Automatic Understanding of Image and Video Advertisements (Supplementary File)

Zaeem Hussain Mingda Zhang Xiaozhong Zhang Keren Ye
Christopher Thomas Zuha Agha Nathan Ong Adriana Kovashka
Department of Computer Science
University of Pittsburgh
{zaeem, mzhang, xiaozhong, yekeren, chris, zua2, nro5, kovashka}@cs.pitt.edu

In this document, we include more information and statistics about our collected data, for both the image and video datasets. We also include additional experimental results and method explanations (from Sec. 15 onward).

1. Keywords used for image search

We compiled the following list of keywords, and used it to search for image advertisements. Each query string contained a keyword from this list followed by either the word 'ads' or 'advertisements'.

– Food:
- Restaurants
  McDonald's, Burger King, KFC, Wendy's, Five Guys Famous Burgers and Fries, Whataburger, In-N-Out Burger, Carl's Jr., Hardee's, Jack-in-the-Box, White Castle, Arby's, Chick-fil-A, Popeyes Chicken & Biscuits, Dunkin' Donuts, Krispy Kreme, Tim Hortons, Qdoba, Chipotle, Baja Fresh, Taco Bell, El Pollo Loco, Bruegger's Bagels, Panera Bread, Au Bon Pan, Cinnabon, Auntie Anne's, Quizno's Classic Subs, Subway, Jimmy John's, Pizza Hut, Dominos, Papa John's, Little Caesars, Boston Market, Sonic Drive-In, Long John Silver's, Sbarro, Panda Express, Applebee's
- Ice cream
  Dairy Queen, Baskin-Robin's, TCBY, Ben & Jerry's, Cold Stone Creamery, Blue Bell, Haagen-Dazs, Breyers, Klondike, Drumstick, Skinny Cow
- Chocolate
  3 Musketeers, 100 Grand Bar, Aero, Almond Joy, Baby Ruth, Butterfinger, Clark Bar, Nestle Crunch, Dove Bar, Heath bar, Hershey bar, KitKat, Krackel, Lindor, Mars Bar, Milky Way, Oh Henry!, PayDay, Reese's Peanut Butter Cup, Rolo, York Peppermint Pattie, Sky Bar, Snickers, Take 5, Toblerone, Twix, Whatchamacallit, Wonka Bar, Mars, Hershey, Cadbury, Nestle, Necco
- Cookies
  Chips Ahoy!, Girl Scout Cookies, Pepperidge Farms, Oreo, Nilla, Nabisco, Keebler, Wheat Thins, Triscuits, Saltines
- Chips
  Lays, Pringles, Doritos, Cheetos, Tostitos
- Gum
  Wrigley, Fruit Stripe, Bubble Yum, Juicy Fruit, Chiclets, Trident, Bazooka
- Nuts
  Planters, Wonderful Pistachios, Emerald
- Condiments
  ketchup, mustard, Heinz, mayonnaise

– Drinks:
- Soda
  Coca Cola, Pepsi, Mountain Dew, A&W, Mug Root Beer, Crush, Fanta, Sprite, 7-Up, Canada Dry, RC Cola

- Alcohol
  vodka, Absolut, whiskey, wine, beer, Coors Lite, Miller Lite, BRB
- Water
  Aquafina, Evian, Perrier, Dasani, Nestle Waters, Deer Park Natural Spring Water
- Coffee
  Eight O'Clock, Starbucks, Maxwell House, Folgers, Keurig
- Energy drinks
  Monster, Red Bull, Four Loko, Five-Hour Energy
- Juice
  Minute Maid, Florida's Natural, Sunny D, Capri Sun
- Chocolate drinks
  Milo, Ovaltine, Nesquik

– Cars:
  Nissan, Kia, Audi, Subaru, Honda, Chevrolet, Porsche, Toyota, Ford, Rolls Royce, Mitsubishi, Hyundai, Oldsmobile, Jaguar, Volvo, Mercedes Benz, General Motors, Mazda, BMW, Volkswagen, Hummer, Tesla, Lincoln, Jeep, Land Rover

– Electronics:
- Phones
  Samsung, LG, iPhone, Motorolla, Ericsson, Pantech, Nokia, Google Nexus, BlackBerry, HTC, Siemens, Palm Pilot, Alcatel
- Service providers
  Comcast, Dish, DirecTV, Google Fiber, Time Warner Cable, Verizon, T-mobile, Sprint, Cingular, AT&T, Nextel, Bell South, Pacific Bell, Virgin Mobile, Boost Mobile, Cricket Wireless
- Computers
  Dell, HP, Acer, Asus, Lenovo, Sony VAIO, Packard-Bell, Gateway, IBM, Apple, Compaq, Wang Laboratories, Microsoft
- Televisions
  Sony, LG, Panasonic, Sanyo, Itachi, Samsung, RCA, Vizio, Toshiba, Sharp, Magnavox, Westinghouse, JVC, General Electric

– Financial institutions:
- Insurance
  Nationwide, Farmer's, Northwestern Mutual, Prudential Insurance Company of America, Progressive Insurance, E-surance, MetLife, Highmark, AETNA, United American Insurance Company, UnitedHealthcare, Delta Dental, Allstate, GEICO, Wells Fargo
- Banks
  CITIbank, Bank of America, Wells Fargo, Capital One, First Niagara, PNC Bank, BNY Mellon, HSBC, USAA

– Travel:
- Airlines
  United, American, Delta, Frontier, Southwest, Spirit, Etihad, Emirates, Singapore, Thai, Qatar, Turkish
- Vacation, toursim, resort, cruise, car rentals, train

– Sports:
- Equipment
  football, soccer, baseball, golf, basketball, hockey, ski, surfing, watersports, sailing, snowboard, ice skating, gymnastics, bowling, curling, volleyball, tennis, squash, swimming, diving, cross country, cross country skiing, cricket, marathon, triathalon, rowing, kayaking, fencing, martial arts, ping pong/table tennis, badminton, equestrian, polo, water polo, shooting, archery, cycling, speed skating, track and field

– Cosmetics:
  Estee Lauder, Maybelline, Cover Girl, L'Oreal, Neutrogena, Oil of Olay, Physician's Formula, Avon, Burt's Bees, Lancome, Chanel, Clinique, Almay, Benefit, Nars, Urban Decay, Dior, Iman, Dermablend, ShiSeido, Revlon, Max Factor, Kiehl's, Armani, Laura Mercier, Bobbi Brown, Clarins, Givenchy, Sephora, Elizabeth Arden, The Body Shop, Mac, Origins, Nivea, Smashbox, Bareminerals, Stila

– Clothing:

Levi's, Lucky, Madewell, J. Crew, Gap, American Eagle, Bebe, Loft, Ann Taylor, Tommy Hilfiger, Ralph Lauren, The Limited, LaCoste, True Religion, Kate Spade, Tory Birch, BCBG, Land's End, LL. Bean, Talbots, Lane Bryant, Calvin Klein, Anne Klein, Nike, Reebok, Under Armor, Children's Palace, Gymboree, Carter's, Coach(they have clothing), Burberry, Guess, Wrangler, Vera Wang, Lee, Adidas, Zara, Uniqlo, Columbia Sportswear, North Face, Converse, New Balance, Eddie Bauer, Van Heusen, Fubu, Kenneth Cole, Ecko, Swatch, Prada, Speedo, Versace, Hugo Boss, Gucci, Chanel, Gloria Vanderbilt, Izod, Fruit of the Loom, Hanes, Jockey, Puma, Abercrombie and Fitch, Old Navy, H and M, Urban Outfitters, Converse, Armani, Brooks Brothers, J.S. Bank Clothiers, Sean John, Victoria's Secret, DKNY, Aerostaple, Liz Claiborne, Arizona, Hollister, Diesel, Timberland, Jessica Simpson, Banana Republic, Fila, Petite Sophisticates, Cache, Delia's, Esprit, Club Monaco, Burton, Osh Kosh Bgosh, LuluLemon, Athleta, Eileen Fisher, Maidenform, Eileen West, Bally's

– Publics service announcements (PSA's):
  - Environment, nature, animal rights, PETA
  - Domestic violence, human rights
  - Safe driving, safety
  - Self-esteem, bullying, cuberbullying
  - Smoking, healthcare

2. Examples of advertisement and non-adverstisement images

Here are the instructions that were given to MTurkers for selecting whether an image was an advertisement or not.

– In this task, you will answer the question: "Is this image an advertisement? You should answer yes if you think this image could appear in print media as an advertisement WITHOUT any changes or additions to the image."

– The image can be an advertisement if it is advertising a commercial product/service or if it is conveying a public service message. To answer yes, the image must be a self contained advertisement. In other words, the image should contain a direct or implied persuasive message in visual or text form. Hence, in this task, 'stills' from video commercials, images that may be cropped from advertisements, personal photographs (for personal ads or otherwise), and pictures of products with no accompanying advertisement flair, are not considered advertisements. For example, a well-photographed picture of a car without any accompanying visual or textual message for persuading a reader to buy that car should be given the answer "no."

– If the image is too small or blurry for you to make out what the content of the image is, select "no." Do not select "no" solely on the basis that the text of the advertisement is unreadable.

– For most of the images in this question, the correct answer should be obvious. There is no need to attempt to infer whether or not a given image could be a part of an advertisement; it either is or is not an advertisement.

Fig. 1 shows some images, from the pool of noisy images collected, which were shown as examples to MTurk workers for completing the above task. Fig. 2 shows examples of ad/not ad annotations by the workers.

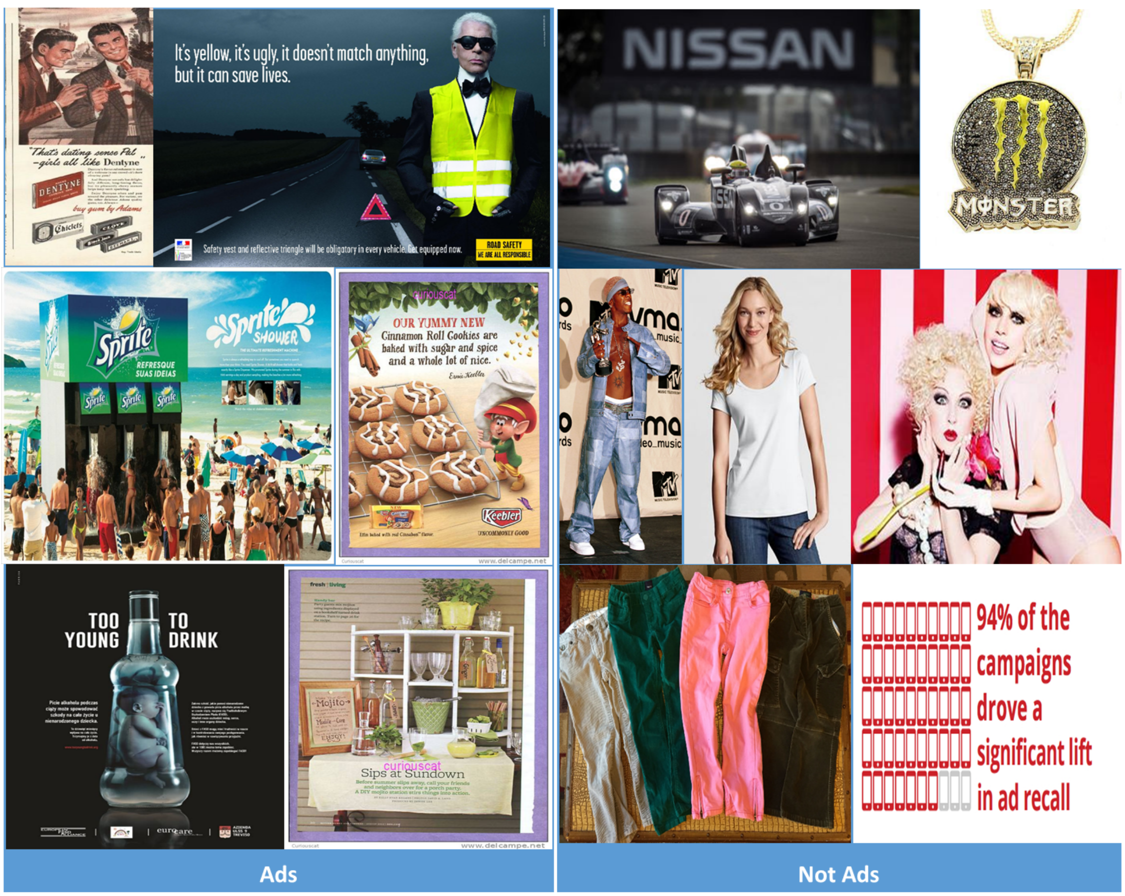


Figure 1. Some of the examples of 'ads' and 'not ads' shown to MTurkers.

