Persuasion Strategies in Advertisements: Dataset, Modeling, and Baselines

Yaman Kumar Singla[1,2,3], **Rajat Jha**[1], **Arunim Gupta**[1], **Milan Aggarwal**[2], **Aditya Garg**[1],
Ayush Bhardwaj[1], **Tushar**[1], **Balaji Krishnamurthy**[2], **Rajiv Ratn Shah**[1], and **Changyou Chen**[3]

[1] IIIT-Delhi, [2] Adobe Media and Data Science Research (MDSR), [3] State University of New York at Buffalo

Abstract

Modeling what makes an advertisement persuasive, *i.e.*, eliciting the desired response from consumer, is critical to the study of propaganda, social psychology, and marketing. Despite its importance, computational modeling of persuasion in computer vision is still in its infancy, primarily due to the lack of benchmark datasets that can provide persuasion-strategy labels associated with ads. Motivated by persuasion literature in social psychology and marketing, we introduce an extensive vocabulary of persuasion strategies and build the first ad image corpus annotated with persuasion strategies. We then formulate the task of persuasion strategy prediction with multi-modal learning, where we design a multi-task attention fusion model that can leverage other ad-understanding tasks to predict persuasion strategies. Further, we conduct a real-world case study on 1600 advertising campaigns of 30 Fortune-500 companies where we use our model's predictions to analyze which strategies work with different demographics (age and gender). The dataset also provides image segmentation masks, which labels persuasion strategies in the corresponding ad images on the test split. We publicly release our code and dataset[1].

1   Introduction

Marketing communications is the mode by which companies and governments inform, remind, and persuade their consumers about the products they sell. They are the primary means of connecting brand with consumers through which the consumer can know what the product is about, what it stands for, who makes it, and can be motivated to try it out. To introduce meaning into their communication, marketers use various rhetorical devices in the form of persuasion strategies such as **emotions** (*e.g.*, Oreo's "Celebrate the Kid Inside", humor by showing Ronald McDonald sneaking into the competitor Burger King's store to buy a burger), **reasoning** (*e.g.*, "One glass of Florida orange juice contains 75% of your daily vitamin C needs"), **social identity** (*e.g.*, Old Spice's "Smell like a Man"), and **impact** (*e.g.*, Airbnb showing a mother with her child with the headline "My home is funding her future")[2]. Similarly, even for marketing the same product, marketers use different persuasion strategies to target different demographies (see Fig. 1,§5). Therefore, recognizing and understanding persuasion strategies in ad campaigns is vitally important to decipher viral marketing campaigns, propaganda, and enable ad-recommendation.

Studying rhetorics of this form of communication is an essential part of understanding visual communication in marketing. Aristotle, in his seminal work on rhetoric, underlining the importance of persuasion, equated studying rhetorics

Figure 1: Different persuasion strategies are used for marketing the same product (footwear in this example). The strategies are in red words and to be defined by us in the paper.

with the study of persuasion[3] (Rapp 2002). While persuasion is studied extensively in social science fields, including marketing (Meyers-Levy and Malaviya 1999; Keller, Lipkus, and Rimer 2003) and psychology (Hovland, Janis, and Kelley 1953; Petty and Cacioppo 1986), computational modeling of persuasion in computer vision is still in its infancy, primarily due to the lack of benchmark datasets that can provide representative corpus to facilitate this line of research. In the limited work that has happened on persuasion in computer vision, researchers have tried to address the question of which image is more persuasive (Bai et al. 2021) or extracted low-level features (such as emotion, gestures, and facial displays), which indirectly help in identifying persuasion strategies without explicitly extracting the strategies themselves (Joo et al. 2014). On the other hand, decoding persuasion in textual content has been extensively studied in natural language processing from both extractive, and generative contexts (Habernal and Gurevych 2016; Chen and Yang 2021a; Luu, Tan, and Smith 2019). This forms the motivation of our work, where we aim to identify the persuasion strategies used in visual content such as advertisements.

The systematic study of persuasion began in the 1920s with the media-effects research by Lasswell (1971), which was used as the basis for developing popular models of persuasion, like the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986), Heuristic Systematic Model (HSM) (Chaiken 1980), and Hovland's attitude change approach (Hovland, Janis, and Kelley 1953). These models of persuasion posit a dual process theory that explains attitude

---

[1] https://midas-research.github.io/persuasion-advertisements/
[2] Refer Appendix:Fig:1 for seeing these ads

[3] "Rhetoric may be defined as the faculty of discovering in any particular case all of the available means of *persuasion*" (Rapp 2002)

Figure 2: Persuasion strategies in advertisements. Marketers use both text and vision modalities to create ads containing different messaging strategies. Different persuasion strategies are constituted by using various rhetorical devices such as slogans, symbolism, colors, emotions, allusion.

and behavior change (persuasion) in terms of the following major factors: stimuli (messages), personal motivation (the desire to process the message), capability of critical evaluation, and cognitive busyness. These factors could be divided into cognitive, behavioral, and affective processes of attitude change. In this work, we build on these psychological insights from persuasion models in sociology and marketing and study the message strategies that lead to persuasion. We codify, extend, and unify persuasion strategies studied in the psychology and marketing literature into a set of 20 strategies divided into 9 groups (see Fig. 2, Table 1): *Authority and Credibility*, *Social Identity and Proof*, where cognitive indirection in the form of group decisioning and expert authority is used for decisions, *Value and Impact Formulation* where logic is used to explain details and comparisons are made, *Reciprocity*, *Foot in the door*, *Overcoming Resistance* where social and cognitive consistency norms are harnessed to aid decision-making, *Scarcity*, *Anthropomorphism* and *Emotion* where information is evaluated from the lenses of feelings and emotions. In addition to introducing the most extensive vocabulary for persuasion strategies, we make a superset of persuasion strategies presented in the prior NLP works, which introduced text and domain-specific persuasion tactics, thus making large-scale understanding of persuasion across multiple contexts comparable and replicable.

Constructing a large-scale dataset containing persuasion strategies labels is time-consuming and expensive. We leverage active learning to mitigate the cost of labeling fine-grained persuasion strategies in advertisements. We first introduce an attention-fusion model trained in a multi-task fashion over modalities such as text, image, and symbolism. We use the action-reason task from the Pitts Ads dataset (Hussain et al. 2017) to train the model and then annotate the raw ad images from the same dataset for persuasion strategies based on an entropy based active learning technique.

To sum up, our contributions include:
1. We construct the largest set of generic persuasion strategies based on theoretical and empirical studies in marketing, social psychology, and machine learning literature.
2. We introduce the first dataset for studying persuasion

strategies in advertisements. This enables initial progress on the challenging task of automatically understanding the messaging strategies conveyed through visual advertisements. We also construct a prototypical dataset containing image segmentation masks annotating persuasion strategies in different segments of an image.
3. We formulate the task of predicting persuasion strategies with a multi-task attention fusion model.
4. We conduct extensive experiments on the released corpus, showing the effect of different modalities on identifying persuasion strategies, correlation between strategies and topics and objects with different strategies. We also conduct a real-world study on 1600 advertising campaigns of 30 Fortune-500 companies, where we used our model's predictions to analyze which strategies work with different demographics (age and gender).

2   Related Work

How do messages change people's beliefs and actions? The systematic study of persuasion has captured researchers' interest since the advent of mass influence mechanisms such as radio, television, and advertising. Work in persuasion spans across multiple fields, including psychology, marketing, and machine learning.

