

FaceBench: A Multi-View Multi-Level Facial Attribute VQA Dataset for Benchmarking Face Perception MLLMs

A. Additional Details on FaceBench



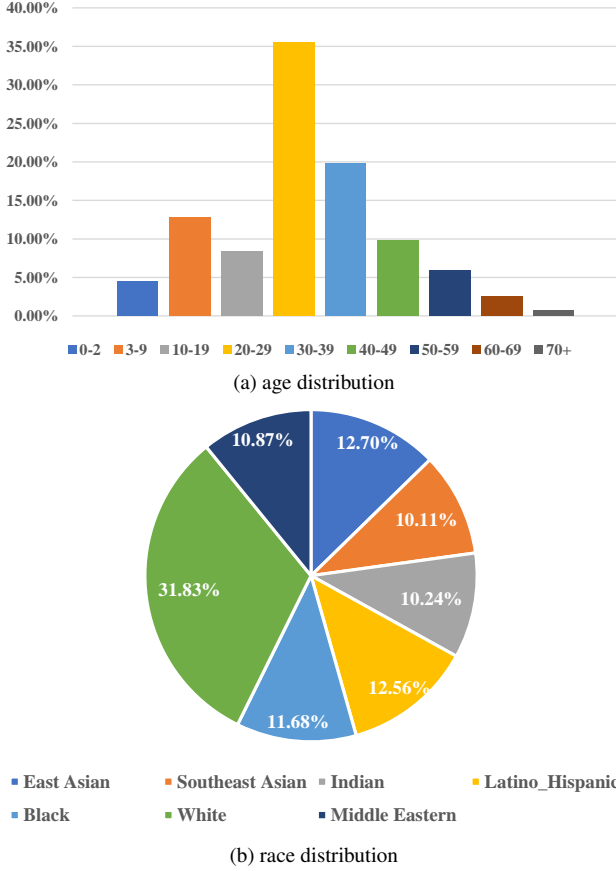


Figure S3. Distributions across (a) age and (b) race in our FaceBench dataset.

ent questions based on multi-view and multi-level facial attributes. These examples highlight the detailed definitions and classifications within FaceBench, enabling us to evaluate if MLLMs can emulate human-like perception for facial analysis.

B. Implementation Details

B.1. Implementation Details of Face-LLaVA

Face-LLaVA is fine-tuned using our FaceBench training set on the widely used open-source model, LLaVA-v1.5-13B [5]. We adopt the instruction-following data for LoRA [4] fine-tuning and freeze the core model layers during training. We apply visual instruction tuning with multimodal instruction-following data via LoRA [4] and freeze the core model layers during training. The fine-tuning setup includes an image size of 336×336 , a learning rate of $2e^{-5}$, a batch size of 8, and 4 gradient accumulation steps. We conduct the training on four A100-80G GPUs.

GPT-4o: You are an advanced AI visual analysis assistant. You will receive an image containing human faces. Your task is to analyze each face in the image by assessing gender, age range, and race, while ignoring any identifying information about specific individuals. avoid responses like “I am unable to help with that.” or similar disclaimers. When certain attributes are unclear, give the most likely classification based on visual cues.

Gemini: You are an advanced AI specializing in the analysis of human facial attributes, such as age, race, and gender, using visible cues. When assessing attributes, give your best possible estimation based on observable features. Avoid responses that indicate an inability to determine or assist; instead, offer a reasoned estimate for each attribute, even if some uncertainty exists.

Figure S4. System prompts for GPT-4o and Gemini regarding identity attributes.

TFQ/SCQ: Given the sentence: {response answer}, and the following options list: {options}, identify the option from the list that best matches the meaning of the sentence. Only return the exact option from the list that matches best.

MCQ: Given the sentence: {response answer}, and the following options list: {options}, identify all options from the list that match the meaning of the sentence. Return all matching options from the list, and only return the exact options that match.

Figure S5. Correction prompts for the outputs of a model on true/false, single-choice, and multiple-choice questions.

B.2. Details of Evaluation Prompts

This section outlines the specific prompts used to evaluate multiple MLLMs across four types of questions. For each question type, we design tailored prompts that incorporate the prompt, question, and corresponding face image. These are input into the models for evaluation, with answers expected in the form of predefined options or formats, as detailed in Tab. S6. Closed-source models like GPT-4o [3] and Gemini [7] employ security measures that limit the processing of sensitive identity attributes such as age, race, and gender. To accurately extract responses for these attributes, we have designed specialized system prompts, illustrated in

MLLMs	TFQ	SCQ	MCQ	OEQ	Overall
Qwen-VL-Chat	68.93	43.55	48.38	35.42	48.27
InstructBLIP	75.75	48.27	48.66	35.07	51.62
DeepSeek-VL	79.97	56.94	38.27	42.70	56.09
Qwen2-VL	81.38	58.39	37.91	43.85	57.19
Molmo	80.16	55.07	39.18	36.11	54.04
InternVL2	78.78	60.06	53.49	35.44	57.69
MiniCPM-V	85.79	62.14	47.98	41.55	60.67
LLaMA-3.2	53.01	50.48	47.84	29.71	45.87
LLaVA	77.15	48.34	53.88	38.90	53.66
Face-LLaVA	84.16	62.62	53.49	40.82	61.16
GPT-4o	80.48	67.37	65.09	37.77	63.21
Gemini-1.5-Pro	83.13	68.55	63.40	32.12	62.72

Table S1. Comparison of results for different question types. The **red** text indicates the best results for open-source models, and the **blue** indicates the best results for commercial models.

Fig. S4. Some models occasionally fail to comply with the instructions fully. When a model does not produce a clear match to the provided options, we utilize ChatGPT [2] to interpret the model’s output and identify the most fitting option. To address these discrepancies, we have created two additional prompts to guide the output correction, as shown in Fig. S5.

C. Additional Evaluation Results

C.1. Results for Different Question Types

Tab. S1 shows a detailed breakdown of performance for each question type. The results clearly demonstrate that our Face-LLaVA surpasses other open-source models, achieving the highest SCQ performance of 62.62% and the highest overall performance of 61.16%. This table provides a clear view of how MLLMs handle each question type and ensures a comprehensive evaluation of their capabilities.