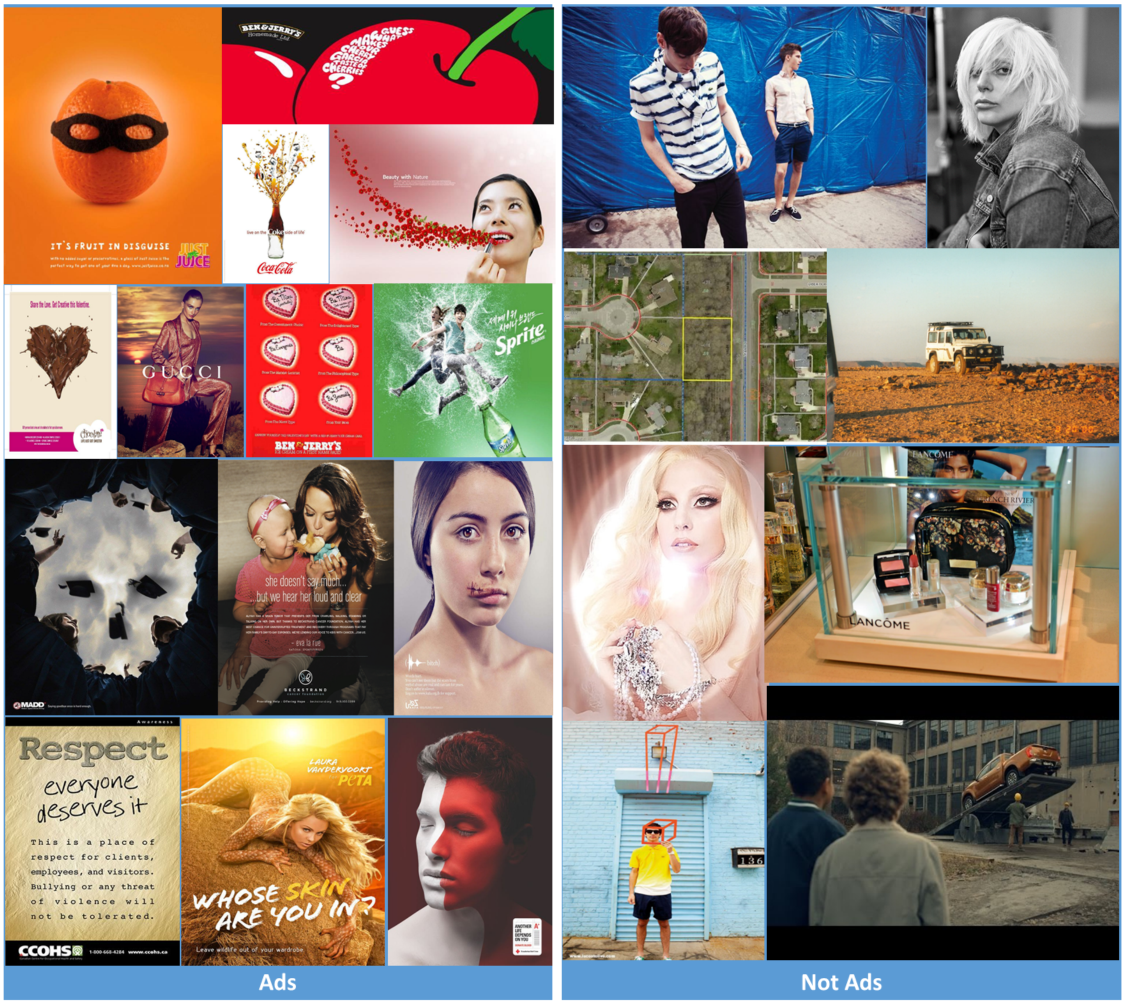


Figure 2. Some examples of 'ads' and 'not ads' as selected by MTurkers.

We labeled images as ads/not ads using a two-stage process, as described in the main text. In the first phase with 21,945 noisy images, we gave the workers only two options (ads vs. not an ad) per image. Each image was annotated by 4 workers. If at least 3 workers voted that the image was an ad, we would consider the image to be an ad. In the larger study with about 63,000 images from the ResNet, we presented the workers with 3 options: not an ad, straightforward ad, or an ad that requires non-literal interpretation. This time the number of workers per image was 5. The image was considered not an ad if at least 3 workers selected the first option. Otherwise, the more common of the latter two options was selected, and the image was considered an ad. If there was a tie between the last two options, with 2 votes each, the image was considered to be an ad that required non-literal interpretation (i.e. symbolism).

3. Topics of advertisement images

Below is the complete list of the topics of advertisements which MTurkers were asked to choose from, for each ad image. They were also given a final 'Other' option in which they were asked to write the topic of the ad, if it was not already present in the provided list.

- Restaurants, cafe, fast food
- Chocolate, cookies, candy, ice cream
- Chips, snacks, nuts, fruit, gum, cereal, yogurt, soups
- Seasoning, condiments, ketchup
- Pet food
- Alcohol
- Coffee, tea
- Soda, juice, milk, energy drinks, water
- Cars, automobiles (car sales, auto parts, car insurance, car repair, gas, motor oil, etc.)
- Electronics (computers, laptops, tablets, cellphones, TVs, etc.)
- Phone, TV and internet service providers
- Financial services (banks, credit cards, investment firms, etc.)
- Education (universities, colleges, kindergarten, online degrees, etc.)
- Security and safety services (anti-theft, safety courses, etc.)
- Software (internet radio, streaming, job search website, grammar correction, travel planning, etc.)
- Other services (dating, tax, legal, loan, religious, printing, catering, etc.)
- Beauty products and cosmetics (deodorants, toothpaste, makeup, hair products, laser hair removal, etc.)
- Healthcare and medications (hospitals, health insurance, allergy, cold remedy, home tests, vitamins)
- Clothing and accessories (jeans, shoes, eye glasses, handbags, watches, jewelry)
- Baby products (baby food, sippy cups, diapers, etc.)
- Games and toys (including video and mobile games)
- Cleaning products (detergents, fabric softeners, soap, tissues, paper towels, etc.)
- Home improvements and repairs (furniture, decoration, lawn care, plumbing, etc.)
- Home appliances (coffee makers, dishwashers, cookware, vacuum cleaners, heaters, music players, etc.)
- Vacation and travel (airlines, cruises, theme parks, hotels, travel agents, etc.)
- Media and arts (TV shows, movies, musicals, books, audio books, etc.)
- Sports equipment and activities
- Shopping (department stores, drug stores, groceries, etc.)
- Gambling (lotteries, casinos, etc.)
- Environment, nature, pollution, wildlife
- Animal rights, animal abuse
- Human rights
- Safety, safe driving, fire safety
- Smoking, alcohol abuse
- Domestic violence
- Self esteem, bullying, cyber bullying
- Political candidates (support or opposition)
- Charities
- Unclear
- Other:

We got a total of 64,340 images annotated. Out of those, 5,660 images were annotated by 5 different workers, while the rest were annotated by 3 workers each. The final topic for an image was chosen by taking the majority label out of 3 (or 5) votes. In case of ties, which occurred in 7.4% images, a topic was chosen randomly from the available votes. We computed the agreement with the majority vote for each image, by dividing the number of majority vote annotations by the total number of annotations for each image. The average of this over the 64,340 images was 85.2%.

Fig. 3 shows the distribution of topics. We see that most common are clothing ads, automobile ads, and beauty ads.

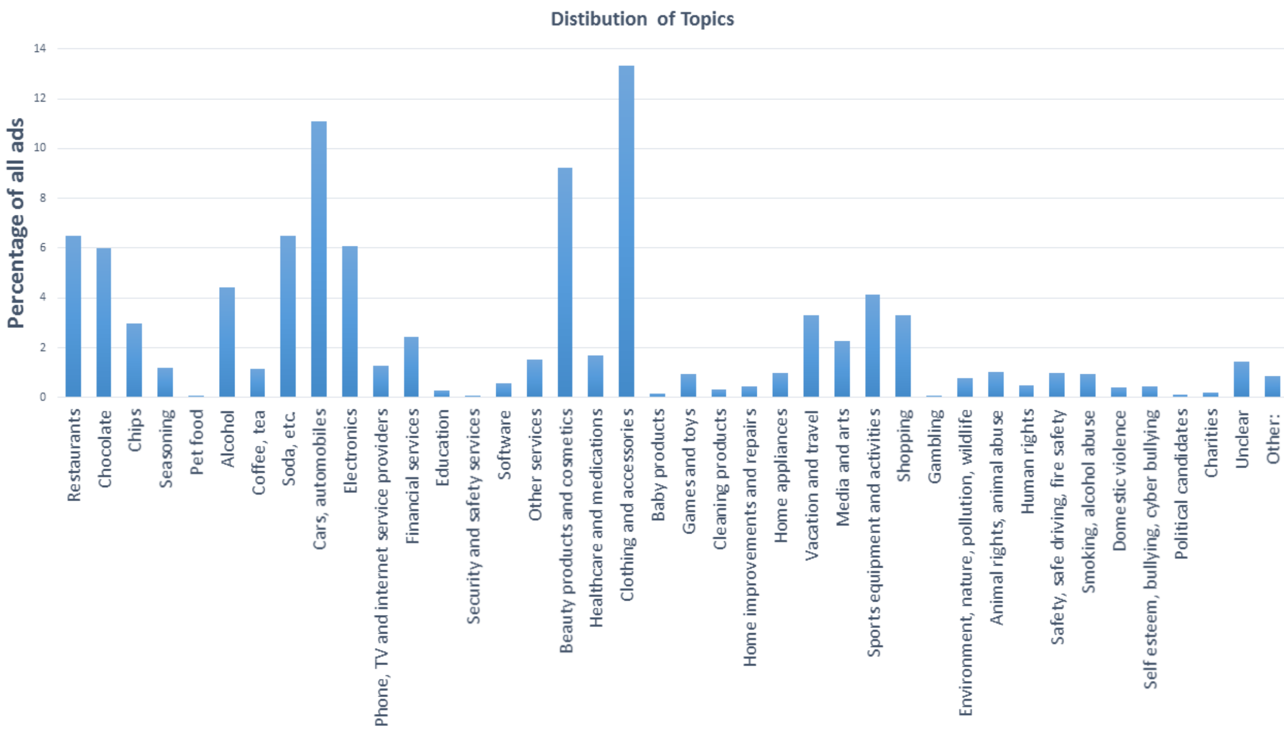


Figure 3. Frequency of each topic as a fraction of all ads.