Persuasion in Marketing and Social Psychology:
Sociology and communication science has studied persuasion for centuries now starting from the seminal work of Aristotle on rhetoric. Researchers have tried to construct and validate models of persuasion. Due to space constraints, while we cannot cover a complete list of literature, in Table 1, we list the primary studies which originally identified the presence and effect of various persuasion tactics on persuadees. We build on almost a century of this research and crystallize them into the persuasion strategies we use for annotation and modeling. Any instance of (successful) persuasion is composed of two events: (a) an attempt by the persuader, which we term as the persuasion strategies, and (b) subsequent uptake and response by the persuadee (Anand et al. 2011; Vakratsas and Ambler 1999). In this work, we study (a) only while leaving (b) for future work. Through-

Table 1: The generic taxonomy of persuasive strategies, their definitions, examples, and connections with prior work. **SPM**: Representative literature from Social Psychology and Marketing (including Consumer Psychology), **ML**: Representative literature from Machine Learning

Group	*Strategy*	Definition	Representative Prior Work
Authority and Credibility	*Guarantees*	Guarantees reduce risk and people try out such products more often.	**SPM**:(Aronson, Turner, and Carlsmith 1963; Milgram and Gudehus 1978; Cialdini and Cialdini 2007; Milgram 1963; McGinnies and Ward 1980; Giffin 1967; Petty and Cacioppo 1986) **ML**:(Anand et al. 2011; Iyer and Sycara 2019; Wachsmuth et al. 2017; Chen and Yang 2021a; Durmus and Cardie 2018)
	Authority	Authority indicated through expertise, source of power, third-party approval, credentials, and awards	
	Trustworthiness	Trustworthiness indicated honesty and integrity of the source through tropes like years of experience, "trusted brand", numbers and statistics	
Social Identity and Proof	*Social Identity*	*Normative* influence, which involves conformity with the positive expectations of "another", who could be "another person, a group, or one's self" (includes self-persuasion, fleeting attraction, alter-casting, and exclusivity)	**SPM**:(Deutsch and Gerard 1955; Petty, Wegener, and Fabrigar 1997; Wood 2000; Cialdini and Goldstein 2004; Levesque and Pons 2020) **ML**: (Anand et al. 2011; Iyer and Sycara 2019; Rosenthal and Mckeown 2017; Yang et al. 2019; Zhang, Litman, and Forbes-Riley 2016)
	Social Proof	*Informational influence* by accepting information obtained from others as evidence about reality, *e.g.*, customer reviews and ratings	
Reciprocity	*Reciprocity*	By *obligating* the recipient of an act to repayment in the future, the rule for reciprocation begets a sense of future obligation, often unequal in nature	**SPM**:(Regan 1971; Cialdini and Cialdini 2007; Clark 1984; Clark and Mills 1979; Clark, Mills, and Powell 1986) **ML**:(Anand et al. 2011; Iyer and Sycara 2019; Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014; Chen and Yang 2021a; Shaikh et al. 2020)
Foot in the door	*Foot in the door*	Starting with small requests followed by larger requests to facilitate compliance while maintaining *cognitive coherence.*	**SPM**: (Freedman and Fraser 1966; Burger 1999; Cialdini and Cialdini 2007) **ML**:(Chen and Yang 2021b; Wang et al. 2019; Vargheese, Collinson, and Masthoff 2020)
Overcoming Resistance	*Overcoming Resistance*	Overcoming resistance (reactance) by postponing consequences to the future, by focusing resistance on realistic concerns, by forewarning that a message will be coming, by acknowledging resistance, by raising self-esteem and a sense of efficacy.	**SPM**:(McGuire and Papageorgis 1961; Knowles and Linn 2004; McGuire 1964) **ML**:{None}
Value and Impact Formulation	*Concreteness*	Using concrete facts, evidence to appeal to the logic of consumers	**SPM**:(Lee, Keller, and Sternthal 2010; Furnham and Boo 2011; Wegener et al. 2001; Tversky and Kahneman 1974; Strack and Mussweiler 1997; Bhattacharya and Sen 2003) **ML**:(Yang et al. 2019; He et al. 2018; Durmus and Cardie 2018; Zhang, Culbertson, and Paritosh 2017; Wang et al. 2019; Longpre, Durmus, and Cardie 2019)
	Anchoring and Comparison	A product's value is strongly influenced by what it is compared to.	
	Social Impact	Emphasizes the importance or bigger impact of a product	
Scarcity	*Scarcity*	People assign more value to opportunities when they are less available. This happens due to psychological reactance of losing freedom of choice when things are less available or they use availability as a cognitive shortcut for gauging quality.	**SPM**:(Brehm 1966; Lynn 1991; Rothman et al. 1999; Tversky and Kahneman 1985) **ML**:(Yang et al. 2019; Chen and Yang 2021a; Shaikh et al. 2020)
Anthropomorphism	*Anthropomorphism*	When a brand or product is seen as human-like, people will like it more and feel closer to it.	**SPM**:(Fournier 1998; Levesque and Pons 2020; Epley, Waytz, and Cacioppo 2007) **ML**:{None}
Emotion	*Amazed Fashionable Active,Eager Feminine Creative Cheerful Further Minor*	Aesthetics, feeling and other non-cognitively demanding features used for persuading consumers	**SPM**:(Hibbert et al. 2007; Petty and Cacioppo 1986; Petty, Cacioppo, and Schumann 1983) **ML**:(Yang et al. 2019; Tan et al. 2016; Hidey et al. 2017; He et al. 2018; Durmus and Cardie 2018; Zhang, Culbertson, and Paritosh 2017; Wachsmuth et al. 2017)
Unclear	*Unclear*	If the ad strategy is unclear or it is not in English	

out the rest of the paper, when we say persuasion strategy, we mean the former without considering whether the persuasion was successful or not.

Persuasion in Machine Learning: Despite extensive work in social psychology and marketing on persuasion, most of the work is qualitative, where researchers have looked at a small set of messages with various persuasion strategies to determine their effect on participants. Computational modeling of persuasion is still largely lacking. In the limited work in computational modeling of persuasion, almost all of it is concentrated in the NLP literature, with only very few works in computer vision. Research on persuasion in NLP under the umbrella of argumentation mining is broadly carried out from three perspectives: extracting persuasion tactics, studying the effect of constituent factors on persuasion, and measurement of persuasiveness nature of content. Anand et al. (2011); Stab and Gurevych (2014); Tan et al. (2016); Chen and Yang (2021b) are some examples of research studies which annotate persuasive strategies in various forms of persuader-persuadee interactions like discussion forums, social media, blogs, academic essays, and de-

bates. We use these and other studies listed in Table 1 to construct our vocabulary of persuasion strategies in advertisements. Other studies focus on factors such as argument ordering (Shaikh et al. 2020; Li, Durmus, and Cardie 2020), target audience (Lukin et al. 2017), and prior beliefs (El Baff et al. 2020) for their effect in bringing about persuasion. Studies such as Althoff, Danescu-Niculescu-Mizil, and Jurafsky (2014); Raifer et al. (2022); Wei, Liu, and Li (2016) also try to measure persuasiveness and generate persuasive content.

As one of the first works in the limited work in the computer vision domain, Joo et al. (2014) introduced syntactical and intent features such as facial displays, gestures, emotion, and personality, which result in persuasive images. Their analysis was done on human images, particularly politicians, during their campaigns. Joo et al. (2014)'s work on political campaigners is more restrictive than general product and public-service advertisements. Moreover, they deal with low-level features such as gestures and personality traits depicted through the face, which are important for detecting persuasion strategies but are not persuasion
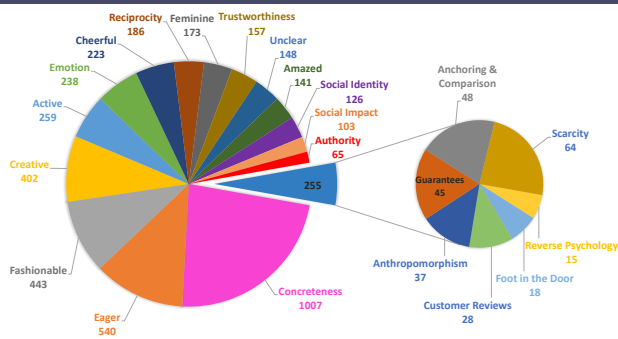
Figure 3: Distribution of Persuasion Strategies. The top-3 strategies are Concreteness, Eager, and Fashionable.

