/Culture, Economic Consequences.

2 Related Work

Media framing is related to many factors such as word choice, the presentation of background information, and the emphasis on certain actors. The subtle nature of news framing can influence the opinion of readers in a certain way without them even noticing it, hence its analysis has many implications, e.g., it has been used to understand why important public affairs issues such as gun violence are polarizing (Liu et al., 2019), how media manipulation strategies are conducted (Field et al., 2018), or how framing is used to perpetuate racial biases (Drakulich, 2015). In Journalism, most "framing

[1] https://cloud.google.com/vision/docs/detecting-web

[2] We include the dataset in the supplementary material and the multimodal annotation is available for download through the GVFC dataset website: https://derrywijaya.github.io/GVFC.html.

analyses” have been done manually hence they are not scalable (Hamborg et al., 2019).

In natural language processing (NLP), automated frame detection focuses on predicting frames from news texts. Some of these methods rely on topic models (Nguyen et al., 2013). Naderi and Hirst (2017) devised various deep neural networks (LSTMs, BiLSTMs, or GRUs) for predicting frames at the sentence level in the Media Frame Corpus (MFC) (Card et al., 2015). The most recent method for predicting news frames in headlines is the work of Liu et al. (2019); Akyürek et al. (2020). They fine-tuned BERT (Devlin et al., 2018) using focal loss (Lin et al., 2017) to predict frames of headlines. They also released a framing benchmark dataset, the Gun Violence Frame Corpus (GVFC), of news links and headlines with frame annotations, related to gun violence in the U.S. They show that fine-tuning BERT for frame prediction using news headlines results to significantly higher accuracy than previous methods.

In this paper, we use GVFC articles’ lead images and perform experiments with a rich set of unimodal and multimodal information to predict frames. Although images and text have been used together as multimodal inputs to improve performance in other NLP tasks such as machine translation (Specia et al., 2016; Hewitt et al., 2018; Caglayan et al., 2019; Yao and Wan, 2020; Khani et al., 2021) or in vision-language tasks such as multilingual image retrieval or captioning (Kim et al., 2020; Burns et al., 2020; Rasooli et al., 2021), the use of images and text in tandem to *automatically* analyze framing i.e., computational multimodal framing has never been explored–all previous works in multimodal framing have been conducted manually (Messaris and Abraham, 2001; Coleman and Wu, 2015; Dan, 2017; Powell et al., 2015; Wessler et al., 2016). Given the growing importance of visual journalism and the contribution of images to media framing which suggest that images may be able to help interpret text, our work is the first to conduct computational multimodal framing analysis, which will enable scalable multimodal framing analysis.

3 Dataset

Our multimodal version of GVFC contains news headlines and their corresponding lead images, news URLs, and the entire news text. The lead images are either the pictures shown at the top of news articles or the editor-picked thumbnails that are shown in news services such as Google News (Fig. 1). Using Brandwatch Consumer Research[3] we analyze 3,000 news headlines, of which 1,300 are are annotated with 9 major frames, e.g., politics, gun control/regulation, mental health, race/ethnicity, etc. (Table 1), that exhaustively cover the discussion of the U.S. “gun violence” issue in communication research. In this paper, we further annotate each lead image with a binary relevance label that indicates whether the image is consistent with the frame associated to the headline. We also annotate the central subject (S) of the image using one of 16 categories that often imply certain frames, e.g., suspect (often implies mental health), politician (often implies politics), company logos (often implies economic consequence), etc. (see Appendix for the full listing), and an additional race/ethnicity (RE) label which can take one of the following 3 values: 1) racial/ethnic minority groups, 2) hate groups, or 3) none of the above.

The details of the visual annotation codebook used to train the coders are given in the Appendix and Supplementary Material. The coders’ agreement on how to apply the codes is measured with inter-coder reliability (ICR). High ICR values (above 90% agreement or 0.70 Krippendorff α (Krippendorff, 2018)) imply that two or more coders consistently categorized the content similarly, signaling a high validity of the results. In our dataset, ICR was met on all variables: Subject (90% agreement, 0.88 α), Race/Ethnicity (91% agreement, 0.64 α) and Relevance (88% agreement, 0.75 α). The number and ratio of relevant images per frame are shown in Table 1. Easy and hard to classify examples and their features are provided in Table 2 and described in detail in §4. We affirm we have the right to use the collected dataset in the way we are using it[4], i.e. the article headline and URL, as well as their annotations and image-derived visual and textual features; and we bear responsibility in case of a violation of rights or terms of service. Researchers can use the article URLs to retrieve images and full texts of the articles.

4 Experiments

We experiment with unimodal (§4.1) and multimodal (§4.2) information obtained from each article and its lead image. We train and report 4-fold

[3] https://www.brandwatch.com

[4] We have confirmed and received approval from Brandwatch Consumer Research whom we obtain the data from.

News Frame	# Articles	# Relevant Images (%)
Politics	373	241 (65%)
Public Opinion	237	147 (62%)
Gun Control/Regulation	215	93 (43%)
School/Public Space Safety	137	68 (50%)
Economic Consequences	80	46 (58%)
Race/Ethnicity	114	34 (30%)
Mental Health	65	28 (43%)
2nd Amendment/Gun Rights	38	13 (34%)
Society/Culture	41	4 (10%)
Overall	1,300	674 (52%)

Table 1: Gun violence frames in our dataset, the number of articles with headlines and lead images, and the number of lead images annotated as relevant to the frame with the percentage indicated in brackets. The news frames are ordered by the number of relevant images from highest to lowest.

cross validation frame prediction accuracies (Table 3) for all articles in our dataset (*All Articles*), and for the subset of articles with relevant images (*Articles with Relevant Images*). We also perform image-to-frame relevance classification (§4.3).

