

Supplementary Material

In this supplementary, we provide:

- dataset/annotation details and prompts;
- extended SEA analyses and qualitative results;
- extra comparisons on classifiers, commonsense extraction, and annotators;
- user study details.

1. Dataset Analysis Details

1.1. Per-category class list

CommonSketch spans 300 object classes grouped into 14 high-level categories. Each category aggregates semantically related concepts to support both category-level and class-level analyses in SEA. Tab. 1 reports the complete breakdown, listing the class count and full class names for every category. This category taxonomy was used consistently in dataset construction, commonsense element extraction, and all evaluation protocols.

1.2. Per-class element list

Tab. 6 provides the full per-class commonsense element lists extracted by GPT-4o [15] and GPT-OSS [1], which form CommonSketch’s class-wise commonsense database and are used by SEA for element-level presence checking and abstraction analysis. We also replicate the extraction with open-source MLLMs (GPT-OSS 20B, Qwen-2.5 32B [31], Mistral 7B [19], and Llama 3 8B [12]) to assess reproducibility, but omit the full Qwen, Mistral, and Llama lists here to avoid an overly long table; all model-specific lists will be released with the CommonSketch dataset. We report elements in their original surface forms (including casing and hyphen/underscore variants) without normalization to preserve raw model outputs.

Table 2. **Per-class element list (sample).** Sample rows from Tab. 6 are shown to illustrate our per-class commonsense annotations. Elements are listed as extracted: black denotes overlap between GPT-4o and GPT-OSS, red indicates GPT-4o-only elements, and blue indicates GPT-OSS-only elements. The 4o/OSS column reports the number of elements extracted by each model.

Category	Class	4o/OSS	Elements
animal	alpaca	13/11	body, ears, eyes, head, legs, tail, feet, fleece, hooves, muzzle, neck, nostrils, smile, fur_lines, motion_lines, mouth, nose, whisker_lines
	ant	11/9	abdomen, antennae, head, legs, mandibles, thorax, compound eyes, jointed legs, mouth, petiole, stinger, body, eyes, segment_lines
	bat	14/12	body, ears, eyes, head, mouth, nose, tail, wings, feet, fingers, fur, legs, teeth, wing membrane, claws, fur_lines, motion_lines, wing_vein_lines

Table 1. **Classes by category.** For each category, we provide the class count and the full set of class labels used in the dataset.

Category	#Classes	Classes
animal	61	alpaca, ant, bat, bee, bird, boar, butterfly, camel, cat, caterpillar, chameleon, cow, crab, crocodile, deer, dog, dolphin, dragonfly, duck, elephant, feather, fish, flamingo, frog, giraffe, goose, hedgehog, hippopotamus, horse, jellyfish, kangaroo, koala, lion, lobster, mole, monkey, mouse, octopus, owl, panda, parrot, peacock, penguin, rabbit, rooster, scorpion, seahorse, shark, sheep, sloth, snail, snake, spider, squid, squirrel, swan, tiger, turtle, whale, zebra
body part	7	ear, eye, foot, hand, mouth, nose, tooth
clothing	8	belt, bowtie, crown, flip flops, hat, shoe, sock, t shirt
container	9	backpack, basket, bucket, envelope, mailbox, present, purse, suitcase, wine bottle
electronic device	22	alarm clock, calculator, camera, cell phone, charger, computer, fan, headphones, ipod, keyboard, laptop, megaphone, microphone, microwave, oven, radio, robot, satellite, telephone, television, toaster, walkie talkie
food	28	apple, asparagus, banana, bread, broccoli, cake, carrot, cookie, cupcake, donut, garlic, grapes, hamburger, hot dog, ice cream cone, lollipop, mushroom, noodle, onion, peanut, pear, pineapple, pizza, pretzel, pumpkin, sandwich, strawberry, watermelon
furniture	25	bathtub, bed, book, calendar, candle, ceiling fan, chandelier, couch, crayon, door, drawer, fireplace, floor lamp, hourglass, lantern, light bulb, map, marker, paintbrush, paper clip, pencil, stairs, table, toilet, vase
icon	13	angel, diamond, dragon, jack o lantern, mermaid, mona lisa, patrick star, santa claus, skull, snowman, sponge bob, stop sign, teddy bear
musical instrument	11	bell, cello, clarinet, drums, guitar, harp, piano, saxophone, trombone, trumpet, violin
nature	14	bamboo, beach, bush, cactus, cloud, clover, dandelion, flower, leaf, moon, palm tree, rainbow, sun, tree
sports equipment	14	barbell, baseball, baseball bat, basketball, dumbbell, golf club, helmet, parachute, roller skate, skateboard, snorkel, soccer ball, table tennis, tennis racket
structure	28	arch of triumph, barn, bench, big ben, bridge, campfire, castle, church, eiffel tower, fence, ferris wheel, fire hydrant, fountain, hospital, house, igloo, leaning tower of pisa, lighthouse, moai stone, pyramids of giza, roller coaster, skyscraper, sphinx, statue of liberty, stonehenge, streetlight, traffic light, windmill
tool	37	axe, bandage, binoculars, boomerang, bottlecap, broom, cannon, comb, compass, drill, fork, frying pan, grenade, hammer, key, knife, ladder, lighter, matches, mug, pipe, rake, rifle, saw, scissors, screwdriver, shovel, spoon, stethoscope, sword, syringe, teapot, tent, toothbrush, toothpaste, umbrella, wine glass
vehicle	23	airplane, ambulance, bicycle, blimp, bulldozer, bus, canoe, car, cruise ship, flying saucer, helicopter, hot air balloon, motorcycle, pickup truck, rocket, sailboat, space shuttle, submarine, tractor, train, truck, van, wheel

1.3. Additional Sketch Examples



Figure 1. **Additional sketch examples from CommonSketch.** The first three rows are dedicated to the animal category, showing a broad slice of its classes. The fourth row groups representative sketches from body_part, clothing, and container. Each remaining row corresponds to one category (electronic_device, food, furniture, icon, musical_instrument, nature, sports_equipment, structure, tool, vehicle) in order.

1.4. Element Frequency Analysis

We provide statistics on how commonsense elements are distributed across classes and categories. For each element, we treat its presence as a binary attribute at the class level and compute both its global frequency across all classes and its category-wise frequency within each category. To quantify how strongly an element is associated with a category relative to its overall prevalence, we use a lift score, following its standard use in association rule mining for interpretability and comparability [2, 20]:

$$\text{lift}(e, c) = \frac{P(e | c)}{P(e)} = \frac{n(e, c)/n(c)}{n(e)/N},$$

where $n(e, c)$ is the number of classes in category c that contain element e , $n(c)$ is the number of classes in category c , $n(e)$ is the number of classes containing e across the full dataset, and N is the total number of classes. Using this formulation, we rank elements within each category by lift to identify elements that are relatively overrepresented in that category compared with the dataset-wide base rate. Tab. 3 reports, for each semantic category, the three elements with the highest lift. To avoid unstable estimates from extremely rare elements, we include only elements with $n_{\text{in.cat}} \geq 3$. Complementarily, Tab. 4 reports the most frequent elements in the full dataset, ranked by the number of classes in which each element appears. This table summarizes the global prevalence of recurring visual components independently of category-level enrichment.

1.5. Cross-Dataset Sketch Quality Comparison

To contextualize the sketch quality of CommonSketch relative to widely used benchmarks, we compare it with QuickDraw [13] and TU-Berlin [8] using a shared evaluation protocol. We select 14 classes, one from each CommonSketch category, restricting the comparison to classes present in all three datasets. For each sketch, we compute the *Probability of the Ground Truth Class* using three classification models, CLIP [28], OpenCLIP [17], and CoCa [33], and average their outputs to obtain a recognizability score. Fig. 6 shows kernel density estimates (KDEs) of these scores for each class, allowing comparison at the distribution level rather than through a single summary statistic. Higher and more concentrated densities indicate more consistently recognizable sketches, whereas broader or lower-probability distributions suggest greater ambiguity. We also show representative sketches near the *mode* of each dataset’s KDE to provide a visual reference for the typical quality level. Across the 14 shared classes, CommonSketch exhibits competitive or stronger modes overall, supporting the quality of our data collection pipeline and its suitability for SEA-based abstraction and element-level analysis.

Table 3. **Category-wise frequent elements.** For each category, we list the three elements with the highest lift. Columns report the per-category class count $n_{\text{in.cat}}$, within-category coverage p_{cat} , and lift. Only elements with $n_{\text{in.cat}} \geq 3$ are included.

Category	Rank	Element	$n_{\text{in.cat}}$	p_{cat}	Lift
animal	1	abdomen	4	0.07	4.92
animal	2	antennae	6	0.10	4.92
animal	3	beak	11	0.18	4.92
container	1	handle	5	0.56	4.76
container	2	body	6	0.67	1.74
electronic device	1	display	4	0.18	13.64
electronic device	2	microphone	4	0.18	13.64
electronic device	3	screen	5	0.23	13.64
food	1	slice	5	0.18	10.71
food	2	bite mark	3	0.11	10.71
food	3	seeds	4	0.14	8.57
furniture	1	tip	4	0.16	4.36
furniture	2	frame	3	0.12	3.60
furniture	3	base	5	0.20	2.40
icon	1	arms	4	0.31	6.59
icon	2	mouth	4	0.31	2.10
icon	3	legs	5	0.38	1.96
musical instrument	1	bridge	3	0.27	27.27
musical instrument	2	bell	4	0.36	21.82
musical instrument	3	keys	3	0.27	20.45
nature	1	leaves	3	0.21	16.07
nature	2	leaf	3	0.21	8.04
nature	3	base	3	0.21	7.14
sports_equipment	1	grip	3	0.21	8.04
structure	1	tower	4	0.14	10.71
structure	2	flag	3	0.11	6.43
structure	3	windows	3	0.11	5.36
tool	1	angle	3	0.08	8.11
tool	2	edge	3	0.08	8.11
tool	3	rivet	3	0.08	8.11
vehicle	1	exhaust pipe	7	0.30	13.04
vehicle	2	bumper	4	0.17	13.04
vehicle	3	cabin	4	0.17	13.04

1.6. Element Annotation Details

We provide further details on the refinement of commonsense elements and the strict criteria applied during the annotation process, supplementing the methodologies described in CommonSketch.

