EmoSet: A Large-scale Visual Emotion Dataset with Rich Attributes (Supplementary Material)

Jingyuan Yang[1], Qirui Huang[1], Tingting Ding[1], Dani Lischinski[2], Daniel Cohen-Or[1,3], Hui Huang[1*]
[1]Shenzhen University [2]The Hebrew University of Jerusalem [3]Tel Aviv University
{jingyuanyang.jyy, qrhuang2021, dingt6616, danix3d, cohenor, hhzhiyan}@gmail.com

1. Construction Details

1.1. Emotion Keywords

Table 2 presents the entire list of emotion keywords, with a total number of 810 words. Based on three widely-used English dictionaries, we first find the synonyms of each emotion category. For example, "amusement" is first synonymized to "entertainment, delightful, delectation, enjoy, pleasure, pleased, pleasant, entrancement, ravishment, cheerful, laughter, mirthful, hilarity, merry, glad, recreation, extravaganza" and then enlarged to the current 58 words. Since there are more positive emotions on the Internet compared with the negative ones, we use more emotion keywords to collect negative images, *i.e.*, 302 (positive) and 508 (negative).

1.2. Emotion Attributes

Here we show the formulation of brightness and colorfulness.

$$brightness = \frac{1}{XY}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} I_v(x,y), \quad (1)$$

which is the average pixel intensity of image.

$$colorfulness = 0.02 \times \log(\frac{\sigma_\alpha^2}{|\mu_\alpha|^{0.2}}) \times \log(\frac{\sigma_\beta^2}{|\mu_\beta|^{0.2}}), \quad (2)$$

where $\alpha = R-G, \beta = 0.5(R+G)-B$ are opponent spaces and , $\sigma_\alpha^2, \sigma_\beta^2, \mu_\alpha, \mu_\beta$ represent the variance and mean values along these two opponent color axes.

1.3. Human Annotation

In EmoSet, there is a 3.3 million automatic labeled set and a 118,102 human labeled one, namely EmoSet-3.3M and EmoSet-118K. We first group a number of 86 people and test them with empathy quotient test and FI accuracy test, where 60 of them are qualified. In empathy quotient test, questions are like "Other people tell me I am good at understanding how they are feeling and what they are thinking." For each affective image, we invite 10 workers to annotate it. In Figure 1, images of different number of votes are presented. When reaching only 4 votes, images are often incorrectly labeled with emotions, where a deer will not arouse *sadness* and a singer will not evoke *anger*. It is obvious when images reach 10 votes, all the annotators agree with automatic label given by machines, where images are presented with strong emotions. However, since emotions are subjective and ambiguous, we set 7 votes as threshold to distinguish correctly labeled images and incorrectly labeled ones.

*Corresponding author

2. Dataset Statistics

EmoSet comprises images retrieved from both social networks and photography community, resulting to diverse image types, as shown in Figure 2. We can see images on the left side are very lifelike, which often appear in Twitter, Flickr, Instagram, and other social networks. While images on the right side are full of photography skills, taken by professional photographers.

We further present each emotion with the top-3 attribute values, where image-text pairs are provided for a better understanding. For example, when viewing scene types like stage, wave and football field, we may experience excitement. Conversely, object classes like caterpillar, snake, or waste container, may bring people a feeling of disgust. The numbers in brackets indicate the degree to which a certain emotion is evoked when viewing a specific attribute. We find that some attribute values are strongly related to emotions, including cemetery-sadness (1.00), surfboard-excitement (0.99), skull-fear (0.99), carousel-amusement (0.97), and snowy mountain-awe (0.92). Besides, we can also filter out some emotion-irrelevant ones, like happy-fear (0.01), plaza-anger (0.23), and driving car-awe (0.32). The image-text pairs and statistics are highly consistent with human cognition, indicating that some attribute values are indeed strongly related to emotions. Once a certain attribute value appears, the image is much more likely to evoke the corresponding emotion.

In Table 3, we show the top-10 attribute values of scene type, object class and human action for different emotions. As there are only ten values in brightness and colorfulness and six values in facial expressions, we don't show top-10 values on these attributes. The top-10 scene attribute values in amusement are: Carousel, Amusement park, Delicatessen, Water park, Park, Toyshop, Lawn, Playground, Restaurant, Playroom, which bring people a pleasure feeling. Oppositely, the top-10 objects in disgust are: Caterpillar, Waste container, Snake, Plastic bag, Reptile, Invertebrate, Worm, Marine invertebrates, Toilet, Beetle, making people feel sick.

We further visualize six emotion attributes with different values, to show the emotion-attribute relationships, where red denotes the background color of images with positive emotion and blue for the negative ones. As shown in Figure 3, ranging from 0.1 to 0.9, each brightness value is presented by nine images. Similarly, Figure 4 shows affective images with different values of colorfulness. Obviously, the brighter the image, the more colorful the image, the more positive emotion the image will evoke. In Figure 5, we choose "beach, lawn, stage, mountain snowy" to represent positive emotions and "corridor, ruin, butchers shop, slum" for negative ones. For object class, Figure 6 presents positive emotions with object values "doll, coffee, rabbit, surfboard" and negative emotions with "lion, caterpillar, skull, riffle". These images indicate that there are indeed some emotion-related attribute values, which are decisive to visual emotions. In view of *empathy*, viewers may have a similar feeling with the people inside the images. We give some examples in Figure 7 on different facial expressions, according to different emotion categories. Notably, though all positive emotions in EmoSet are mapped to "happy", there are still some subtle differences between different emotions. Besides, we also represent some images with different actions, where "dining, walking the dog, yoga, side kick" relate to positive emotions while "extinguishing fire, smoking, pumping fist, digging" the negative ones. With human's appearance in the image, we are more likely to evoke a strong feeling.

3. Network Structure

Each attribute branch comprises multiple downsampling blocks and a resolution-maintaining block that reduces the number of channels to 2048, as shown on Table 1. Each downsampling block has three components: a 1×1 convolution with 256 output channels, a 3×3 convolution with 256 output channels, and an avgpool layer. The number of downsampling blocks is determined by the size of the input feature map, as these blocks aims at the same output size, *i.e.*, 7×7. We conduct experiments using various commonly-used backbones such as AlexNet, VGG-16, ResNet-50, and DenseNet-121. We standardize the input image size to 224×224 for each backbone and utilize three different sizes of the middle feature map (*i.e.*, 56×56, 28×28, and 14×14) to represent low-level, mid-level, and high-level feature maps, respectively. The three feature maps then serve as inputs for corresponding attribute branches.