4.1 Unimodal

In this set of experiments we predict news frames from only one mode of information, either image-derived or article-derived. For raw images, we use ResNet-50 (He et al., 2016) for both image representation and frame classification (RESNET-50). We use BERT (Vaswani et al., 2017) to represent text from the article headlines (BERT HEADLINE), the image Web Entity API tags (BERT API), or the automatically generated captions from images (BERT CAPTION). We also experiment with other article-derived information in the form of text: the headline concatenated with the automatically generated extractive summary (BERT HEADLINE + SUMMARY), or with the first three sentences of the article (BERT HEADLINE + 3SENTENCES), a typical baseline for extractive summarization. For all text-form information derived from the article (irrespective of whether the text was extracted from the image, headline or body of the article), we follow the state-of-the-art methodology for frame detection based on news headlines (Liu et al., 2019), which constitutes our baseline (BERT HEADLINE). Specifically, we use the text as input into BERT's pre-trained base uncased model and fine-tune the model to predict the frames of the articles over 25 different random seeds to avoid the fine-tuning instability due to the small dataset size (Devlin et al., 2019; Dodge et al., 2020; Mosbach et al., 2021). In all our models, the number of epochs is 10, the batch size is 4, and the learning rate is 2e-5.

Using Image-derived *Visual* Features (RESNET-50). We predict news frames based only on the raw lead images of the news articles. We use the ResNet-50 model (He et al., 2016), pre-trained on ImageNet (Deng et al., 2009), and replace the output layer of the original ResNet-50 network with a flattened layer of 512 nodes followed by a dropout layer with a 0.5 dropout rate to the frame classification (9-nodes) layer. All images are scaled to 224 × 224 pixels and are normalized based on the mean and standard deviation of ImageNet.

Using Image-derived *Visual* Annotations (SRE). We create a 19-length feature vector for each image obtained from the Subject, Race/Ethnicity (SRE) annotations of the image that indicate the human coders' background knowledge of the image's central Subject, and its connection to Race/Ethnicity. We train a logistic regression frame classifier with this feature vector as input.

Extracting Image-derived *Textual* Features: Google Web Entity API Tags (BERT API). Here, frames are predicted based only on the Google Web Entity tags of lead images. Web Entity detection is a Google cloud service that reads an image as input and returns a ranked list of web entities as tags. For each image, we form a "sentence" by concatenating the top-10 Web Entity tags returned for the image.

Extracting Image-derived *Textual* Features: Image Captions (BERT CAPTION). We follow Tran et al. (2020) to generate captions for the lead images of news articles. The model introduced in the paper consists of different encoders generating representations for each modality (article text, images, faces, and objects), and a Transformer as the decoder attending over text, images, image faces and objects. It uses Byte-Pair-Encoding, breaking sequences into subwords and then merging common sequences into larger words. This leads to better generalization and prediction of out-of-vocabulary words and names, and ultimately to linguistically rich captions for images accompanying each news article. We follow all default settings and parameters suggested in the paper, and use RoBERTa (Liu et al., 2020) as the article text encoder, a ResNet-152 (Dauphin et al., 2017) pretrained on ImageNet as the image encoder, MTCNN (Zhang et al., 2016) as the face detector, and YOLOv3 (Redmon and Farhadi, 2018) as the object detector, with the latter two operating as the specialized image face and object attention modules, respectively. All representations