Disambiguation of Symbolic and Anatomical Features.

During the refinement of raw elements generated by LLMs, a key challenge was distinguishing between realistic physical attributes and stylized or anthropomorphic depictions common in sketch representations. For instance, while a butterfly biologically possesses a *proboscis*, it is frequently depicted in sketches with a simple, smiling *mouth*. To accommodate both anatomical accuracy and sketch-style ab-

Table 4. **Global element frequency ranking.** We report the 20 elements that appear in the largest number of classes across all 300 classes.

Element	$n_{\text{classes_global}}$	Rank	Element	$n_{\text{classes_global}}$	Rank
body	115	1	neck	24	11
head	67	2	feet	20	12
legs	59	3	stem	15	13
eyes	57	4	water	14	14
tail	52	5	wings	14	15
mouth	44	6	arms	14	16
handle	35	7	teeth	13	17
ears	29	8	nostrils	12	18
base	25	9	claws	12	19
nose	24	10	tip	11	20

straction, we separated these features into distinct annotation categories. This distinction was similarly applied to other classes such as ant, bee, and octopus. This separation ensures that the model is evaluated on what is actually drawn rather than what is biologically expected.

Strict Criteria for Visual Presence. To establish ground truth labels, we applied clear guidelines to determine the presence of an element, operating under the principle that visual existence takes precedence over biological facts.

Certain textural elements, such as *fur* for animals or *feathers* for birds, are intrinsically present in the real-world subjects. However, in our annotation, these elements were marked as present only if the sketch explicitly contained additional strokes or shading to depict texture. If an animal was drawn with a simple, smooth outline without specific textural details, attributes like fur or feathers were labeled as absent, even if they naturally belong to the subject. This strict criterion enables us to distinguish whether the model relies on actual visual evidence or merely depends on background knowledge.

2. SEA: Sketch Evaluation metric for Abstraction efficiency

2.1. Basic properties of SEA

Notation. We denote by E the number of class-defining commonsense elements that are available for a given class, by V the number of visual elements that are actually rendered in a sketch, and by $P \in [0, 1]$ the predicted probability of the ground-truth class from the recognition model. We consider the domain

$$E \in \mathbb{N}_{\geq 1}, \quad 0 < V \leq E, \quad 0 < P < 1,$$

and write $v = V/E$ for the visual ratio. Intuitively, E captures semantic coverage, v measures how much of the avail-

able visual information is expressed, and P measures recognizability.

In implementation, we apply a small numerical clipping $V \leftarrow \min(\max(V, 0), E)$ and $P \leftarrow \min(\max(P, \varepsilon), 1 - \varepsilon)$ with $\varepsilon \approx 10^{-6}$, but this does not affect the analytical properties discussed below.

SEA formulation. Given (E, V, P) , the SEA score $S(E, V, P)$ is defined as

$$S(E, V, P) = \tanh(\alpha Z(E, V, P)),$$

where $\alpha > 0$ is a fixed scale parameter and

$$Z(E, V, P) = \text{reward}(E, V, P) - \text{penalty}(E, V, P).$$

The reward term encourages efficient abstraction, while the penalty term suppresses both excessive visual expression and low recognizability. In other words, the reward favors sketches that remain recognizable with minimal visual expression, and the penalty discourages sketches that either use too much visual detail or fail to be recognized.

We first define the visual ratio

$$v = \frac{V}{E}.$$

The *economy of expression* is given by

$$u(v) = \log\left(\frac{1 + \delta}{v + \delta}\right),$$

where $\delta > 0$ is a small numerical constant. This term is positive and monotonically decreasing in v : it is large when the sketch uses few visual elements (v small) and approaches zero as $v \rightarrow 1$. Thus, for fixed recognizability, sketches with more visual representation efficiency (smaller v) receive higher u .

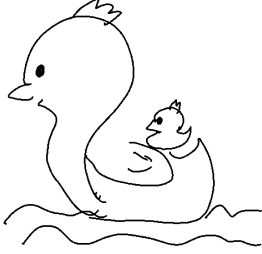
The *centered gate* compares visual expression and recognizability:

$$g(P, v) = \tanh\left(\frac{1}{2}\beta \log\frac{P + \delta}{v + \delta}\right),$$

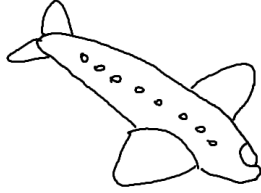
where $\beta > 0$ controls the sharpness of the transition. We have $g(P, v) = 0$ when $v = P$; if $v < P$ (the sketch is more recognizable than its level of visual expression would suggest), then $g(P, v) > 0$ and the reward is amplified; if $v > P$ (the sketch is more detailed than necessary for its recognizability), then $g(P, v) < 0$ and the reward is attenuated. When v and P are well aligned, the gate remains near zero and does not dramatically amplify or suppress the reward.

The reward term combines economy of expression, the centered gate, and recognizability:

$$\text{reward}(P, v) = P^\gamma u(v) g(P, v),$$



(a) A detailed sketch with relatively high visual ratio and moderate recognizability.



(b) A visually sparse sketch with low recognizability despite a moderate visual ratio.

Example	Visual ratio (v)	Probability (P)	SEA score (w/ g)	SEA score (w/o g)
(a) Detailed sketch	0.69	0.63	-0.43	+0.17
(b) Low- P sketch	0.60	0.18	-0.93	+0.03

Figure 2. **Effect of the gate function in SEA scoring.** Two representative cases illustrate how the gate term changes the final SEA score. The gated formulation assigns substantially lower scores in both cases, showing that it penalizes overly detailed or weakly recognizable sketches more strongly.

where $\gamma > 0$ controls how strongly the reward is guided by recognizability. As a result, sketches with high recognizability P , strong economy of expression u , and a positive centered gate g receive large positive reward.

The reward term is designed to (i) reward lower v with diminishing returns via a log-ratio, (ii) penalize misalignment between v and P using the centered gate $g(\cdot)$, and (iii) use a bounded, differentiable \tanh gate to prevent extreme values from dominating. Fig. 2 shows that removing $g(\cdot)$ inflates scores in both (a) detailed and (b) low- P cases, confirming the necessity of $g(\cdot)$. Also, SEA quantifies excessive detail at the element level as over-expression rather than identifying specific strokes as responsible.

The penalty term consists of two parts:

$$\text{penalty}(P, v) = \lambda v^\eta (1 - P)^k + \tau (1 - P)^r,$$

where $\lambda > 0$ scales the penalty on excessive visual expression, $\eta \in (0, 1)$ controls the curvature with respect to v (reducing sensitivity at very high usage), $k > 0$ and $r > 0$ control the sensitivity to low recognizability, and $\tau > 0$ sets the base penalty for low P regardless of v . The first term penalizes sketches that use many visual elements (v large) when P is low, while the second term ensures that sketches with very low recognizability are penalized even if they are visually sparse. Thus, the penalty discourages both excessive visual expression and failure to be recognized. Sec. 2.3 provides an ablation over these hyperparameters.

Boundedness. By construction, SEA is a bounded score. For any admissible (P, v) , the inner quantity $Z(P, v)$ is real-

valued, and $\alpha > 0$ is a constant. Since the hyperbolic tangent satisfies $-1 < \tanh(x) < 1$ for all real x , it follows that

$$-1 < S(P, v) = \tanh(\alpha Z(P, v)) < 1$$

for all admissible (P, v) . This boundedness makes SEA directly comparable across classes and datasets and avoids scale issues when aggregating scores.

Continuity and differentiability. We now formalize the smoothness of SEA. On the domain $E \geq 1$, $0 < V \leq E$, and $0 < P < 1$, we have $v \in (0, 1]$ and $P \in (0, 1)$. Since $\delta > 0$, the arguments $v + \delta$ and $P + \delta$ are strictly positive. The logarithm, powers $x \mapsto x^\gamma$, $x \mapsto x^k$, $x \mapsto x^r$, and the hyperbolic tangent are all smooth on $(0, \infty)$ (and on \mathbb{R} for \tanh). Therefore $u(v)$ and $g(P, v)$ are smooth in (V, P) , and so are the reward and penalty. The function $Z(E, V, P)$ is a sum of these smooth terms, hence smooth in (V, P) , and $S(E, V, P) = \tanh(\alpha Z(E, V, P))$ is a smooth composition of smooth functions. In particular, $u(E, V)$, $g(P, v)$, reward(E, V, P), and penalty(E, V, P) are continuous in (E, V, P) and differentiable in (V, P) , and the same holds for $Z(E, V, P)$ and $S(E, V, P)$.

Smoothness is beneficial when SEA is used not only as an evaluation metric but also as a reward or critic in optimization, e.g., for training sketch generative models.

Extreme cases. We summarize the qualitative behavior of SEA in several extreme regimes, which follow directly from the definitions of u , g , the reward, and the penalty.

First, consider *unrecognizable sketches*, where $P \rightarrow 0$. As P approaches zero, the factor P^γ in the reward term tends to zero, so the reward vanishes. At the same time, $(1 - P)^k$ and $(1 - P)^r$ both approach 1, so the penalty converges to $\lambda v^\eta + \tau > 0$. Thus, $Z(E, V, P)$ tends to $-(\lambda v^\eta + \tau)$, and $S(E, V, P)$ moves toward the lower end of its range. This reflects the design choice that sketches not recognized by the classifier are treated as abstraction failures regardless of visual detail.

Second, consider *efficient abstraction*, where $P \rightarrow 1$ and $v \ll 1$. As P approaches one, both $(1 - P)^k$ and $(1 - P)^r$ tend to zero, so the penalty vanishes. For very small v , the economy of expression $u(E, V) = \log((1 + \delta)/(v + \delta))$ is large and positive. Since P is close to one, we typically have $P > v$, implying $\log((P + \delta)/(v + \delta)) > 0$ and hence $g(P, v) > 0$. Consequently, the reward becomes large and positive, $Z(E, V, P) > 0$, and $S(E, V, P)$ moves toward the upper end of its range. This corresponds to sketches that remain highly recognizable while using very few strokes.