Table 1: The network structure of attribute module.

Attribute level	Input size	Downsampling blocks	Resolution-maintaining block
Low-level	56x56	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ avgpool \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 2048 \\ avgpool \end{bmatrix} \times 1$
Mid-level	28x28	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ avgpool \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 2048 \\ avgpool \end{bmatrix} \times 1$
High-level	14x14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ avgpool \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 2048 \\ avgpool \end{bmatrix} \times 1$

Specifically, we use $\mathbf{v}_0$ to represent the output of backbone branch, and $\mathbf{v}_i \in \mathbb{R}^{2048}, i \in \{low, mid, high\}$ to denote the output of each attribute branch. We then confuse the backbone branch with three attribute branches, to jointly predict visual emotions:

$$\mathbf{v}_{emo} = concat(\mathbf{v}_0, \mathbf{v}_{low}, \mathbf{v}_{mid}, \mathbf{v}_{high}), \tag{3}$$

$$\mathcal{L}_{emo} = \mathcal{L}_{CE}(y_{emo}, p_{emo}), \tag{4}$$

$$p_{emo}\left(j \mid \mathbf{v}_{emo}, \mathbf{W}_{emo}\right) = \frac{\exp\left(\mathbf{w}_j \mathbf{v}_{emo}\right)}{\sum_{j=1}^{C_{emo}} \exp\left(\mathbf{w}_j \mathbf{v}_{emo}\right)}, \tag{5}$$

$$\mathcal{L}_i = \mathcal{L}_{CE}(y_i, p_i), \tag{6}$$

$$p_i\left(j \mid \mathbf{v}_i, \mathbf{W}_i\right) = \frac{\exp\left(\mathbf{w}_j \mathbf{v}_i\right)}{\sum_{j=1}^{C_i} \exp\left(\mathbf{w}_j \mathbf{v}_i\right)}, \tag{7}$$

where both loss functions are built with the Cross-Entropy (CE) loss, for multi-class prediction tasks. The overall loss function consists of four parts, namely, an emotion prediction loss and three attribute prediction losses:

$$\mathcal{L} = \lambda_{emo}\mathcal{L}_{emo} + \sum_i \lambda_i \mathcal{L}_i. \tag{8}$$

In order to maintain a relatively balanced weight between emotion and attribute, as well as different attributes, we set $\lambda_{emo} = 1$, and $\lambda_i = 1/3, i \in \{low, mid, high\}$ in our experiments.

Table 2: Emotion keywords.