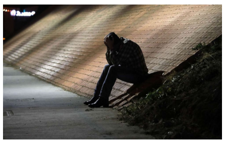
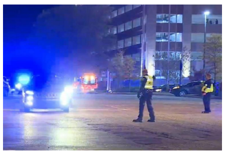
Image	Description
	An article with **relevant** image in a **frame with many examples** (potentially **easy** to classify) ***Frame***: School/Public Space Safety ***Headline***: To Defend Against School Shootings, Massachusetts District Is Passing Out Emergency Buckets With Hammer, Rope ***API tag***: Classroom, School, Harry S Truman, High School, Active shooter, Lockdown, Campus, Student ***Caption***: A school shooting victim in Brockton, Mass., last month. ***3sentences***: U.S.: More than 1,000 blue buckets were assembled to be passed out to classrooms in the Brockton, Massachusetts, school district, filled with curated items aimed at saving lives in the event of an emergency, including a school shooting. The Brockton school district partnered with the mayor's office, the Brockton Police Department and a local Lowe's to put together buckets filled with four items to help defend classrooms. Each blue five-gallon bucket contains a wooden wedge, a one-pound hammer, a 50-foot length of rope and a roll of duct tape, according to The Enterprise. ***Summary***: U.S.: More than 1,000 blue buckets were assembled to be passed out to classrooms in the Brockton, Massachusetts, school district. The buckets can be used for emergency bathroom situations. Mayor Bill Carpenter applauded the decision to put the buckets in the classrooms.
	An article with **irrelevant** image in a **frame with many examples** (potentially **harder** to classify). ***Frame:*** School/Public Space Safety ***Headline***: Mass shootings 'increasing' and pose 'most serious threat' in US, expert says, ***API tag:*** Thousand Oaks shooting, Borderline Bar Grill, Mass shooting, California Bar, Police officer ***Caption***: A gunman at the scene of the shooting at a country bar in Sacramento. ***3sentences***: Mass shootings 'increasing' and pose 'most serious threat' in US, expert says At least 59 people have been killed as a result of mass shootings this year. Deadliest mass shootings of 2018 in the U.S. Mike Nelson/EPA via Shutterstock. ***Summary***: At least 59 people have been killed as a result of mass shootings this year. There have been at least six mass shootings in the U.S. this year, according to the U.S., at least 10 mass shootings have been linked to mass shootings at a California bar.
	An article with **relevant** image in a **frame with few examples** (potentially **harder** to classify). ***Frame:*** Mental Health ***Headline***: Accused Fredericton shooter will undergo psych assessment ***API tag:*** Car, Job, Vehicle, Staff, Capilar y Corporal ***Caption***: Matthew Vincent Raymond Murder Officer Suspect Broward Police officer Arrest warrant Criminal charge Suspect Murder. ***3sentences***: The Fredericton man accused of killing four people in August will be sent for a psychiatric assessment. Judge Julian Dickson ordered the assessment Wednesday to determine if Matthew Vincent Raymond, 48, is fit to stand trial on four counts of first-degree murder. Raymond is charged in the Aug. 10 shooting deaths of Fredericton police constables Robb Costello, 45, and Sara Burns, 43, and civilians Donnie Robichaud, 42, and Bobbi Lee Wright, 32. ***Summary***: A judge orders the assessment to determine if Matthew Vincent Raymond, 48, is fit to stand trial on four counts of first-degree murder. Raymond is charged in the Aug. 10 shooting deaths of Fredericton police constables Robb Costello, 45, and Sara Burns, 43, and civilians Donnie Robichaud, 42, and Bobbi Lee Wright, 32. Arguments about who should conduct the assessment. The assessment is expected to be completed before Dec. 4, when Raymond is due back in court.
	An article with **irrelevant** image in a **frame with few examples** (potentially **hardest** to classify). ***Frame***: Race/Ethnicity ***Headline***: Alabama mall shooting: Family of black man killed by police officer on Thanksgiving hires civil rights lawyer ***API tag***: Shooting of Emantic Fitzgerald Bradford Jr., Alabama Shooting of Michael Brown, News, Shopping Centre, Breaking news, Street light, Television show, Street Broken Horses ***Caption***: A police officer at the scene of the shooting at the Riverchase Galleria in Birmingham, Ala., on Friday. ***3sentences***: Emantic Fitzgerald Bradford Jr 's family has employed Benjamin Crump who previously represented the families of shooting victims Trayvon Martin and Michael Brown to also represent them: WVTM The family of a 21-year-old black man who was shot by a police officer at shopping centre in Alabama on Thanksgiving has hired a national civil rights lawyer to represent them. ***Summary***: Emantic Fitzgerald Bradford Jr was fatally shot by a police officer at the shopping centre in Alabama on thanksgiving has hired a national civil rights lawyer. Police initially said a hoover police officer who was responding to reports of gunfire at a shopping mall confronted an armed man running away from the scene and fatally shot him. The shots responsible for injuring an 18-year-old man and a 12-year-old girl, but investigators have since said they believe he did not firesle the gunman is still at large. Mr Crump said Bradford was a veteran who was licensed to carry a concealed firearm. Bradford's family said they are working with our legal team to determine.

Table 2: Examples of articles, their images and image- and article-derived textual features that are potentially easy or hard to classify using the multimodal approach.

obtained from the individual encoders are fed into a four block Transformer decoder, which employs a multi-head multi-modal attention mechanism and generates byte-pair encoded tokens, that are finally concatenated to form the caption.

Extracting Article-derived *Textual* Features: Summary (BERT SUMMARY) We automatically extract the summary of the article, following Liu and Lapata (2019), that uses BERT to represent sentences, and inter-sentence Transformer layers on top of the BERT encoder to classify whether a sentence should be in the extractive summary.

4.2 Multimodal

In this set of experiments, we predict news frames using multiple modes of information derived from both the image and the article.

Using Image-derived *Visual* and *Textual* Features and Article-derived *Textual* Features (RESNET-50 + BERT HEADLINE, RESNET-50 + BERT HEADLINE + API, RESNET-50 + BERT HEADLINE + CAPTION). Here, frames are predicted with multiple input modalities (visual *and* textual features). We follow a simple concat fusion approach, which allows us to build a modular pipeline, obtain the text and visual representations from their respective modules, and use them to predict the frame class. Specifically, we use RESNET-50 representations of the raw images and, as suggested by Devlin et al. (2019) for best performance, representations of the text obtained from the concatenation of the contextual embeddings of the last four layers of BERT, which has been fine-tuned for frame classification, as inputs to our multimodal 3-layer feed forward fully connected classifier neural network that we *then* train jointly with RESNET-50. We use the AdamW optimizer, cross entropy loss, and "no improvement in validation accuracy over 5 epochs" as the stopping criterion.

Using Image-derived *Visual* Annotations and Article-derived *Textual* Features (BERT HEADLINE + SRE). We concatenate BERT representation of the headline with the SRE feature vector and train jointly with fine-tuning BERT using a 1-layer feed forward fully connected classifier neural network added on top of BERT.

Using Image-derived *Textual* Features and Article-derived *Textual* Features (BERT HEADLINE + API, BERT HEADLINE + CAPTION). Here we concatenate article headlines and image-derived textual features (API tags or captions) as input to fine-tune BERT for frame classification.

4.3 Relevance

We use the article headline (BERT HEADLINE) and the image-derived features (BERT API, BERT CAPTION), the SRE annotations (SRE), and their combinations (BERT HEADLINE + API, BERT HEADLINE + CAPTION, RESNET-50 + BERT HEADLINE, RESNET-50 + BERT HEADLINE + API, RESNET-50 + BERT HEADLINE + CAPTION) to predict the relevance of the images to the frames of their headlines. We perform a 4-fold cross-validation binary classification for relevance prediction, with the same BERT and ResNet-50 architectures and hyperparameters as before. Accuracies with uni- and multimodal information sources are reported in Table 4. To mimic the relevance annotation process of our coders, who are given the labeled frame of the headline to decide whether the lead image is relevant to it, we provide our relevance prediction model with headline frame. We concatenate the frame label to the input of the top performing models of Table 4 and their combinations, and report accuracies in Table 5.