Figure 4: Image with a segmentation mask depicting the strategies *Emotion:Cheerful*, *Emotion:Eager* and *Trustworthiness*. Appendix:Fig:9 contains more such examples.

strategies themselves. Recently, Bai et al. (2021) studied persuasion in debate videos where they proposed two tasks: debate outcome prediction and intensity of persuasion prediction. Through these tasks, they predict the persuasiveness of a debate speech, which is orthogonal to the task of predicting the strategy used by the debater. Other similar works which discuss persuasiveness of images and videos are (Joo, Steen, and Zhu 2015; Siddiquie, Chisholm, and Divakaran 2015).

3 Persuasion Strategy Corpus Creation

To annotate persuasion strategies on advertisements, we leverage raw images from the Pitts Ads dataset. It contains 64,832 image ads with labels of topics, sentiments, symbolic references (*e.g.* dove symbolizing peace), and reasoning the ad provides to its viewers (see Appendix:Fig:8 for a few examples). The dataset had ads spanning multiple industries, products, services, and also contained public service announcements. Through this, they presented an initial work for the task of understanding visual rhetoric in ads. Since the dataset already had a few types of labels associated with the ad images, we used active learning on a model trained in a multi-task learning fashion over the reasoning task introduced in their paper. We explain the model and then the annotation strategy followed in §4.

To commence training, we initially annotated a batch of 250 ad images with persuasion strategies defined in Table 1. We recruited four research assistants to label persuasion strategies for each advertisement. Definitions and examples of different persuasion strategies were provided, together with a training session where we asked annotators to annotate a number of example images and walked them through any disagreed annotations. To assess the reliability of the annotated labels, we then asked them to annotate the same 500 images and computed Cohen's Kappa statistic to measure inter-rater reliability. We obtained an average score of 0.55. The theoretical maximum of Kappa given the unequal distribution is 0.76. In such cases, Cohen (1960) suggested that one should divide kappa by its maximum value $\Bbbk/\Bbbk_{max}$, which comes out to be 0.72. This is substantial agreement. Further, to maintain the labeling consistency, each image was double annotated, with all discrepancies resolved by an intervention of the third annotator using a majority voting.

The assistants were asked to label each image with no more than 3 strategies. If an image had more than 3 strategies, they were asked to list the top-3 strategies according to the area covered by the pixels depicting that strategy. In to-

tal, we label 3000 ad-images with their persuasion strategies; and the numbers of samples in train, val and test split are 2500, 250 and 250 resp.[4] Fig. 3 presents the distribution of persuasion strategies in the dataset. It is observed that concreteness is the most used strategy in the dataset, followed by eager and fashionable. The average number of strategies in an ad is 1.49, and the standard deviation is 0.592. We find that scarcity (92.2%), guarantees (91.1%), reciprocity (84.4%), social identity (83.3%), and cheerful (83%), are the top 5 strategies, which occur in groups of 2 or 3. We observe that the co-occurrence of these strategies is due to the fact that many of them cover only a single modality (*i.e.*, text or visual), leaving the other modality free for a different strategy. For example, concreteness is often indicated by illustrating points in text, while the visual modality is free for depicting, say, emotion. See Fig:9 in the Appendix for an example, where the image depicting *Authority* also has concreteness strategy in it. Similarly, feminine emotion is also depicted in Fig. 1, along with concreteness.

Next, we calculate the Dice correlation coefficient statistics for pairs of co-occurring persuasion strategies. The top-5 pairs are eager-concreteness (0.27), scarcity-reciprocity (0.25), eager-cheerful (0.19), amazed-concreteness (0.17), and eager-reciprocity (0.17). We find that these correlation values are not particularly high since marketers seldom use *common pairings* of messaging strategies to market their products. The visual part mostly shows eager strategy in ads; therefore, we find that the text modality becomes free to show other strategies. That is why primarily text-based concreteness, cheerfulness, and reciprocity strategies are present in the text modality with the visual-based eager strategy. On the other hand, primarily vision-based amazement, eagerness, and scarcity (short-text) strategies co-occur with text-based reciprocity and concreteness. Fig. 1 shows an example.

Next, we calculate the correlation between image topics and objects present with persuasion strategies. We see that the emotion:feminine and emotion:fashionable strategies are most often associated with beauty products and cosmetics (corr=0.4256, 0.2891). This is understandable since most beauty products are aimed at women. We see that the fast-food and restaurant industries often use eagerness as their messaging strategy (corr = 0.5877,0.3470). We find that the presence of humans in ads is correlated

---

[4] Appendix:Table:1 shows the detailed distribution of number of strategies in ads

with the concreteness strategy[5] (corr=0.3831). On the other hand, vehicle ads use emotion:amazed and concreteness (corr=0.5211,0.2412)[6]. Similar to a low correlation in co-occurring strategies, we find that product segments and their strategies are not highly correlated. This is because marketers use different strategies to market their products even within a product segment. Fig. 1 shows an example in which the footwear industry (which is a subsegment of the apparel industry) uses different strategies to market its products. Further, for a batch of 250 images, we also label segmented image regions corresponding to the strategies present in the image. These segment masks were also double-annotated. Fig. 4 presents an example of masks depicting parts of the image masked with different persuasion strategies in a drink advertisement.

4 Modeling: Persuasion Strategy Prediction

The proposed Ads dataset $\mathcal{D}$ annotated with the persuasion strategies comprises of samples where each sample advertisement a_i is annotated with a set of annotation strategies S_i such that $1 \leq |S_i| \leq 3$. The unique set of the proposed persuasion strategies $\mathcal{P}$ is defined in Table 1. Given a_i, the task of the modeling is to predict the persuasion strategies present in the input ad. As we observe from Fig. 2, advertisements use various rhetoric devices to form their messaging strategy. The strategies thus are in the form of multi-modalities, including images, text and symbolism. To jointly model the modalities, we design an attention fusion multi-modal framework, which fuses multimodal features extracted from the ad, *e.g.*, the ad image, text present in the ad extracted through the OCR (Optical Character Recognition), regions of interest (ROIs) extracted using an object detector, and embeddings of captions obtained through an image captioning model (see Fig. 5). The information obtained through these modalities are firstly embedded independently through their modality specific encoders followed by a transformer-based cross-attention module to fuse the extracted features from different modalities. The fused embeddings from the attention module are then used as input for a classifier that predicts a probability score for each strategy $p \in \mathcal{P}$. The overall architecture of the proposed model is illustrated in Fig.5. In the following, we describe each step in the prediction pipeline in detail.

4.1 Feature Extractors

In order to capture different rhetoric devices, we extract features from the image, text, and symbolism modalities.

Image Feature: We use the Vision Transformer (Dosovitskiy et al. 2020) (ViT) model for extracting image features from the entire input image. The model resizes the input image to size 224×224 and divides it into patches of size 16×16. The model used has been pre-trained on the ImageNet 21k dataset. We only use the first output embedding, which is the CLS token embedding, a 768 dimension tensor, as we only need a representation of the entire image. Then,

a fully connected layer is used to reduce the size of the embedding, resulting in a tensor of dimension 256.