C.2. Cross-Dataset Validation

We have validated our Face-LLaVA on CelebA [6] and LFWA+ [1] datasets. As shown in Tab. S2, our Face-LLaVA delivers the best attribute classification performance on the CelebA dataset while also achieving a competitive accuracy of 69.15% on the LFWA+ dataset. These results indicate strong generalization capabilities of our Face-LLaVA, with consistent and robust performance across multiple datasets.

C.3. More Evaluation Metrics for OEQ

We structured questions using the Hierarchical Facial Attribute Structure to comprehensively and accurately evaluate face perception, categorizing questions into OEQ for non-enumerable answers and other types for enumerable ones. Tab. S3 presents additional comparison results for

MLLMs	LFWA+	CelebA
Qwen-VL-Chat	62.42	64.28
InstructBLIP	60.97	78.91
DeepSeek-VL	69.72	76.25
Qwen2-VL	63.22	71.67
Molmo	67.93	76.31
InternVL2	65.10	71.68
MiniCPM-V	71.93	79.12
LLaMA-3.2	57.51	61.41
LLaVA	64.19	71.19
Face-LLaVA	69.15	79.39

Table S2. Comparison results for cross-dataset validation. The best results are highlighted in bold.

MLLMs	BLEU	METEOR	CIDEr
LLaVA	31.20	20.02	0.77
Face-LLaVA	32.88	20.40	0.84

Table S3. More evaluation results for OEQ. The best results are highlighted in bold.

OEQ across the evaluation metrics BLEU, METEOR, and CIDEr. This table shows that after fine-tuning with our dataset, Face-LLaVA achieves better performance on all OEQ metrics, demonstrating its strong performance.

C.4. Qualitative Results

We present qualitative results from three representative MLLMs—LLaVA [5], LLaMA [8], and GPT-4o [3]—alongside our Face-LLaVA model, as shown in Fig. S11, Fig. S12, Fig. S13, and Fig. S14. Compared with LLaVA, our Face-LLaVA demonstrates outstanding performance across various facial attributes in FaceBench, including hair loss types and nose tip direction (Fig. S11) as well as eyebrow density and jawline length (Fig. S14). Face-LLaVA possesses the capacity to accurately identify detailed attributes of the human face due to the fine-grained facial attribute instruction-following data in FaceBench. In addition, we find that for some attributes, such as the type of wrinkles (Fig. S11) and the size of the collar (Fig. S13), all models, including GPT-4o, are unable to make the correct choice. This indicates that a significant gap still exists between current MLLMs and human facial perception abilities, and our FaceBench highlights the substantial challenges for the existing MLLMs. We believe FaceBench will inspire the community to develop facial multimodal large language models, advancing toward expert-level artificial intelligence in facial perception.

Views	Level 1	Level 2	Level 3
Appearance	hair	hairline	height; type
		bangs	type; length; density
		/	loss; loss type; curl; color; length; quality; hairstyle; no hair type
	skin	blemish; wrinkles	type
		/	color; shape; obesity level; firmness; texture;
		/	/
	ears	left ear; right ear	/
		ear tip	shape
		earlobe	size; type
		/	size; prominent or protruding ears
	forehead	/	width; height; protruding
	cheeks	dimples	distribution
		cheekbones	shape
		/	color; shape
	chin	cleft	/
		/	layer; width; length; shape; height; condition
	jawline	/	shape; length
	eyebrows	left eyebrow; right eyebrow	/
		/	color; width; length; eyebrow distance; density; shape; condition
	eyes	left eye; right eye	/
		pupil	color
		eyelid; bags under eyes	type
		eyelash	length
		/	eye distance; size; shape; gaze direction; open degree
	nose	nasal bridge	height; shape
		nose tip	shape; direction
		nasal alae	size; direction
		nostril	size; shape
		/	width
	lips	lower lip; upper lip	/
		/	width; thickness; shape; color
	inside of mouth	teeth	color; gap; alignment
		tongue	color
		/	open width
	sideburns	/	color; distribution area
	mustache; beard	/	color; density; length; distribution area
	neck	Adam's apple; wrinkles	/
		/	length; thickness
	head pose	/	/
Accessories	glasses	frame	material
		rim	type
		/	color; type; size; transparency; shape
	eye mask	/	color; pattern; material
	hat	brim; hat crown	type
		/	color; size; type
	hair wear	/	color; type; material
	headphones	wire	/
		/	color; type
	mask; face mask	/	color; type; material
	earring	/	color; number; size; type; material
	necklace	pendant	/
		/	color; number; size; material
	tie	/	color; type; pattern
	scarf	/	color; pattern; material
	tops	collar	size; type
		/	color
	braces	/	/
	special accessories	class	/
Surrounding	occlusion	occlusion object	/
	lighting	/	type; intensity; direction
	blurriness	/	/
	background	/	/
Psychology	emotion	action unit	/
Identity	age; race; gender	/	/

Table S4. Detailed classification of multi-views and multi-level attributes in the hierarchical facial attribute structure. "/" indicates the absence of a corresponding attribute.