4. Sentiments of advertisement images

Below is the complete list of the sentiments of advertisements which MTurkers were asked to choose from, for each ad image. MTurkers were allowed to select multiple sentiments for a single image.

- **Active** (energetic, adventurous, vibrant, enthusiastic, playful)
- **Afraid** (horrified, scared, fearful)
- **Alarmed** (concerned, worried, anxious, overwhelmed)
- **Alert** (attentive, curious)
- **Amazed** (surprised, astonished, awed, fascinated, intrigued)
- **Amused** (humored, laughing)
- **Angry** (annoyed, irritated)
- **Calm** (soothed, peaceful, comforted, fullfilled, cozy)
- **Cheerful** (delighted, happy, joyful, carefree, optimistic)
- **Confident** (assured, strong, healthy)
- **Conscious** (aware, thoughtful, prepared)
- **Creative** (inventive, productive)
- **Disturbed** (disgusted, shocked)
- **Eager** (hungry, thirsty, passionate)
- **Educated** (informed, enlightened, smart, savvy, intelligent)
- **Emotional** (vulnerable, moved, nostalgic, reminiscent)
- **Empathetic** (sympathetic, supportive, understanding, receptive)
- **Fashionable** (trendy, elegant, beautiful, attractive, sexy)
- **Feminine** (womanly, girlish)
- **Grateful** (thankful)
- **Inspired** (motivated, ambitious, empowered, hopeful, determined)
- **Jealous**
- **Loving** (loved, romantic)
- **Manly**
- **Persuaded** (impressed, enchanted, immersed)
- **Pessimistic** (skeptical)
- **Proud** (patriotic)
- **Sad** (depressed, upset, betrayed, distant)
- **Thrifty** (frugal)
- **Youthful** (childlike)

We got a total of 30,340 images annotated, out of which 5,660 were annotated by 5 workers per image, while the rest were annotated by 3 workers each. For images that were annotated by 5 workers, any sentiment with at least 2 votes was considered, and for images which were annotated by 3 workers each, every vote for a sentiment was considered sufficient to indicate the presence of that sentiment in that ad.

Fig. 4 shows the distribution of some popular topics on the more common sentiments, based on 24,660 images. We see that beauty product ads evoke the fashionable and feminine sentiments, while cars, alcohol, and sports ads inspire manly sentiments. We also see that PSAs across the board inspire more of the following sentiments than commercial product ads: alarm, alertness, anger, consciousness, disturbance, feeling educated, emotion, empathy and sadness. The active sentiment is predominant in sports ads, environment ads inspire consciousness, while restaurant and alcohol ads inspire eagerness. Domestic abuse ads are most emotional and inspire the most sadness.

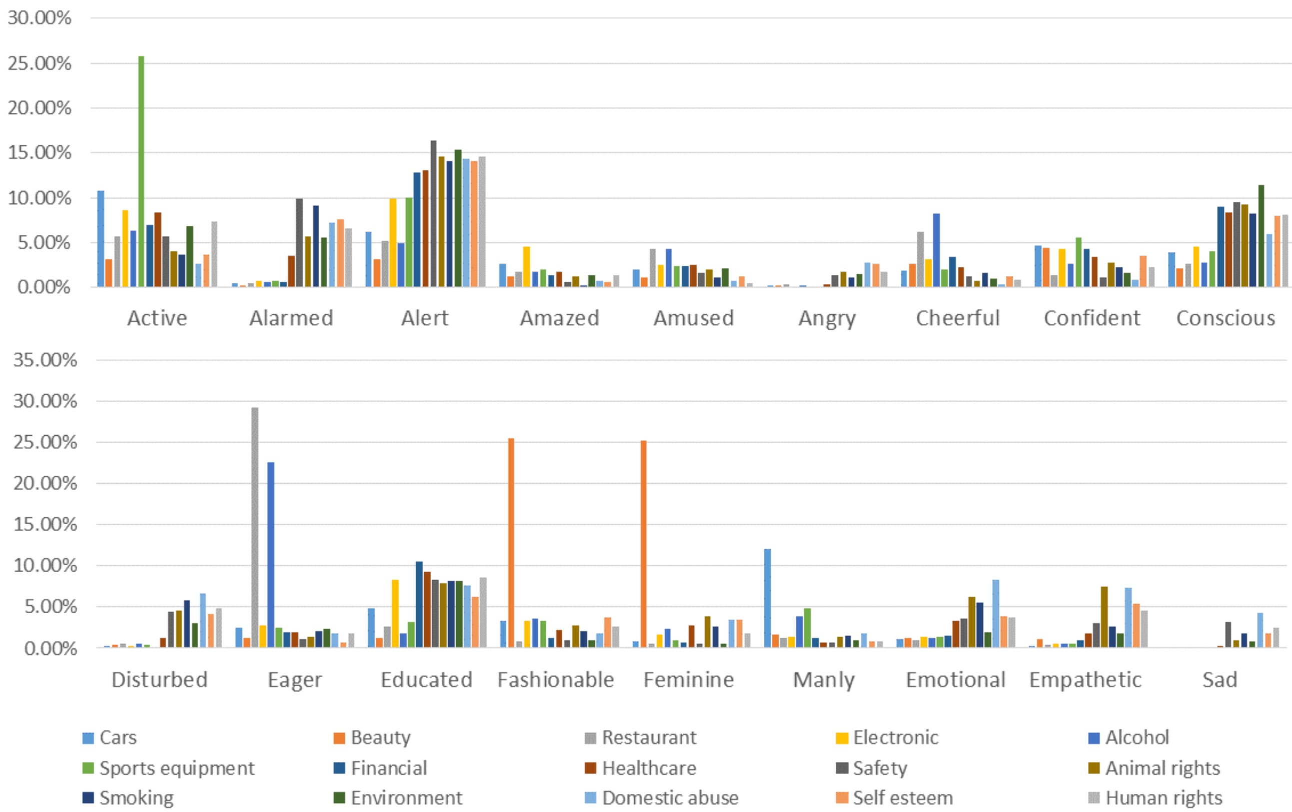


Figure 4. Distribution of several topics over common sentiments.

5. Question-answer examples for images

Fig. 5 shows the responses to the ”What should you do, according to this ad?” and ”Why, according to this ad, should you take this action?” questions.

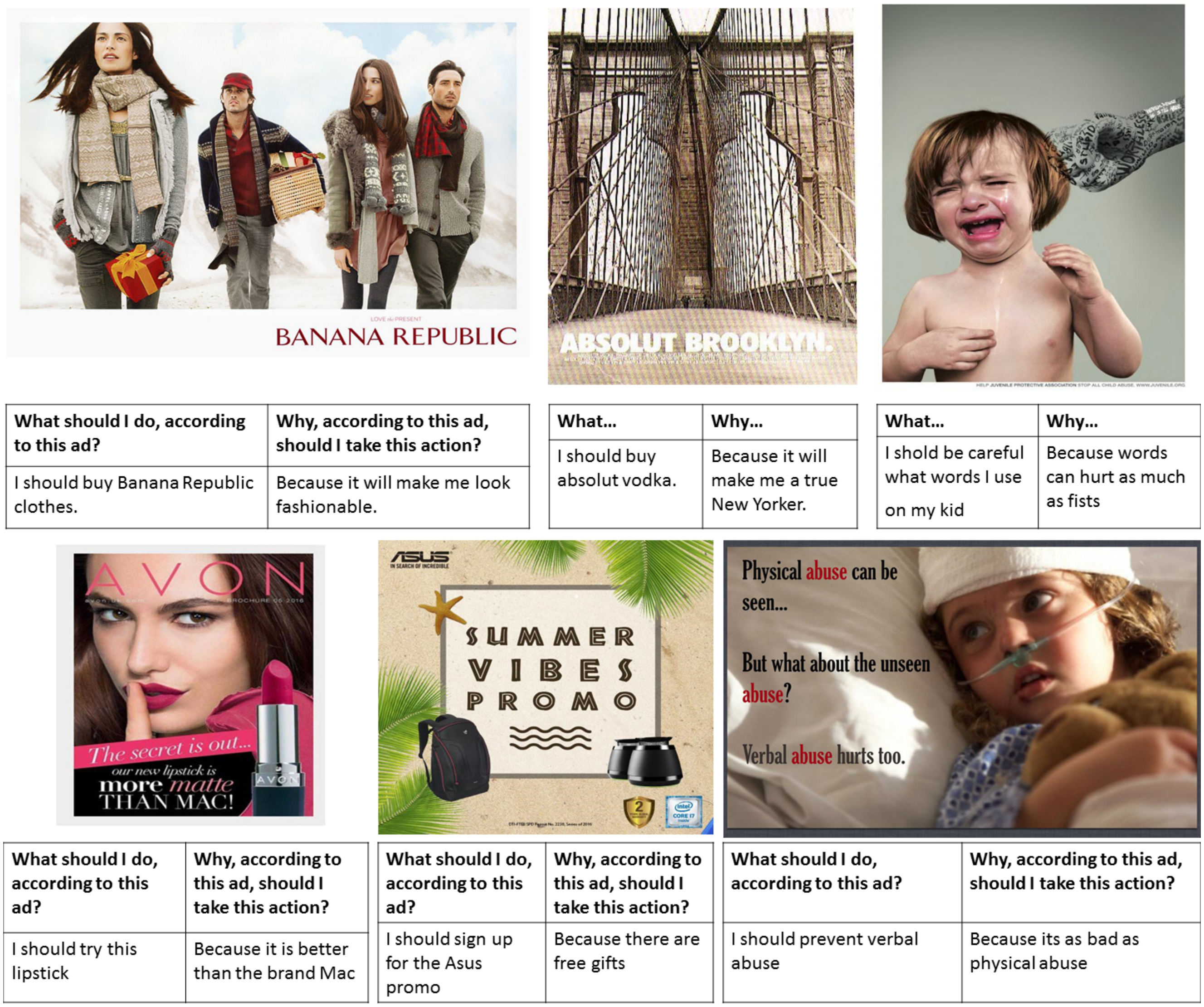


Figure 5. Some images with their questions and answers.

Tab. 1 shows the common words occurring in response to the “What should I do...” and “Why should I...” questions for 35 of the 38 topics. The common words for education, travel and smoking ads are shown in the main paper. We see that overall the words “buy,” “want” and “go” are common. However, we also see some topic specific words such as “drink” and “refreshing” for beverages, “play” and “fun” for games and toys, “support” and “kids” for charities, etc.