Finally, consider *over-detailed sketches*, where $v \rightarrow 1$. When the sketch renders nearly all available elements, the visual ratio v approaches one, and the economy of expres-

sion $u(E, V)$ approaches

$$\log\left(\frac{1+\delta}{1+\delta}\right) = 0,$$

so the reward becomes small. In contrast, the first term of the penalty approaches $\lambda(1-P)^k$, which is strictly positive for any $P < 1$, and the second term $\tau(1-P)^r$ is also non-negative. Therefore, $Z(E, V, P)$ decreases and $S(E, V, P)$ is reduced even when P is high. This reflects our design choice that overly detailed sketches exhibit lower abstraction and should therefore receive lower SEA scores.

Together, these properties show that SEA is a bounded, smooth scoring function that rewards semantically efficient sketches with high recognizability and minimal visual expression, while penalizing both unrecognizable and unnecessarily detailed drawings. In the next section, we analyze how these qualitative behaviors are enforced through monotonicity and consistency constraints on the partial derivatives of $S(E, V, P)$.

2.2. Analysis of SEA under constraint conditions

Constraint conditions. In this section, we formalize the global constraints that SEA is designed to satisfy in the interior of the domain. The formulation reflects four main conditions: recognizability monotonicity, visual representation efficiency, the failure region, and the efficient-abstraction region.

First, SEA is designed to be monotone with respect to recognizability. When the amount of visual content V and the semantic capacity E are fixed within a typical range, increasing the recognizability score P should not reduce the SEA score. This reflects the principle that sketches with higher recognizability, and thus stronger semantic alignment, should not be penalized.

Second, the metric promotes visual representation efficiency. When a sketch exhibits low recognizability, increasing the visual ratio v should not increase the score; additional strokes cannot compensate for semantic failure. Conversely, when recognizability is sufficiently high, SEA encourages moderate visual usage. The score increases with visual representation only up to an optimal usage level $v^*(E, P)$, after which excessive detail is penalized. This ensures that SEA rewards sketches that maintain recognizability using only the essential visual elements.

The third condition defines the failure region: all sketches whose recognizability falls below a threshold P_{fail} must receive non-positive SEA scores, regardless of visual representation. This ensures that a sketch that fails to convey its semantic identity is treated as an abstraction failure, even if it uses very few strokes.

The final condition defines the efficient-abstraction region. Sketches that achieve high recognizability with economical visual information should receive strictly posi-

tive SEA scores. Thus, for recognizability values above a threshold P_{good} and visual ratio below an efficiency boundary v_{eff} , the score must be positive. This region captures the central goal of SEA: rewarding sketches that preserve semantic content with minimal visual expression.

Together, these four conditions govern the global behavior of SEA, determining how recognizability, visual usage, and semantic efficiency interact to produce the final score.

Derivative analysis. To understand how these constraints arise from the functional form of SEA, we analyze the partial derivatives of the score, where $v = V/E$. Recall that

$$\begin{aligned} S(E, V, P) &= \tanh(\alpha Z(E, V, P)), \\ Z(E, V, P) &= \text{reward}(E, V, P) - \text{penalty}(E, V, P), \end{aligned}$$

with

$$\begin{aligned} u(E, V) &= \log\left(\frac{1+\delta}{v+\delta}\right), \\ g(P, V, E) &= \tanh\left(\frac{1}{2}\beta \log\frac{P+\delta}{v+\delta}\right), \\ \text{reward}(E, V, P) &= P^\gamma u(E, V) g(P, V, E), \\ \text{penalty}(E, V, P) &= \lambda v^\eta (1-P)^k + \tau(1-P)^r. \end{aligned}$$

Because the hyperbolic tangent is strictly increasing, the sign of the derivatives of S matches that of Z . The derivative with respect to recognizability is

$$\frac{\partial S}{\partial P} = \alpha(1 - \tanh^2(\alpha Z)) \frac{\partial Z}{\partial P}.$$

An explicit expansion of $\partial Z/\partial P$ shows that it consists of two positive penalty-related terms and two reward-related terms. Under the hyperparameter setting used in our experiments, the reward-related terms are non-negative over the classifier’s operating range, while the penalty-related terms remain strictly positive for all $0 < P < 1$. Numerical evaluation over $P \in [0.1, 0.99]$, $v \in [0.05, 1.0]$, and $E \in \{4, 8, 16, 32\}$ confirms that $\partial Z/\partial P \geq 0$ throughout this region. Consequently, $\partial S/\partial P \geq 0$, establishing recognizability monotonicity in practice.

For the derivative with respect to the visual ratio v , we obtain

$$\frac{\partial S}{\partial v} = \alpha(1 - \tanh^2(\alpha Z)) \frac{\partial Z}{\partial v},$$

where $\partial Z/\partial v$ decomposes into reward-related terms involving $\partial u/\partial v$ and $\partial g/\partial v$, and a penalty-derived term $\lambda \eta v^{\eta-1} (1-P)^k$. When recognizability is low, the factor P^γ suppresses the reward derivatives, leaving the penalty derivative dominant and strictly negative, which yields $\partial S/\partial v \leq 0$. This enforces the principle that additional strokes do not help a sketch that is already semantically unrecognizable.

When recognizability is high, P^γ is large enough for the reward derivatives to compete with the penalty derivative. For small v , the efficiency term $u(E, V)$ is large and positive, and $g(P, v)$ is typically positive because recognizability exceeds usage. Hence $\partial Z/\partial v > 0$ and the score increases with visual usage. As v grows, u decreases and the penalty derivative increases, eventually causing $\partial Z/\partial v$ to become negative. This produces an interior optimum $v^*(E, P)$ where the score is maximized, and ensures that excessively detailed sketches are penalized.

Taken together, these derivative properties show that the SEA formulation, combined with its chosen hyperparameters, enforces the global constraint conditions described above. The behavior observed in the extreme cases of Section S.1 extends across the full interior of the domain, ensuring coherent and consistent scoring of sketches according to abstraction efficiency.

2.3. Ablation Studies on Hyperparameters

Default hyperparameter setting. The SEA score is determined by a compact set of hyperparameters that control the scale, sharpness, and relative strength of the reward and penalty components. Throughout all main-paper experiments, we use the following default values:

$$\alpha = 2.2, \beta = 8.0, \lambda = 1.0, \eta = 0.8, k = 2.3, \\ \tau = 0.4, r = 1.7, \gamma = 1.7, \delta = 10^{-6}.$$

This configuration produces a stable and interpretable scoring surface. Efficiently abstracted sketches, which show high recognizability with minimal visual representation, tend to obtain positive scores. In contrast, sketches that lack recognizability or contain unnecessary visual details tend to obtain negative scores. These hyperparameters therefore define the overall operating regime of SEA and serve as the baseline for subsequent ablation studies.

One-dimensional sweeps. To illustrate how SEA responds to variations in visual representation, we perform one-dimensional sweeps over the normalized visual representation v . Since v is normalized by the number of available elements, the choice of E does not affect the qualitative shape of the SEA curve. We therefore fix $E = 10$ for all sweeps.

Fig. 3 shows SEA scores for $v \in [0.05, 1.0]$ at recognizability levels $P = 0.3, 0.5, 0.8$. The curves exhibit three characteristic regimes. When recognizability is low ($P = 0.3$), the score decreases monotonically as v increases, indicating that additional visual detail does not compensate for low recognizability. At moderate recognizability ($P = 0.5$), SEA remains nearly flat for small v and then gradually declines as the sketch becomes more detailed. When recognizability is high ($P = 0.8$), the score

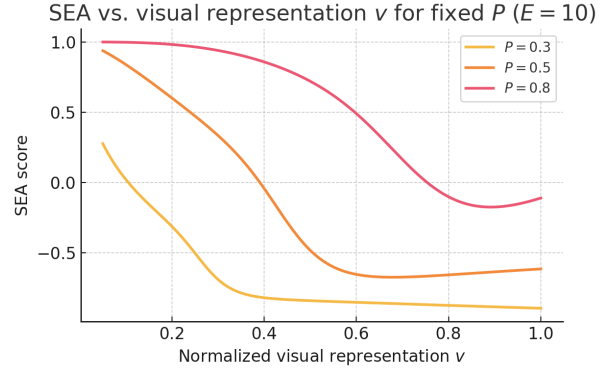


Figure 3. **SEA vs. normalized visual representation.** For fixed $E = 10$, the SEA score decreases with v at low P , shows a mild plateau then decline at moderate P , and peaks at moderate v when P is high, illustrating SEA’s preference for efficient abstraction.

first increases, reaches a maximum at a moderate level of visual representation, and then decreases as excessive detail is added. These trends highlight SEA’s preference for efficient abstraction: highly recognizable sketches with minimal visual representation achieve higher scores, whereas unrecognizable or overly detailed sketches are penalized.

Hyperparameter-wise 2D heatmap analysis. To examine SEA over the joint space of visual ratio v and recognizability P , we generate two-dimensional heatmaps under different hyperparameter settings. Since $v = V/E$, the qualitative structure of the SEA surface does not depend on the absolute value of E . For consistency, all visualizations in Fig. 4 use $E = 10$. Fig. 4 presents a 5×3 grid, where each row varies one hyperparameter group. The center column shows the default setting, the left column decreases the parameter, and the right column increases it. This layout provides a comparison of how each component reshapes the score surface over (v, P) .

The first row shows the effect of the scale parameter α . When α is reduced, the heatmap becomes smoother, with broader contour bands and more gradual transitions between positive and negative scores. Increasing α has the opposite effect: the surface becomes dominated by the saturated extremes of ± 1 , and intermediate contours collapse toward the decision boundaries. Thus, α controls the contrast and saturation of the outer tanh activation.