Regions of Interest (RoIs) from Detected Objects and Captions: Ad images contain elements that the creator deliberately chooses to create *intentional impact* and deliver some *message* in addition to the ones that occur *naturally* in the environment. Therefore, it is important to identify the composing elements of an advertisement to understand the creator's intention and the ad's message to the viewer. We detect and extract objects as regions of interest (RoIs) from the advertisement images. We get the RoIs by training the single-shot object detector model by Liu et al. (2016) on the COCO dataset (Lin et al. 2014). We compare it with the recent YOLOv5 model (Redmon et al. 2016). We also extract caption embeddings to detect the most important activity from the image using a caption generation mode. We compare DenseCap (Yang et al. 2017) and the more recent BLIP (Li et al. 2022) for caption generation.

OCR Text: The text present in an ad presents valuable information about the brand, such as product details, statistics, reasons to buy the product, and creative information in the form of slogans and jingles that the company wants its customers to remember and thus making it helpful in decoding various persuasion strategies. Therefore, we extract the text from the ads and use it as a feature in our model. We use the Google Cloud Vision API for this purpose. All the extracted text is concatenated, and the size is restricted to 100 words. We pass the text through a BERT model (Devlin et al. 2019) and use the final CLS embedding as our OCR features. Similar to image embeddings, an FC layer is used to convert embeddings to 256 dimensional vectors. The final embedding of the OCR is a tensor of dimension 100×256.

Symbolism: While the names of the detected objects convey the names or literal meaning of the objects, creative images often also use objects for their symbolic and figurative meanings. For example, an upward-going arrow represents growth or the north direction or movement towards the upward direction depending on the context; similarly, a person with both hands pointing upward could mean danger (*e.g.*, when a gun is pointed) or joy (*e.g.*, during dancing). In Fig. 2, in the creative Microsoft ad, a symbol of a balloon is created by grouping multiple mice together. Therefore, we generate symbol embeddings to capture the symbolism behind the most prominent visual objects present in an ad. We use the symbol classifier by Hussain et al. (2017) on ad images to find the distribution of the symbolic elements present and then convert this to a 256 dimension tensor.

4.2 Cross-Modal Attention

To capture the inter-dependency of multiple modalities for richer embeddings, we apply a cross-modal attention (CMA) layer (Frank, Bugliarello, and Elliott 2021) to the features extracted in the previous steps. Cross-modal attention is a fusion mechanism where the attention masks from one modality (*e.g.* text) are used to highlight the extracted features in another modality (*e.g.* symbolism). It helps to link and extract common features in two or more modalities since common elements exist across multiple modalities, which complete and reinforce the message conveyed in the ad.

---

[5]see Appendix:Fig:7 for a few examples
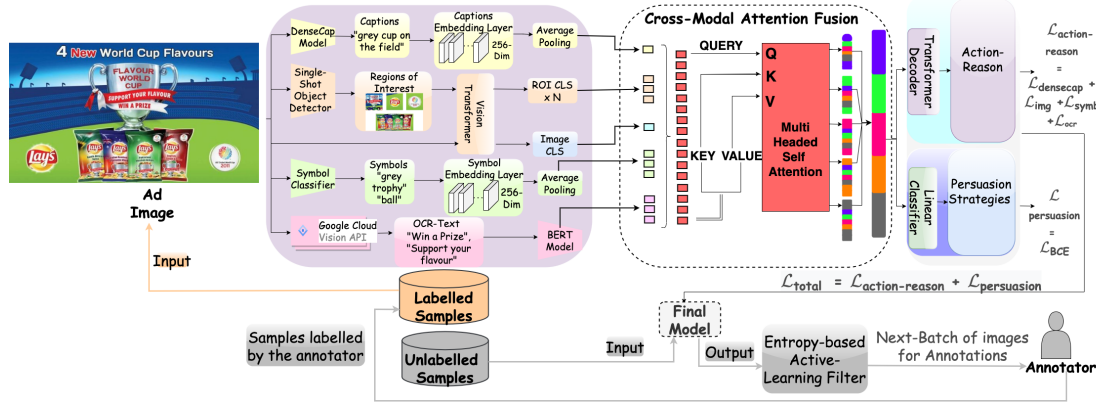[6]See Appendix:Fig:2 for detailed correlations.

Figure 5: Architecture of the Persuasion Strategy Prediction model. To capture the different rhetoric devices, we extract features for the image, text, and symbolism modalities and then apply cross-modal attention fusion to leverage the interdependence of the different devices. Further, the model trains over two tasks: persuasion strategies and the reasoning task of action-reason prediction.

For example, the pictures of the silver cup, stadium, and ball, words like "Australian", "Pakistani", and "World Cup" present in the chips ad shown in Fig. 5 link the idea of buying *Lays* with supporting one's country's team in the World Cup. Cross attention can also generate effective representations in the case of missing or noisy data or annotations in one or more modalities (Frank, Bugliarello, and Elliott 2021). This is helpful in our case since marketing data often uses implicit associations and relations to convey meaning.

The input to the cross-modal attention layer is constructed by concatenating the image, RoI, OCR, caption, and symbol embeddings. This results in a 114×256 dimension input to our attention layer. The cross-modal attention consists of two layers of transformer encoders with a hidden dimension size of 256. The output of the attention layer gives us the final combined embedding of our input ad. Given image embeddings E_i, RoI embeddings E_r, OCR embeddings E_o, caption embeddings E_c and symbol embeddings E_s, the output of the cross-attention layer E_{att} is formulated as:

$$\text{Enc}(X) = \text{CMA}([E_i(X), E_r(X), E_o(X), E_c(X), E_s(X)]),$$

where $[\ldots, \ldots]$ is the concatenation operation.

4.3 Persuasion Strategy Predictor

This module is a persuasion strategy predictor, which processes the set of feature embedding Enc(X) obtained through cross-modality fusion. Specifically, Enc(X) is passed through a self-attention layer as:

$$o_1 = softmax(\text{Enc}(X) \otimes W_{self-attn})^{\intercal} \otimes \text{Enc}(X) \quad (1)$$

where Enc(X) is of the dimension 114×256, $W_{self-attn} \in \mathcal{R}^{256 \times 1}$, $\otimes$ denote tensor multiplication and o_1 denotes the output of self attention layer, which is further processed through a linear layer to obtain $o_{|\mathcal{P}|}$ to represent the logits for each persuasion strategy. We apply sigmoid over each output logit such that the ith index of the vector after applying sigmoid denotes p_i - the probability with which i^{th} persuasion strategy is present in the ad image. Our choice of using sigmoid over softmax is motivated by the fact that multiple persuasion strategies can be present simultaneously in an ad image. Consequently, the entire model is trained in an end-to-end manner using binary cross-entropy loss $\mathcal{L}_s$ over logit for each strategy:

$$\mathcal{L}_s = [-y_i \log(p_i) - (1 - y_i) \log(1 - p_i)] \quad (2)$$

where, y_i is 1 if i^{th} persuasion strategy is present in the ad and 0 otherwise. It can be observed in Table 2 that our model achieves an accuracy of 59.2%, where a correct match is considered if the strategy predicted by the model is present in the set of annotated strategies for a given ad. Further, we perform several ablations where we exclude each modality while retaining all the other modalities. We note that for each modality, excluding the modality results in a noticeable decrease in accuracy, with significant decreases observed when excluding DenseCap ($\sim 3.6\%$) and OCR ($\sim 4.4\%$). Further, we observe that using DenseCap for obtaining caption embeddings, and SSD for object detection works better than BLIP and YOLOv5, respectively (see Table 3). We also explore using focal loss (Lin et al. 2017) in place of cross-entropy loss to handle class imbalance but observed that it led to degradation instead of improvements (top-1 acc. of 56.4% *vs* 59.2% using cross-entropy).