Restaurants, etc.		Chocolate, etc.		Chips, etc.		Ketchup, etc.		Pet food	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
eat	good	buy	good	buy	good	buy	good	cat	cat
buy	food	eat	delicious	eat	make	heinz	taste	buy	good
go	can	candy	make	gum	like	ketchup	food	food	like
pizza	delicious	chocolate	like	chips	taste	use	make	feed	dog
burger	get	cream	taste	chew	flavor	eat	like	friskies	want
Coffee, tea		Soda, etc		Cars		Electronics		Phone, TV, etc.	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
coffee	coffee	drink	good	buy	car	buy	good	buy	get
buy	good	buy	make	car	good	phone	great	use	free
drink	like	water	refreshing	drive	like	use	make	get	want
starbucks	make	milk	drink	get	great	get	like	phone	phone
maxwell	get	pepsi	like	ford	make	computer	want	service	good
Security services		Software		Beauty products		Healthcare		Clothing	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
buy	want	use	help	buy	make	buy	help	buy	make
use	fun	buy	want	use	look	use	make	wear	look
sign	help	ibm	make	makeup	skin	go	want	shop	sexy
take	dangerous	get	ads	wear	beautiful	get	good	shoes	like
up	employees	microsoft	like	perfume	good	health	health	clothes	attractive
Baby products		Games and toys		Cleaning products		Home improvements		Home appliances	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
buy	baby	buy	fun	buy	clean	buy	make	buy	make
baby	keep	play	game	use	make	use	look	westinghouse	good
use	healthy	video	like	soap	good	shop	good	ge	better
product	best	game	play	product	stains	paint	home	use	great
formula	help	get	looks	detergent	like	company	like	refrigerator	food
Media and arts		Sports		Shopping		Gambling		Environment	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
watch	like	buy	fun	shop	good	play	win	use	environment
buy	want	go	make	buy	sale	go	money	buy	help
go	fun	watch	want	go	deals	use	fun	environment	good
see	good	use	good	store	prices	want	gambling	want	nature
movie	looks	want	like	foods	great	casino	take	nature	want
Human rights		Safety		Domestic violence		Self esteem		Political candidates	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
support	people	drive	dangerous	domestic	violence	bullying	people	vote	help
rights	help	drink	safe	violence	help	buy	help	support	care
human	rights	buy	want	abuse	domestic	use	bullying	hillary	people
amnesty	want	use	get	against	women	go	self	clinton	change
want	human	wear	save	help	want	want	want	ad	against
Alcohol		Financial		Animal Rights		Other Services		Charities	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
drink	good	use	help	fur	animal	use	help	buy	help
buy	make	bank	want	animals	want	buy	make	donate	support
beer	like	get	money	wear	fur	insurance	want	cookies	save
vodka	drink	insurance	make	support	help	get	good	support	kids
absolut	great	buy	get	peta	cruel	go	need	want	good

Table 1. Common words in responses to "What should I do, according to the ad?" and "Why, according to the ad, should I do it?" questions for the image dataset.

6. Ads strategies

We submitted 4,000 ad images for annotation of strategies on MTurk. Each image was annotated by 5 workers, who could select multiple strategies for an image. We kept all strategies that got at least 2 out of 5 votes. Below we provide further description of each strategy. These descriptions were part of the instructions given to MTurkers to help them identify the strategies. Fig. 6 shows more examples of each strategy.

Understanding physical processes or direction

This includes examples where there is an implicit process or motion which is assumed to be happening at the moment when the image is taken. Think about physical occurrences like gravity or forces, or properties of the 3D world (like being under/over/pointed at) which affect the world or participants in some ways (e.g. a car hitting a face deforms it, sharpness causes injury, eating causing parts of the food to disappear), etc.

Symbolism and physical allegories

Ads contain limited space, so sometimes they require the viewer to make inferences outside of the content of the ad. For example, an ad might show a dove and assume the viewer will associate this dove with the concept of "peace". Similarly, a lemon might symbolize freshness, or blood might symbolize injury/death. We refer to these mappings from a set of pixels to a non-visual concept as "symbols". In this same category we also group what we refer to as "physical allegories". By that we mean objects that are meant to look like other objects. Unlike "atypical objects", these physical allegories say more that an object serves the role of another object, or is placed or arranged in the same way as another object, rather than saying that one object IS another object.

Atypical objects

For example, a combination of pieces of trash in the form of a deer is made to look like a deer.

Contrast

These ads juxtapose two objects and show that these objects have contrasting qualities, or show contrasting situations. For example, an ad might show two tree branches, one with and one without a bird on it, in an attempt to convince you to prevent animals from dying off.

Transfer of qualities

These ads juxtapose two objects that are expected to have similar properties. For example, an ad that shows a lady drinking out of a tall and slim coke can is saying "If you drink this beverage, you will also become thin." Similarly, a classy actor eating an ice-cream makes the ice-cream appear classy.

References to culture or celebrities

These ads require that you are familiar with the meme or cultural reference that they are portraying. For example, an ad might expect you to know about the apple that Adam and Eve ate from, and that it symbolizes sin. Similarly, an ad might expect you to know what the hand gesture of devil's horns symbolizes (it is often used in rock/metal culture). Alternatively, an ad might require that you know who a certain person is (a celebrity), since a celebrity advertising a product makes it seem attractive.

Surprise or humor

These ads show a surprising twist compared to how you expect things to occur in the real world. For example, lungs might be filled with cigarette butts, or a grandmother might be riding a bicycle in dangerous fashion.

Human shown experiencing product

Some ads attempt to excite the viewer about the product by showing a human experiencing the product, and implying that the product has qualities such as e.g. deliciousness.

Product qualities described in literal fashion (Straightforward)

In this group are ads that do not appeal to allegories or symbolism, nor show atypical objects, require understanding of physical processes, or involve surprise, humor, or contrast (i.e. these are your typical not-so-creative ads). Many of these ads show the product in the center, as being attractive, beautiful, stylish, delicious, etc. In some of these, the qualities of the product are either obvious without any symbolism or allegory, or are described in text.

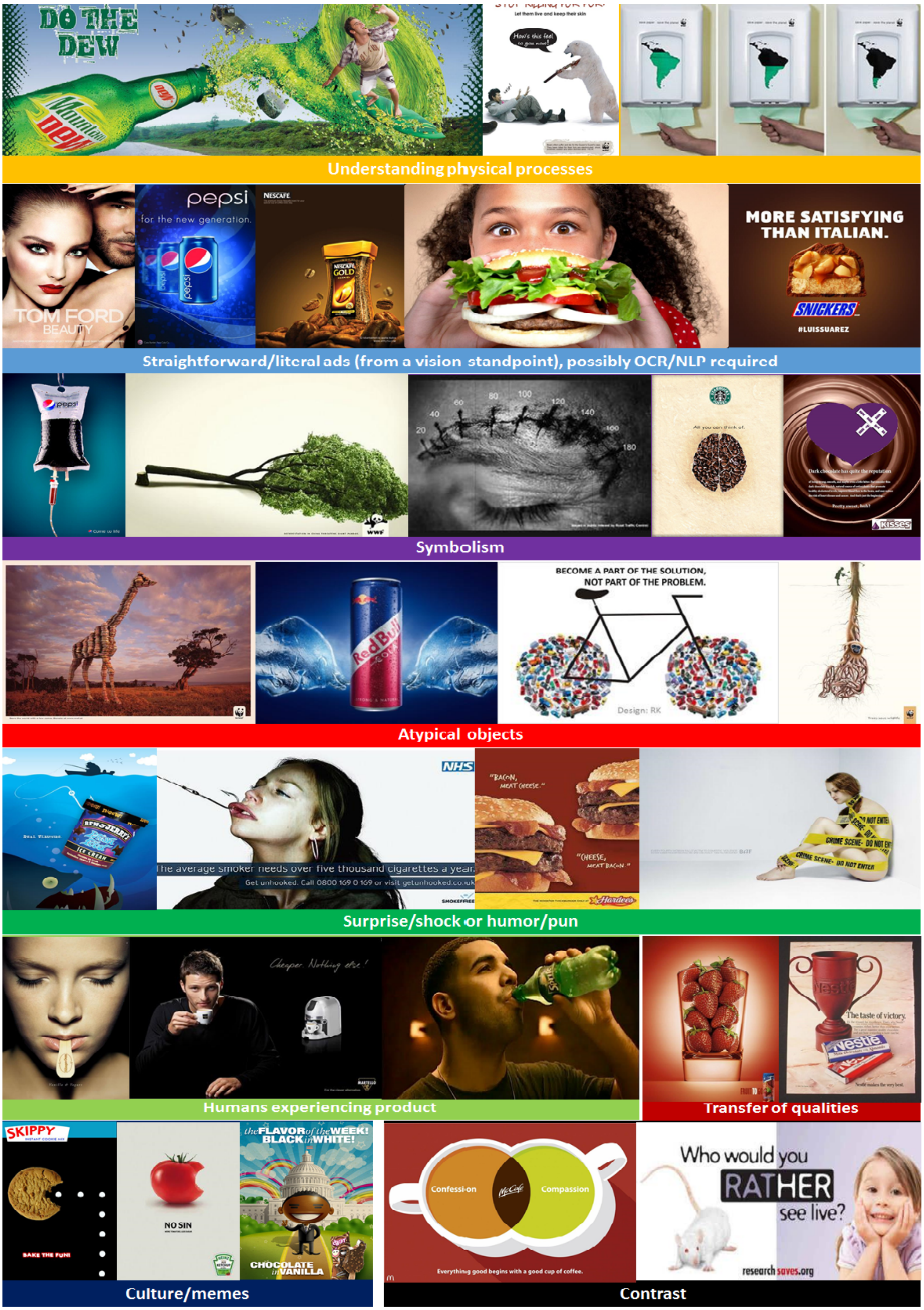


Figure 6. Examples of ads grouped by strategy or visual understanding required for decoding the ad.