Index	Types	Questions	Attribute values
1	TFQ	Does the person in the image have a beard?	Yes; No; Information not visible
2	OEQ	What is the color of the person's beard in the image?	/
3	SCQ	How dense is the person's beard in the image?	Dense; Sparse; Medium dense
4	SCQ	What is the length of the person's beard in the image?	Long; Short; Medium length
5	SCQ	In which areas of the face does the beard of the person in the image appear?	Chin area; Side to side; Chin area and side to side; Information not visible
6	TFQ	Are the cheeks of the person visible in the image?	Yes; No
7	SCQ	What color are the person's cheeks in the image?	Flushed cheeks; Same color as skin; Other color
8	SCQ	What is the shape of the person's cheeks in the image?	Full cheeks; Sharp cheeks; Concave cheeks; Flat cheeks
9	SCQ	What is the shape of the person's cheekbones in the image?	Low cheekbones; High cheekbones; Medium height cheekbones
10	TFQ	Does the person in the image have dimples?	Yes; No; Information not visible
11	TFQ	Is the person's chin visible in the image?	Yes; No
12	TFQ	Does the person in the image have a cleft chin (a vertical indentation in the center)?	Yes; No; Information not visible
13	SCQ	How many layers does the person's chin have in the image?	Single chin; Double chin; Information not visible
14	SCQ	What is the width of the person's chin in the image?	Broad chin; Narrow chin; Medium wide chin; Information not visible
15	SCQ	What is the length of the person's chin in the image?	Long chin; Short chin; Medium length chin; Information not visible
16	SCQ	What is the shape of the person's chin in the image?	Round chin; Square chin; Pointy chin; Information not visible
17	SCQ	What is the height of the person's chin in the image?	Protruding chin; Receding chin; Vertical chin; Information not visible
18	SCQ	What is the condition of the person's chin in the image?	Tighten; Expand; Natural
19	TFQ	Are the person's ears visible in the image?	Yes; No;
20	SCQ	Judging by the left and right orientation of the person in the image, is the person's left ear visible in the image?	Fully visible; Partially visible; Not visible
21	SCQ	Judging by the left and right orientation of the person in the image, is the person's right ear visible in the image?	Fully visible; Partially visible; Not visible
22	SCQ	What is the shape of the person's ear tip in the image?	Square ear; Pointed ear; Narrow ear; Round ear; Information not visible
23	SCQ	What is the size of the earlobes of the people in the image?	Big earlobes; Small earlobes; Medium earlobes; Oversized earlobes; Information not visible
24	SCQ	What type of earlobe does the person in the image have?	Free earlobes; Attached earlobes; Partially attached earlobes; Information not visible
25	TFQ	Are the person's ears prominent or protruding in the image?	Yes; No; Information not visible
26	SCQ	How large are the person's ears in the image?	Small; Large; Medium; Oversized; Information not visible
27	TFQ	Are the person's eyebrows visible in the image?	Yes; No
28	SCQ	Judging by the left and right orientation of the person in the image, is the person's left eyebrow visible in the image?	Fully visible; Partially visible; Not visible
29	SCQ	Judging by the left and right orientation of the person in the image, is the person's right eyebrow visible in the image?	Fully visible; Partially visible; Not visible
30	SCQ	What is the color of the person's eyebrows in the image?	Black; Brown; Blonde; Red or auburn; Gray; White; Other
31	SCQ	How wide are the eyebrows of the person in the image?	Wide; Narrow; Medium; Information not visible
32	SCQ	How long are the eyebrows of the person in the image?	Long; Short; Medium; Information not visible
33	SCQ	What is the distance between the person's eyebrows in the image?	Wide; Narrow; Medium; Information not visible
34	SCQ	What is the density of the person's eyebrows in the image?	Thin; Bushy; Medium density
35	SCQ	What is the shape of the person's eyebrows in the image?	S-shaped eyebrows; Hard angled eyebrows; Soft angled eyebrows; Straight eyebrows; Rounded eyebrows; Messy eyebrows; Other; Information not visible
36	SCQ	What is the condition of the person's eyebrows in the image?	Cocked; Scrunched together; Natural state
37	TFQ	Are the person's eyes visible in the image?	Yes; No
38	SCQ	Judging by the left and right orientation of the person in the image, is the person's left eye visible in the image?	Fully visible; Partially visible; Not visible
39	SCQ	Judging by the left and right orientation of the person in the image, is the person's right eye visible in the image?	Fully visible; Partially visible; Not visible
40	SCQ	What type of eyelids does the person in the image have?	Single eyelid/monolid; Double eyelid; Triple eyelid; Other; Information not visible
41	SCQ	What is the length of the person's eyelashes in the image?	Long; Short; Medium; Information not visible
42	SCQ	What is the distance between the eyes of the person in the image?	Wide; Narrow; Medium; Extra wide; Information not visible

Index	Types	Questions	Attribute values
43	SCQ	What is the size of the people's eyes in the image?	Big; Small; Medium; Oversized; Information not visible
44	SCQ	What is the color of the person's pupils in the image?	Amber; Blue; Brown; Green; Grey; Hazel; Red-violet; Heterochromatic; Other; Information not visible
45	SCQ	What is the shape of the person's eyes in the image?	Round eyes; Almond eyes; Upturned eyes; Downturned eyes; Hooded eyes; Protruding eyes; Information not visible
46	SCQ	Judging by the left and right orientation of the person in the image, what is the direction of the gaze of the eyes in the image?	Upper; Under; Left; Right; Front; Information not visible
47	SCQ	What type of eye bags does the person in the image have?	None; Puffiness; Dark circles; Puffiness and dark circles; Information not visible
48	SCQ	How open are the eyes of the people in the image?	Fully opened eyes; Fully closed eyes; Squinting eyes
49	TFQ	Is the person's facial skin visible in the image?	Yes; No
50	SCQ	What is the skin tone of the face in the image?	Pale white skin; White skin; Light brown skin; Moderate brown skin; Dark brown skin; Deeply pigmented dark brown to black skin
51	SCQ	What is the shape of the person's face in the image?	Oval face; Round face; Square face; Oblong face; Heart-shaped face; Diamond-shaped face
52	SCQ	What is the obesity level of the face in the image?	Fat; Thin; Medium
53	SCQ	What is the skin firmness of the face in the image?	Firm; Loose; Sagging
54	SCQ	What type of skin texture is visible in the person shown in the image?	Dry skin; Oily skin; Normal skin; Sensitive skin
55	TFQ	Does the face in the image show any wrinkles?	Yes; No
56	MCQ	What type of wrinkles are present on the person's face in the image?	Forehead lines; Frown lines; Bunny lines; Crow's feet; Brow droop; Vertical lip lines; Mouth frown; Tear throughs; Nasolabial folds; Mental crease
57	TFQ	Does the face in the image have any obvious blemishes?	Yes; No
58	TFQ	What type of blemishes are present on the face in the image?	Scar; Acne; Birthmarks; Pigmentary blemishes/Age spots or sun spots; Other
59	TFQ	Is the person's forehead visible in the image?	Yes; No
60	SCQ	What is the width of the person's forehead in the image?	Wide; Narrow; Medium wide; Extra wide; Information not visible
61	SCQ	What is the height of the person's forehead in the image?	Low; High; Very high; Medium high; Information not visible
62	TFQ	Does the person in the image have a forehead that protrudes forward?	Yes; No; Information not visible
63	TFQ	Is the person's hair visible in the image?	Yes; No
64	TFQ	Does the person in the image have hair loss?	Yes; No; Information not visible
65	SCQ	What type of hair loss is evident in the person shown in the image?	Bald top only; Bald sides only; Bald top and sides
66	SCQ	How curly is the person's hair in the image?	Straight; Slightly curly/Wavy hair; Medium curly/Curly hair; Tightly curly/Coily hair
67	OEQ	What is the hair color of the person shown in the image?	/
68	TFQ	Does the person in the image have bangs?	Yes; No; Information not visible
69	SCQ	What type of bangs does the person in the image have?	Blunt bangs; Side bangs; Curtain bangs; Other
70	SCQ	What is the length of the person's bangs in the image?	Short (above eyebrows); Medium (near eyebrows); Long (below eyebrows)
71	SCQ	How dense are the bangs on the person in the image?	Wispy bangs; Dense bangs; Medium dense bangs
72	SCQ	What is the height of the person's hairline in the image?	Low hairline; Medium high hairline; High hairline; Information not visible
73	SCQ	What type of hairline does the person in the image have?	Widow's peak; Triangular hairline; Uneven hairline; Bell-shaped hairline; Straight hairline; Receding hairline or M-shape; Information not visible
74	SCQ	What is the length of the person's hair in the image?	Buzz cut; Ear length (Pixie); Chin length (Bob); Shoulder length; More than shoulder length; Information not visible
75	SCQ	What is the hair quality of the person in the image?	Dry hair; Oily hair; Normal hair
76	SCQ	What hairstyle does the person in the image have?	Down style; Updo style; Braid style; Ponytail style; Information not visible
77	SCQ	Is the person's hair obscured or bald in the image?	Bald; Occlusion
78	SCQ	Based on the person's orientation in the image, which direction is the person's head oriented in the image?	Up; Down; Left; Right; Forward
79	TFQ	Is the person's mouth open in the image?	Yes; No; Information not visible
80	SCQ	How wide is the person's mouth open in the image?	Lips slightly parted; Mouth open about 1/4; Mouth open about 1/2; Mouth open about 3/4; Mouth completely open
81	TFQ	Are the teeth of the people in the image visible?	Yes; No