The second row examines the gate sharpness parameter β , which determines how sharply SEA separates under-drawn and over-drawn sketches around $P \approx v$. With smaller β , the transition becomes broad and diffuse, producing a wide intermediate band. At the default setting, the boundary is clear but not overly sharp, whereas larger β makes it razor-thin and causes the score to change more abruptly across the boundary. This confirms that β mainly

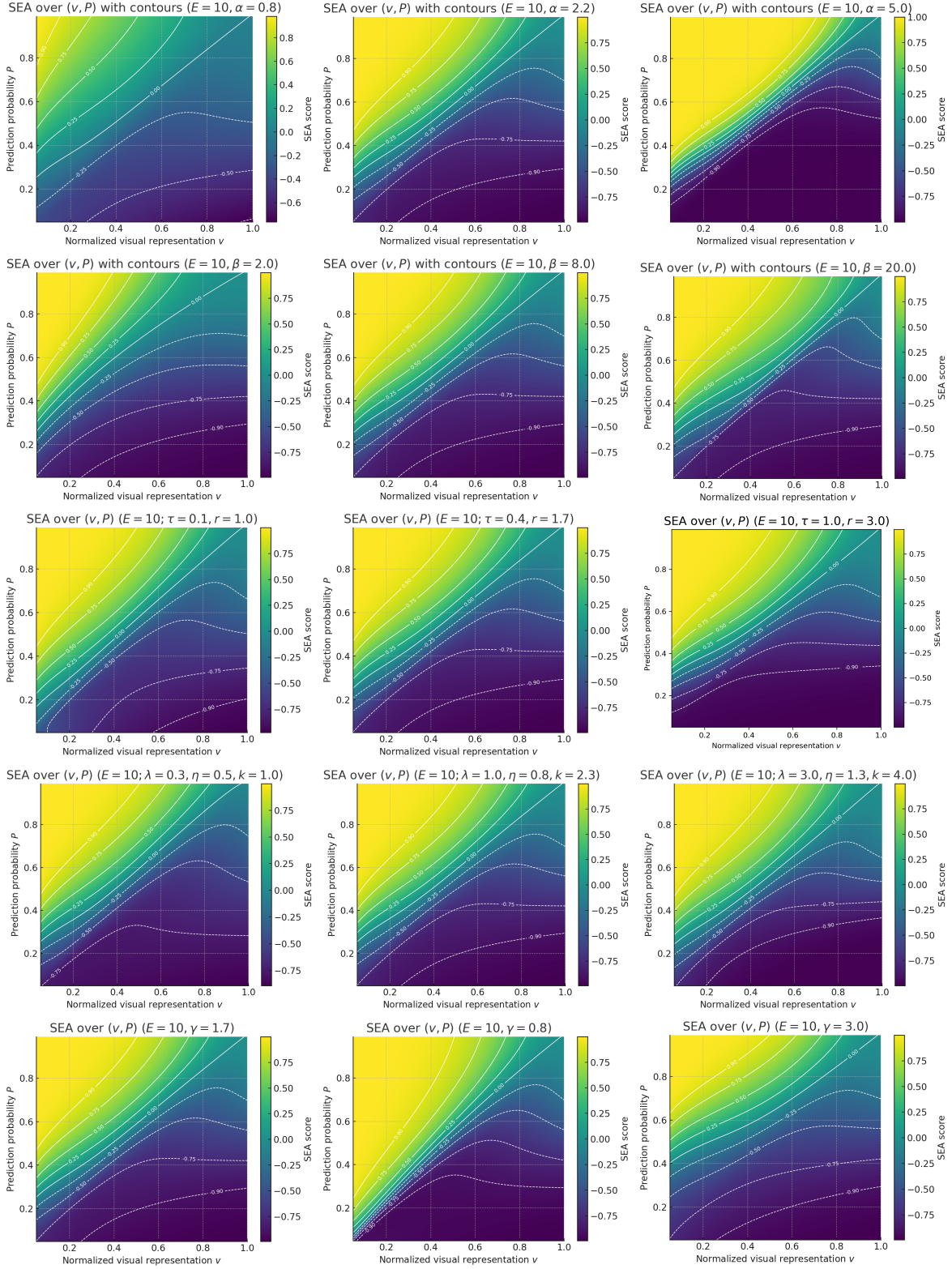


Figure 4. **Effect of SEA hyperparameters on the score surface:** each row varies a single parameter (left: decreased, center: default, right: increased), illustrating how α , β , visual representation efficiency penalties (λ , η , k), base penalties (τ , r), and recognizability guidance γ reshape the (v, P) score landscape.

controls the sharpness of the consistency gate.

The third row shows how the visual representation efficiency penalty, governed by (λ, η, k) , changes the over-detailed region. Weakening the penalty makes the upper-right region less negative and shifts positive contours rightward, allowing more high- v sketches to remain in the efficient region. In this case, $\eta < 1$ softens the tail penalty as $v \rightarrow 1$, and smaller k makes the low- P failure region thinner. Strengthening the penalty shifts the contours leftward, sharply reducing the permissible visual representation; here, $\eta > 1$ steepens the decline near $v = 1$, while larger k expands the failure region upward. Overall, λ controls penalty strength, η its curvature, and k the severity of low-recognizability penalization.

The fourth row examines the base penalty parameters τ and r , which control how strongly SEA penalizes low recognizability regardless of visual representation. Smaller values yield a milder failure region, allowing moderately low P to remain near zero. Larger values expand the negative region across the bottom of the heatmap, enforcing a clearer failure regime for insufficient recognizability.

The final row shows the effect of the recognizability guidance parameter γ . Lower values distribute the reward across recognizability levels, enlarging the efficient-abstraction region for moderate P . Higher values concentrate the reward near $P \approx 1$, reducing the positive region and increasing sensitivity to recognizability differences.

Overall, Fig. 4 shows how each hyperparameter shapes the SEA landscape. Adjusting $\alpha, \beta, (\lambda, \eta, k), (\tau, r)$, and γ systematically expands or contracts the failure, efficient, and over-detailed regions. These comparisons provide a direct visual interpretation of SEA’s core principle: sketches with high recognizability and efficient visual representation should be rewarded, whereas unrecognizable or overly detailed sketches should not.

2.4. SEA as a Training Critic

In this work we primarily use SEA as an evaluation metric. Given a set of sketches produced by a generative model or drawn from a reference dataset, we compute $S(E, V, P)$ for each sketch and aggregate the scores to compare different generation methods, training regimes, or datasets. In this setting, SEA plays a similar role to existing scalar metrics such as FID or CLIP-based similarity [14, 28]: it is applied post-hoc to fixed samples and does not directly affect the training dynamics of the generator.

SEA can also be used as a reward or critic for learning. For a sampled sketch s , we can define a scalar reward

$$R(s) = S(E(s), V(s), P(s)),$$

and consider an objective of the form

$$\max_{\phi} \mathbb{E}_{s \sim \pi_{\phi}} [R(s)],$$

where π_{ϕ} is a sketch generator parameterized by ϕ . When the computation of $E(s), V(s), P(s)$ involves non-differentiable components such as a VQA model or a discrete classifier, one may combine SEA with standard techniques for learning from non-differentiable rewards, e.g., policy-gradient methods such as REINFORCE [30], stop-gradient tricks, or continuous relaxations such as the Gumbel–Softmax estimator [18, 25] for discrete stroke decisions. In cases where parts of the pipeline are differentiable (e.g., CLIP-based recognizability), gradients can be back-propagated through those components while treating the remaining terms as a black-box reward.

In summary, the experiments and analyses in this paper focus on SEA from an evaluation perspective and use it only to assess models and datasets. Using SEA as a generation reward or training critic is therefore left as an extension and a promising direction for future work.

3. Detailed Analysis on SEA Components

We disentangle SEA’s three components (commonsense elements, visual representation, and zero-shot prediction) through staged qualitative analyses. Sec. 3.1 varies the zero-shot backbone (CLIP vs. OpenCLIP) and presents SEA-scored examples on SEVA and CommonSketch using each model’s prediction probabilities. Sec. 3.2 varies the commonsense database (GPT-4o vs. GPT-OSS) and shows the same qualitative SEA diagnostics on both datasets. In all settings, visual representation is measured with two annotators, GPT-4o and Qwen2.5-VL. Across these controlled swaps, SEA scores track abstraction efficiency in a stable way, indicating robustness to changes in both the classifier and the commonsense source.

3.1. CLIP vs. OpenCLIP on Classification

We fixed the classifier to CLIP when computing and analyzing SEA, since CLIP exhibited the strongest alignment with human judgments on SEVA [27]. We additionally evaluated two classifiers, OpenCLIP [17] and CoCa [33]. Supplementary qualitative examples on SEVA and CommonSketch using OpenCLIP are provided in Figs. 7 to 9. For this comparison, we re-estimated model–human alignment by leveraging human responses from the sketch classification questions in our user study and matching them against each model’s predictions. As shown in Tab. 5, CoCa achieves the highest top-1 accuracy, whereas OpenCLIP demonstrates the strongest correlations with human assessments. Accordingly, in the subsequent analysis we adopt OpenCLIP as an alternative zero-shot classifier and conduct a qualitative comparison between SEA computed with OpenCLIP and with CLIP.

Table 5. Comparison of performance on Top-1 accuracy and correlation between Human assessment.

Model	Top-1 Acc	Spearman's ρ	Kendall's τ	Pearson's r
Human	0.952	–	–	–
CLIP	0.794	0.369	0.284	0.496
OpenCLIP	0.912	0.534	0.435	0.785
CoCa	0.941	0.518	0.426	0.650

3.2. GPT-4o vs. GPT-OSS on Commonsense

We examine whether the proprietary model GPT-4o can be replaced by an open-weights alternative for commonsense extraction, while keeping the remainder of the SEA pipeline unchanged. We therefore extract commonsense elements using multiple open-source LLMs, including GPT-OSS, Qwen-2.5, Llama 3, and Mistral, and compare their extraction behavior on the 14 classes shown in Fig. 6. Among these candidates, GPT-OSS shows the closest agreement with GPT-4o, both in the distribution of extracted element counts and in the overall class-wise extraction tendency. We therefore adopt GPT-OSS as the open-weights extractor for feasibility verification. Using this setting, we compute SEA scores with commonsense elements derived from GPT-OSS and compare them against those obtained with GPT-4o.