7. List of symbols

Following is a list of the 221 most common symbols we obtained after pruning. The list is sorted based on the frequency of occurrence of these symbols in descending order, i.e. the more commonly occurring symbols appear earlier in the list.

danger, fun, nature, beauty, death, sex, health, natural, adventure, environment, power, sexy, food, love, violence, fresh, strength, energy, abuse, speed, safety, sports, travel, fashion, entertainment, excitement, healthy, youth, technology, family, happiness, hunger, strong, protection, injury, desire, delicious, art, humor, freedom, refreshing, happy, pain, clean, style, cool, comfort, vacation, luxury, sex appeal, variety, freshness, unique, hot, smoking, different, fitness, craving, fast, life, quality, active, tough, music, relaxation, sexuality, classic, alcohol, flavorful, sexual, cold, wild, innovation, dangerous, harmful, class, christmas, romance, fear, innocence, seduction, light, friendship, tasty, party, change, accident, elegance, athletic, harm, destruction, attraction, flavor, celebration, unhealthy, taste, pollution, wealth, sport, imagination, simplicity, physical abuse, exotic, car, nutrition, creativity, togetherness, powerful, adventurous, flight, sadness, outdoors, lust, choices, beautiful, stylish, rugged, new, animal cruelty, scary, exciting, enjoyment, work, spicy, attractive, old, milk, education, animal abuse, action, clothing, toughness, thirst, indulgence, heat, candy, surprise, smooth, safe, drinking, space, gift, water, time, purity, home, growth, future, dirty, chocolate, care, big, animal, individuality, holiday, exercise, drink, color, anger, unity, simple, relax, money, coolness, confidence, broken, success, intelligence, fancy, culture, competition, suicide, heart, coffee, strange, royalty, peace, messy, joy, funny, innovative, domestic violence, determination, creative, odd, beer, sweet, pure, performance, flying, fantasy, exploration, diversity, damage, satisfaction, history, childhood, awareness, war, size, play, mystery, fire, easy, convenience, control, communication, choice, celebrity, athleticism, winter, support, shoes, fame, break, animals, small, risk, options, help, halloween

8. Topics in videos

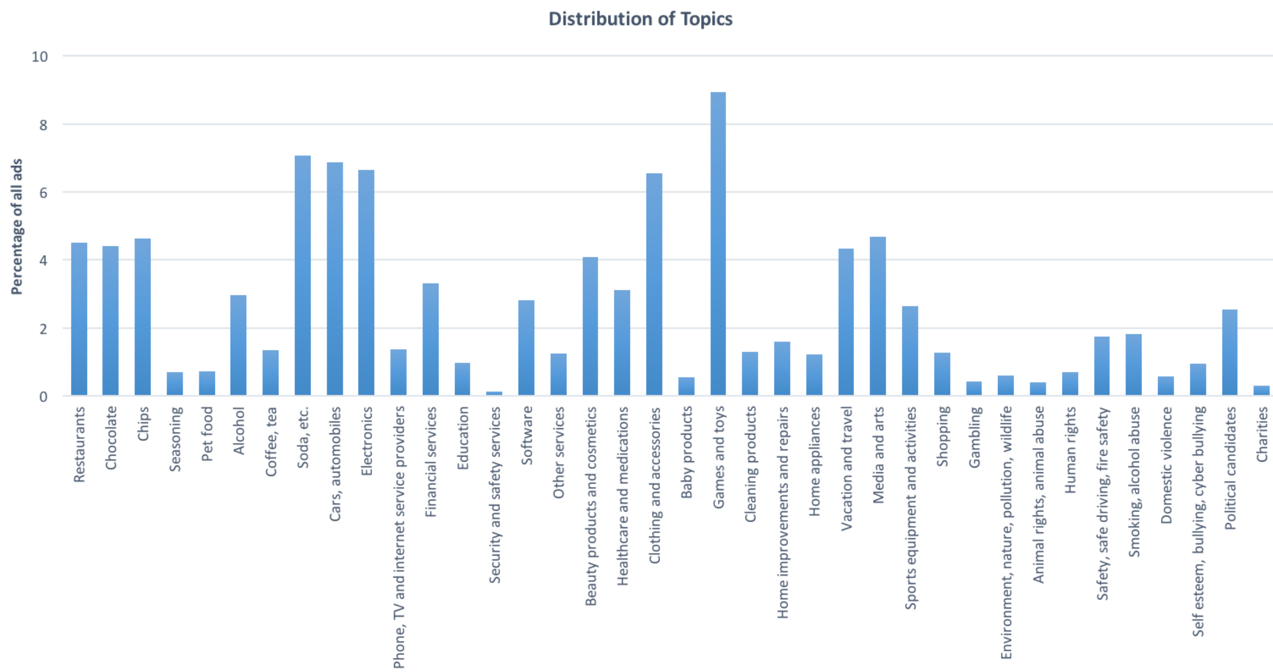


Figure 7. Fraction of each topic out of all ads in the video dataset.

Fig. 7 shows the distribution of topics from our video dataset. Each video was annotated by 5 MTurkers and we picked the most frequently selected topic as the annotation for the video. If there was a tie, we randomly selected one annotation.

From the topic distribution, we can tell that "games and toys" contains the most videos in our dataset, and the number (8.9%) is much higher than that in the image dataset (0.95%). The reason is that videos are more appropriate for advertising computer or mobile phone games, since the graphics and the game experience can hardly be expressed in static images. A similar trend is also observed in topics such like "political candidates" and "healthcare and meditations", and the reason is that themes in these ads are relatively complicated (political views, benefits of medical system, etc.) thus video ads demonstrate more advantages in delivering such messages due to the extended time and space. We also observed that many ads in our video dataset are in "Clothing and accessories", "Cars, automobiles", and "Beauty products and cosmetics", which agrees with what we have seen in our image ads dataset.

9. Sentiments in videos

Fig. 8 shows the distribution of several representative topics over common sentiments. Each video is assigned to 5 different MTurkers and they can choose multiple sentiments for each video. We picked the most frequent sentiment as the one for the video. If there is a tie, a random sentiment from the most frequent candidates is chosen. In this figure we show the most frequent sentiments in all topics as well as the topics containing most diverse sentiments. Therefore the selected topics and sentiments differ a little bit from the image dataset.

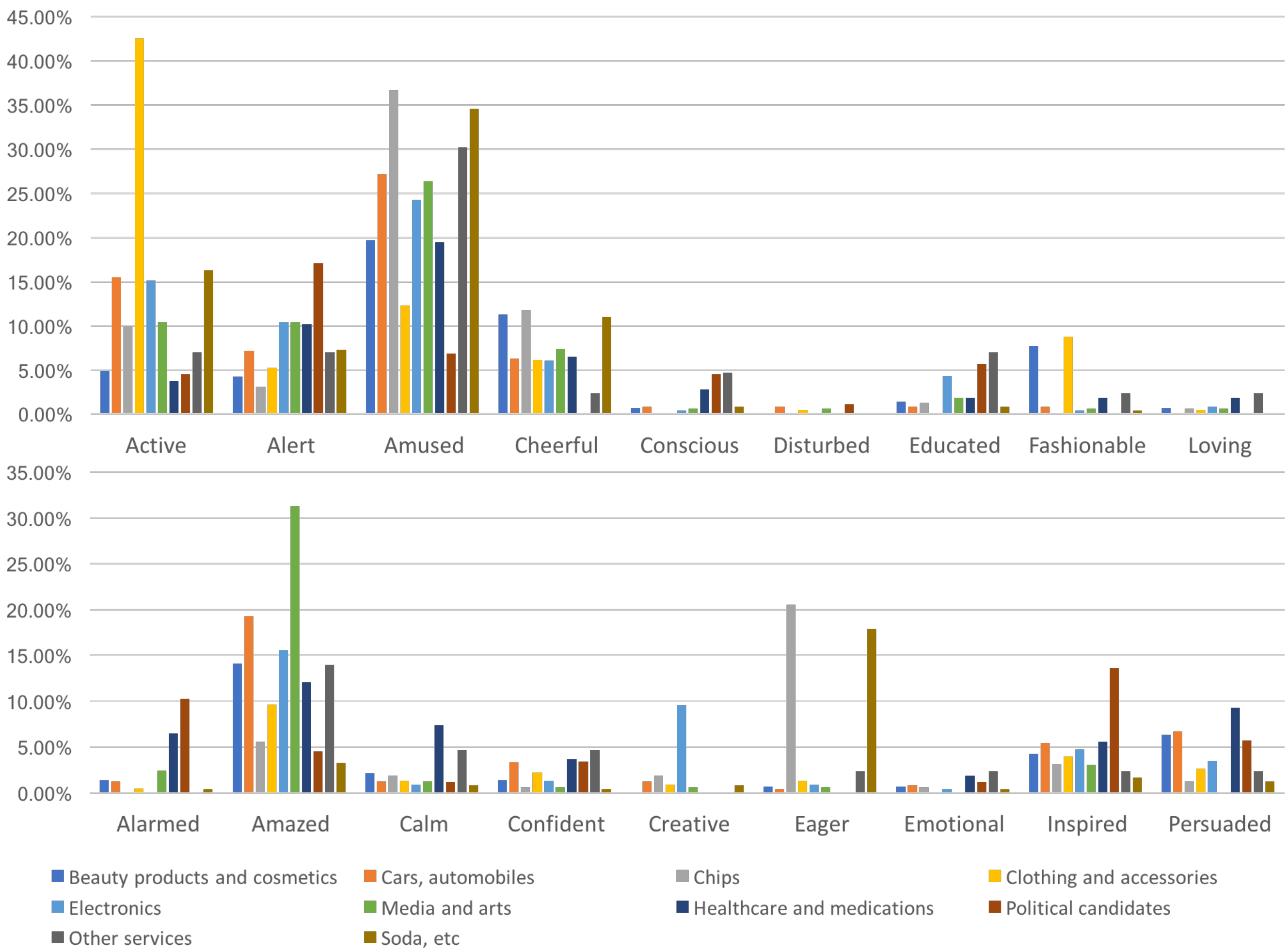


Figure 8. Distribution of several representative topics over common sentiments in the video dataset.

From the figure we observe some interesting trends. "Amused" is a very common sentiment in ads covering a variety of topics, but in "Political candidates" it is rarely used. The reason is likely that ads treat political issues in serious fashion. However, in the image dataset, not many ads try to spread "Amused" sentiments. We believe the reason can be attributed to the limited space in static images. This observation also verifies our assumption that using humor and excitements as hints might be very helpful in understanding video ads.

In "Clothing and accessories", "Active" is more frequently used because many clothing brands targeted for teenagers are promoting their energetic and adventurous life styles. We also observe that "Eager" is most related to food or drink ad topics, such as "Chips" and "Soda". This observation agrees with our daily experience that most food ads are trying to make the audience feel hungry and thirsty. Another example is the feeling of "Fashionable", as both image ads and video ads about "Beauty products" and "Clothing and accessories" are trying to deliver such sentiment.

10. Inter-annotator agreement in videos

In Fig. 9 we show some statistics about the inter-annotator agreement over topic and sentiment annotations in video. For this figure we only analyzed the annotations for the 2,528 videos where we ask annotators to choose among provided options (rather than write free-form text). We exclude low-quality videos where people cannot find a meaningful topic. From the left chart in Fig. 9 we can see that on 88% of videos, at least 3 of 5 annotators agree on the topic, which is reliable agreement.