Index	Types	Questions	Attribute values
82	SCQ	What is the color of the person's teeth in the image?	White; Yellow; Brown; Gray; Other
83	TFQ	Does the person in the image have a noticeable gap between the teeth?	Yes; No
84	TFQ	Are the teeth in the image aligned?	Yes; No
85	TFQ	Does the person in the image have braces on the teeth?	Yes; No
86	TFQ	Is the person's tongue visible in the image?	Yes; No
87	SCQ	What is the color of the person's tongue in the image?	Yellow; Ruddy; Pale; Other
88	TFQ	Is the person's jawline visible in the image?	Yes; No
89	SCQ	What is the shape of the person's jawline in the image?	Weak jawline; Strong jawline
90	SCQ	What is the length of the person's jawline in the image?	Long; Short; Medium
91	TFQ	Are the person's lips visible in the image?	Yes; No
92	SCQ	Is the lower lip of the person in the image visible?	Fully visible; Partially visible; Not visible
93	SCQ	Is the upper lip of the person in the image visible?	Fully visible; Partially visible; Not visible
94	SCQ	What is the width of the person's lips in the image?	Wide lips; Narrow lips; Medium width lips; Information not visible
95	SCQ	What is the thickness of the person's lips in the image?	Full lips; Thin lips; Heavy upper lip; Heavy lower lip; Information not visible
96	SCQ	What is the shape of the person's lips in the image?	Round lips; Downward-turned lips; Bow-shaped lips; Heart-shaped lips; Uneven lips; Information not visible
97	OEQ	What is the color of the person's lips in the image?	/
98	TFQ	Does the person in the image have a mustache?	Yes; No; Information not visible
99	OEQ	What is the color of the person's mustache in the image?	/
100	SCQ	How dense is the person's mustache in the image?	Dense; Sparse; Medium dense
101	SCQ	What is the length of the person's mustache in the image?	Long; Short; Medium length
102	TFQ	Are the mustache and beard connected in the image?	Yes, mustache connected to beard; No, mustache isolated; Information not visible
103	TFQ	Is the person's neck visible in the image?	Yes; No
104	SCQ	How long is the person's neck in the image?	Long; Short; Medium length; Information not visible
105	SCQ	How thick is the person's neck in the image?	Thin; Thick; Medium thickness; Information not visible
106	TFQ	Does the person in the image have an Adam's apple?	Yes; No; Information not visible
107	SCQ	What type of neck wrinkles does the person in the image have?	No neck wrinkles; Mild neck wrinkles; Deep neck wrinkles
108	TFQ	Is the person's nose visible in the image?	Yes; No
109	SCQ	What is the height of the nasal bridge of the person in the image?	Low nasal bridge; High nasal bridge; Medium height nasal bridge; Information not visible
110	SCQ	What is the shape of the nasal bridge on the person in the image?	Straight nasal bridge; Convex nasal bridge; Concave nasal bridge; Information not visible
111	SCQ	What is the width of the person's nose in the image?	Wide nose; Narrow nose; Medium width nose; Information not visible
112	SCQ	What is the shape of the tip of the person's nose in the image?	Pointed nose tip; Round nose tip; Flat nose tip; Information not visible
113	SCQ	In which direction is the tip of the nose tilted in the image?	Upturned tip; Horizontal tip; Downturned tip; Information not visible
114	SCQ	What is the size of the nasal alae in the person shown in the image?	Small; Large; Medium; Information not visible
115	SCQ	In which direction is the nasal alae tilted in the image?	Upturned nasal alae; Horizontal nasal alae; Downward nasal alae; Information not visible
116	SCQ	Are the nostrils of the people visible in the image?	Fully visible; Partially visible; Not visible
117	SCQ	What is the size of the person's nostrils in the image?	Small; Large; Medium
118	SCQ	What is the shape of the person's nostrils in the image?	Oval nostrils; Round nostrils; Teardrop nostrils; Other
119	TFQ	Does the person in the image have sideburns?	Yes; No; Information not visible
120	OEQ	What is the color of the person's sideburns in the image?	/
121	TFQ	Are the sideburns and beard connected in the image?	Yes, sideburns connected to beard; No, sideburns isolated; Information not visible
122	TFQ	Is the person in the image wearing any earrings?	Yes; No; Information not visible
123	OEQ	What color is the earring that the person in the image is wearing?	/
124	SCQ	Judging by the left and right orientation of the person in the image, how many earrings is the person in the image wearing on the left ear?	Zero; One; Multiple; Information not visible
125	SCQ	Judging by the left and right orientation of the person in the image, how many earrings is the person in the image wearing on the right ear?	Zero; One; Multiple; Information not visible
126	SCQ	What is the size of the earrings the person is wearing in the image?	Small; Large; Medium; Information not visible
127	SCQ	What kind of earrings is the person wearing in the image?	Stud earrings; Hoop earrings; Drop earrings; Other
128	SCQ	What material is the earring worn by the person in the image made of?	Gold; Silver; Fabric; Plastic; Pearls; Gemstones; Other
129	TFQ	Is the person in the image wearing an eye mask?	Yes; No