Emotion	Keyword	Number
Amusement	amusement, amusing, amused, amuse, amuses, amusingly, entertainment, entertain, entertainer, entertaining, delight, delightful, delighting, delightfully, delectation, enjoyment, enjoy, enjoying, enjoyable, pleasure, pleasureful, pleasurable, pleasurably, pleasing, pleased, pleases, pleasant, pleasantly, pleasantness, pleasance, entrancement, ravishment, cheer, cheered, cheering, cheers, cheerful, laughter, laughing, laugh, mirth, mirthful, mirthfully, mirthfulness, hilarity, hilarious, hilariously, merry, merrily, merriness, merriment, glad, gladden, gladly, gladness, recreation, recreational, extravaganza	58
Awe	awe, awed, awing, awes, awestruck, wonder, wonders, wondered, wondering, wonderment, reverence, reverenced, reverencing, reverent, reverently, revere, revered, revering, reveres, venerate, venerated, venerates, venerating, venerable, veneration, inspire, inspired, inspires, inspiring, inspiration, respect, respected, respecting, respects, respectful, respectfully, respectfulness, hallowed, hallowedly, hallowedness, exalted, exaltedly, exaltedness, amaze, amazes, amazed, amazing, amazingly, amazement, astonish, astonished, astonishing, astonishes, astonishment, impress, impressed, impressing, impresses, impressive, impressively, impressiveness, marvel, marveled, marveling, marvels, astound, astounded, astounding, astounds, astoundingly, startle, startled, startling, startles, surprise, surprised, surprising, surprises, surprisingly, worship, worshiped, worshiping, worships	83
Contentment	contentment, contented, contentedly, contentedness, happy, happier, happiest, happily, happiness, satisfaction, satisfactory, satisfactorily, satisfied, satisfying, fulfilment, fulfillment, fulfill, fulfilled, fulfilling, fulfills, fulfils, need, needed, needing, needs, expect, expected, expectedly, expecting, expects, expectable, expectably, expectedness, expectancy, expectant, expectantly, expectation, longing, longingly, complacency, complacence, smug, smugly, smugness, gloat, gloated, gloating, gloats, gloatingly, peace, peaced, peacing, peaceful, peacefully, peacefulness, peaceable, peaceably, ease, easing, eased, eases, comfort, comforted, comforting, comforts, comfortable, comfortably, gratification, gratify, gratified, gratifying, gratifies, gratifier, gratifyingly, serenity, serenities, serenely, sereneness, equanimity, equanimous, equanimously, repletion, replete, repletely, repleteness, warmth	86
Excitement	excitement, excite, excited, exciting, excitedly, excitation, pride, glee, gleeful, gleefully, exhilaration, exhilarate, exhilarated, exhilarating, fervor, fervour, fervency, lively, cheerly, joy, joys, joyousness, joyfulness, joyful, joyfully, joyous, joyously, rouse, rousing, agitate, agitated, agitating, agitates, agitatedly, agitative, agitation, agitator, passion, thrill, thrilled, thrilling, thriller, adventure, adventured, adventuring, adventures, adventurer, adventurous, adventurously, enthusiasm, enthusiastic, enthusiastically, enthusiast, flurried, flurry, flurrying, flurries, furore, furor, commotion, elate, elates, elated, elating, elation, kick, kicks, kicked, kicking, nightlife, show, frisson, frissons, hysteria, hysterical	75
Anger	anger, angered, angering, angers, angry, angrily, choler, choleric, cholerically, ire, ireful, irefully, irefulness, ireless, grievance, fury, furious, rage, raging, raged, rages, wrath, wrathful, infuriation, infuriate, infuriated, infuriating, infuriates, enrage, enraged, enraging, enrages, enragement, enragingly, umbrage, offend, offensive, offended, offending, offense, offence, indignation, indignant, indignance, indignantly, outrage, outraged, outrages, outraging, outrageous, outrageously, outrageousness, dudgeon, irascibility, irascible, irascibleness, irascibly, annoyance, annoy, annoyed, annoying, annoyingly, chafe, chafed, chafing, chafes, vex, vexation, pique, piqued, piquing, piques, irritate, irritated, irritating, irritates, aggravation, aggravate, aggravated, aggravating, exasperation, exasperate, exasperated, exasperating, harassment, harass, harassed, harassing, harasses, torment, tormented, tormenting, displeasure, displeased, displease, resent, resentful, resentfully, resentment, antagonistic, antagonist, antagonism, provoke, provoked, provoking, provokes, hassle, hassled, hassling, burn, burnt, burned, burning, burns, explode, exploding, explodes, explosion, fume, fumed, fuming, seethe, seethed, seething, aggressive, aggressively, aggressiveness, aggression	128
Disgust	disgust, disgusts, disgusting, disgusted, disgustedly, disgustedness, dislike, disliked, disliking, dislikes, dislikeable, dislikable, distaste, distasted, distasting, distastes, distasteful, distastefully, distastefulness, flush, revolt, revolted, revolting, revolts, revulsion, vomit, vomited, vomiting, vomits, repel, repelled, repelling, repels, sicken, sickening, sickens, sickly, sickness, sickeningly, abhorrence, abhorrency, abhor, abhorrent, abhorred, abomination, abominate, detestation, execration, loathing, loathingly, loathe, loathsome, odium, repugnance, repugnant, repugnancy, repugnantly, repulse, repulsion, repulsive, repulsively, nausea, nauseate, nauseated, nauseating, nauseates, nauseation, nauseatingly, hatred, hate, hateful, hatefully, hatefulness, hated, hates, hating, averse, aversion, aversely, averseness, antipathy	81
Fear	fear, feared, fearing, fears, fearful, fearfulness, fearfully, horror, horrors, horrible, horribly, horrid, horrific, horrify, horrifying, afraid, scared, scare, scaring, scares, frightened, frighten, frightening, frightens, frightful, fright, frighted, frighting, frights, panic, panicked, panicking, panics, panicky, terror, terrorism, terrorist, affright, dread, dreadful, dreadfully, dreadfulness, anxious, anxiously, anxiousness, anxiety, apprehensive, apprehensively, apprehensiveness, apprehension, alarmed, alarming, alarmingly, dismay, dismayed, dismaying, dismays, consternation, agitation, agitate, agitated, shiver, shivered, shivering, shivers, chill, chilled, chilling, quiver, quivered, quivering, shudder, shuddered, shuddering, worry, worried, worrying, worries, concern, concerns, concerned, concerning, trouble, troubled, troubling, troubles, troublesome, uneasy, unease, uneasily, uneasiness, tremor, tremors, tremorous,qualm, qualms, consternation, trepidation, trepidatious, timid, timidly, timidity, timidness, craven, cravenly, cravenness, funk, funks, funked, funking, creep, crept, creeping, creepy, attack, attacked, attacking, attacks, timorous, timorousness, timorously, intimidate, intimidated, intimidate, intimidation	125
Sadness	sad, sadness, sadly, sadden, depress, depressed, depressing, depression, sorrow, sorrows, sorrowful, sorrowfully, unhappy, unhappily, unhappiness, bitter, bitterly, bitterness, doleful, dolefully, dolefulness, mourn, mourned, mourning, mournful, mournfully, mournfulness, melancholy, melancholic, melancholily, pensive, pensively, pensiveness, wistful, wistfully, wistfulness, tired, tiredness, tiring, tiresome, tiresomely, tiresomeness, despair, despaired, despairing, desperate, desperately, desperation, deplorable, deplorably, distress, distressed, distressing, distresses, lamentable, lamentably, pity, pitiful, pitifully, pitifulness, piteous, piteously, piteousness, pitiable, pitiably, sorry, sorrily, sorriness, gloom, gloomy, gloomily, gloominess, grieve, grieved, grieving, grievously, grievousness, dismal, dismally, dismalness, sombre, sombrely, sombreness, sombrous, glum, glumly, glumness, dejected, dejectedly, dejection, downcast, grief-stricken, tear, tearful, tearfully, tearfulness, lugubrious, lugubriously, lugubriousness, pensive, pensively, pensiveness, disconsolate, disconsolately, disconsolation, heavy-hearted, cheerless, cheerlessly, cheerlessness, lachrymose, woebegone, low-spirited, triste, tristful, tristfully, tristfulness, tragic, tragical, tragically, touching, touchingly, moving, movingly, upset, upsetting, upsets, upsettingly, disastrous, disastrously, disastrousness, pathetic, pathetically, poignant, poignance, poignancy, poignantly, harrowing, heart-rending, terrible, terribleness, terribly, unfortunate, unfortunately, miserable, miserably, misery, miseries, heartbreak, heartbreaking, heartbreaker, wretched, wretchedly, wretchedness, desolate, desolated, desolating, desolation, desolately, dispirit, dispirited, dispiriting, dispirited, dispiritedly, downhearted, spiritless, dolorous, dolorously, plaintive, plaintively, rueful, ruefully, woeful, woefully, joyless	174

Table 3: Top-10 attribute values of scene type, object class and human action for different emotions.

