

Appendices

A. Hyperparameters

All the trainings were done on 8 Nvidia A100 GPUs.

- BrandVLM training: Batch size 1024, Learning rate $2e-5$, Number of epochs 1.5, AdamW Optimizer, Cosine learning rate scheduler with warmup ratio 0.03
- BrandDM training
 1. PixArt- Σ : Batch size 1024, Learning rate $2e-5$, Number of epochs 40, CAME optimizer, Constant learning rate scheduler
 2. SDXL: Batch size 1024, Learning rate $1e-6$, Number of epochs 20, AdamW Optimizer, Constant learning rate scheduler
 3. SDXL-IPAdapter: Batch size 32, Learning rate $1e-6$, Number of epochs 20, AdamW Optimizer, Constant learning rate scheduler
- Brand-Classifier training (Used for evaluation): Batch size 32, Learning rate $1e-3$, Number of epochs 80, AdamW Optimizer, 42K parameters, training strategy – to obtain (brand style attribute labels, brand) pairs during training, for each brand’s train images (forming correct pairings), we sample an equal number of images from other brands of the same industry sector and pass the brand as input to get incorrect pairings
- BrandCLIP training: OpenAI’s ViT-L-14-336 base model, Batch size 128, Learning rate $1e-3$, Number of epochs 5, AdamW Optimizer, 0.1 weight decay

B. BrandVLM performance comparison

Model version	Verbalisation evaluation			Embedding evaluation	
	BLEU score (\uparrow)	IoU score (\uparrow)	BAS (\downarrow)	MSE (\downarrow)	Cosine similarity (\uparrow)
LLaVA-1.5-13B	0.0788	0.1859	16.1064	-	-
BrandVLM	0.4569	0.5499	3.3752	0.5102	0.5598

Table 3. Different VLM performances for generating image style characteristics (verbalization) and brand-relevant image style embeddings on the full test set. The non-finetuned baseline does not predict the style embeddings. The IoU score is the Intersection over Union of the style labels predicted by the VLM and style labels present in the ground truth verbalizations. BAS refers to the Brand Alignment Score as explained in Section 5.1.

C. Prompt used for training BrandVLM

A marketer from company `company` which belongs to the sector `companySector` wants to create a social media post containing an image for marketing purposes. The following information about the social media post is available :

- (1) Social media post text: `tweetText`
- (2) Date of posting: `tweetTimeStamp`
- (3) Number of likes on a scale of 0 to 100 that the social media post is expected to receive: `tweetLikesPercentile`
- (4) Image description: `imCaption`
- (5) Image tags: `imKeywords`.

Now, considering the company’s visual identity and the above information, predict the colors and tones describing the image that the marketer will use in the social media post from the lists given below. Also predict the spatial coverage ratios (with respect to the total image area) of the colors and tones that will be used.

- **Allowed colors:** [Beige, Black, Blue, Bright_Green, Brown, Cream, Cyan, Dark_Blue, Dark_Brown, Dark_Gray, Dark_Green, Dark_Pink, Dark_Red, Emerald, Gold, Gray, Green, Khaki, Lavender, Light_Blue, Light_Green, Lilac, Magenta, Maroon, Mud_Green, Mustard, Off_White, Olive, Orange, Pink, Plum, Purple, Red, Royal_Blue, Silver, Tan, Turquoise, Violet, White, Yellow]

- **Allowed tones:** [warm, neutral, cool]

Now, for each of the following 8 visual attributes, predict at most 3 labels which should be the most prominent in the image out of the given list of labels.