Index	Types	Questions	Attribute values
130	OEQ	What color is the eye mask that the person in the image is wearing?	/
131	SCQ	What pattern is on the eye mask that the person in the image is wearing?	Plaid; Solid; Floral; Striped; Graphic; Other
132	SCQ	What is the material of the eye mask worn by the person in the image?	Cotton; Fabric; Lace; Fur; Other
133	TFQ	Is the person in the image wearing a face mask?	Yes; No
134	OEQ	What color is the face mask that the person in the image is wearing?	/
135	SCQ	What type of face mask is the person in the image wearing?	Skin care mask; Gas mask; Decorative mask; Sun protection mask; Other
136	SCQ	What material is the face mask worn by the person in the image made of?	Fabric; Plastic; Other
137	TFQ	Is the person in the image wearing glasses?	Yes; No
138	OEQ	What color are the glasses that the person in the image is wearing?	/
139	SCQ	What type of glasses are the people in the image wearing?	VR glasses; Eyeglasses; Sunglasses; Swimming goggles; Protection goggles; Other
140	SCQ	What size are the glasses worn by the people in the image?	Small; Medium; Large; Oversized
141	SCQ	How transparent are the lenses of the glasses shown in the image?	Opaque; Transparent; Transparent with color
142	SCQ	What shape are the glasses worn by the person in the image?	Oval glasses; Round glasses; Square glasses; Aviator glasses; Cat eye glasses; Browline glasses; Geometric glasses; Other
143	SCQ	What type of rim do the glasses shown in the image have?	Full-rimmed; Semi-rimless; Rimless; Other; Information not visible
144	SCQ	What material are the frames of the glasses shown in the image made of?	Plastic eyeglass frames; Metal eyeglass frames; Other
145	TFQ	Is the person in the image wearing any hair accessories?	Yes; No
146	OEQ	What color is the hair accessory that the person in the image is wearing?	/
147	SCQ	What type of hair accessories is the person in the image wearing?	Hairpins; Headbands; Head scarf; Wreath; Crown; Veil; Other
148	SCQ	What material are the hair accessories that the person in the image is wearing made of?	Fabric; Jewelry; Metal; Plastic; Wooden; Other
149	TFQ	Is the person in the image wearing a hat?	Yes; No
150	OEQ	What color is the hat worn by the person in the image?	/
151	SCQ	What is the size of the hat worn by the person in the image?	Small; Large; Medium
152	SCQ	What type of the hat is worn by the person in the image?	Casual hat; Sports hat; Safety helmet; Graduate hat; Professional hat; National hat; Other
153	SCQ	What type of brim does the hat worn by the person in the image have?	Brimless hat; Front brim hat; Full brim hat; Other
154	SCQ	What type of crown does the hat worn by the person in the image have?	High hat top; Pointed hat top; Flat hat top; No hat top; Other; Information not visible
155	TFQ	Is the person in the image wearing headphones?	Yes; No; Information not visible
156	OEQ	What color are the headphones that the person in the image is wearing?	/
157	SCQ	What type of headphones is the person in the image wearing?	In-ear; Over-ear; Other
158	SCQ	Are the headphones in the image wired or wireless?	Wired headphones; Wireless headphones; Information not visible
159	TFQ	Is the person in the image wearing a mask?	Yes; No
160	OEQ	What color is the mask that the person in the image is wearing?	/
161	SCQ	What kind of mask is the person in the image wearing?	Medical/Surgical mask; Daily use mask; Oxygen mask; Other
162	SCQ	What material is the mask worn by the person in the image made of?	Fabric; Plastic; Other
163	TFQ	Is the person in the image wearing a necklace?	Yes; No; Information not visible
164	OEQ	What color is the necklace worn by the person in the image?	/
165	SCQ	How many necklaces is the person in the image wearing?	One necklace; Multiple necklaces (e.g., 2 or more)
166	SCQ	Is the person in the image wearing a necklace with a pendant?	Yes; No; Information not visible
167	SCQ	What is the size of the necklace worn by the person in the image?	Large; Small; Medium; Information not visible
168	SCQ	What material is the necklace worn by the person in the image made of?	Gold; Silver; Fabric; Plastic; Pearls; Gemstones; Other
169	TFQ	Is the individual in the image wearing a scarf?	Yes; No; Information not visible
170	OEQ	What is the color of the scarf worn by the person in the image?	/
171	SCQ	What pattern does the scarf worn by the person in the image have?	Plaid; Solid; Floral; Striped; Graphic; Other
172	SCQ	What material is the scarf worn by the person in the image made of?	Cotton; Fabric; Lace; Fur; Other
173	TFQ	Is the person in the image wearing any special accessories?	Yes; No
174	OEQ	What special accessories are the people in the image wearing?	/
175	TFQ	Is the person in the image wearing a tie?	Yes; No; Information not visible
176	OEQ	What is the color of the tie worn by the person in the image?	/
177	SCQ	What type of tie is the person in the image wearing?	Bowtie; Necktie; Other; Information not visible
178	SCQ	What pattern does the tie worn by the person in the image have?	Plaid; Solid; Floral; Striped; Graphic; Other
179	TFQ	Is the person in the image wearing any tops?	Yes; No; Information not visible
180	OEQ	What color is the top the person in the image is wearing?	/