Emotion	Attribute	Top-10 Attribute Values
Amusement	Scene	Carousel, Amusement park, Delicatessen, Water park, Park, Toyshop, Lawn, Playground, Restaurant, Playroom
	Object	Christmas tree, Doll, Tomato, Wine, Cake, Juice, Snack, Fast food, Plate, Billboard
	Human action	Decorating the christmas tree, Belly dancing, Dining, Eating ice cream, Cheerleading, Flying kite, Paragliding, Water sliding, Tasting beer, Crawling baby
Awe	Scene	Mountain snowy, Valley, Waterfall, Sky, Lake, Ocean, Mountain, Canyon, Coast, Church
	Object	Castle, Airplane, Eagle, Sheep, Bee, Lighthouse, Maple, Skyscraper, Falcon, Horse
	Human action	Answering questions, Driving car, Skiing, Sailing, Paragliding, Wrestling, Rock climbing, Ice climbing, Yoga, Walking the dog
Contentment	Scene	Lawn, Beach, Park, Pond, Lake, Field, Pasture, Ocean, Forest path, Swimming pool
	Object	Goose, Coffee, Duck, Pillow, Coffee cup, Bench, Mug, Couch, Saucer, Rabbit
	Human action	Baby waking up, Petting cat, Reading book, Walking the dog, Texting, Yoga, Petting animal (not cat), Using computer, Writing, Crawling baby
Excitement	Scene	Stage, Wave, Football field, Athletic field, Stadium, Discotheque, Orchestra pit, Park, Ski slope, Arena
	Object	Surfboard, Guitar, Paddle, Football, Snowboard, Personal flotation device, Bicycle helmet, Drum, Canoe, Ski
	Human action	Water skiing, Playing ice hockey, Side kick, High kick, Kitesurfing, Tapping guitar, Snowboarding, Shooting goal (soccer), Swinging legs, Catching or throwing baseball
Anger	Scene	Boxing ring, Plaza, Forest, Tree farm, Lawn, Bullring, Field, Crosswalk, Hayfield, Forest path
	Object	Lion, Leopard, Tiger, Jaguar, Brown bear, Crocodile, Bear, Weapon, Lynx, Traffic light
	Human action	Extinguishing fire, Finger snapping, Pumping fist, Punching bag, Reading newspaper, Answering questions, Shaking head, Using computer, Pushing cart, Brushing teeth
Disgust	Scene	Landfill, Underwater, Butchers shop, Delicatessen, Alley, Market, Slum, Boardwalk, Campsite, Crosswalk
	Object	Caterpillar, Waste container, Snake, Plastic bag, Reptile, Invertebrate, Worm, Marine invertebrates, Toilet, Beetle
	Human action	Garbage collecting, Smoking, Unloading truck, Holding snake, Cleaning toilet, Sweeping floor, Carving pumpkin, Digging, Eating cake, Extinguishing fire
Fear	Scene	Corridor, Staircase, Forest, Alley, Ruin, Corn field, Cottage, House, Natural history museum, Forest path
	Object	Skull, Lantern, Stairs, Pumpkin, Bust, Weapon, Bronze sculpture, Lizard, Rifle, Scorpion
	Human action	Holding snake, Scuba diving, Texting, Bandaging, Changing wheel, Finger snapping, Driving car, Smoking, Drinking beer, Archery
Sadness	Scene	Cemetery, Mausoleum, Ruin, Stable, Corridor, Staircase, Doorway, Forest, Slum, House
	Object	Bust, Bronze sculpture, Handbag, Monkey, Bench, Stairs, Curtain, Picture frame, Fedora, Couch
	Human action	Massaging persons head, Smoking, Digging, Blowing nose, Tasting beer, Fixing hair, Dining, Arranging flowers, Driving car, Brushing hair

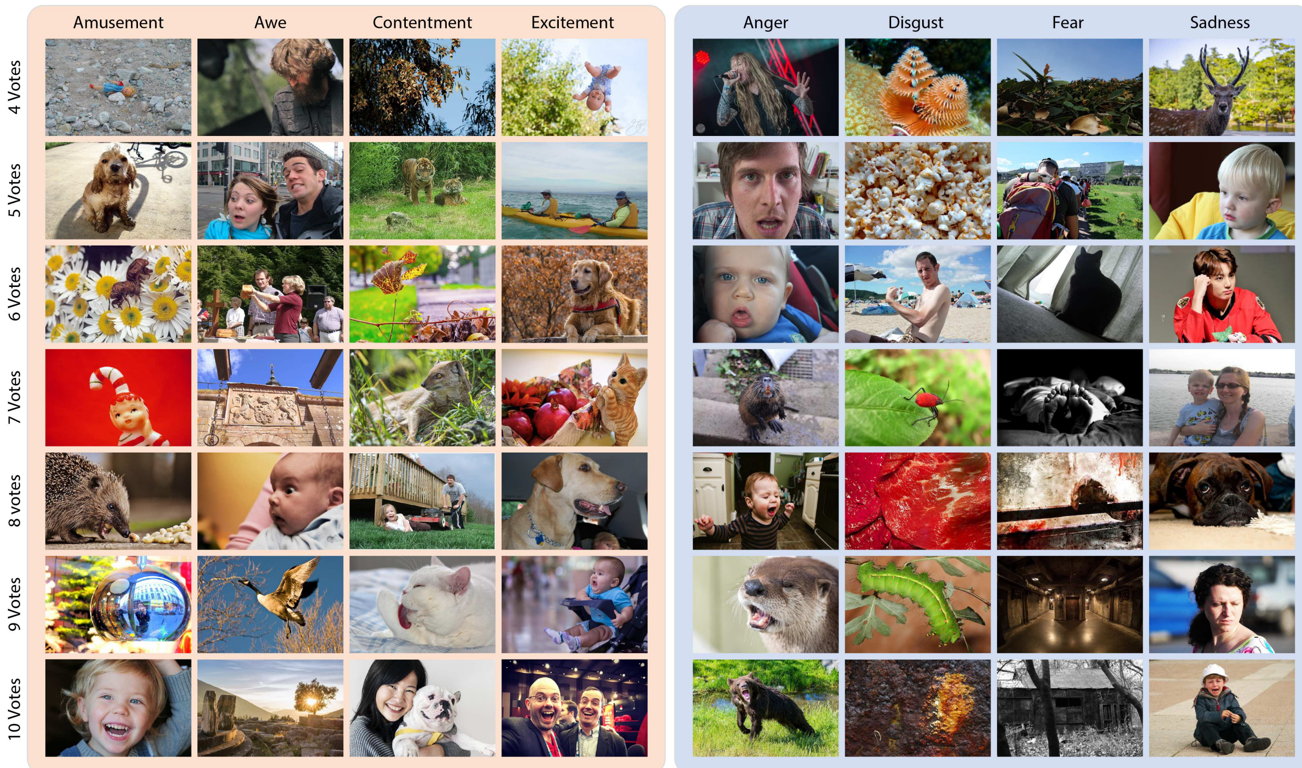

Figure 1: Images in EmoSet voted by different number of people, where the higher the number, the more agreements people may reach in emotion.