- (1) **Image Lighting:** [bright, dark, moderate, studio, natural, soft, hard, light glare, vignette, colored, light on subject]
- (2) **Perspective:** [bird eye view, worm eye view, fish eye view, panorama view, centered composition, rule of third, altered perspective, framed image, high angle photo, low angle photo, vertical composition, corner shot, point of view shot, audience perspective]
- (3) **Image Background:** [solid, pattern, gradient, background as frame, textured, wood, blurred, transparent, bright, dark, light]
- (4) **Color Palette:** [grayscale, monotone, two tone, bright colors, pastel colors, complementary colors, analogous colors, inverted colors, galaxy colors, aquatic colors, sunset colors, autumnal colors]
- (5) **Photography Genre:** [architectural, candid, staged, portrait, selfie, group, product, fashion, beauty, bridal, interior, street, landscape, sky, still-life, action, underwater, botanical, historical, amateur, abstract, live

```

stage]
(6) Concept: [illustration, photorealism,
typography, vintage, graphic design,
cartoon, incomplete art, wave pattern,
text heavy]
(7) Depth: [wide angle shot, mid shot,
close up shot, macro shot, motion blur,
radial blur, gaussian blur, fully
focused subject, unfocused subject,
partly focused subject, bokeh effect,
isolated focal point, multiple focal
points, bright focal point, dark focal
point, shallow depth of field]
(8) Image Effects: [short exposure, long
exposure, neutral density filter,
artificial shadow, silhouette, pixelated
image, vanishing point, negative space,
motion capture, cut-out, symmetric,
asymmetry, low saturation, high
saturation, low contrast, high contrast]
If the main subject of this image contains
a human, then for each of the following
7 attributes, predict at most 3 labels
out of the provided list which should be
relevant to the subject in the image,
otherwise predict 'Not applicable' for
each of these attributes.
(1) Hair Style: [short, covered, wavy,
loose, varied, straight, neat, ponytail,
casual, tied back, flowing, curly, updo
, pulled back, braided, Not applicable]
(2) Facial Expression: [engaged, content,
focused, neutral, joyful, relaxed,
contemplative, Not applicable]
(3) Clothing Style: [casual, athletic,
formal, business, swimwear, business
casual, traditional, protective,
beachwear, costume, form fitting, Not
applicable]
(4) Clothing Color Palette: [neutral,
colorful, vibrant, monochrome, earthy,
pastel, muted, Not applicable]
(5) Posing: [standing, seated, holding,
leaning, active, reclined, walking,
stretching, dynamic, running, relaxed,
confident, Not applicable]
(6) Gaze: [forward, downward, sideways,
away, upward, outward, engaged, Not
applicable]
(7) Visible Body section: [upper body, full
body, hand only, lower half, close up,
midsection, full back, head shot, Not
applicable]
Answer properly in JSON format with the
following keys - "colors_and_tones", "
image_lighting", "perspective", "
image_background", "color_palette", "
photography_genre", "concept", "depth",

```

```

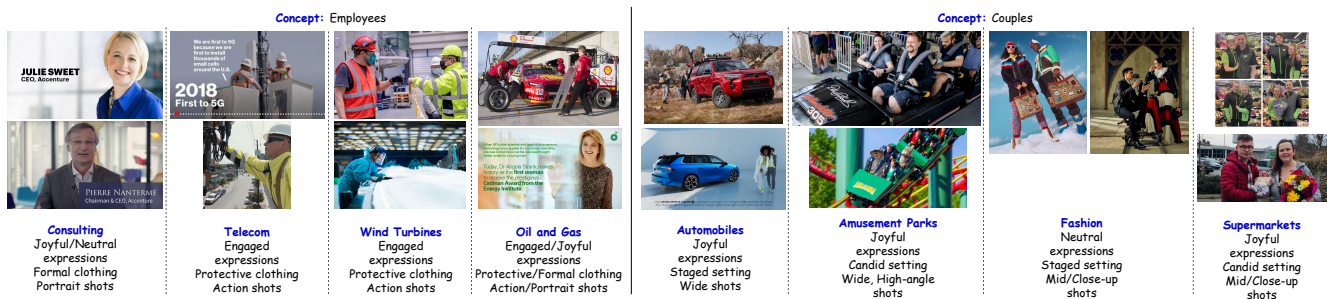
"image_effects", "hair_style", "
facial_expression", "clothing_style", "
clothing_color_palette", "posing", "gaze
", "visible_body_section". The values of
the JSON should be in a dictionary for
colors_and_tones and a list for all
other keys. Do not include any other
information in your answer.