In the experiment, only the commonsense extraction stage is switched from GPT-4o to GPT-OSS; all other components remain fixed, including CLIP as the classifier and the same VLM annotators for visual representations. This controlled setup allows us to isolate the effect of replacing the commonsense extractor without conflating it with changes in recognizability estimation or visual-element annotation. Fig. 5 summarizes the quantitative agreement. When GPT-4o is used as the VQA annotator, SEA scores computed with GPT-OSS closely match the baseline, achieving a concordance correlation coefficient (CCC) of 0.86. This agreement remains after switching to Qwen, with a CCC of 0.812. In both cases, scatter points concentrate near the identity line ($y = x$), indicating that GPT-OSS preserves the same semantic ranking patterns as GPT-4o regardless of annotator choice.

Qualitative results in Figs. 7 to 9 corroborate this trend. Across both SEVA and CommonSketch, sketches are ordered within each class, with the score increasing from left to right. Low-scoring examples are typically hard to recognize and therefore receive negative SEA values together with low prediction probabilities. Mid-range examples contain sufficient detail to support recognition, yielding high visual representations and prediction probabilities that correspond to a reasonable abstraction regime. High-scoring examples maintain strong recognizability despite comparatively lower visual representations, reflecting superior abstraction efficiency in which minimal depiction suffices for reliable classification. This same progression is preserved

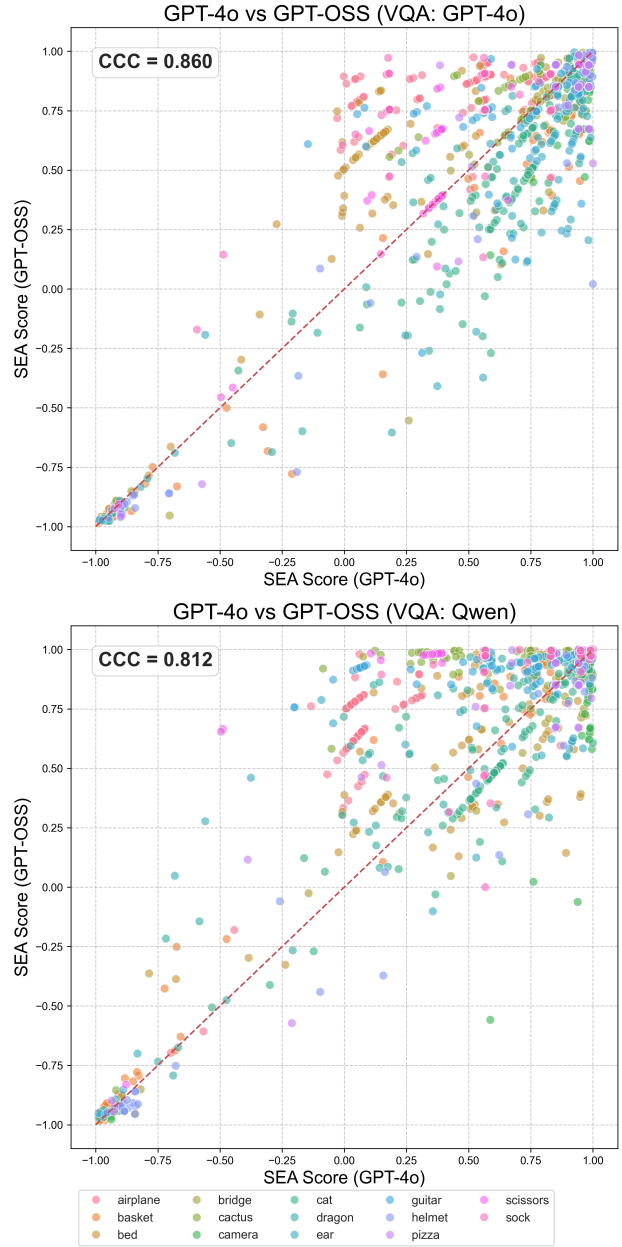


Figure 5. Comparison of SEA scores obtained using GPT-4o and GPT-OSS for commonsense extraction. The top and bottom panels show results with GPT-4o and Qwen as VQA models, respectively. The red dashed line indicates the identity line ($y = x$).

when GPT-OSS is used for commonsense extraction, indicating that the resulting score differences do not alter the qualitative ordering of sketches within each class. These results show that SEA tracks visual abstraction efficiency in a stable and interpretable manner, and that its qualitative ordering and quantitative scores remain consistent even when the LLM used for commonsense element extraction is replaced.

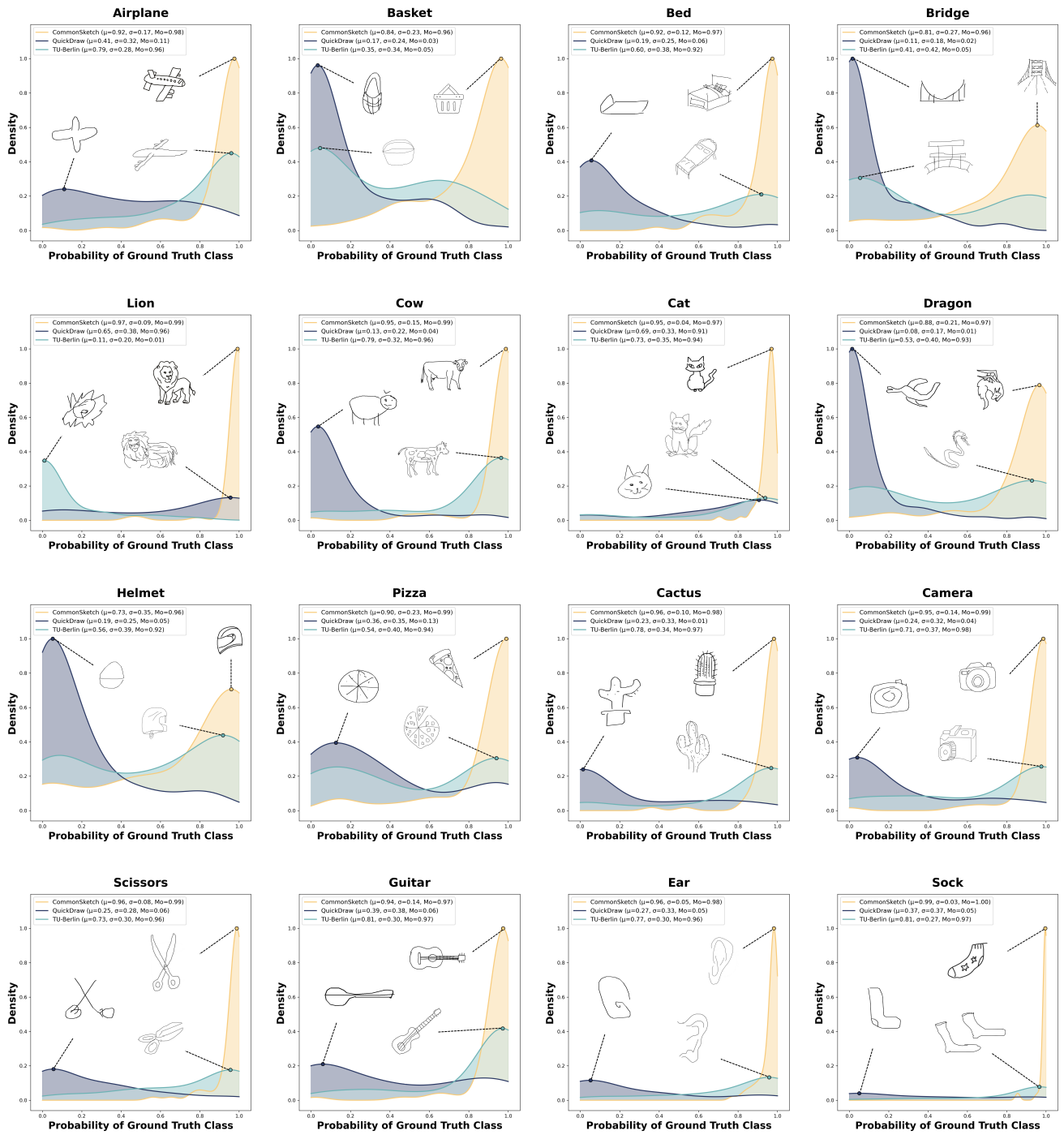


Figure 6. **Cross-dataset sketch quality comparison.** KDE plots show the *Probability of the Ground Truth Class* for 14 classes shared across CommonSketch, QuickDraw, and TU-Berlin, with one class chosen from each of the 14 categories. For each sketch, class probability is computed separately with CLIP, OpenCLIP, and CoCa, and the three values are then averaged to obtain a model-agnostic recognizability score. The distributions highlight differences in typical recognizability and variance across datasets, providing a fine-grained view beyond single-number quality metrics. Representative sketches at the *mode* of each KDE are displayed to link the quantitative trends to visually typical samples.