4.4 Multi Task Learning

One of the key opportunities for our persuasion strategies data labeling and modeling task was the presence of additional labels already given in the base Pitts Ads dataset. In that, Hussain et al. (2017) had given labels about the reasoning task. For the reasoning task, the annotators were asked to provide answers in the form "I should [Action] because [Reason]." for each ad. In other words, they asked the annotators to describe *what the viewer should do and why*, according to the ad. Similar to the reasoning task, persuasion strategies provide various cognitive, behavioral, and affective reasons to try to elicit the motivation of the ad viewers towards their products or services. Therefore, we hypothesize that these natural language descriptions of *why the viewers should follow* the ad will be informative in inferring the ad's persuasion strategy.

We formulate obtaining action-reason statement as a sequence generation task where the model learns to generate a sentence $Y^g = (y_1^g, \ldots, y_T^g)$ of length T conditioned on advertisement X by generating the sequence of tokens present in the action-reason statement. To achieve this, we use a transformer decoder module that attends on the features $\text{Enc}(X)$ as shown in Fig. 5. The annotated action-reason statement is used to train the transformer decoder as an auxiliary task to strategy prediction through the standard teacher forcing technique used in Seq2Seq framework.

Please refer to Appendix for more architectural details about the action-reason generation branch. As shown in Table 2, generating action-reason as an auxiliary task improves the strategy prediction accuracy by 2.8%.

4.5 Active Learning

We use an active learning method to ease the large-scale label dependence when constructing the dataset. As in every active learning setting, our goal is to develop a learner that selects samples from unlabeled sets to be annotated by an oracle. Similar to traditional active learners (Gilad-Bachrach, Navot, and Tishby 2005; Lewis and Catlett 1994), we use uncertainty sampling to perform the sample selection. In doing so, such function learns to score the unlabeled samples based on the expected performance gain they are likely to produce if annotated and used to update the current version of the localization model being trained. To evaluate each learner, we measure the performance improvements, assessed on a labeled test set at different training dataset sizes.

At every learning step t, a set of labeled samples L_t is first used to train a model f_t. Then, from an unlabeled pool $U_t = D - L_t$, an image instance a is chosen by a selection function g. Afterwards, an oracle provides temporal ground-truth for the selected instance, and the labeled set L_t is augmented with this new annotation. This process repeats until the desired performance is reached or the set U_t is empty.

In our implementation, we instantiate the active learning selection function as the entropy of the probability distribution predicted by the model over the set of persuasion strategies for a given ad image instance a. Formally, $g = -\sum_{i=1}^{|\mathcal{P}|} p_i^n * \log(p_i^n)$, where p_i^n denotes the normalized probability with which i^{th} persuasion strategy is present in a as per the model prediction. The normalized probability p_i^n is estimated as $p_i / \sum_{j=1}^{|\mathcal{P}|} p_j$. Intuitively, ad samples with high entropy selection values indicate that the model trained on limited data has a higher degree of confusion while predicting the persuasion strategy since it is not decisively confident about predicting few strategies. Hence, we rank the unlabeled ad images in the decreasing order of difficulty according to the corresponding values of the entropy selection function and select the top-k ads in the subsequent batch for annotation followed by training. As shown in Fig. 6, we set k to be 250 and analyze the effect of incrementally introducing new samples selected through active learning. It can be seen that both top-1 and top-3 accuracy increases with the addition of new training data. We stop at the point when 2500 training samples are used since the model performs reasonably well with a top-1 and top-3 strategy prediction accuracy of 59.2% and 84.8% (see Fig. 6).

5 Case Study

As an interesting application of the model developed for predicting persuasion strategies, we analyze the advertising campaigns of 30 Fortune-500 companies across 8 sectors. We extract 1600 ads using the Facebook Ads API[7]. We predict the persuasion strategies for these advertisements and

---

[7]https://www.facebook.com/ads/library/api/

Models	Top-1 Acc.	Top-3 Acc.
Our Model	**59.2**	**84.8**
w/o DenseCap	55.6	80.8
w/o Symbol	58.8	81.6
w/o DenseCap & Symbol	55.2	80.8
w/o OCR	54.8	82
w/o Symbol, OCR & DenseCap	58	78.8
w/o Action-Reason Task	56.4	80.4
Random Guess	6.25	18.75

Table 2: Effect of different Modalities and Tasks on the accuracy and performance of the strategy prediction task.

Model Used	Top-1 Accuracy	Top-3 Accuracy	Recall
Model with DenseCap & SSD	59.2	84.8	74.59
Model with BLIP & YOLOv5	58.4	83.8	71.58

Table 3: Comparison of caption and object detection models. We noticed that BLIP while being more recent and trained on a larger dataset, generates more informatory captions for background objects which DenseCap successfully ignores.

correlate them with attributes returned by the API, such as demographic (age and sex), company sector, and marketing spend on the ad.
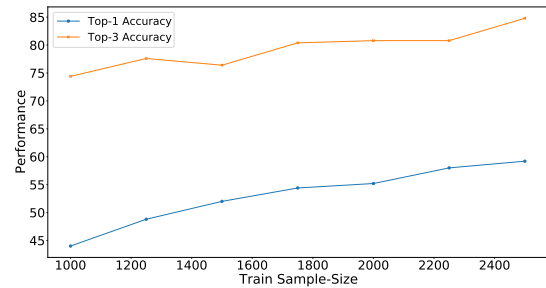

Figure 6: Incremental effect of introducing new data through active learning; Results for prediction of persuasion strategies on the test set

Sector: We find that the healthcare sector uses cheerfulness strategy, heavy industries use concreteness, the food sector uses eager, and apparel sector uses fashionable and active strategies (Appendix:Fig:3). Interestingly, the technology sector does not significantly correlate with any particular strategy. **Age:** At the same time, the strategies most shown to different age-groups are 13-17:Fashionable, 18-24 and 25-34: Authority, 35-44: Feminine, 45-54 and 55-64: Social Identity, and 65+: Trustworthiness (Appendix:Fig:4). **Gender:** Unsurprisingly, feminine strategy was mostly shown to females while "unknown" gender did not have significant correlation with anything. Interestingly, the authority strategy was shown the most to the male gender (Appendix:Fig:5).

6   Conclusion and Future Work

What does an advertisement say which makes people change their beliefs and actions? With limited works, the computational study of rhetoric of this all-pervasive form of marketing communication is still in its infancy. In this paper, based on the well-developed social psychology and marketing literature, we develop and release the largest vocabulary of persuasion strategies and labeled dataset. We develop a multi-task attention-fusion model for predicting strategies. We do a case study, correlated demographic and spend level information with strategies on 1600 Fortune-500 company ads. Future work can investigate how audience factors further affect persuasion strategies. Additionally, we would like to investigate what role do strategies play in viral marketing and how to generate advertisements given strategies and brands.

7   Acknowledgments

References

Althoff, T.; Danescu-Niculescu-Mizil, C.; and Jurafsky, D. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 12–21.

Anand, P.; King, J.; Boyd-Graber, J.; Wagner, E.; Martell, C.; Oard, D.; and Resnik, P. 2011. Believe me—we can do this! Annotating persuasive acts in blog text. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Aronson, E.; Turner, J. A.; and Carlsmith, J. M. 1963. Communicator credibility and communication discrepancy as determinants of opinion change. *Journal of Psychopathology and Clinical Science*, 67(1).

Bai, C.; Chen, H.; Kumar, S.; Leskovec, J.; and Subrahmanian, V. 2021. M2p2: Multimodal persuasion prediction using adaptive fusion. *IEEE Transactions on Multimedia*.

Bhattacharya, C. B.; and Sen, S. 2003. Consumer–company identification: A framework for understanding consumers' relationships with companies. *Journal of marketing*, 67(2): 76–88.

Brehm, J. W. 1966. A theory of psychological reactance.

Burger, J. M. 1999. The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and social psychology review*, 3(4): 303–325.

Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5).

Chen, J.; and Yang, D. 2021a. Weakly-Supervised Hierarchical Models for Predicting Persuasive Strategies in Good-faith Textual Requests. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14): 12648–12656.