5 Discussion of Results

Despite the challenges of a highly nuanced multi-class frame identification and an intrinsically imbalanced dataset, we achieve a high prediction accuracy of up to 87% for *Articles with Relevant Images*, and 82.4% for *All Articles* (Table 3).

It is instructive to examine the utility of article- and image-derived features, and a fusion of all in cases where the lead image is relevant to the article headline. We observe that contextual information derived from the image, in the form of API tags or a caption, along with the article headline, can drive the article perspective more clearly. The headline + API tags combination provides the best performance (87%), compared to image-only (43%), the SRE image annotation (81.2%), or the the article headline (83%) which is a strong unimodal baseline. Even when considering examples with irrelevant images, adding API to headlines does not hurt performance and is comparable to using headlines alone, which is unlike SRE whose performance drops significantly for articles with irrelevant images as these annotations are designed with relevant, i.e., frame-implying images in mind. Furthermore, SRE requires training another model to produce these annotations automatically, which is not trivial as they capture real-world knowledge of the subjects in the image, e.g., whether the person in the image is a politician (cuing *politics* frame) or a gun activist/NRA representative (cuing *2nd Amendment*). These findings, namely that API tags, captions, or SRE yield higher accuracy than the raw image alone, indicate the importance of the contextual or background knowledge of the lead image in driving the news frame This strongly suggests that the highly nuanced task of frame prediction is challenging using images in isolation. Our findings also confirm previous observations that training with multiple input modalities, e.g., both visual and textual inputs is hard as each modality may generalize differently and hence underperform when trained jointly (Wang et al., 2020).

In terms of relevance prediction, the performance is highest for models supplied with frame labels, mimicking the relevance annotation process. Given a headline and a frame, our method can correctly predict the relevance of an image to the frame with 74% accuracy using the image's API tag. Without frame labels, however, the accuracy of the top-performing method drops to 68.1% and is based on SRE only. While several SRE categories are strong indicators for certain frames, e.g., the presence of demonstrators suggests the *public opinion* frame, the use of SRE at inference time necessitates the training of another model for predicting these annotations, which is not trivial.

Examining the content of the API tags and captions, we observe that API tags have significantly more proper nouns (71% to 29% of all words in

Method	ResNet-50	SRE	BERT headline	BERT API	BERT Caption	BERT headline + API	BERT headline + Caption	BERT headline + Summary	BERT headline + 3sentences	ResNet-50 + BERT headline	ResNet-50 + BERT headline + API	ResNet-50 + BERT headline + Caption	BERT headline + SRE
All Articles													
	9.3	49.2	81.9	47	48.1	82	82	82.4	81.8	13.8	13.7	12.2	81.5
Articles with Relevant Images													
	42.8	81.2	83	72.1	72.5	87	84.6	83.1	83.1	49.7	65.3	63.8	83.2

Table 3: Overall micro accuracy of our methods for **frame classification** for *All articles* and *Articles with Relevant Images*. BERT HEADLINE is the baseline we compare to (Liu et al., 2019)

Method	BERT headline	SRE	BERT API	BERT Caption	BERT headline + API	BERT Headline + Caption	ResNet-50 + BERT headline	ResNet-50 + BERT headline + Caption	ResNet-50 + BERT headline + API
	62	68.1	59.3	62.5	65.8	65.7	55.5	60	58.3

Table 4: Overall micro accuracy of our methods for image **relevance classification** (without frame label) for *All articles*.

Method:	BERT headline + API tag + Frame	SRE + Frame	SRE + BERT headline + API tag + Frame
	74.2	71.0	72.0

Table 5: Overall micro accuracy of our methods for image **relevance classification with frame** for *All articles*.

tags/captions) and named entities (53% to 31%) than captions. On the other hand, captions have more common nouns and verbs. Since the performance of frame prediction for articles with relevant images is higher when using headline and API compared to headline and caption, this suggests that frames can be directly cued by lexical items such as proper nouns or named entities, e.g., politicians' names cue politics frame (Mendelsohn et al., 2021). Since models may lack real-world knowledge required to identify these, especially when there is insufficient text evidence, e.g., when using headlines, API tags that provide this background knowledge from images facilitate frame prediction.

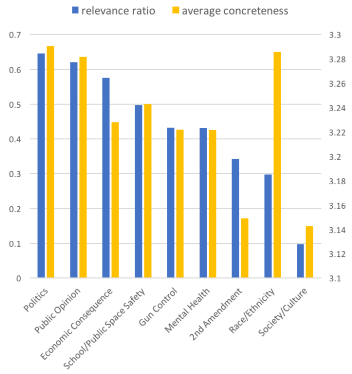


Figure 2: Frame relevance ratio and average concreteness.

One possible cause for why predicting or even deciding on image relevance is challenging (e.g., in our dataset roughly half of the lead images are irrelevant) may be related to the nature of frames and images of U.S. gun violence coverage. Although frames or perspectives are abstract concepts, some frames such as "Society/Culture", which focuses on *society-wide factors* related to gun violence[5], are by nature more abstract and thus harder to convey through images than more concrete frames, such as "Politics", which focuses on the political issues around guns and can be expressed more easily via images of politicians. As the ability of images to usefully represent a word is strongly dependent on how concrete or abstract the word is (Gilhooly and Logie, 1980; Friendly et al., 1982), a measure of frame concreteness or the ease of identifying tangible concepts and mental images that arise in correspondence to the frame, should *relate* to the ease of expressing frames via images or the ratio of relevant images for frames (Table 1).