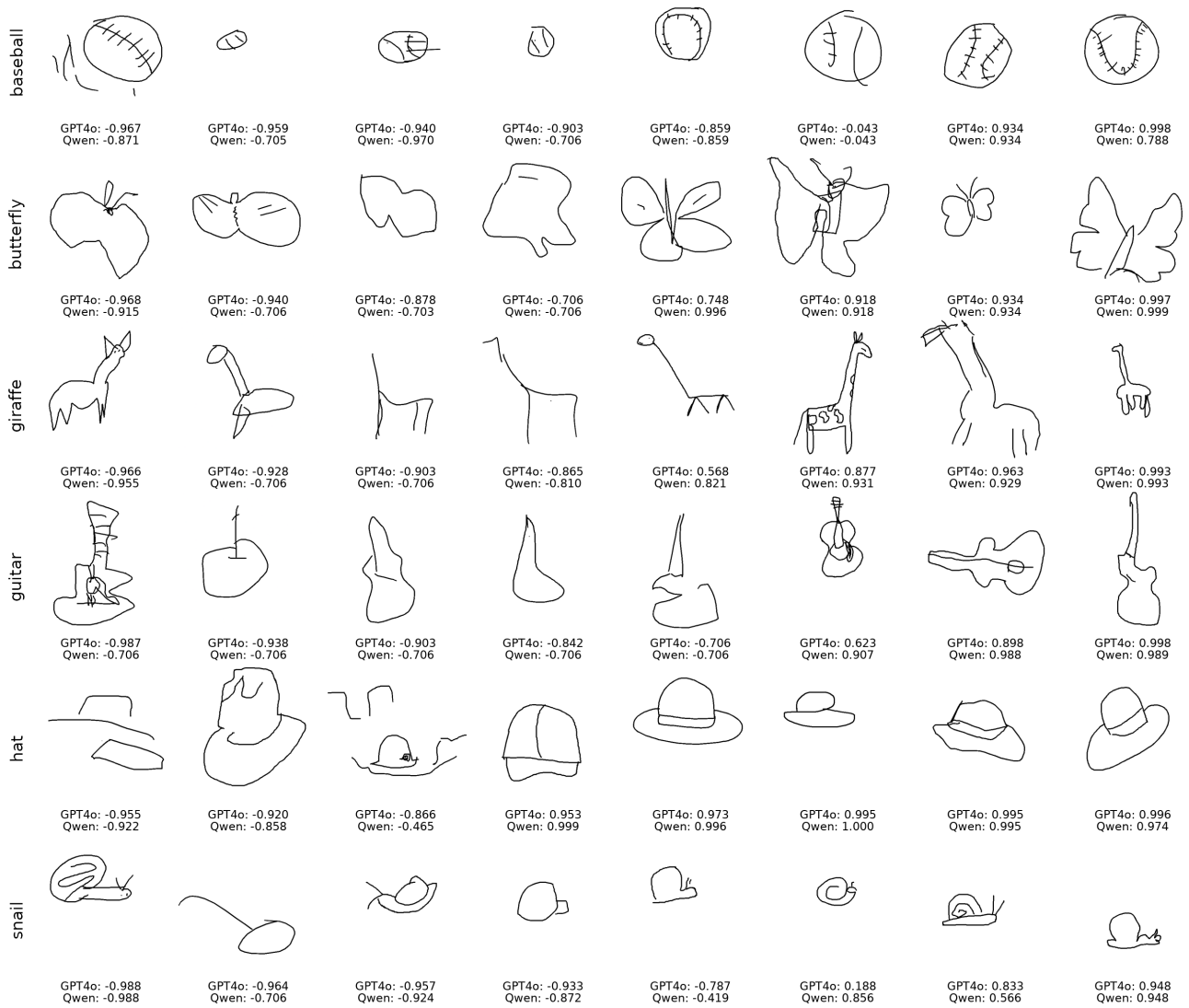


Figure 7. **Qualitative examples by SEA score on SEVA.** Six SEVA classes are shown (baseball, butterfly, giraffe, guitar, hat, snail), with eight sketches per class selected to span a low-to-high SEA range for visual inspection. All sketches display four SEA variants computed under OpenCLIP as a classifier: *GPT4o* (GPT-4o database + GPT-4o annotations), *Qwen* (GPT-4o database + Qwen annotations).



Figure 8. **Qualitative examples by SEA score on CommonSketch (set 1).** Seven CommonSketch classes are shown (airplane, basket, bed, bridge, cactus, camera, cat). All sketches display four SEA variants computed under OpenCLIP as a classifier: *GPT4o* (GPT-4o database + GPT-4o annotations), *Qwen* (GPT-4o database + Qwen annotations).



Figure 9. Qualitative examples by SEA score on CommonSketch (set 2). Seven additional CommonSketch classes are shown (dragon, ear, guitar, helmet, pizza, scissors, sock). All sketches display four SEA variants computed under OpenCLIP as a classifier: *GPT4o* (GPT-4o database + GPT-4o annotations), *Qwen* (GPT-4o database + Qwen annotations).

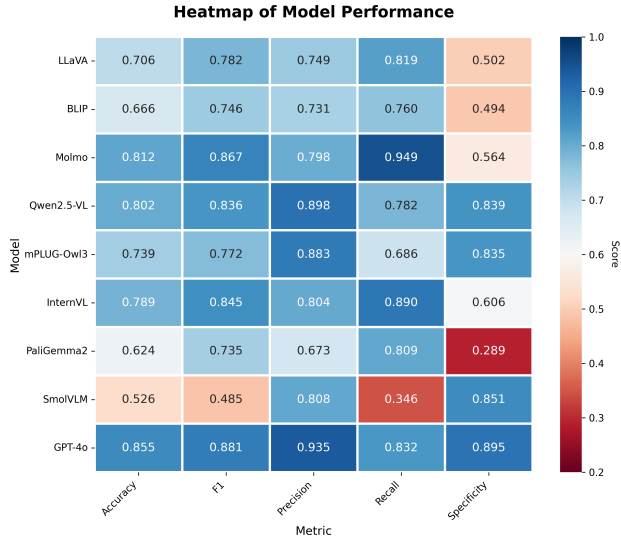


Figure 10. **Comprehensive performance evaluation of nine VLMs.** The heatmap visualizes model performance across five key metrics. Higher scores are in blue, and lower scores in red.

3.3. VLM Comparison on Annotation

We benchmark nine VLMs as element-presence annotators for SEA. Fig. 10 summarizes overall Accuracy, F1, Precision, Recall, and Specificity, and Fig. 11 reports category-wise accuracy across the 14 CommonSketch groups. GPT-4o is the strongest and most consistent annotator across categories. Among open-source models, Molmo [6] and Qwen2.5-VL [3] perform best: Molmo achieves the highest open-source accuracy but is recall-heavy, whereas Qwen2.5-VL exhibits a more balanced precision-recall profile with higher specificity, and follows GPT-4o’s per-category trends most closely. InternVL3 [35] and mPLUG-Owl3 [32] remain competitive but show larger category-to-category fluctuations, while LLaVA [24] and BLIP [23] perform substantially lower overall. PaliGemma2 [29] often over-predicts elements which mean lower specificity, whereas SmolVLM [26] tends to under-predict, lower recall.

Performance also depends on category. Annotation is relatively easier for categories with clear and repeatedly visible part structure (e.g., animal, structure, sports_equipment), and more difficult for categories with high intra-class shape variation or subtle defining cues (e.g., clothing, container). Fig. 10 further indicates that maintaining both precision and specificity is important for SEA, since false positives can inflate the visual-representation term. Overall, these results support GPT-4o as the primary annotator and motivate Qwen2.5-VL as the strongest open-source substitute, while highlighting categories where annotator choice can most affect SEA scores.

4. User Study Details

We report the human evaluation protocol, including the survey interface, question format, abstraction rating guidelines, and sampling strategy. The study was approved by an Institutional Review Board (IRB).

4.1. User Study Interface and Instructions

We conducted the study on a custom web-based survey platform. Fig. 12 presents the interface: participants first saw a short introduction and consent page (Fig. 12a), then answered a classification question for a single sketch (Fig. 12b), and finally rated abstraction for a set of four sketches from the same class (Fig. 12c). During classification, a progress indicator was shown, and after submission participants were informed only whether the response was correct, without revealing the true label, to limit learning effects across questions.

For classification, we used a four-option multiple-choice format to reduce burden given the long session length. Because our model benchmarks span roughly 430 possible classes across CommonSketch, QuickDraw, and TU-Berlin, we additionally provided an optional free-response field so that participants could enter an alternative label when none of the four candidates matched their judgment.

For abstraction scoring, participants used a continuous 0–4 slider with interval-specific guidelines aligned to SEA:

- 0–1: Abstraction failed; the target is hard to infer.
- 1–2: Incomplete abstraction; some cues exist but the target remains unclear.
- 2–3: Good abstraction; the target is clear but with noticeable detail.
- 3–4: Excellent abstraction; the target is clear despite strong simplification.

Sampling and study composition. To keep the session manageable, we evaluated 88 sketches in total. The core comparison set was drawn from CommonSketch, QuickDraw, and TU-Berlin: we selected one shared class from each of the 14 CommonSketch categories and sampled four sketches per class using quartiles of the model score distribution to cover a range of abstraction levels.

We additionally tested generalization in two settings. First, for class-level out-of-distribution evaluation, we used SEVA sketches and selected four classes with four sketches each, following SEVA’s predefined abstraction levels. Second, for domain-shift evaluation, we used pictogram-style sketches from Art Pictogram [4] and Flaticon [5, 7, 9–11, 16, 21, 22, 34], again selecting four classes with four sketches per class.

The survey was presented in three blocks: (i) the core cross-dataset comparison set, (ii) SEVA OOD samples, and (iii) pictogram-domain samples. Within each block, sketch