Chen, J.; and Yang, D. 2021b. Weakly-supervised hierarchical models for predicting persuasive strategies in good-faith textual requests. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 12648–12656.

Cialdini, R. B.; and Cialdini, R. B. 2007. *Influence: The psychology of persuasion*, volume 55. Collins New York.

Cialdini, R. B.; and Goldstein, N. J. 2004. Social influence: Compliance and conformity. *Annual review of psychology*, 55(1): 591–621.

Clark, M. S. 1984. Record keeping in two types of relationships. *Journal of personality and social psychology*, 47(3).

Clark, M. S.; and Mills, J. 1979. Interpersonal attraction in exchange and communal relationships. *Journal of personality and social psychology*, 37(1).

Clark, M. S.; Mills, J.; and Powell, M. C. 1986. Keeping track of needs in communal and exchange relationships. *Journal of personality and social psychology*, 51(2): 333.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1): 37–46.

Deutsch, M.; and Gerard, H. B. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3): 629.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 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)*. Association for Computational Linguistics.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Durmus, E.; and Cardie, C. 2018. Exploring the Role of Prior Beliefs for Argument Persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1035–1045. New Orleans, Louisiana: Association for Computational Linguistics.

El Baff, R.; Wachsmuth, H.; Al Khatib, K.; and Stein, B. 2020. Analyzing the Persuasive Effect of Style in News Editorial Argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3154–3160. Online: Association for Computational Linguistics.

Epley, N.; Waytz, A.; and Cacioppo, J. T. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4): 864.

Fournier, S. 1998. Consumers and their brands: Developing relationship theory in consumer research. *Journal of consumer research*, 24(4): 343–373.

Frank, S.; Bugliarello, E.; and Elliott, D. 2021. Vision-and-Language or Vision-for-Language? On Cross-Modal Influence in Multimodal Transformers. *arXiv preprint arXiv:2109.04448*.

Freedman, J. L.; and Fraser, S. C. 1966. Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, 4(2): 195.

Furnham, A.; and Boo, H. C. 2011. A literature review of the anchoring effect. *The journal of socio-economics*, 40(1): 35–42.

Giffin, K. 1967. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological bulletin*, 68(2): 104.

Gilad-Bachrach, R.; Navot, A.; and Tishby, N. 2005. Query by committee made real. *Advances in neural information processing systems*, 18.

Habernal, I.; and Gurevych, I. 2016. What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1214–1223.

He, H.; Chen, D.; Balakrishnan, A.; and Liang, P. 2018. Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637*.

Hibbert, S.; Smith, A.; Davies, A.; and Ireland, F. 2007. Guilt appeals: Persuasion knowledge and charitable giving. *Psychology & Marketing*, 24(8): 723–742.

Hidey, C.; Musi, E.; Hwang, A.; Muresan, S.; and McKeown, K. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, 11–21.

Hovland, C. I.; Janis, I. L.; and Kelley, H. H. 1953. Communication and persuasion.

Hussain, Z.; Zhang, M.; Zhang, X.; Ye, K.; Thomas, C.; Agha, Z.; Ong, N.; and Kovashka, A. 2017. Automatic understanding of image and video advertisements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1705–1715.

Iyer, R. R.; and Sycara, K. 2019. An unsupervised domain-independent framework for automated detection of persuasion tactics in text. *arXiv preprint arXiv:1912.06745*.

Joo, J.; Li, W.; Steen, F. F.; and Zhu, S.-C. 2014. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 216–223.

Joo, J.; Steen, F. F.; and Zhu, S.-C. 2015. Automated facial trait judgment and election outcome prediction: Social dimensions of face. In *Proceedings of the IEEE international conference on computer vision*, 3712–3720.

Keller, P. A.; Lipkus, I. M.; and Rimer, B. K. 2003. Affect, framing, and persuasion. *Journal of Marketing Research*, 40(1): 54–64.

Knowles, E. S.; and Linn, J. A. 2004. *Resistance and persuasion*. Psychology Press.

Lasswell, H. D. 1971. *Propaganda technique in world war I*. MIT press.

Lee, A. Y.; Keller, P. A.; and Sternthal, B. 2010. Value from regulatory construal fit: The persuasive impact of fit between consumer goals and message concreteness. *Journal of Consumer Research*, 36(5): 735–747.

Levesque, N.; and Pons, F. 2020. The Human Brand: A systematic literature review and research agenda. *Journal of Customer Behaviour*, 19(2): 143–174.

Lewis, D. D.; and Catlett, J. 1994. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, 148–156. Elsevier.

Li, J.; Durmus, E.; and Cardie, C. 2020. Exploring the Role of Argument Structure in Online Debate Persuasion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8905–8912. Online: Association for Computational Linguistics.

Li, J.; Li, D.; Xiong, C.; and Hoi, S. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *ICML*.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755. Springer.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; and Berg, A. C. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*, 21–37. Springer.

Longpre, L.; Durmus, E.; and Cardie, C. 2019. Persuasion of the Undecided: Language vs. the Listener. In *Proceedings of the 6th Workshop on Argument Mining*.

Lukin, S.; Anand, P.; Walker, M.; and Whittaker, S. 2017. Argument Strength is in the Eye of the Beholder: Audience Effects in Persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 742–753. Valencia, Spain: Association for Computational Linguistics.

Luu, K.; Tan, C.; and Smith, N. A. 2019. Measuring online debaters' persuasive skill from text over time. *Transactions of the Association for Computational Linguistics*, 7: 537–550.

Lynn, M. 1991. Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology & Marketing*, 8(1): 43–57.

McGinnies, E.; and Ward, C. D. 1980. Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, 6(3): 467–472.

McGuire, W. J. 1964. Inducing resistance to persuasion. Some contemporary approaches. *CC Haaland and WO Kaelber (Eds.), Self and Society. An Anthology of Readings, Lexington, Mass.(Ginn Custom Publishing) 1981, pp. 192-230.*

McGuire, W. J.; and Papageorgis, D. 1961. The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal Psychology*, 62(2).

Meyers-Levy, J.; and Malaviya, P. 1999. Consumers' processing of persuasive advertisements: An integrative framework of persuasion theories. *Journal of marketing*, 63(4_suppl1): 45–60.

Milgram, S. 1963. Behavioral study of obedience. *The Journal of abnormal and social psychology*, 67(4): 371.

Milgram, S.; and Gudehus, C. 1978. Obedience to authority.

Petty, R. E.; and Cacioppo, J. T. 1986. The elaboration likelihood model of persuasion. In *Communication and persuasion*, 1–24. Springer.

Petty, R. E.; Cacioppo, J. T.; and Schumann, D. 1983. Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of consumer research*, 10(2): 135–146.

Petty, R. E.; Wegener, D. T.; and Fabrigar, L. R. 1997. Attitudes and attitude change. *Annual review of psychology*, 48(1): 609–647.

Raifer, M.; Rotman, G.; Apel, R.; Tennenholtz, M.; and Reichart, R. 2022. Designing an Automatic Agent for Repeated Language–based Persuasion Games. *Transactions of the Association for Computational Linguistics*, 10: 307–324.

Rapp, C. 2002. Aristotle's rhetoric.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Regan, D. T. 1971. Effects of a favor and liking on compliance. *Journal of experimental social psychology*, 7(6): 627–639.

Rosenthal, S.; and Mckeown, K. 2017. Detecting influencers in multiple online genres. *ACM Transactions on Internet Technology (TOIT)*, 17(2): 1–22.

Rothman, A. J.; Martino, S. C.; Bedell, B. T.; Detweiler, J. B.; and Salovey, P. 1999. The systematic influence of gain-and loss-framed messages on interest in and use of different types of health behavior. *Personality and Social Psychology Bulletin*, 25(11): 1355–1369.