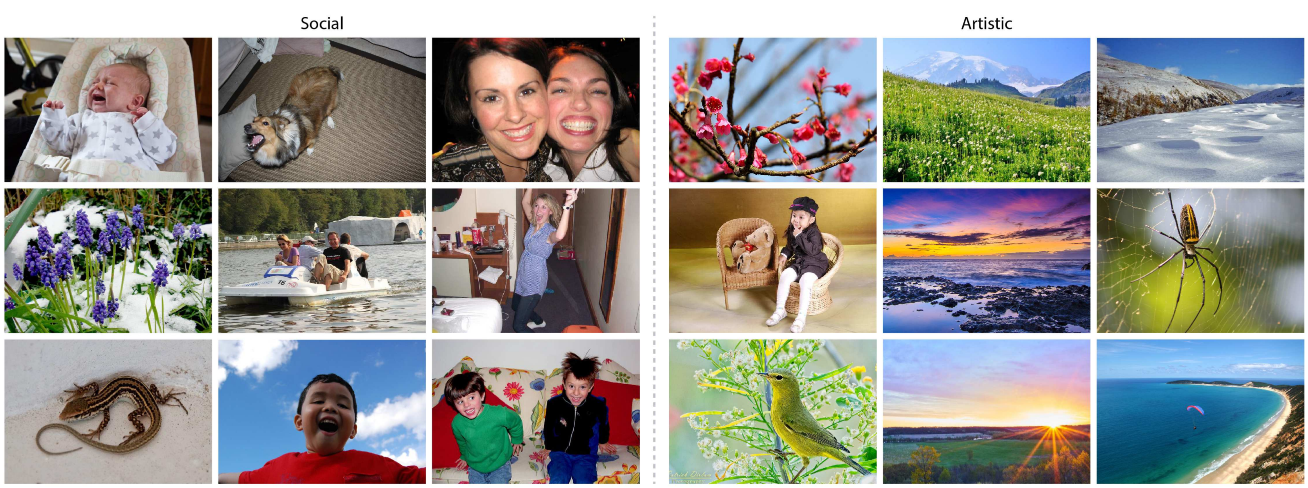

Figure 2: Diverse image types in EmoSet, covering both social images uploaded by internet users and artistic photos taken by professional photographers.

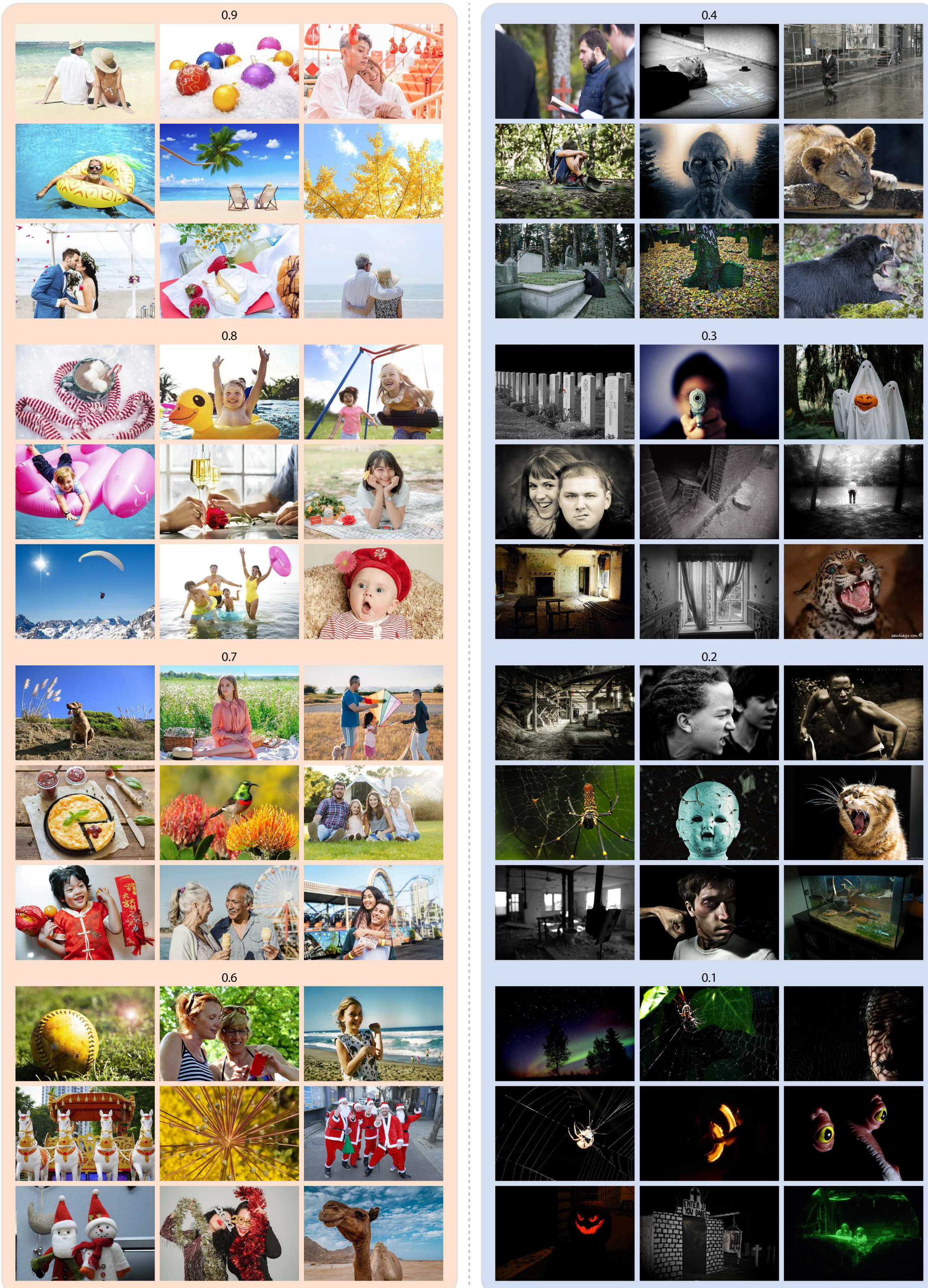

Figure 3: Visualizations on Emotion-Brightness, where the higher the brightness value, the more positive emotion the image may evoke.