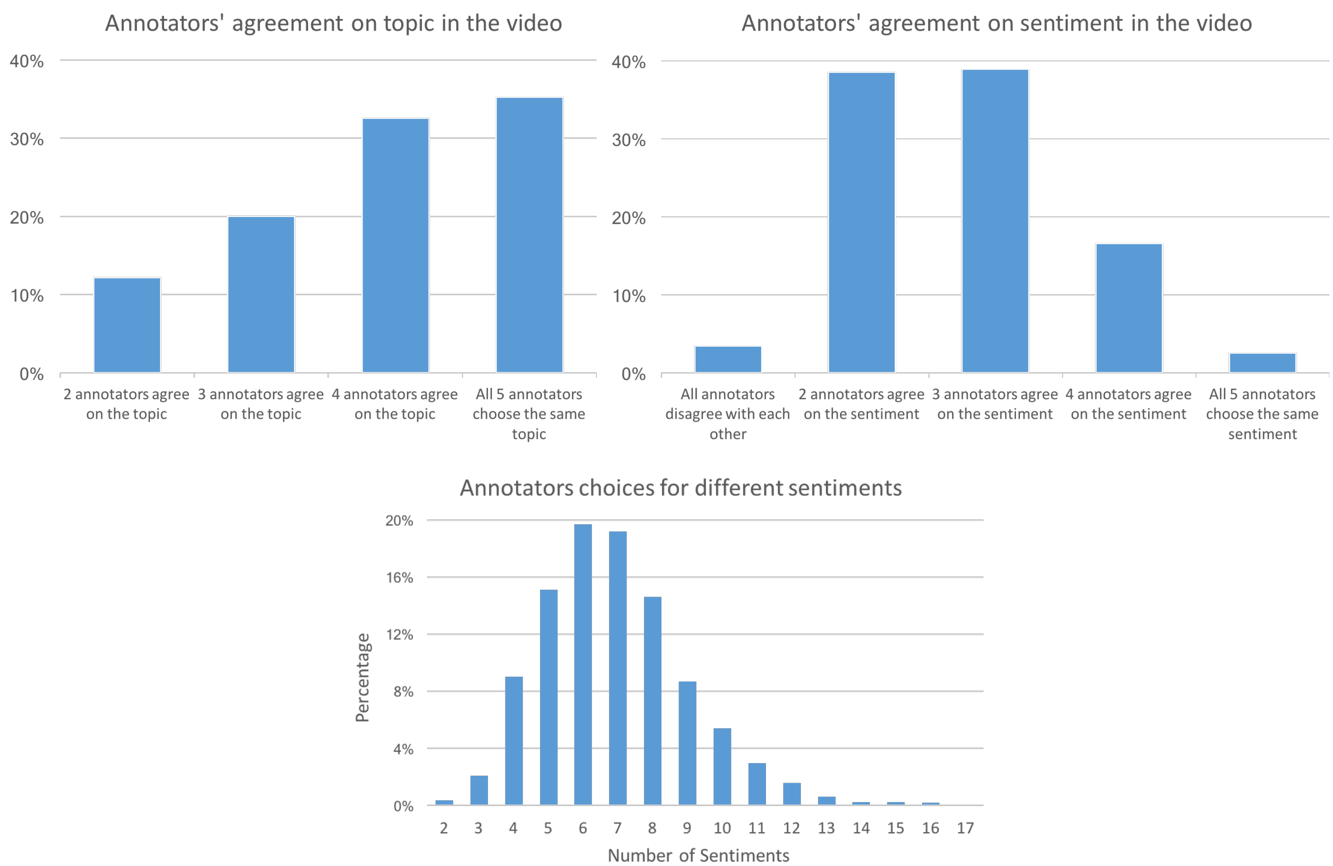


Figure 9. Inter-annotator agreement on topic and sentiment in videos.

In contrast, annotating sentiment is more ambiguous. From the right chart, we notice that annotators have more diverse opinions on sentiments within videos. Specifically, only in 3% do all 5 annotators agree on the sentiments being delivered. However it is also uncommon (3%) that people all disagree on the sentiment in a video. Recall that annotators can mark multiple sentiments per video, hence agreement is harder to accomplish. To gain further insight into the sentiment annotations, in the bottom chart in Fig. 9 we show the unique number of sentiments marked per video. Since every video is assigned to 5 annotators and every annotator can choose multiple sentiments, there can be a large number of unique marked statements. The smaller the number, the more agreement the annotators have on the video's understanding. Nearly 40% videos deliver 6 to 7 different sentiments to their audiences, and over 90% videos contain less than 10 sentiments.

11. Question-answer examples in videos

Tab. 2 shows the common words occurring in response to the “What should I do” and “Why should I do it” questions for each topic in the video dataset.

When comparing the results from video dataset and image dateset, we notice that the top 5 common words are extremely similar for the same topics. This makes sense if we consider that the goal for image ads and video ads are all the same (promoting products or services in viewers), thus it is not surprising that people have similar responses to these questions. We also notice that a large part of the words in the table are verbs and the topic word itself appears frequently in the responses.

Restaurants, etc.		Chocolate, etc.		Chips, etc.		Ketchup, etc.		Pet food	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
eat	food	buy	good	buy	good	buy	make	buy	dog
buy	good	eat	make	eat	make	use	taste	food	cats
restaurant	eat	candy	taste	gum	eat	product	good	dog	love
food	burger	chocolate	chocolate	chips	delicious	food	food	cat	food
fast	delicious	bar	eat	chew	taste	sauce	use	pet	pet
Coffee, tea		Soda, etc		Cars		Electronics		Phone, TV, etc.	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
coffee	coffee	drink	drink	buy	car	buy	phone	service	service
buy	make	buy	make	car	drive	phone	features	use	offer
drink	drink	soda	good	insurance	good	use	take	phone	internet
tea	good	milk	refreshing	purchase	vehicle	camera	use	buy	phone
use	help	product	give	get	get	get	pictures	internet	fast
Security services		Software		Beauty products		Healthcare		Clothing	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
buy	home	use	help	buy	make	buy	help	buy	make
insurance	check	service	use	product	look	use	health	shoes	wear
system	use	app	make	use	beautiful	get	care	clothing	help
blink	easy	website	business	purchase	skin	condoms	need	wear	clothes
home	dangerous	buy	easy	beauty	hair	health	effective	brand	shoes
Baby products		Games and toys		Cleaning products		Home improvements		Home appliances	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
buy	baby	game	game	buy	clean	buy	home	buy	make
diapers	parents	buy	fun	use	clothes	home	sleep	use	use
baby	diapers	play	play	product	stains	use	make	purchase	clean
product	keep	video	exciting	detergent	effective	furniture	products	water	cooking
cup	make	download	like	laundry	make	store	help	machine	food
Media and arts		Sports		Shopping		Gambling		Environment	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
watch	show	buy	help	shop	business	lottery	money	buy	environment
show	entertaining	sports	sports	store	need	play	win	stop	destroy
movie	fun	product	make	buy	offer	tickets	back	use	help
buy	funny	watch	active	use	shopping	buy	get	support	energy
see	movie	brand	play	retail	retail	gambling	tickets	polluting	polluting
Human rights		Safety		Domestic violence		Self esteem		Political candidates	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
support	right	drive	accident	violence	violence	bullying	bullying	vote	candidate
rights	help	safe	life	domestic	domestic	stop	hurt	candidate	people
women	women	drink	save	abuse	help	cyber	help	support	voting
marriage	want	speed	drive	stop	stop	watch	stop	bernie	make
equality	world	safety	kill	help	abuse	stand	people	sanders	work
Alcohol		Financial		Animal Rights		Other Services		Smoking	
What?	Why?	What?	Why?	What?	Why?	What?	Why?	What?	Why?
buy	drink	use	help	adopt	animals	use	help	smoking	smoking
beer	beer	bank	money	animals	need	service	get	stop	bad
drink	good	service	make	pet	help	company	make	drink	health
alcohol	make	invest	investing	dog	home	get	use	drugs	kill
brand	taste	card	use	shelter	make	website	business	quit	life

Table 2. Common words in responses to “What should I do, according to the ad?” and “Why, according to the ad, should I do it?” questions for the video dataset.

12. Funny/exciting annotations in videos

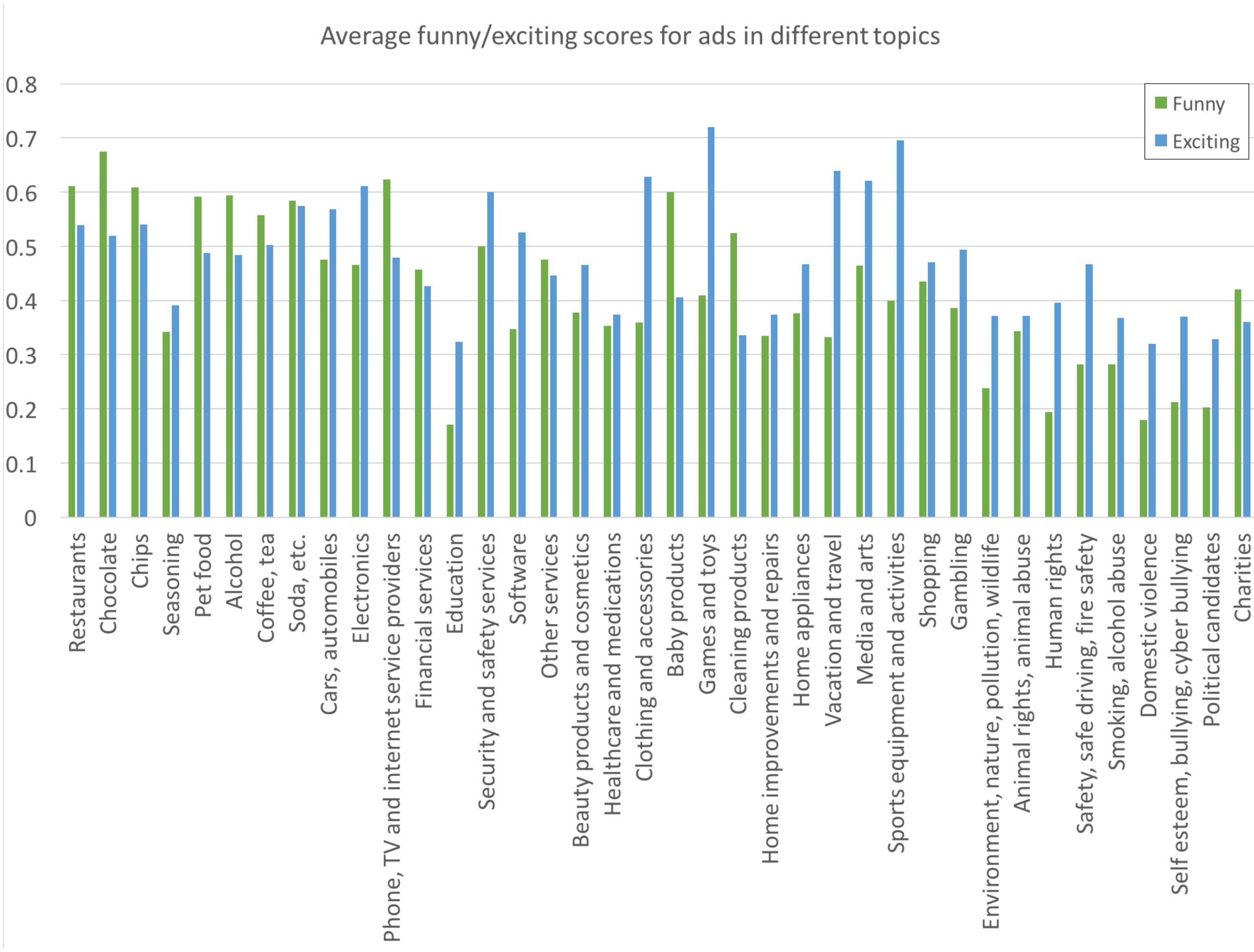


Figure 10. Average funny/exciting scores for ads in different topics.