Index	Types	Questions	Attribute values
181	SCQ	What is the collar size of the top the person in the image is wearing?	Oversized; Large; Medium; Small; Information not visible
182	SCQ	What type of collar is on the top the person is wearing in the image?	Scoop neckline; Collar; V-neckline; High neckline; Square neckline; Off-shoulder top; Suspender top; Other; Information not visible
183	SCQ	How blurry is the face in the image?	Low blurriness; High blurriness; Medium blurriness
184	SCQ	Based on the person's orientation in the image, what direction is the light coming from?	Back lighting (Lighting coming from behind the face); Front lighting (Lighting coming from in front of the face); Upward lighting (Lighting coming from below the face); Downward lighting (Lighting coming from above the face); Left-side lighting (Lighting coming from the left side of the face); Right-side lighting (Lighting coming from the right side of the face); Information not visible
185	SCQ	What is the lighting intensity of the face in the image?	Normal brightness; Excessive brightness; Insufficient brightness; Uneven lighting
186	SCQ	What type of lighting is present in the image?	Natural lighting; Artificial lighting; Information not visible
187	TFQ	Is the human face in the image obscured or partially obstructed by any objects?	Yes; No
188	OEQ	What object is obscuring or partially obscuring the face in the image?	/
189	OEQ	What does the surrounding environment look like in the image? Please describe the environment surrounding the person in the image in one sentence, and infer the person's behavior based on that environment.	/
190	SCQ	What is the most likely age range of the individuals in the image?	0-2; 3-9; 10-19; 20-29; 30-39; 40-49; 50-59; 60-69; more than 70
191	SCQ	To which racial or ethnic group do the individuals in the image most likely belong?	Black; White; Latino Hispanic; Indian; East Asian; Middle Eastern; Southeast Asian
192	SCQ	What is the individual's gender in the image?	Male; Female
193	SCQ	Which emotion is most likely conveyed by the facial expression in the image?	Surprise; Fear; Disgust; Happiness; Sadness; Anger; Neutral
194	SCQ	Which facial action units are present in the image?	AU1 (Inner Brow Raiser); AU2 (Outer Brow Raiser); AU4 (Brow Lowerer); AU5 (Upper Lid Raiser); AU6 (Cheek Raiser); AU7 (Lid Tightener); AU9 (Nose Wrinkler); AU10 (Upper Lip Raiser); AU12 (Lip Corner Puller); AU14 (Dimpler); AU15 (Lip Corner Depressor); AU16 (Lower Lip Depressor); AU17 (Chin Raiser); AU18 (Lip Pucker); AU19 (Tongue Visible); AU20 (Lip stretcher); AU22 (Lip Funneler); AU23 (Lip Tightener); AU24 (Lip Pressor); AU25 (Lips part); AU26 (Jaw Drop); AU27 (Mouth Stretch); AU28 (Lip Suck); AU29 (Jaw Thrust); AU35 (Cheek Suck); AU43 (Eyes Closed); No Facial Action Units

Table S5. Detailed definitions of questions and attribute values in the hierarchical facial attribute structure. "/" represents an open-ended answer.

Prompt example for true/false and single-choice question

<face image>

Question: What is the skin tone of the face in the image?

Options:

Pale white skin

White skin

Light brown skin

...

Please directly select the appropriate option from the given choices based on the image.

Please return the answer directly without additional explanation.

Prompt example for multiple-choice question

<face image>

Question: What type of wrinkles are present on the person's face in the image?

Options:

Forehead lines

Frown lines

Bunny lines

...

Please select one or more appropriate options from the given choices based on the image.

Please return the answer directly without additional explanation.

Prompt example for open-ended question

<face image>

Question: What is the hair color of the person shown in the image?

Answer the question using a single word or phrase.

Please return the answer directly without additional explanation.

Table S6. Prompt examples for true/false, single-choice, multiple-choice, and open-ended questions.

 <p>What type of earlobe does the person in the image have?</p> <p>Free earlobes Attached earlobes Partially attached earlobes Information not visible</p>	 <p>What is the color of the person's pupils in the image?</p> <p>Amber Blue Brown Green Grey Hazel Red-violet Heterochromatic Other Information not visible</p>	 <p>What is the individual's gender in the image?</p> <p>Male Female</p>
 <p>What color are the headphones that the person in the image is wearing?</p> <p>Black</p>	 <p>What is the shape of the person's ear tip in the image?</p> <p>Square ear Pointed ear Narrow ear Round ear Information not visible</p>	 <p>To which racial or ethnic group do the individuals in the image most likely belong?</p> <p>Black White Latino_Hispanic Indian East Asian Middle Eastern Southeast Asian</p>
 <p>What is the shape of the person's eyebrows in the image?</p> <p>S-shaped eyebrows Hard angled eyebrows Soft angled eyebrows Straight eyebrows Rounded eyebrows Messy eyebrows Other Information not visible</p>	 <p>What is the shape of the person's face in the image?</p> <p>Oval face Round face Square face Oblong face Heart-shaped face Diamond-shaped face</p>	 <p>Which facial action units are present in the image?</p> <p>AU1 (Inner Brow Raiser) AU2 (Outer Brow Raiser) AU4 (Brow Lowerer) AU5 (Upper Lid Raiser) AU6 (Cheek Raiser) AU7 (Lid Tightener) AU9 (Nose Wrinkler) AU10 (Upper Lip Raiser) AU12 (Lip Corner Puller) AU14 (Dimpler) AU15 (Lip Corner Depressor)</p>
 <p>How transparent are the lenses of the glasses shown in the image?</p> <p>Opaque Transparent Transparent with color</p>	 <p>What is the most likely age range of the individuals in the image?</p> <p>0-2 3-9 10-19 20-29 30-39 40-49 50-59 60-69 more than 70</p>	 <p>Which emotion is most likely conveyed by the facial expression in the image?</p> <p>Surprise Fear Disgust Happiness Sadness Anger Neutral</p>

Figure S6. VQA examples from FaceBench, encompassing multi-views and multi-level attributes, with correct answers highlighted in red.