To measure concreteness of frames and test this hypothesis, we trained a regression network that takes as input a word's vector representation as extracted and concatenated from the last 4 layers of the pre-trained BERT model and outputs its concreteness measure between 1 (most abstract) to 5 (most concrete). We use a dataset created by Brysbaert et al. (2014), which contains human evaluations of concreteness for 39,954 English words, to train and evaluate the network, achieving a high 0.95 Pearson's Correlation between our concreteness predictions and the ground-truth measures. The concreteness of a frame is then measured as the average concreteness of non named-entity words in its headlines (we treat named-entities as having a concreteness measure of 5). As seen in Figure 2, no frame has a high (> 4) average concreteness.

We observe, however, that some of the more concrete frames have higher ratios of their images

[5] Definitions of gun violence frames taken from the publicly available GVFC codebook and dataset https://derrywijaya.github.io/GVFC.html

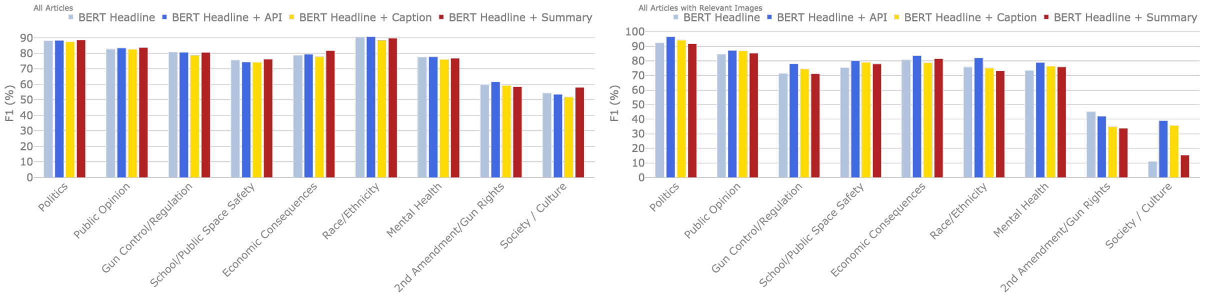


Figure 3: Per Frame F1 score of the best performing frame prediction methods for (a) *All Articles* (left), (b) *Articles with Relevant Images* (right).

Error Type	**Description**	**Examples**
Plausible Interpetation	Predicted frames can be appropriate labels.	***Example Frame:*** *Mental Health* ***Headline:*** Florida shooter a troubled loner with **white** supremacist ties • Model erroneously predicted "Race/Ethnicity": **"white"** ***+ API tag:*** Nikolas Cruz Stoneman Douglas High School shooting Marjory Stoneman Douglas Murder Mass shooting AR-15 style rifle Suspect Student • Model correctly predicted "Mental Health": **"Suspect Student"**
Inferring frames not explicitly cued in text	Predicted frames capture an author's intention without sufficient text evidence.	***Example Frame:*** *2nd Amendment/Gun Rights* ***Headline:*** The NRA Versus the Constitution • Model erroneously predicted "Gun Control": **"NRA"** ***+ API tag:*** Pennsylvania Obergefell v. Hodges Supreme Court of the United States Concealed carry Rights Reciprocity Act of 2017 Constitution of the United States • Model correctly predicted "2nd Amendment": **"Rights"**,**"Constitution"**
Missing necessary contextual knowledge	Frames can be directly cued by lexical items (e.g. politicians' names cue Politics frame), yet the model lacks real-world knowledge required to identify those.	***Example Frame:*** *Politics* ***Headline:*** In closed-door meeting, Roskam brings pro-gun rights teens to talk gun violence prevention with students • Model erroneously predicted "2nd Amendment/Gun Rights": "**pro-gun**", "**rights**" ***+ API tag:*** Peter Roskam Republican Party Democratic Party United States Congress Illinois Member of Congress 2018 United States elections United States House of Representatives Presidency of Donald Trump House Committee on Ways and Means • Model correctly predicted "Politics": using lexical cues of politician and party names and political terms: **"Trump"**, **"Republican"**, **"Democratic"**, **"Congress"**
Overgeneralizing highly-correlated features along with (long-distance dependencies)	Highly correlated words and phrases that do not directly cue frames, and are used in different contexts.	***Example Frame:*** *Race/Ethnicity* ***Headline:*** Lawyers call US gun charges for Mexican man **"vindictive"** • Model erroneously predicted "Mental Health": **"vindictive"** ***+ API tag:*** Shooting of Kate Steinle Acquittal Murder San Francisco Homicide Jury Death Defendant Illegal immigration Manslaughter • Model correctly predicted "Race/Ethnicity": better context

Table 6: Framing classification common error types, their definition and examples indicating the prediction error and how additional features in our top-performing methods of BERT HEADLINE + API can drive correct predictions.

annotated as relevant, e.g., "Politics", which is the most concrete frame, has 65% of its images annotated as relevant compared to just 10% for "Society/Culture", the least concrete frame. In fact, for most frames, a higher average concreteness implies a higher image relevance ratio. Exceptions to this are "Economic Consequences" and "Race/Ethnicity". Although more words in the former are abstract (e.g., *sales, demand, supply*), it is relatively easy to identify relevant images for economic consequence: e.g. *company logos, gun stores*. On the other hand, although more words in "Race/Ethnicity" may be concrete: e.g., people/organization names, ethnic minority group names, or hate group names, it is harder to find relevant images for this frame in news articles. This may be due to editors in mainstream media, and photographers or journalists, withholding certain imagery from readers for fear of causing offence or shock, or for fear that a part of their audience may abandon the publication altogether (Ritchin, 2014). Thus, although we find that frame concreteness is related to image relevance ratio (Pearson correlation of 0.69), there may be other factors that influence the choice of images for news articles that are beyond relevance to frames.