Fig. 10 shows the average funniness and excitement scores for each topic in our video dataset. Specifically, we considered an annotator's choice of "funny" or "exciting" as a score of 1, and "not funny" or "not exciting" as 0, then averaged all annotations within a topic to obtain the score for that topic. From Fig. 10 we can tell that ads in "Games and toys" and "Sports equipment and activities" frequently use exciting elements in their videos. Similarly, food related topics tend to use humor to impress their viewers, as seen in the higher scores in "Chocolate", "Restaurants", and "Chips". We can also tell that in serious topics, such as "Education", "Domestic violence", and "Political candidates", neither funny nor exciting strategies are frequently used as these ads aim to gain trust from their audiences.

13. Duration of videos

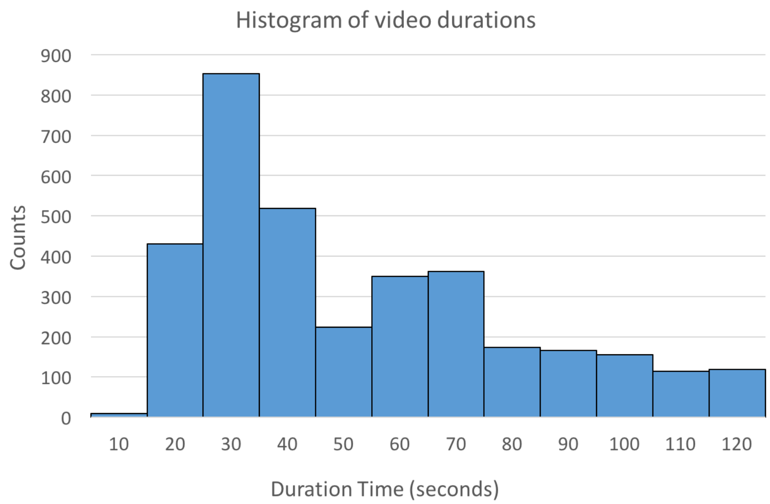


Figure 11. Histogram of video durations.

Fig. 11 shows a histogram of the video durations in our collected dataset. During data collection we excluded all video ads that longer than 120 seconds. The average duration for all 3,477 videos is 49 seconds. Note that nearly one fourth of the ads (24.5%) in our dataset are from 20 seconds to 30 seconds, and 68.6% of all videos are shorter than 1 minute.

14. Interface for video ads collection

As shown in Fig. 12, before the MTurkers accept the HITs we provide six sample videos with corresponding answers to each question. To give MTurkers a better understanding of our desired responses, we provide both qualified sample answers as well as unacceptable sample answers. We pick sample videos covering various topics, including cars, games, safety PSA, political campaign ads, etc.

Here are some samples for you. Please make sure you have watched all the videos and understood all the sample answers before you accept the HIT.

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6

- **Is this advertisement in English?**
  Acceptable Answer — **Yes (English)**
- **Is this advertisement funny?**
  Acceptable Answer — **Yes**
- **Is this advertisement exciting?**
  Acceptable Answer — **No**
- **What is being advertised in this video?**
  Acceptable Answer — **Cars and Automobiles**
  Unacceptable Answer — **Subaru** (Do not use brand name)
  Dogs
  Cellphones
- **What emotions is the video aiming to make the viewers experience?**
  Acceptable Answer — **Delighted**
  Excited
  Unacceptable Answer — **Sympathetic**
- **What does the video try to persuade the viewers to do?**
  Acceptable Answer — **Buy this car**
  Unacceptable Answer — **Adopt a dog**
- **According to the video, why should the viewers do this?**
  Acceptable Answer — **Because it is pet-friendly.**
  Because I can drive with my dogs in this car.
  Unacceptable Answer — **Because it is good/funny.** (Not specific to this video)
  Because my dogs can drive this car for me. (Incorrect understanding of the video)

Figure 12. We show MTurkers six sample videos as well as both acceptable and unacceptable answers for these sample videos.

Fig. 13 demonstrates how our HIT interface evolved. In early stage questionnaires we ask MTurkers to write free-form responses to answer the questions about the video. After we collected enough responses, we compiled a representative list for common topics (and sentiments as well) in video advertisements. In later HITs MTurkers just needed to mark the options from our list. If they felt that the video did not belong to any existing options, they could write a free-form response.

Early-stage free-form questionnaire:

3. What is being advertised in this video?

4. How do you feel after watching this advertisement? In other words, what emotions the video is aiming to make you experience?
Please use a few words or phrases to describe your feelings. Please do NOT write a full sentence for this question.
We understand that same video can provoke different feelings in different viewers, but there are some feelings that are extremely unlikely (e.g. negative feelings for an happy video).
Please answer this question from a general viewer's pespective and do NOT treat this question as a subjective opinion survey.

Multiple choice selections questionnaire:

5. What is being advertised in this video?

If the product being advertised belongs to one of the categories listed below, please **just mark the option**. Otherwise, you can write your own answer.

- **Restaurants, Cafe, Fast food**
- **Chips, Snacks, Nuts, Fruit, Gum, Cereal, Yoghurt, Soups**
  general food should also be put in this category
- **Pet food**
- **Coffee, Tea**
- **Sports equipment and Activities**
- **Phone, TV and Internet service providers**
- **Education**
  universities, colleges, kindergartens, online degrees, etc.
- **Baby products**
  baby food, sippy cups, diapers, etc.
- **Games and toys**
  video games, mobile games, etc.
- **Shopping**
  department stores, drug stores, groceries, etc.
- **Beauty products and cosmetics**
  deodorant, toothpaste, makeup, hair product, laser hair removal, etc.
- **Clothing and accessories**
  jeans, shoes, eye glasses, handbags, watches, jewelry, etc.
- **Vacation and travel**
  airlines, cruises, theme parks, hotels, travel agents, etc.
- **Home improvements and repairs**
  furniture, decoration, lawn care, plumbing, etc.
- **Software**
  Internet radio, streaming, job search website, grammar correction apps, travel planning, etc.
- **Chocolate, Cookies, Candy, Ice cream**
- **Seasoning, Condiments, Ketchup**
- **Soda, Juice, Milk, Water, Energy drinks**
- **Alcoholic drinks**
  vodka, rum, beer, etc.
- **Electronics**
  computers, laptops, tablets, cellphones, TVs, etc.
- **Financial services**
  banks, credit cards, investment firms, etc.
- **Security and safety related products**
  anti-theft, safety courses, etc.
- **Other services**
  dating, tax, legal, loan, religious, printing, catering, etc.
- **Gambling**
  lotteries, casinos, etc.
- **Media and arts**
  TV shows, movies, musicals, books, audio books, etc.
- **Healthcare and medications**
  hospitals, health insurance, allergy, cold remedy, home tests, vitamins, etc.
- **Cars and automobiles**
  car sales, auto parts, car insurance, car repair, gas, motor oils, etc.
- **Cleaning products**
  detergents, fabric softeners, soap, tissues, paper towels, etc.
- **Home appliances**
  coffee makers, dishwashers, cookware, vacuum cleaners, heaters, music players, etc.

- **Political candidates** (support or opposition)
- **Animal rights, Animal abuse**
- **Safety, Safe driving, Fire safety**
- **Domestic violence**
- **Charities**
- **Environment, Nature, Pollution, Wildlife**
- **Human rights**
  including rights for minorities, women's rights and LGBT rights
- **Smoking, Alcohol abuse, Drug abuse**
  PSAs relating anti-smoking, drug prevention etc.
- **Self esteem, Bullying, Cyber bullying**

- **Write my own:**

Figure 13. We first collect free-form answers for topics and sentiments questions, and in the later HITs MTurkers just need to choose from the list we compiled.

15. Symbolism prediction model

In Sec. 6.2 of the main text, we use an attention network to determine what symbols are present in an image. Here we include details about this network.

We use 224×224 input images and build both an attention predictor and a feature extractor on top of the feature map from a pre-trained ResNet-50 model. Each 7x7 attention output indicates the probability that the associated region contributes to the final representation. Accordingly, each 7x7 feature extractor output denotes a feature vector associated with the same region. The final representation is a weighted average over the image features using the predicted attention distribution. We use sigmoid cross entropy as our loss function. We include the architecture for the network in Fig. 14. The network achieves F-score of 15.79%.

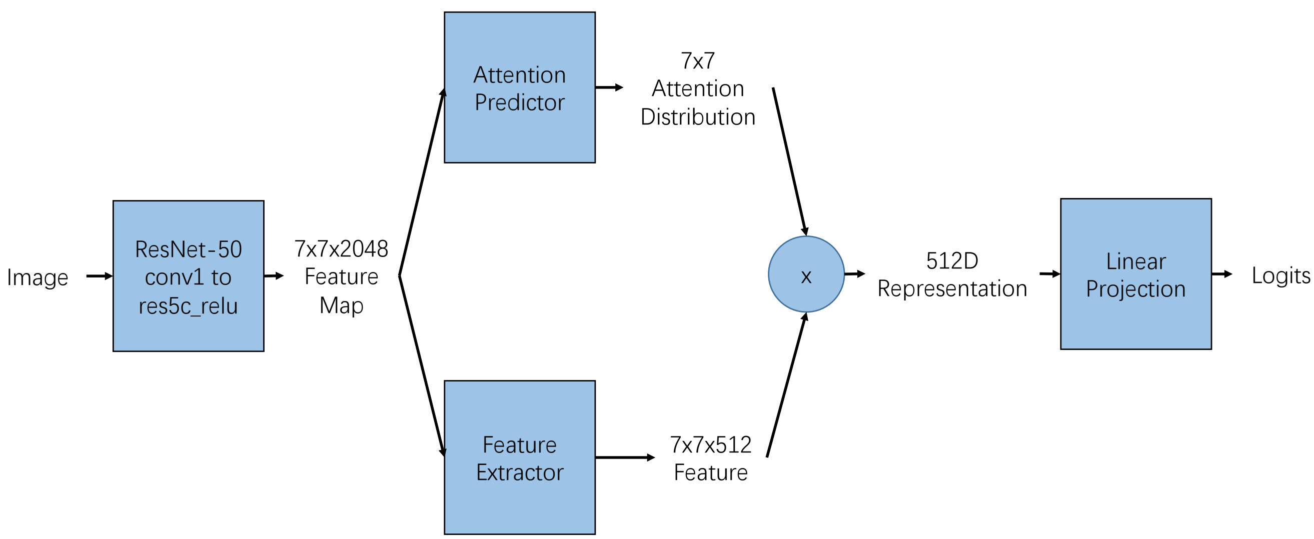


Figure 14. The topology of the attention network for predicting symbols. The feature extractor is constructed by two 1x1x1x1 convolutional layers which output $512D$ feature. We add batch normalization layer after each convolutional layer and we add dropout layer while training. The attention predictor has the same architecture except that an extra linear projection layer mapping $512D$ feature to $1D$ is added, and softmax activation is applied on the final 7×7 output.