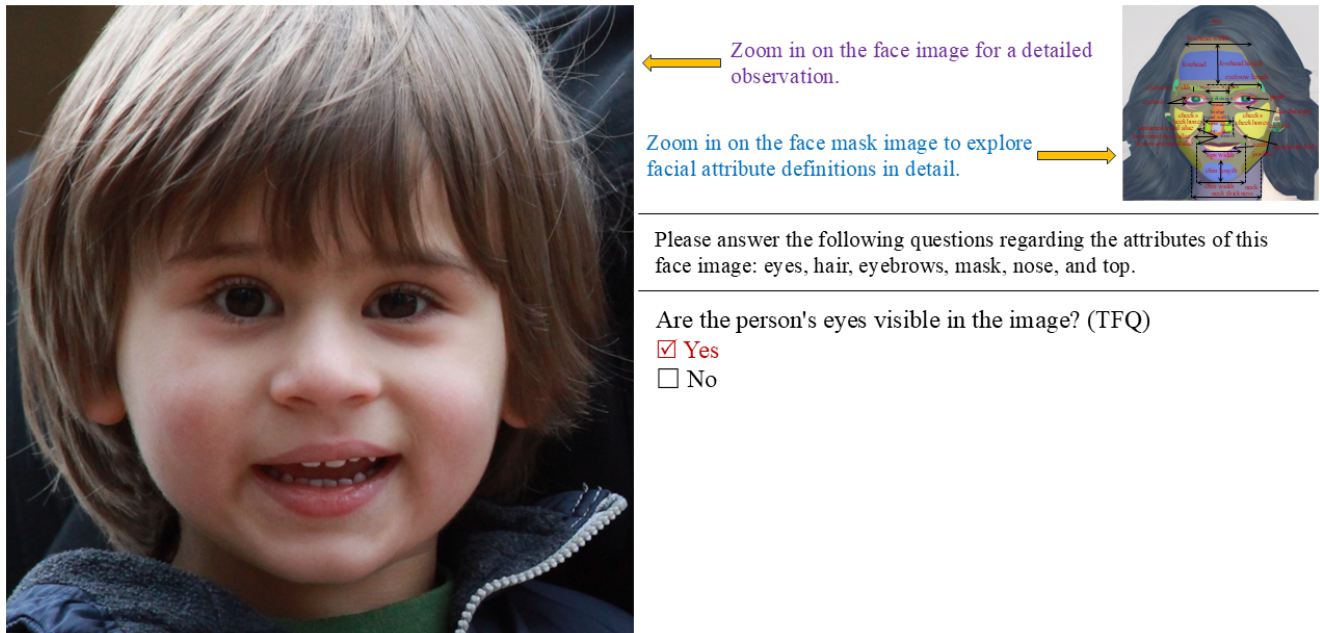


Figure S7. A true/false question example of our annotation system.

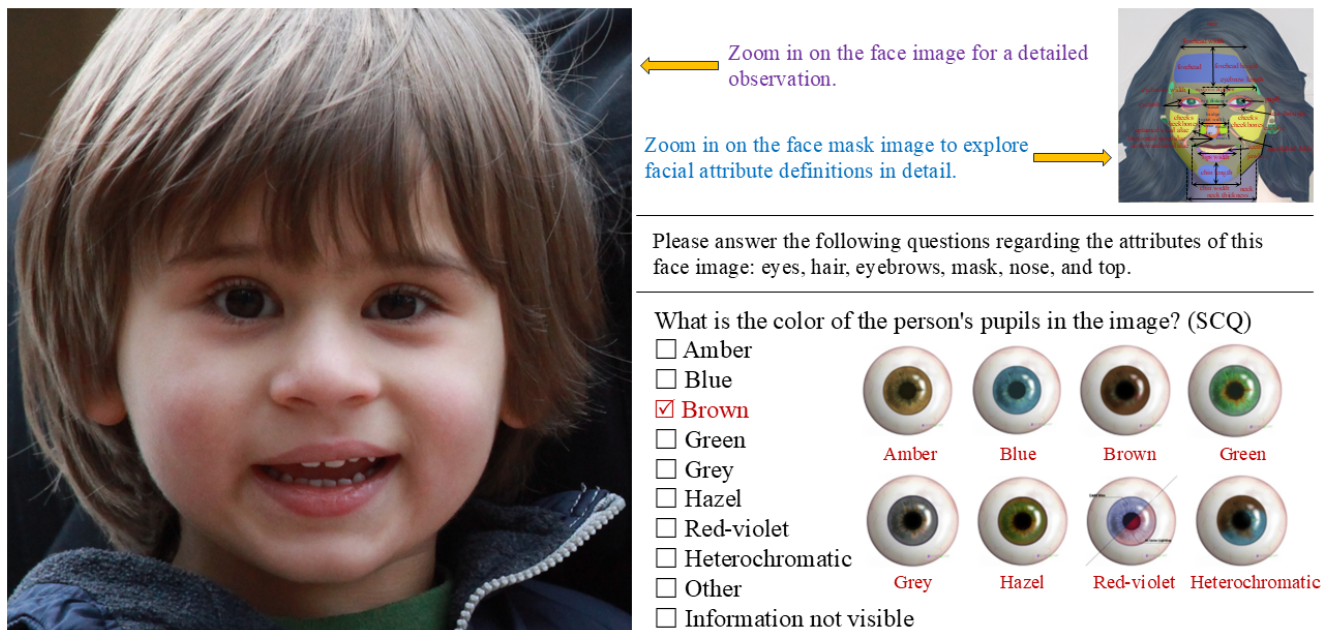

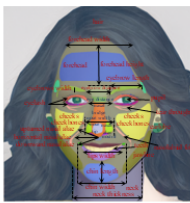


Figure S8. A single-choice question example of our annotation system.



← Zoom in on the face image for a detailed observation.