We also report per frame classification F1 scores for the baseline and our best performing models on *Articles with Relevant Images* and *All Articles*, i.e., when using the article headline alone, or with the the API, the Caption, and the Summary in Figure 3. Performance when using information from articles and images is remarkable for frames with either high image relevance ratios or high concreteness.

In *Articles with Relevant Images*, the frames with the highest image relevance ratio i.e., "Politics" shows an impressive 96.6% F1 score with the headline and the API tags, followed by "Public Opinion" (87.1%); while frames with few relevant images ("2nd Amendment", "Society/Culture"), have a substantially lower F1 scores. On *All Articles*, the inclusion of articles with irrelevant images can hurt performance for frames with high image relevance ratio such as Politics and Public Opinion. However, other frames may benefit from having more examples to learn from. For example, we observe that a low image relevance but a highly concrete frame such as "Race/Ethnicity" benefits significantly from this inclusion, reaching a high F1 score of 90.6% using *All Articles* from 82.1% using only *Articles with Relevant Images* as the model learns more lexical cues, e.g., named-entities from headlines of more articles, including those with irrelevant images. Concreteness may also augment relevance in explaining improved performance for some frames on *All Articles*. We observe high correlations between frame average concreteness and average F1 scores on articles with relevant images (Pearson correlation of 0.93) and on *All Articles* (Pearson correlation of 0.94), which exceed correlations between frame image relevance ratio and average F1 scores (Pearson correlation of 0.81 and 0.67 on articles with relevant images and on *All Articles*, respectively). These findings suggest that concreteness might be worth exploring for frame prediction and use of imagery in the future, in addition to concreteness annotation in framing datasets.

To complete our analysis, we applied the frame prediction error taxonomy proposed by Mendelsohn et al. (2021) to our news framing with image- and article-derived information, to identify and summarize common classification errors in Table 6. We provide specific examples, highlight possible error sources and observe how background information in BERT HEADLINE + API, drives correct predictions, illustrating our previous remarks.

6 Conclusions

We presented the first ever study and dataset on computational multimodal framing. Our results show that image-derived contextual features can be useful for providing missing contextual or background information that can improve frame prediction significantly, particularly for concrete frames or frames with relevant images. We also proposed methods for predicting frame image relevance and for measuring frame concreteness, which we define as the ease of expressing frames via images.

7 Ethical Considerations

Regarding the data we collected i.e., the Gun Violence Frame Corpus, we have made sure that there is no design experiment that was biased toward extracting only articles from a particular ethnic or minority group. We collect articles that had at least one keyword in their headlines from the following list, based on previous literature on gun violence framing analysis as described in Liu et al. (2019). The keywords are "gun", "firearm", "NRA", "2nd amendment", "second amendment", "AR15", "assault weapon", "rifle", "Brady act", "Brady bill", "mass shooting". The articles were retrieved in 2018 from 21 media outlets, from a list of top, in terms of website traffic, U.S. news websites; and synthesizing these lists towards creating one list that contained news sites from the left, center, and right sides of the ideological spectrum based on categories defined in MediaCloud; Pew Research Center (2016); Ad Fontes Media (2019).

Our analysis of the headlines and images gave each racial group's mentions' and portrayals' percentages provided in Table 7. We notice that when racial groups are mentioned in news headlines (which is only in ∼11% of all headlines), they are used to refer to victims of race-related gun violence incidents. Among the mentions, Blacks and Jews are the most common, as the corpus contains articles from 2018 that reported the mass shooting at Pittsburgh Synagogue[6] and several high-profile shootings of black men–widely reported as instances of the controversial and race-related "Stand Your Ground Law"[7], all of which occurred in 2018.

In terms of images, we notice that they are dominated by white people (∼50% of all images). However, the majority of them (∼71% of all white people images) are images of politicians or public figures related to gun laws/debates. There are much less images of victims and perpetrators (only ∼9% of all images each). In terms of victims vs. perpetrators, there are more images of black victims (∼1.8%) than black perpetrators (∼0.6%). The same applies to Asian, while for whites the numbers of victim and perpetrator images are more balanced. Based on our data analysis, in which we saw different coverage in terms of racial groups

[6] https://en.wikipedia.org/wiki/Pittsburgh_synagogue_shooting

[7] https://en.wikipedia.org/wiki/Shooting_of_Markeis_McGlockton

racial group	**headlines**				**images**			
	all	**P**	**PO**	**V**	**all**	**P**	**PO**	**V**
white	0.92	100	-	-	50.6	16	71	13
black	2.15	7	-	93	4.2	14	45	42
white & black	0.69	-	-	-	2.7	3	84	13
asian	-	-	-	-	0.46	16	50	34
white & asian	-	-	-	-	0.006	-	100	-
hispanic	0.07	100	-	-	-	-	-	-
jewish	7.1	2	-	98	-	-	-	-
other	89	-	-	-	-	-	-	-

Table 7: In column "all" we see the percentage (%) of headline mentions or image portrayals of certain racial groups in the 1,300 articles of the GVFC dataset. In columns **P**, **P0**, **V**, we can see the percentage (%) of the people in each of these groups who are either the **P**erpetrator, **PO**litician (or Public Figure), or **V**ictim.

in headlines vs. images, examining the difference between the race of the people mentioned in headlines and the race of those portrayed in the images would be an interesting future research direction.