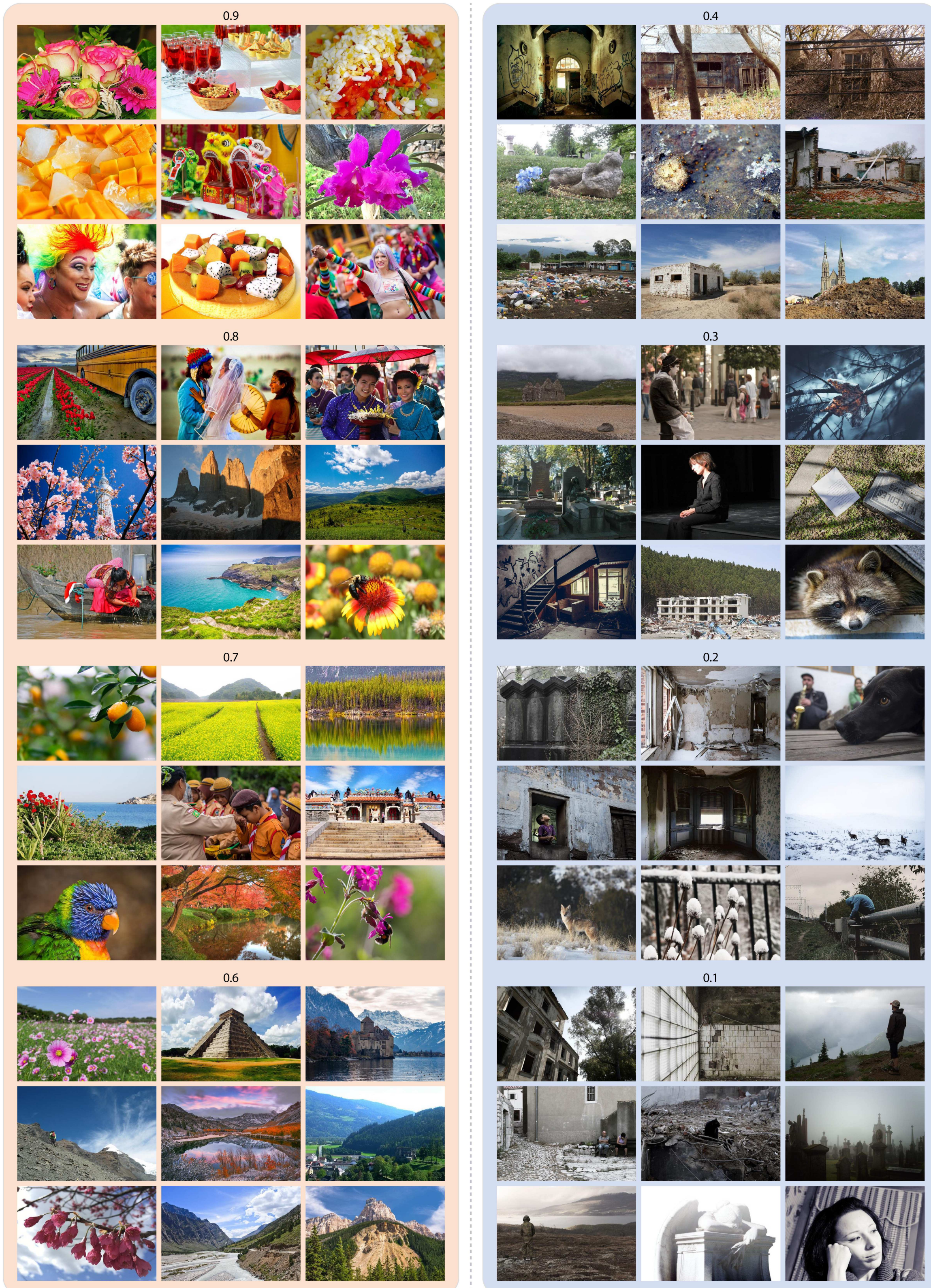

Figure 4: Visualizations on Emotion-Colorfulness, where the higher the colorfulness value, the more positive emotion the image may evoke.

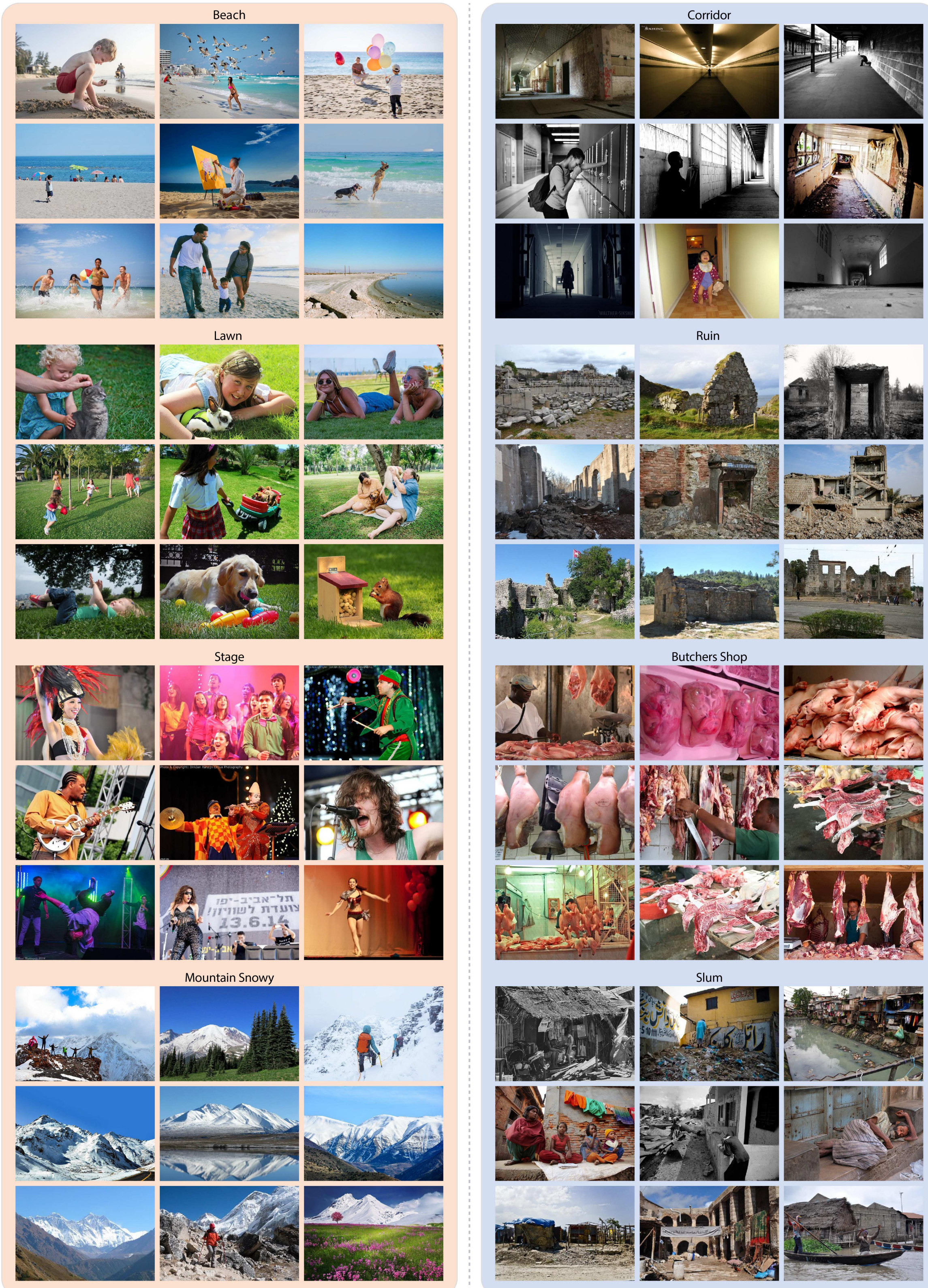

Figure 5: Visualizations on Emotion-Scene type, where some scene types are emotion-related.