Zoom in on the face mask image to explore facial attribute definitions in detail. →



Please answer the following questions regarding the attributes of this face image: skin, hair, eyebrows, mask, hat, and top.

What type of wrinkles are present on the person's face in the image? (MCQ)

- ☒ Forehead lines
- ☐ Frown lines
- ☐ Bunny lines
- ☒ Crow's feet
- ☐ Brow droop
- ☐ Vertical lip lines
- ☒ Mouth frown
- ☒ Tear throughs
- ☒ Nasolabial folds
- ☐ Mental crease

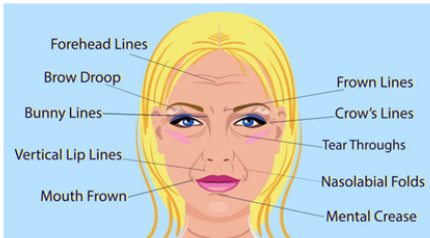

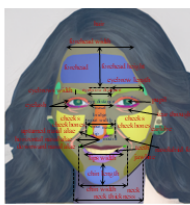


Figure S9. A multiple-choice question example of our annotation system.



← Zoom in on the face image for a detailed observation.

Zoom in on the face mask image to explore facial attribute definitions in detail. →



Please answer the following questions regarding the attributes of this face image: skin, hair, eyebrows, mask, hat, and top.

What color is the hat worn by the person in the image? (OEQ)

white

Figure S10. An open-ended question example of our annotation system.

Comparison of 7660.jpg



Question: What type of wrinkles are present on the person's face in the image?

Options: Forehead lines; Frown lines; Bunny lines; Crow's feet; Brow droop; Vertical lip lines; Mouth frown; Tear troughs; Nasolabial folds; Mental crease

Answer: Forehead lines; Frown lines; Crow's feet; Nasolabial folds; Mouth frown

—— Below are the model predictions ——

LLaVA: Frown lines

LLaMA: Forehead lines

GPT-4o: Forehead lines; Crow's feet; Nasolabial folds

Face-LLaVA: Forehead lines

Question: What type of hair loss is evident in the person shown in the image?

Options: Bald top only; Bald sides only; Bald top and sides

Answer: Bald top only

—— Below are the model predictions ——

LLaVA: Bald top and sides

LLaMA: Bald top and sides

GPT-4o: Bald top only

Face-LLaVA: Bald top only

Question: In which direction is the tip of the nose tilted in the image?

Options: Uprturned tip; Horizontal tip; Downturned tip
Information not visible

Answer: Horizontal tip

—— Below are the model predictions ——

LLaVA: Uprturned tip

LLaMA: Downturned tip

GPT-4o: Horizontal tip

Face-LLaVA: Horizontal tip

Figure S11. An example of qualitative comparison.

Comparison of 60022.png



Question: How transparent are the lenses of the glasses shown in the image?

Options: Opaque; Transparent; Transparent with color

Answer: Transparent

—— Below are the model predictions ——

LLaVA: Opaque

LLaMA: Transparent

GPT-4o: Transparent

Face-LLaVA: Opaque

Question: What type of brim does the hat worn by the person in the image have?

Options: Brimless hat; Front brim hat; Full brim hat; Other

Answer: Front brim hat

—— Below are the model predictions ——

LLaVA: Barely visible brim

LLaMA: Other

GPT-4o: Front brim hat

Face-LLaVA: Full brim hat

Question: What is the height of the nasal bridge of the person in the image?

Options: Low nasal bridge; High nasal bridge; Medium height nasal bridge; Information not visible

Answer: Medium height nasal bridge

—— Below are the model predictions ——

LLaVA: Low nasal bridge

LLaMA: Medium height nasal bridge

GPT-4o: Medium height nasal bridge

Face-LLaVA: Medium height nasal bridge

Figure S12. An example of qualitative comparison.

Comparison of 60072.png



Question: What color are the person's cheeks in the image?

Options: Other color; Flushed cheeks; Same color as skin

Answer: Same color as skin

Below are the model predictions

LLaVA: Flushed cheeks

LLaMA: Flushed cheeks

GPT-4o: Flushed cheeks

Face-LLaVA: Same color as skin

Question: What is the length of the person's bangs in the image?

Options: Short (above eyebrows); Medium (near eyebrows); Long (below eyebrows)

Answer: Medium (near eyebrows)

Below are the model predictions

LLaVA: Short (above eyebrows)

LLaMA: Short (above eyebrows)

GPT-4o: Short (above eyebrows)

Face-LLaVA: Medium (near eyebrows)

Question: What is the collar size of the top the person in the image is wearing?

Options: Oversized; Large; Medium; Small; Information not visible

Answer: Medium

Below are the model predictions

LLaVA: Oversized

LLaMA: Small

GPT-4o: Information not visible

Face-LLaVA: Oversized

Figure S13. An example of qualitative comparison.

Comparison of 62314.png



Question: What is the shape of the person's cheekbones in the image?

Options: Low cheekbones; High cheekbones; Medium height cheekbones

Answer: Medium height cheekbones

Below are the model predictions

LLaVA: Low cheekbones

LLaMA: Medium height cheekbones

GPT-4o: Medium height cheekbones

Face-LLaVA: Medium height cheekbones

Question: What is the density of the person's eyebrows in the image?

Options: Thin; Bushy; Medium density

Answer: Medium density

Below are the model predictions

LLaVA: Bushy

LLaMA: Thin

GPT-4o: Medium density

Face-LLaVA: Medium density

Question: What is the length of the person's jawline in the image?

Options: Long; Short; Medium

Answer: Medium

Below are the model predictions

LLaVA: Long

LLaMA: Short

GPT-4o: Medium

Face-LLaVA: Medium

Figure S14. An example of qualitative comparison.

References

- [1] <https://mmlab.ie.cuhk.edu.hk/projects/celeba.html>. 3
- [2] <https://openai.com/index/chatgpt/>. 3
- [3] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. 2, 3
- [4] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 2
- [5] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024. 2, 3
- [6] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015. 3
- [7] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. 2
- [8] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 3
- [9] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. Datasetgan: Efficient labeled data factory with minimal human effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10145–10155, 2021. 1