Shaikh, O.; Chen, J.; Saad-Falcon, J.; Chau, P.; and Yang, D. 2020. Examining the Ordering of Rhetorical Strategies in Persuasive Requests. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1299–1306. Online: Association for Computational Linguistics.

Siddiquie, B.; Chisholm, D.; and Divakaran, A. 2015. Exploiting multimodal affect and semantics to identify politically persuasive web videos. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 203–210.

Stab, C.; and Gurevych, I. 2014. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*, 1501–1510.

Strack, F.; and Mussweiler, T. 1997. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of personality and social psychology*, 73(3): 437.

Tan, C.; Niculae, V.; Danescu-Niculescu-Mizil, C.; and Lee, L. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, 613–624.

Tversky, A.; and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157): 1124–1131.

Tversky, A.; and Kahneman, D. 1985. The framing of decisions and the psychology of choice. In *Behavioral decision making*, 25–41. Springer.

Vakratsas, D.; and Ambler, T. 1999. How advertising works: what do we really know? *Journal of marketing*, 63(1): 26–43.

Vargheese, J. P.; Collinson, M.; and Masthoff, J. 2020. Exploring susceptibility measures to persuasion. In *International Conference on Persuasive Technology*, 16–29. Springer.

Wachsmuth, H.; Naderi, N.; Hou, Y.; Bilu, Y.; Prabhakaran, V.; Thijm, T. A.; Hirst, G.; and Stein, B. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 176–187.

Wang, X.; Shi, W.; Kim, R.; Oh, Y.; Yang, S.; Zhang, J.; and Yu, Z. 2019. Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5635–5649. Florence, Italy: Association for Computational Linguistics.

Wegener, D. T.; Petty, R. E.; Detweiler-Bedell, B. T.; and Jarvis, W. B. G. 2001. Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, 37(1): 62–69.

Wei, Z.; Liu, Y.; and Li, Y. 2016. Is this post persuasive? Ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 195–200.

Wood, W. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology*, 51(1): 539–570.

Yang, D.; Chen, J.; Yang, Z.; Jurafsky, D.; and Hovy, E. 2019. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. 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)*, 3620–3630.

Yang, L.; Tang, K.; Yang, J.; and Li, L.-J. 2017. Dense captioning with joint inference and visual context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2193–2202.

Zhang, A. X.; Culbertson, B.; and Paritosh, P. 2017. Characterizing online discussion using coarse discourse sequences. In *Eleventh International AAAI Conference on Web and Social Media*.

Zhang, F.; Litman, D.; and Forbes-Riley, K. 2016. Inferring discourse relations from pdtb-style discourse labels for argumentative revision classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2615–2624.

Appendix

1 Definitions

Dice Coefficient : $2 * |X \cap Y| / (|X| + |Y|)$,
where X and Y are two sets; a set with vertical bars on either side refers to the cardinality of the set, i.e. the number of elements in that set; and $\cap$ refers to the intersection of two sets.

Top-1 Accuracy : It is defined as the fraction of images, where the highest predicted strategy is present in the ground-truth strategies.

Top-3 Accuracy : It is defined as the fraction of images, where any of the top-3 highest predicted strategies is present in the ground-truth strategies.

	#ads with 1 strategy	#ads with 2 strategies	#ads with 3 strategies	Avg. strategies	Std. Dev.
Train-Set	1440	905	155	1.486	0.612
Val-Set	132	98	20	1.552	0.639
Test-Set	147	93	10	1.452	0.574
Total	1719	1096	185	1.49	0.592

Table 1: Distribution of test, train, validation, and the total dataset

Figure 1: Various rhetoric strategies used in advertisements
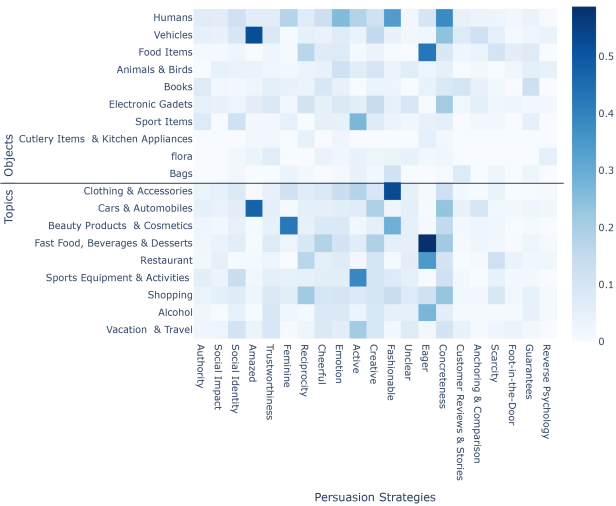
Figure 2: Dice correlation between topics and strategies. Topics are taken from the Pitts Ad dataset and further similar topics are combined to get these values.

Figure 3: Dice correlation between company sectors and strategies. Companies are taken from the Meta advertisement data and further these companies are labeled according to which sector they belong.

Figure 4: correlation between age-group and top 6 strategies on the basis of frequency. Age-groups are taken from the Meta advertisement data.
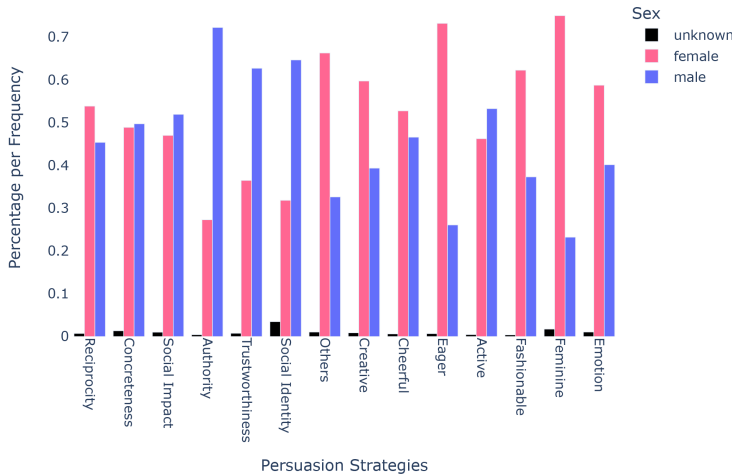
Figure 5: correlation between Gender and strategies.
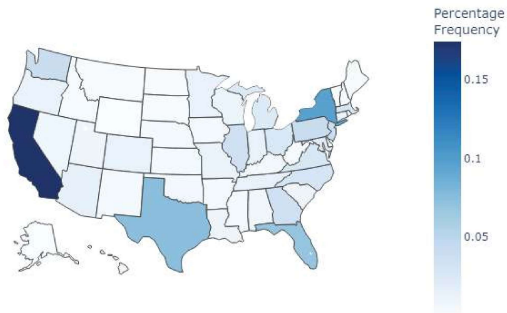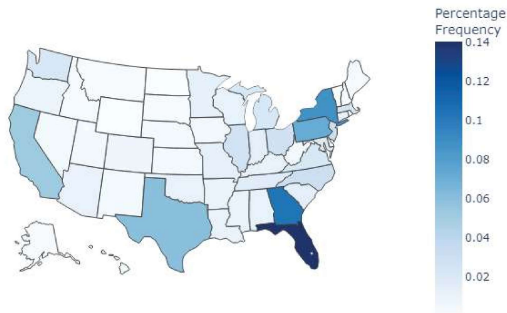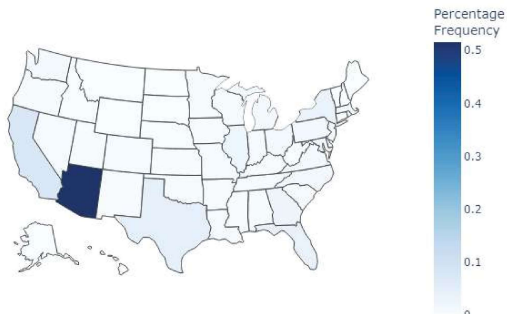
(a) US Heat Maps