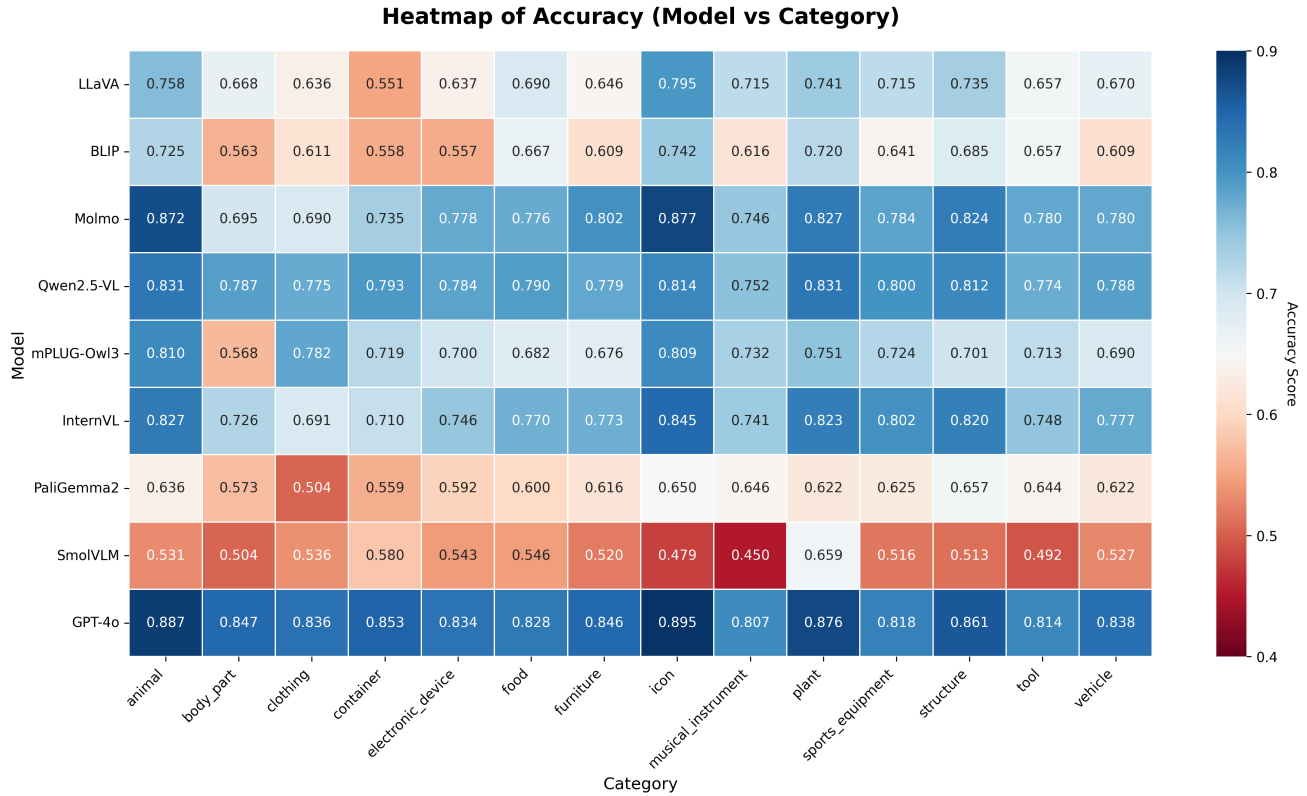


Figure 11. Per-category performance comparison of nine VLMs across the 14 categories of CommonSketch. The heatmap visualizes the accuracy scores, where blue indicates higher performance and red indicates lower performance.

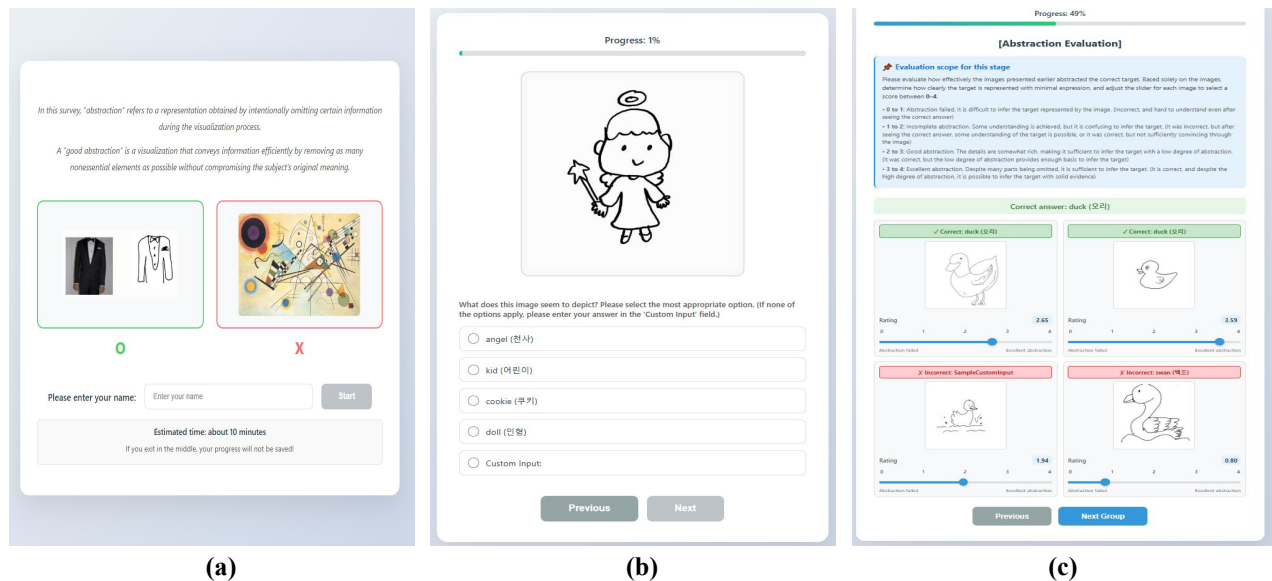


Figure 12. Human survey UI snapshots. (a) Landing page with study overview and consent. (b) Classification interface where participants choose among four candidates, with an optional free-response field for alternative labels. (c) Abstraction rating interface, where participants score each sketch on a continuous 0–4 slider with level-specific guidelines.

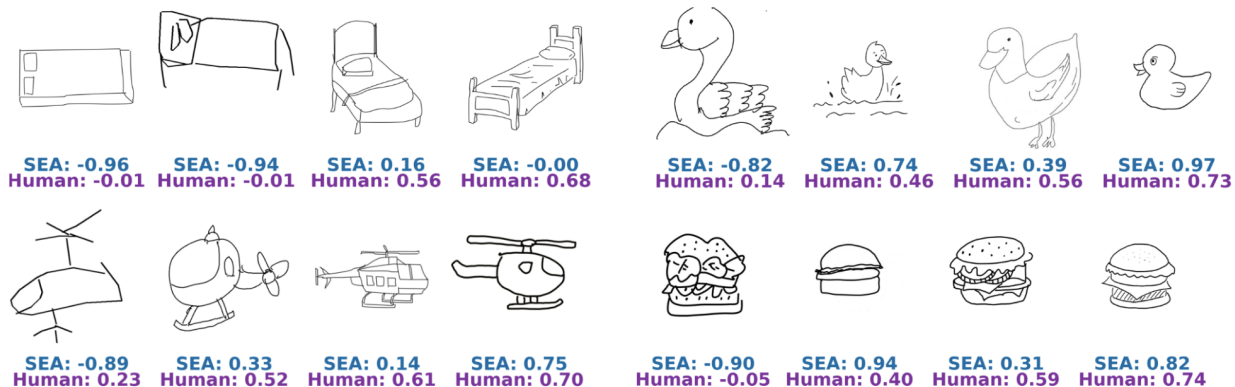


Figure 13. Comparison of human abstraction score distributions and SEA metric scores across four images per class. Density curves represent individual images, with point markers showing their medians and vertical dashed lines indicating SEA scores. Four images in each class are annotated with their SEA scores and mean human abstraction scores.

order was randomized, and the four multiple-choice options were shuffled per question. Extremely unclear or unfinished sketches were excluded and replaced.

Multiple-choice candidates were initialized from the top CLIP predictions and manually filtered to remove obviously unrelated labels. When CLIP did not yield plausible distractors, alternatives were selected from a larger GPT-5 candidate list, while ensuring the correct label was always included among the four options.

4.2. Human-SEA Alignment Analysis

We assess alignment between SEA and human abstraction judgments by comparing their score distributions on the same sketches. Fig. 13 shows four sketches per class with paired SEA and human scores. The two measures agree at the extremes: unrecognizable sketches receive low scores, while recognizable and well-abstracted sketches receive high scores, and class-level ranking trends are largely consistent. A clear gap emerges in the mid-range around a human score of about 0.5, where participants tend to favor drawings that are immediately recognizable even with extra detail, whereas SEA rewards sketches that retain core elements with fewer marks. Overall, SEA matches human intuition for failed versus successful abstraction, but applies a stricter preference for minimal sufficient evidence in borderline cases.

4.3. Generalization on Pictograms

To examine whether the SEA generalizes beyond the sketch domain, we compared SEA with human evaluations. We observed that the model’s generalization ability increased in a manner similar to human scores, which can be seen in Fig. 14. However, some clean line drawings occasionally caused an unexpected drop in classification performance, which can be seen in Fig. 15, leading to outlier metric values. These results suggest that SEA can be meaningfully

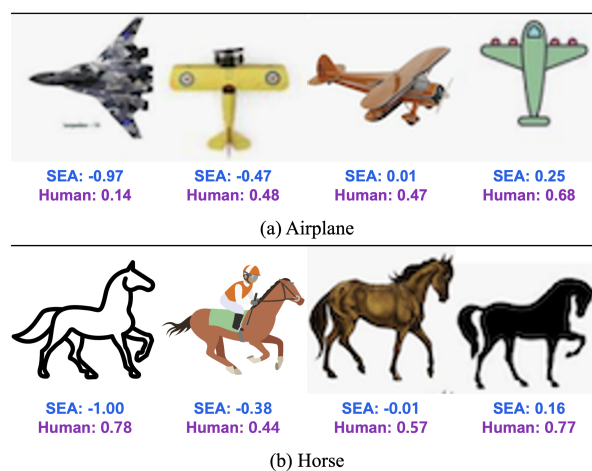


Figure 14. Comparison of human scores and SEA values across pictograms to assess generalization ability.

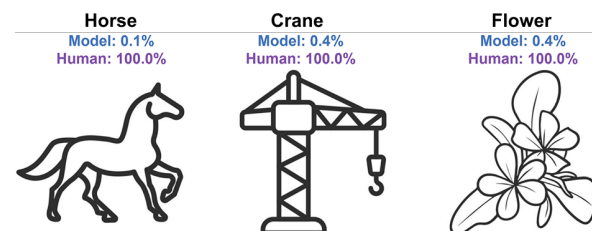


Figure 15. Comparison between classification model performance and human accuracy.

applied to pictographic symbols that share the same underlying class semantics as our sketches, and that the metric exhibits a degree of generalization beyond the original sketch domain, provided that the backbone model has sufficient coverage of the visual style.

5. Prompts for Dataset Construction

5.1. Sketch Validation Cycle

We use **GPT-4o** to generate multiple captions per sketch during the validation cycle. The exact prompt is shown below.

Model: GPT-4o

System message

You are a helpful assistant for generating image captions.

User message

Please describe this image 5 times based on the following format. The input image is provided as a base64-encoded JPEG string.

Output template

"A black line drawing of {{text1}} on a white background."

OR

"A simple drawing of {{text1}} on a white background."

Instructions

- Replace {{text1}} with a detailed description of the image.
- Avoid vague descriptions; focus on clear details such as objects, shapes, and actions.
- The fourth and fifth descriptions must focus on unexplained details in the other descriptions, except for the main object.
- Do not include “{{}}” in the final output.
- Choose the appropriate template based on the complexity of the image.
- Separate each description with \n\n.
- Do not put any numbers or symbols in front of the descriptions.
- Do not use commas (“,”).