Acknowledgments

This work is supported in part by the U.S. NSF grant 1838193, DARPA HR001118S0044 (the LwLL program), and the Department of the Air Force FA8750-19-2-3334 (Semi-supervised Learning of Multimodal Representations). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes. The views and conclusions contained in this publication are those of the authors and should not be interpreted as representing official policies or endorsements of DARPA, the Air Force, and the U.S. Government.

References

Ad Fontes Media. 2019. The media bias chart.

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46(3):904–911.

Andrea Burns, Donghyun Kim, Derry Wijaya, Kate Saenko, and Bryan A Plummer. 2020. Learning to scale multilingual representations for vision-language tasks. In *European Conference on Computer Vision*, pages 197–213. Springer.

Ozan Caglayan, Pranava Madhyastha, Lucia Specia, and Loïc Barrault. 2019. Probing the need for visual context in multimodal machine translation. *arXiv preprint arXiv:1903.08678*.

Helen Caple. 2010. What you see and what you get: The evolving role of news photographs in an Australian broadsheet. *Journalism and Meaning-Making: Reading the Newspaper*, pages 199–220.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Renita Coleman and Denis Wu. 2015. *Image and emotion in voter decisions: The affect agenda*. Lexington Books.

Viorela Dan. 2017. *Integrative framing analysis: Framing health through words and visuals*. Routledge.

Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. 2017. Language modeling with gated convolutional networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 933–941. PMLR.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 248–255.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah Smith.

2020. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *arXiv preprint arXiv:2002.06305*.

Kevin M Drakulich. 2015. Explicit and hidden racial bias in the framing of social problems. *Social Problems*, 62(3):391–418.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3570–3580.

Michael Friendly, Patricia E Franklin, David Hoffman, and David C Rubin. 1982. The toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, 14(4):375–399.

Ken J Gilhooly and Robert H Logie. 1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior research methods & instrumentation*, 12(4):395–427.

Lei Guo, Kate Mays, Yiyan Zhang, Derry Wijaya, and Margrit Betke. 2021. What makes gun violence a (less) prominent issue? a computational analysis of compelling arguments and selective agenda setting. *Mass communication and society*, pages 1–25.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019. Illegal aliens or undocumented immigrants? Towards the automated identification of bias by word choice and labeling. Technical report, University of Konstanz, Germany.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

John Hewitt, Daphne Ippolito, Brendan Callahan, Reno Kriz, Derry Tanti Wijaya, and Chris Callison-Burch. 2018. Learning translations via images with a massively multilingual image dataset. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2566–2576.

Nikzad Khani, Isidora Tourni, Mohammad Sadegh Rasooli, Chris Callison-Burch, and Derry Tanti Wijaya. 2021. Cultural and geographical influences on image translatability of words across languages. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 198–209.

Donghyun Kim, Kuniaki Saito, Kate Saenko, Stan Sclaroff, and Bryan Plummer. 2020. Mule: Multimodal universal language embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11254–11261.

Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision*, pages 2980–2988.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Ro{bert}a: A robustly optimized {bert} pretraining approach.

Wilson Lowrey. 2002. Word people vs. picture people: Normative differences and strategies for control over work among newsroom subgroups. *Mass Communication & Society*, 5(4):411–432.

MediaCloud. 2018. https://mediacloud.org.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. *arXiv preprint arXiv:2104.06443*.

Paul Messaris and Linus Abraham. 2001. The role of images in framing news stories. In *Framing public life*, pages 231–242. Routledge.

Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow. 2021. On the stability of fine-tuning {bert}: Misconceptions, explanations, and strong baselines. In *International Conference on Learning Representations*.

Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, Varna, Bulgaria, September 2 - 8, 2017*, pages 536–542. INCOMA Ltd.

Viet-An Nguyen, Jordan L Ying, and Philip Resnik. 2013. Lexical and hierarchical topic regression. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Allan Paivio, John C Yuille, and Stephen A Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2):1.

Pew Research Center. 2016. Ideological placement of each source's audience.

Thomas E Powell, Hajo G Boomgaarden, Knut De Swert, and Claes H de Vreese. 2015. A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, 65(6):997–1017.

Mohammad Sadegh Rasooli, Chris Callison-Burch, and Derry Tanti Wijaya. 2021. " wikily" neural machine translation tailored to cross-lingual tasks. *arXiv preprint arXiv:2104.08384*.

Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767.

Stephen D Reese, Oscar H Gandy Jr, August E Grant, et al. 2001. *Framing public life: Perspectives on media and our understanding of the social world*. Routledge.

Fred Ritchin. 2014. Why violent news images matter. [time.com;4-September-2014].

Lucia Specia, Stella Frank, Khalil Sima'An, and Desmond Elliott. 2016. A shared task on multimodal machine translation and crosslingual image description. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 543–553.

Alasdair Tran, Alexander Mathews, and Lexing Xie. 2020. Transform and tell: Entity-aware news image captioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13035–13045.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Weiyao Wang, Du Tran, and Matt Feiszli. 2020. What makes training multi-modal classification networks hard? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12695–12705.

Hartmut Wessler, Antal Wozniak, Lutz Hofer, and Julia Lück. 2016. Global multimodal news frames on climate change: A comparison of five democracies around the world. *The International Journal of Press/Politics*, 21(4):423–445.

Shaowei Yao and Xiaojun Wan. 2020. Multimodal transformer for multimodal machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4346–4350.

Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.