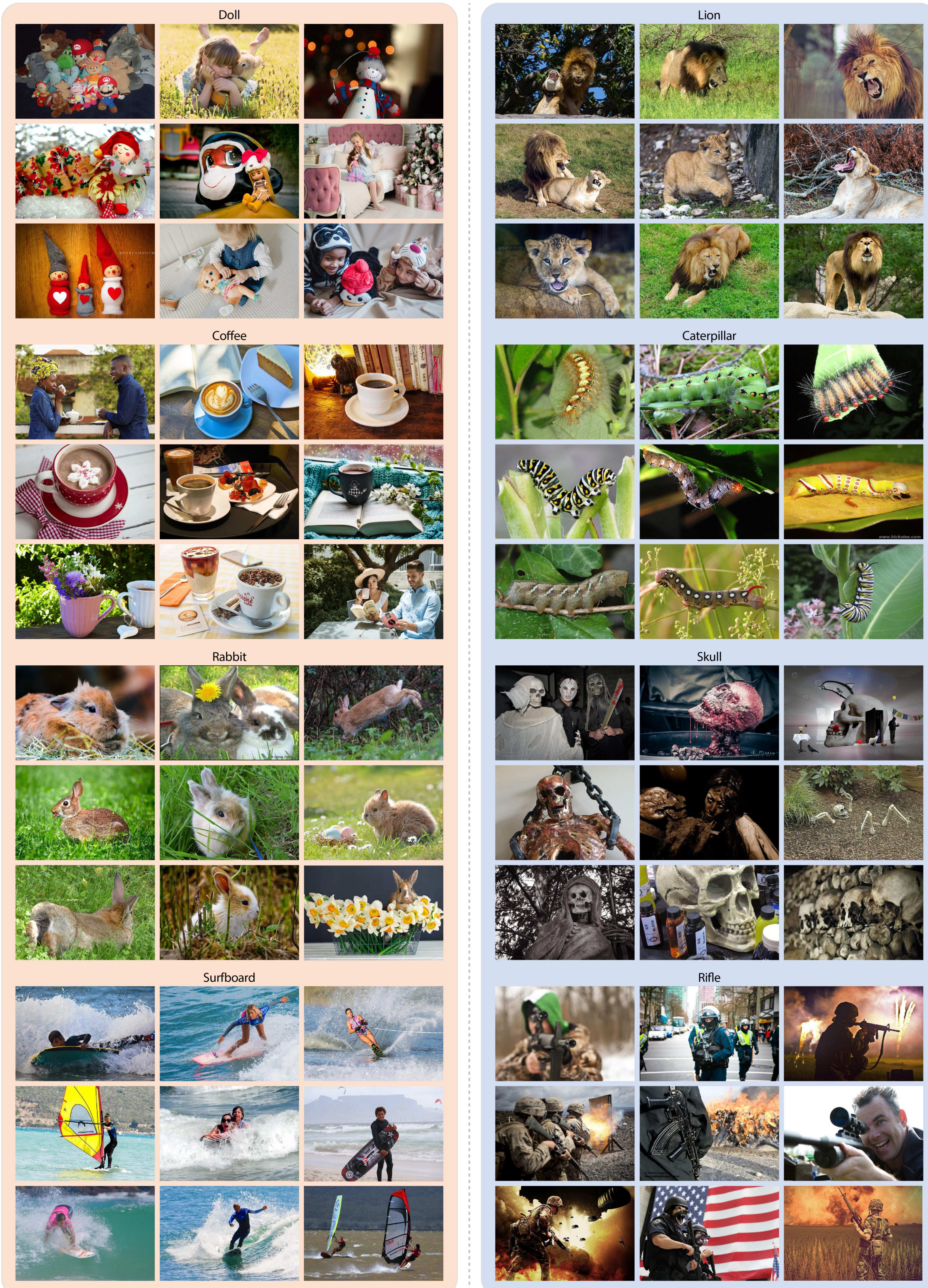

Figure 6: Visualizations on Emotion-Object class, where some object classes are emotion-related.