```

Listing 1. Prompt used while training BrandVLM

Considered Brand Sectors	
Aerospace, Agricultural Heavy Equipment, Airline, Amusement Park, Automobile, Beauty, BioTech, Brewery, Car rental, Construction, Consulting, Consumer goods, Cruise, Defense, Drink, E-commerce, Education, Electronics, Entertainment, Eyecare, Fashion, Finance, Fitness, Food, Footwear, Gaming, Gas, Hardware, Healthcare, Home appliances, Homecare, Hospitality, Insurance, Jewelry, MLM, Networking, NGO, Oil, Parcel service, Pet Supermarket, Petrol station, Petroleum, Pharma, RailRoad, Research, Restaurant, Ride sharing, Satellite, Software, Sports, Supermarket, Telecom, Tires, Tourism, Underwater Diving, Watches, Wind Turbines	
Characteristic	Labels
Image Lighting	Bright, Dark, Moderate, Studio, Natural, Soft, Hard, Light glare, Vignette, Colored, Light on subject
Perspective	Bird eye view, Worm eye view, Fish eye view, Panorama view, Centered composition, Rule of third, Altered perspective, Framed image, High angle photo, Low angle photo, Vertical composition, Corner shot, Point of view shot, Audience perspective
Image Background	Solid, Pattern, Gradient, Background as frame, Textured, Wood, Blurred, Transparent, Bright, Dark, Light
Color Palette	Grayscale, Monotone, Two tone, Bright colors, Pastel colors, Complementary colors, Analogous colors, Inverted colors, Galaxy colors, Aquatic colors, Sunset colors, Autumnal colors
Photography Genre	Architectural, Candid, Staged, Portrait, Selfie, Group, Product, Fashion, Beauty, Bridal, Interior, Street, Landscape, Sky, Still-life, Action, Underwater, Botanical, Historical, Amateur, Abstract, Live stage
Concept	Illustration, Photorealism, Typography, Vintage, Graphic design, Cartoon, Incomplete art, Wave pattern, Text heavy
Depth	Wide angle shot, Mid shot, Close up shot, Macro shot, Motion blur, Radial blur, Gaussian blur, Fully focused subject, Unfocused subject, Partly focused subject, Bokeh effect, Isolated focal point, Multiple focal points, Bright focal point, Dark focal point, Shallow depth of field
Image Effects	Short exposure, Long exposure, Neutral density filter, Artificial shadow, Silhouette, Pixelated image, Vanishing point, Negative space, Motion capture, Cut-out, Symmetric, Asymmetry, Low saturation, High saturation, Low contrast, High contrast
Hair Style	Short, Covered, Wavy, Loose, Varied, Straight, Neat, Ponytail, Casual, Tied back, Flowing, Curly, Updo, Pulled back, Braided
Facial Expression	Engaged, Content, Focused, Neutral, Joyful, Relaxed, Contemplative
Clothing Style	Casual, Athletic, Formal, Business, Swimwear, Business casual, Traditional, Protective, Beachwear, Costume, Form fitting
Clothing Color Palette	Neutral, Colorful, Vibrant, Monochrome, Earthy, Pastel, Muted
Posing	Standing, Seated, Holding, Leaning, Active, Reclined, Walking, Stretching, Dynamic, Running, Relaxed, Confident
Gaze	Forward, Downward, Sideways, Away, Upward, Outward, Engaged
Visible Body Section	Upper body, Full body, Hand only, Lower half, Close up, Midsection, Full back, Head shot

Appendix D. The Brand sectors, Brand Style Characteristics and their corresponding labels considered in our approach. For details please refer to Section 3.



Appendix E. Here, we show more examples of nuanced Brand variations. Different brands from a particular sector can have some common characteristics, such as *Protective clothing style* of *employees* in *Telecom* sector, or *Joyful facial expressions* and *Candid setting* images of *couples* in *Amusement Parks* sector.