Persuasion Strategy: Creative
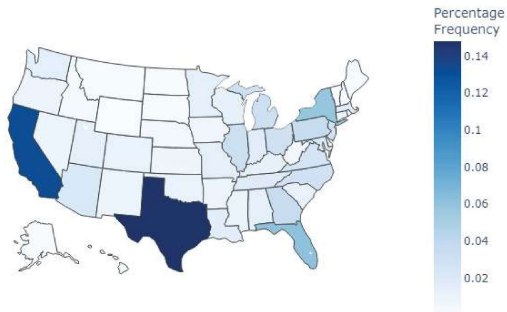
Persuasion Strategy: Concreteness
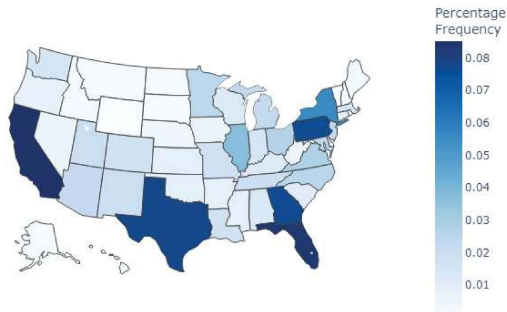
Persuasion Strategy: Authority
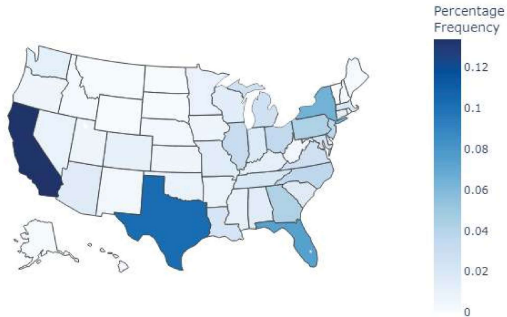
Persuasion Strategy: Active
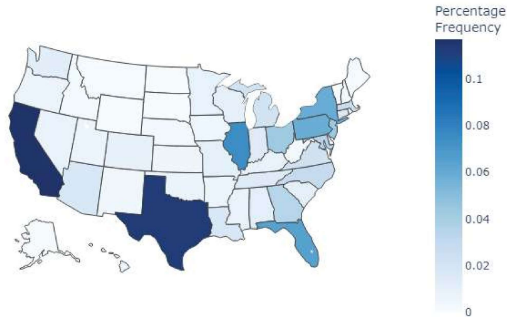
Persuasion Strategy: Cheerful
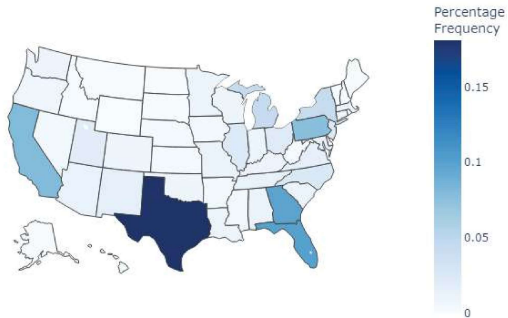
(b) US Heat-Maps
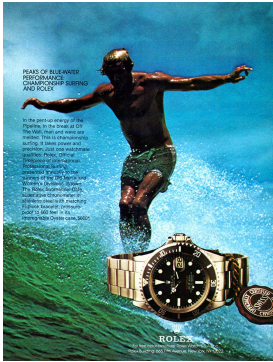
Figure 7: Advertisements containing humans and concreteness

Figure 8: Some samples from the Pitts Ads dataset along with the ground truth and predicted action-reason statement, topic and sentiment.
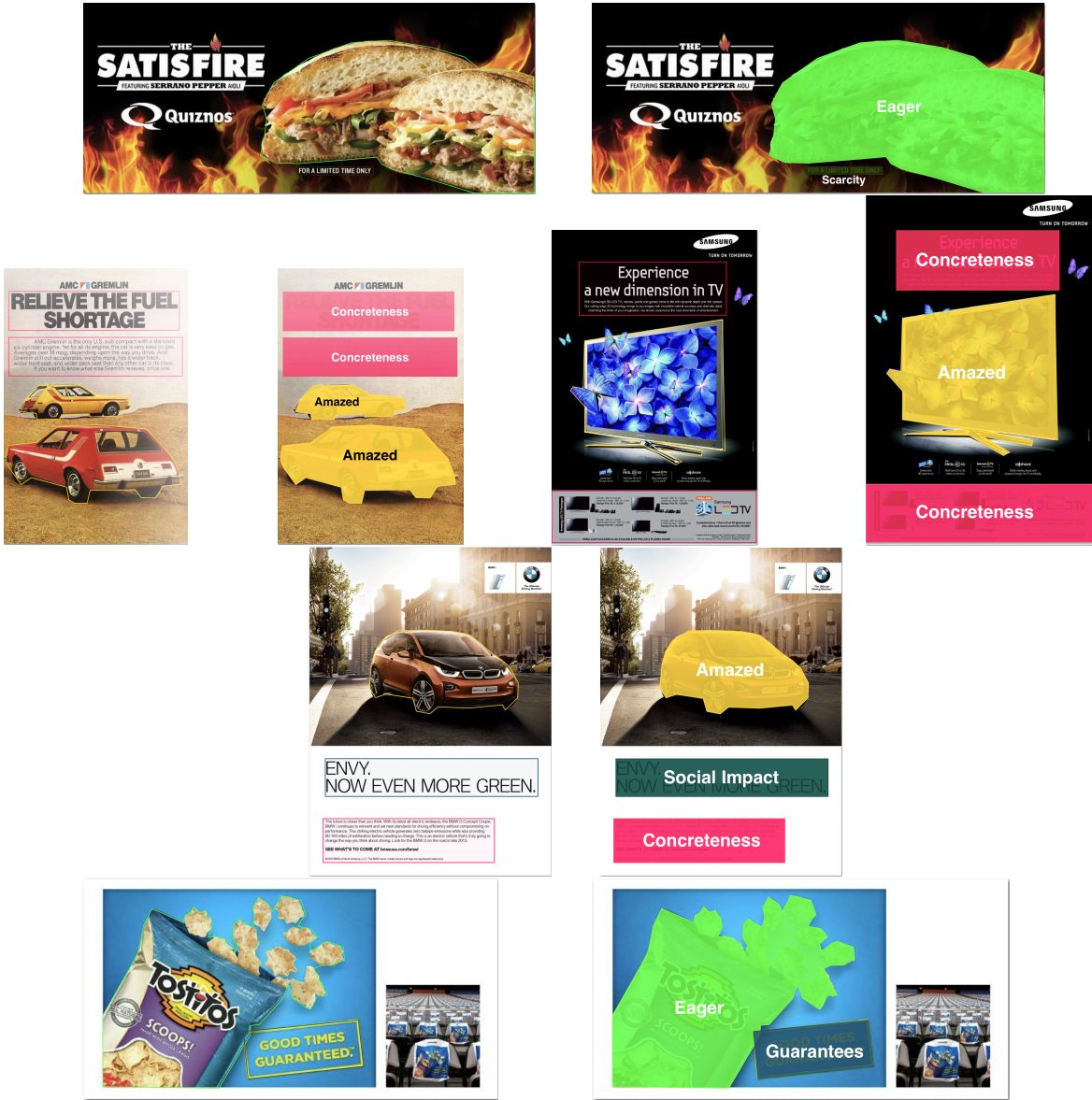
Figure 9: Image with a segmentation mask depicting the strategies *Emotion:Cheerful*, *Emotion:Eager* and *Trustworthiness*.