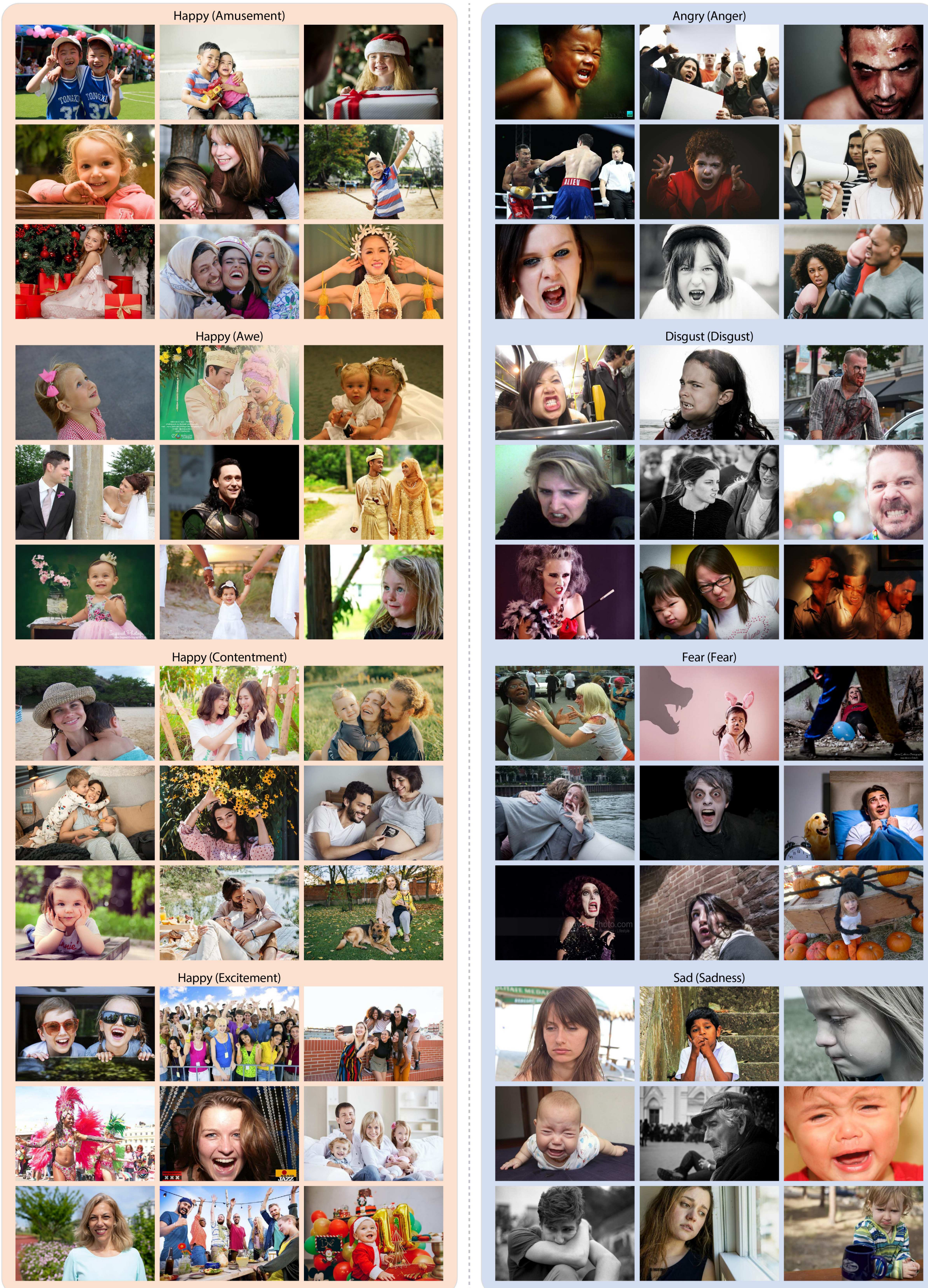

Figure 7: Visualizations on Emotion-Facial expression, where some facial expressions are decisive to emotion evocation.

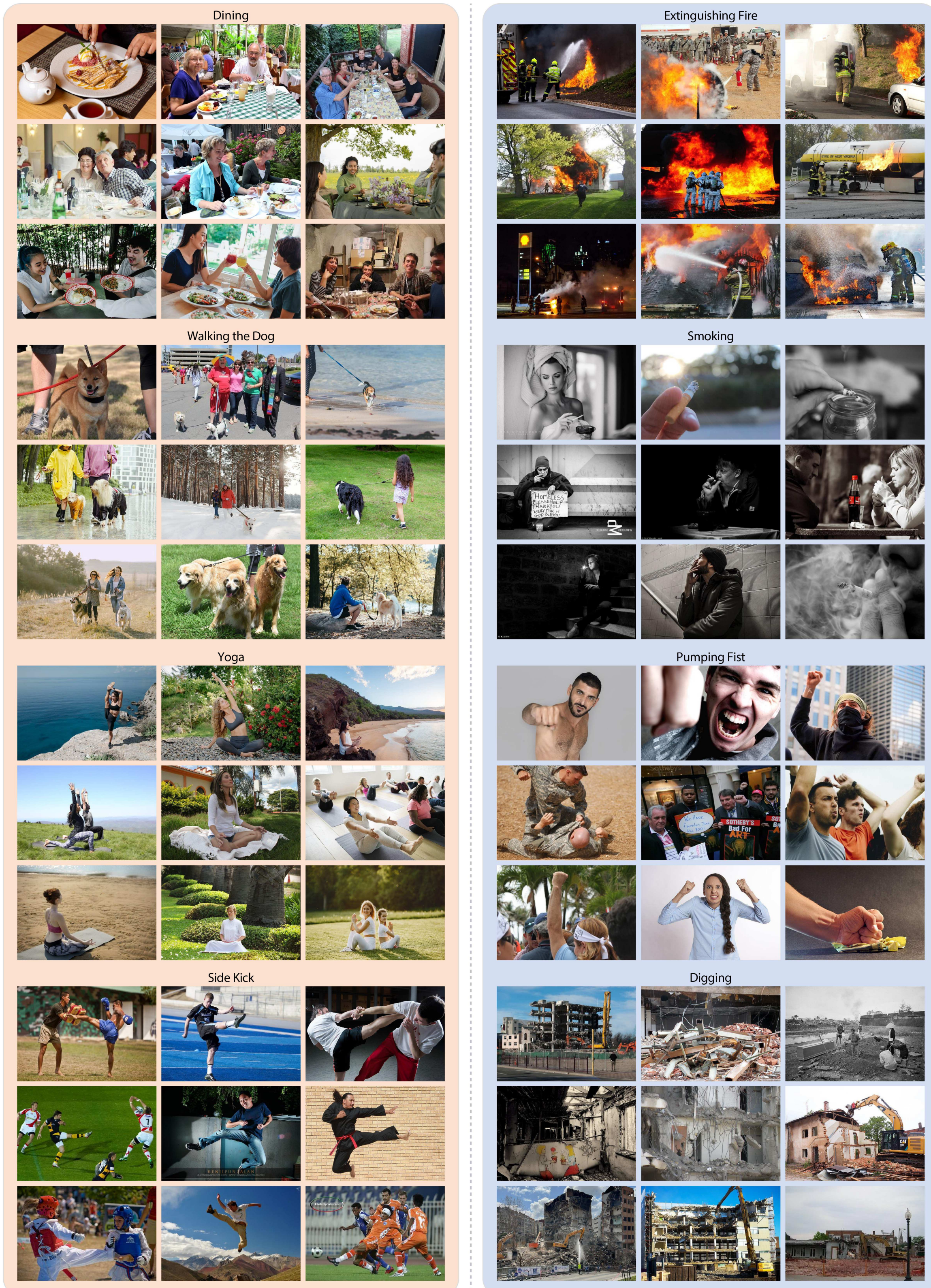

Figure 8: Visualizations on Emotion-Human action, where some human actions are emotion-related.