A Appendix

The first 16 entries in the **SRE** feature vector indicate the central **S**ubject of the image, with each Subject implying certain frame(s):

1. People: Gun shooter/suspect (**Mental Health**)
2. People: Gun hobbyist/activist + gun-related activities with a hand (**Gun Rights**)
3. People: Victim/affected family and friends/bystanders (**Public Opinion**)
4. People: Politicians (**Politics**)
5. People: Law enforcement (e.g., police offers, security guards) (**Public Safety**)
6. Object: Firearm/bullets (can mean anything or **Gun Control** for certain gun images)
7. Object: Gun /hunting gear stores/gun show (**Economic Consequences** or **Gun Rights**)
8. People: Demonstrators/Demonstrations (**Public Opinion)**
9. Object: Protest signs (**Gun Control** or **Gun Rights**)
10. People/mainly object: Memorials objects and people (**Public Opinion**)
11. Object/people: Crime scene/police cars/people during or right after the crisis (episodic frame)
12. Object: Legislative buildings/courthouses (**Gun Control** or **Politics**)
13. Object/people: School/campus/students indicating school/campus (**Public Safety**)
14. NRA objects/NRA representatives (**Gun Rights** or **Economic Consequences**)
15. Object: Company buildings/logos (**Economic Consequence**)
16. Other

The last three entries in the feature vector indicate relevance to **R**ace or **E**thnicity:

17. None
18. Racial/ethnic minority groups /buildings of a specific group (**Ethnicity**) (only if the central subject is from racial/ethnic minority group - not if there is only one or a couple of non-White people in a large crowd)
19. KKK/white supremacy/hate groups (**Ethnicity**)

Frame	Description	Examples
Politics	Political issues around guns and shootings, including: • Political campaigns and upcoming elections (e.g., using guns as a wedge issue or motivating force to get people to the polls) • Fighting between the Democratic and Republican parties, or politicians • Political money contributions from gun lobbies (e.g., NRA) • One political party or one politician's stance on gun violence. • Therefore, as long as the news headline mentions a politician's name, it often indicates the theme of politics. • Often times, the politicians' names or the party names should be mentioned.	"Baltimore students walk out of class to protest gun violence"
Public Opinion	Public's opinion and the community's reactions to gun-related issues, including: (e.g., using guns as a wedge issue or motivating force to get people to the polls) • Public opinion polls related to guns • Protests • One political party or one politician's stance on gun violence. • Mourning victims of gun violence • The public's emotional responses	"How Illinois governor candidates would address gun violence" "Trump warns Dems will 'take away your Second Amendment'" "Lindsey Graham: Both parties will suffer if Congress doesn't act on new gun bill"
Gun Control/Regulation	Issues related to regulating guns through legislation and other institutional measures: • Enforcing and/or expanding background checks • Limiting sale of guns and/or related dangerous equipment (e.g., AR15s, semi-automatic rifles, bump stocks, Huge-capacity ammo) • Increasing age limits on gun purchases • Implementing licensing and gun safety training programs	"GOP lawmaker calls for age restriction on AR-15s" "No bump stocks turned in to Denver police after ban"
School/Public Space Safety	Issues related to institutional and school safety, including: • Awareness and monitoring of "troubled" individuals by law enforcement (e.g., local police, FBI) • Safety measures in schools to prevent or mitigate shootings (e.g., police/safety officers in the school, armed teachers, metal detectors, clear backpacks) • Note that a headline simply mentioning "school shooting" does not necessarily mean it uses this safety measure frame.	"Preschoolers among students required to carry clear backpacks in Texas school district" "Scott wants armed police at Stoneman Douglas after disturbing incidents at Parkland school" "Sales of bulletproof school supplies spike after Florida shooting"
Economic consequences	Financial losses or gains, or the costs involved in gun-related issues, including: • The actual sales of firearms • The financial consequences of gun regulation (e.g., lost tax revenue, or gun manufacturing companies moving to a different state) • The financial state of gun-related lobbying groups (e.g., the NRA) • Federal budget for gun-related programs	"The NRA Is In Deep, Deep Financial Trouble"
Race/Ethnicity	Gun issues related to certain ethnic group(s), including: • Angry, isolated white men as primary perpetrators of domestic gun violence • Immigrants from Mexico bringing in guns from across the border • Muslim "terrorists" • Gun violence in African American communities	"Illegal immigrant acquitted of Kate Steinle's murder faces judge on gun charges" "The disparities in how black and white men die in gun violence, state by state"
Mental Health	Issues related to individuals' mental illnesses or emotional well-being, or the mental health system as a whole, including: • Predicting and preventing mental health breakdowns • Treating mental illness • Creating measures to ensure mentally ill people do not have access to guns • Descriptions of individuals' behavioral / personality traits that indicate instability, impulsivity, anger, etc.	"Gun debate hits home for families dealing with myths about violence, mental illness" "Renewed Debate Over Gun Access, Mental Health" "Las Vegas gunman lost money, became unstable before shooting"
2nd Amendment/Gun Rights	Related to the Constitution, the second amendment, and protection of individual liberty and gun ownership as a right, including: • Meaning of the 2nd amendment • The irrefutability of one's right to own guns • Gun ownership as critical to democracy and protecting oneself	"Membership, interest in gun rights groups soar in the weeks after the Florida high school shooting" "Rapper 'Killer Mike,' NRA host Colion Noir: No guns would turn people into slaves"
Society/Culture	Societal-wide factors that are related to gun violence, including: • Violence in media (e.g., TV/movies and video games) • Social pressures that may incite someone to violence (e.g., cliques/bullying and isolation) • Breakdown in family structures, so there is a lack of familial support and stability • Breakdown in community structures (e.g., religious organizations, other civic-oriented groups), so there is a lack of community support and stability	"There's Not A Single Ounce Of Evidence To Link Mass Shootings To Video Games"

Table 8: News frames' description and Headline examples.