In a subsequent experiment, we attempted to improve symbolism prediction by narrowing down our list of symbols to a shorter list of more distinct and reliable symbols. We cluster symbols based on co-occurrence relations, and we measure the similarity between symbols in the following way. We define the similarity between symbols s_i and s_j to be $Sim(s_i, s_j) = \max(P(s_i|s_j), P(s_j|s_i))$, where $P(s_i|s_j)$ denotes the probability that we observe the presence of symbol i given the presence of symbol j. Given the similarity matrix, we use agglomerative clustering to cluster symbols, and define the similarity between two clusters c_i and c_j to be $Sim(c_i, c_j) = \min\{Sim(x, y) : x \in c_i, y \in c_j\}$. During the clustering process, clusters with similarity greater than or equal to 10% are merged. Finally, we keep clusters that contain more than 200 images to ensure a large enough amount of training data for each cluster. This gives us 53 final clusters. A model trained to distinguish between these 53 symbols achieves 26.84% F-score.

Figure 15 shows examples of some of the data we collected for each symbol (where annotations are grouped into the 53 symbol categories). Note the significant diversity within each cluster.

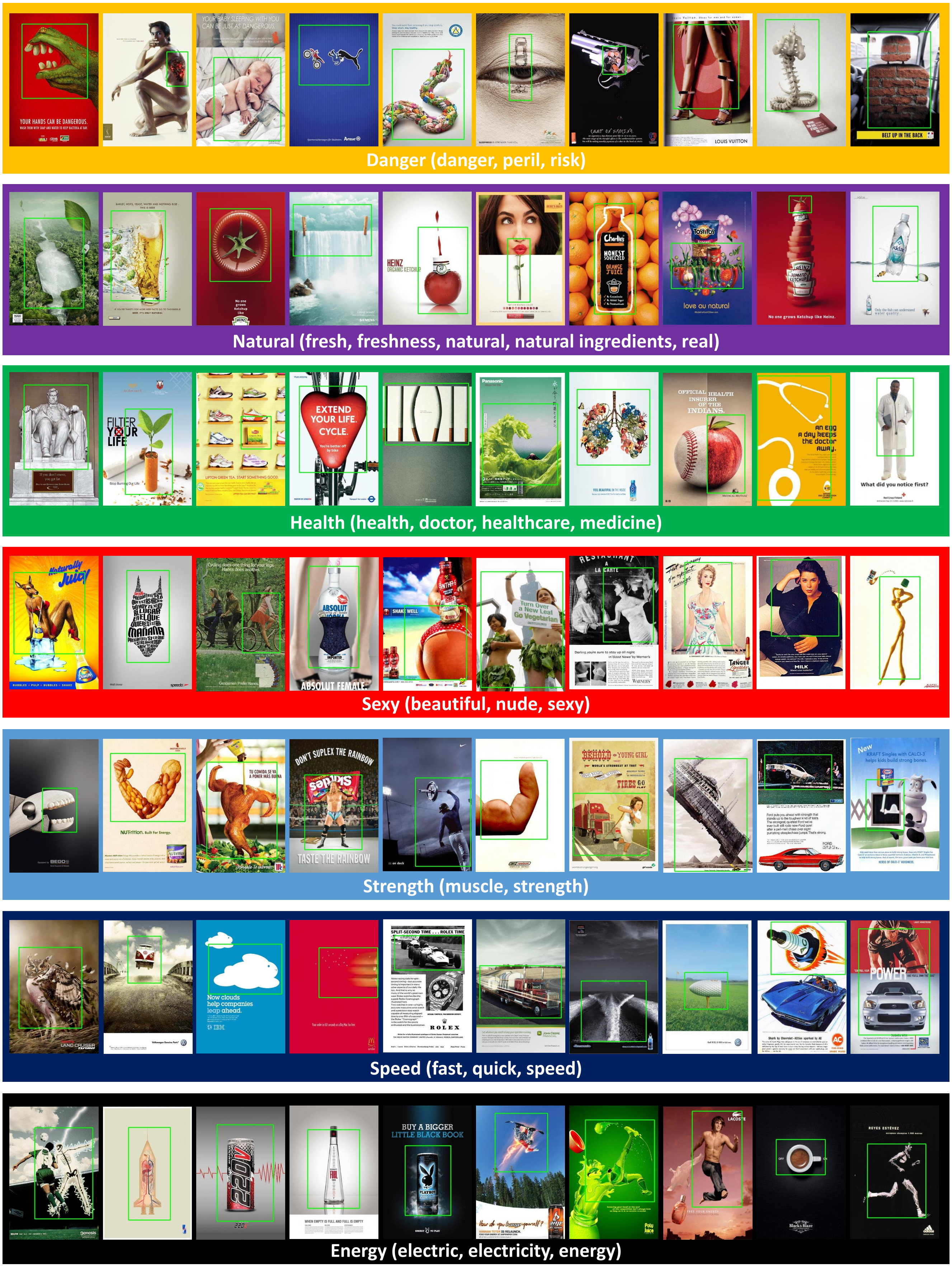


Figure 15. Examples of symbolism labels and bounding boxes from our annotators.

16. Examples of question-answering results and detected symbols

Fig. 17 shows some examples of question-answering results by the baseline and our method capturing symbolism. We measure accuracy as whether the machine-given answer agrees with any of the human-given answers. We also show the symbols detected by our symbol classifier. We show the image, the question selected based on TFIDF, and the ground-truth answers. We show the response from both our method and the baseline.

In (a), we see that while both methods' answers are reasonable, they both disagree with the human-given responses. In (b), we see that our method which can predict what visual content symbolizes "sexiness," can give the right answer, while the baseline gives a reasonable but technically incorrect answer. In (c), we successfully predict "refreshing" as a symbol, but the baseline also correctly outputs that answer. In (d), the baseline is correct and our method is not, but note that "kills" might be a more correct response to "Why should I not buy products made from animals," compared to the given answers. In (e), our system fails to predict "French" as a symbol. In (f), our method likely predicts "taste" because the symbol classifier misinterprets the objects in the image as bowls rather than candles.

In Fig. 16, we see responses for the video question-answering task. In this figure, we show the top-5 responses from the system, with confidence. The second-to-last column shows the ground-truth answers. We see that in (b), the system interprets the visual cues correctly (e.g. "weather"), but fails to capture the rhetoric of the system. Similarly in (c), it correctly detects nature, but fails to put that in context. In (f), the system produces the correct answer likely because it successfully detected the first frame as a "home".

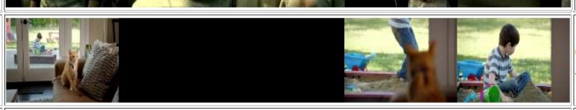
	Video frames	One question and answer	GT single-word answers	Predicted answers
(a)		**Q**: why should you have a cup of soup . **A**: because it is delicious and makes a quick lunch or meal	sandwiches lunch sandwich moves	cats:0.1177 delicious:0.1114 make:0.1024 tasty:0.0801 creative:0.0661
(b)		**Q**: why should you i should follow the australian government . **A**: because aids can kill you .	save kill associate	weather:0.1907 dangerous:0.1693 best:0.0849 warm:0.0744 diverse:0.0653
(c)		**Q**: why should you visit colorado . **A**: because there is a lot to do there .	life alive lot waiting style	environment:0.1351 nature:0.1036 vacation:0.0592 adventures:0.0527 gas:0.0497
(d)		**Q**: why should you download and play the game shown in the ad **A**: because it has neat designs and exciting music to suggest it is a fun experience	fun exciting	fun:0.6555 action:0.6305 game:0.4171 fight:0.3830 exciting:0.3037
(e)		**Q**: why should you do n't smoke . **A**: because it will kill you .	body kill problems	safe:0.3272 kill:0.3055 health:0.1971 could:0.1430 bad:0.1258
(f)		**Q**: why should you adopt a cat from a shelter . **A**: because loving the animals gives a way of taking care	need home animals pets	home:0.4071 faster:0.2473 pets:0.1374 stains:0.1362 useful:0.1312

Figure 16. Predicted answers for video question-answering.

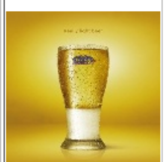
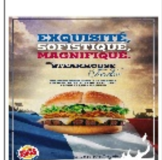
	Image	Symbols	Question	GT-Answers	QA-Baseline	QA+Symbols
(a)		beauty:0.5956 violence:0.1103 youth:0.0972 desire:0.0745 celebrity:0.0727 sexy:0.0719 abuse:0.0657 sex:0.0629 fame:0.0601 physical abuse:0.0562	Why should I buy Maybelline Rocket Volume mascara?	lashes short brush volume voluminous	attractive	attractive
(b)		sex:0.6208 beauty:0.6054 sexy:0.4231 sex appeal:0.2933 seduction:0.2327 desire:0.2305 lust:0.1645 sexuality:0.1441 sexual:0.1256 fashion:0.0819	Why should I wear Vera Wang Perfume?	attractive beautiful sexy feel sexy	smell	sexy
(c)		refreshing:0.1719 alcohol:0.1151 health:0.1076 sports:0.0742 danger:0.0720 cold:0.0662 thirst:0.0654 energy:0.0560 cool:0.0516 death:0.0496	Why should I buy this brand of beer?	refreshing feeling refreshing	refreshing	refreshing
(d)		death:0.2833 animal abuse:0.2755 danger:0.2303 abuse:0.1852 animals:0.1827 violence:0.1326 animal:0.1306 nature:0.1074 strong:0.1008 environment:0.0981	Why should I not buy products made from animals?	hurts dangerous bloody	dangerous	kills
(e)		fresh:0.1829 food:0.1601 hunger:0.1377 flavorful:0.0974 freshness:0.0877 natural:0.0840 delicious:0.0588 fun:0.0553 healthy:0.0545 craving:0.0512	Why should I eat this Burger King Fondue burger?	different politics french	delicious	delicious
(f)		fresh:0.4897 flavorful:0.3733 hot:0.3649 food:0.2781 freshness:0.2522 spicy:0.2360 delicious:0.2173 hunger:0.1925 healthy:0.1822 tasty:0.1805	Why should I celebrate Diwali?	wished entertaining	fun	taste

Figure 17. Predicted symbols and answers for image question-answering.

17. Full-sentence question-answering

Finally, we show an additional quantitative result. Rather than predict a single word out of a vocabulary of 1000 words, or an answer cluster ID out of 30 clusters, here we train a network to predict full-sentence answers for the "Why should I do [Action] according to the ad?" question. We compare a baseline against a method which uses supervision from our topic, sentiment, and symbol annotations, using a variety of machine translation metrics. The baseline uses a 128-dimensional encoding of the question, and a 2048-dimensional encoding of the image (obtained from a Residual Network trained on the ILSVRC2015 1000 classes). Our method replaces this image encoding with a concatenation of three 512-dimensional encodings. These come from a network that fine-tunes the ResNet using three branches, each of which learns to distinguish between our 38 products, 30 sentiments, and 53 symbols, respectively. Both methods then train an LSTM network to generate full-sentence answers to the question, using beam size 1 (other beam sizes produced similar results). The results from both networks are shown in Table 3 below. We see that for most metrics our symbolism-aware method achieves an improvement over the baseline.

Metric	Baseline	Ours
CIDEr	0.2025	0.2207
ROUGE	0.4403	0.4498
METEOR	0.2148	0.2158
BLEU-4	0.2032	0.1945
BLEU-3	0.2684	0.2700
BLEU-2	0.3542	0.3623
BLEU-1	0.4518	0.4639

Table 3. Full-sentence prediction results using a baseline and a method using an image representation based on symbols, topics, and sentiments.