5.2. Commonsense Extraction

We use four language models for commonsense element extraction: **GPT-4o**, **GPT-OSS-20B**, **Qwen-2.5 32B**, **Llama 3 8B**, and **Mistral 7B**. The exact prompts are shown below.

Model: GPT-4o

Instruction prompt

You are a sketch analysis expert. Your task is to extract a structured list of **common visual elements** that are typically included — or **semantically expected** — when humans sketch a given object class.

Use the object class name, along with general visual common sense and knowledge of object structure, to infer **as many relevant visual components as possible**.

Your goal is to produce a **comprehensive and fine-grained breakdown** of visual parts, including:

- core parts,
- minor or optional parts,
- functional attachments,
- repeated units,
- motion-related components (e.g., rotating blades, walking legs),
- relevant environmental or contextual elements.

Even if a part is rarely drawn, include it if it is **semantically meaningful or distinctive** for understanding or sketching the object, and assign lower `importance_score` accordingly.

This output will be used to build a **commonsense database** for sketch abstraction, so prioritize **coverage and interpretability**.

For each visual element, return the following fields:

- `id`: in the form `<class>.<element_name>`
- `name`: the name of the part
- `shape`: geometric form (e.g., circle, triangle, curve)
- `position`: typical relative location in the object
- `count`: usual number (e.g., 1, 2, or “varies”)
- `importance_score`: integer from 1 to 5 (5 = essential; 1 = optional or rare)
- `optional`: true or false

- `description`: what it looks like and why it is relevant

Return the result strictly in the following structure:

```
{
  "class": "{class_name}",
  "total_elements": <number_of_elements>,
  "elements": [
    ...
  ]
}
```

Important general rules:

- Do not include color information.
- Do not include fictional or humorous features.
- Do not include decorative elements unless they are functionally or culturally tied to the class.
- Use consistent, interpretable IDs in the format `<class>.<element_name>`.
- Include both (1) frequently drawn elements and (2) structurally important elements even if rarely drawn.
- Include context or environment features only if they are logically essential to how the object is typically depicted.
- Think about what makes this class visually different from nearby classes and reflect that in part selection.
- Favor over-inclusion: include more elements with appropriately scaled `importance_score`.

Class: `{class_name}`

Models: GPT-OSS 20B, Qwen-2.5 32B, Llama 3 8B, Mistral 7B

Instruction prompt

You are a **Structured Visual Object Analyzer for sketches**. Your job is to output a **single JSON object** describing the visible parts of a sketched object class.

Hard rules (follow strictly):

1. **No environment-only items.**

Do not include background or scene items that are not intrinsic parts of the object (e.g., no clouds for *sun*, no road or buildings for *car*).

2. **Visible-only.**

Include only parts plausibly visible in a typical sketch; exclude hidden internals (e.g., car engine, phone mainboard).

3. **Variants allowed, naming rules apply.**

Common sketch variants replacing or decorating real parts may be included with `"optional": true` (e.g., `human_mouth` on an insect). Do not use the word “stylized”; do not use parentheses or brackets. All names must be in **snake.case** (lowercase, digits allowed, words separated by a single underscore). Examples: `steam_lines`, `wing_vein_lines`, `tail_fan`, `human_mouth`.

4. **Expressive lines/effects.**

Expressive effects (e.g., `airflow_lines`, `motion_lines`, `steam_lines`, `sparkle`) are excluded by default. They may be included only when they represent an essential and commonly used feature of the object’s sketch and must then be marked `"optional": true`. Background-only elements (ground, sky, clouds, water, etc.) must still be excluded.

5. **Merge symmetric or duplicated parts.**

Merge symmetric repeats (e.g., left/right wheels, pairs of legs) into a single element (e.g., `wheels`, `legs`).

6. **Coverage and granularity.**

Produce a rich but concise set of features (recommended 9–16 elements). Prefer coarse-to-mid granularity: split obvious appendages or facial parts (head, arms, legs) instead of using a single `body`. Consider including elements from: (a) anatomy or core shape, (b) facial features, (c) iconic clothing or accessories, (d) explicit surface, texture, or pattern marks (e.g., `seed_dots`, `peel_lines`, `feather_lines`, `shell_pattern`, `fur_lines`), (e) expressive lines only if visibly drawn.

7. **Ground truth first, then variants.**

List physically correct parts first, then common variants or expressive features with `"optional": true`.

8. **Labelability and non-ambiguous features.**

Every feature must be binary labelable (0/1) from the sketch without subjective judgment. Disallow vague descriptors (e.g., `smooth_surface`, `shiny_surface`). Do not describe the absence of texture (e.g., `no_texture`). Prefer positive, observable evidence (lines, dots, edges, explicit patterns), such as `glaze_lines`, `seed_dots`, `crack_lines`, `slice_lines`.

Output format (exact structure):

```

{
  "class": "<object_name>",
  "total_elements": <int>,
  "elements": [
    {
      "id": "<class_name>.<part_name>",
      "name": "<part_name>",
      "optional": <true or false>
    },
    ...
  ]
}

```

Additional guidance:

- "total_elements" must equal the number of objects in "elements".
- Output JSON only (no commentary outside the JSON).
- <part_name> must be snake_case; IDs must be of the form <class_name>.<part_name>.
- Ensure at least 8 elements (prefer 9–16) and reasonable coverage of anatomy, facial features, accessories, surface or pattern, and expressive lines.
- Ensure all features are 0/1 labelable and not environment-only.

Final instruction: Now, provide the structured JSON for the following object: {word}.

5.3. Element Annotation

For element-level commonsense VQA, we use the following vision–language models: **GPT-4o**, **Qwen2.5-VL 7B**, **mPLUG-Owl3 7B**, **InternVL3 8B**, **Molmo 7B**, **PaliGemma2 3B**, **SmolVLM 500M**, **LLaVA 1.5 7B**, and **BLIP**. We list the exact prompts below, grouping models that share the same template.

Models: GPT-4o, Qwen2.5-VL 7B, mPLUG-Owl3 7B

Instruction prompt

<|image|>

You are a strict vision auditor for sketched objects.

Target class: "{class_name}".

Valid elements for this class (use **only** these ids; do not add new keys):

{element_block}

Task: For **each** element id above, return only whether the element is depicted (true/false).

Do not return counts. If ambiguous, use false.

Return only a compact JSON object with element ids as keys and boolean values (true/false).

No prose, no code block, no extra keys.

Example schema (structure only):

```

{
  "element_id.1": true,
  "element_id.2": false,
  ...
}

```

Models: InternVL3 8B, PaliGemma2 3B, SmolVLM 500M

Question template

In this {class_name} image, does this sketch contain a {e}?

Answer exactly “yes” or “no”.

For SmolVLM we wrap the question in the model’s chat template:

```

<|user|>
<image>
In this {class_name} image, does this sketch contain a {e}? Answer exactly 'yes'
or 'no'.
<|end|>
<|assistant|>

```

Model: Molmo-7B-D-0924

Instruction prompt

You are an assistant that analyzes an image of a {category} and answers in JSON format only.

Task: For the given {category} image, decide if each of the following elements is present (1) or not present (0):

[{element_list}].

Return the result strictly as a JSON object in the following format:

```

{
  "{file_name}": {
    {element_lines}
  }
}

```

Do not include explanations or extra text. Output only valid JSON.

Model: LLaVA 1.5 7B

Question template

In this {category} image, is there a {element}?

Answer Yes or No.

Model: BLIP (blip-vqa-capfilt-large)

Question template

In this {category} image, is there a {element}?

Table 6. **Per-class element list.** For each class, we list the commonsense elements extracted by GPT-4o and GPT-OSS. Black indicates elements shared by both models; **red** denotes elements unique to GPT-4o; **blue** denotes elements unique to GPT-OSS. The 4o/OSS column reports the number of elements extracted by GPT-4o and GPT-OSS 20B, respectively.

Category	Class	4o/OSS	Elements
animal	alpaca	13/11	body, ears, eyes, head, legs, tail, feet, fleece, hooves, muzzle, neck, nostrils, smile, fur_lines, motion_lines, mouth, nose, whisker_lines
	ant	11/9	abdomen, antennae, head, legs, mandibles, thorax, compound eyes, jointed legs, mouth, petiole, stinger, body, eyes, segment_lines
	bat	14/12	body, ears, eyes, head, mouth, nose, tail, wings, feet, fingers, fur, legs, teeth, wing membrane, claws, fur_lines, motion_lines, wing_vein_lines
	bee	14/10	abdomen, antennae, head, legs, mouth, stinger, thorax, wings, compound eyes, flower context, flying motion, hairs, mouthparts, stripes, body_stripe_lines, eyes
	bird	15/10	beak, body, feet, head, legs, tail, wings, branch, crest, eyes, feathers, flying pose, standing pose, tail feathers, wing feathers, eye, feather_lines, motion_lines
	boar	13/9	body, ears, head, legs, snout, tail, tusks, brush, eyes, hair, hooves, hump, mane, eye, fur_lines
	butterfly	13/9	antennae, body, eyes, head, legs, proboscis, flight posture, flowers, lower wings, mouth, resting posture, upper wings, wing patterns, tail, wing_pattern, wings
	camel	13/9	body, ears, eyes, head, hump, legs, mouth, neck, tail, feet, fur, nostrils, saddle
	cat	15/14	body, ears, eyes, head, legs, mouth, nose, paws, tail, whiskers, bowtie, claws, fur, fur markings, whisker pads, collar, fur_lines, motion_lines, stripe_lines
	caterpillar	13/10	antennae, body, head, legs, mouth, tail, environment, eye, mandibles, prolegs, segments, setae, spiracles, body_stripes, dorsal_spines, eyes, motion_lines

Continued on next page

Category	Class	4o/OSS	Elements
	chameleon	13/9	body, eyes, head, legs, mouth, tail, branch, crest, feet, gular pouch, nostrils, scaly skin, tongue, claws, prehensile.tail, vertical_eye_slit
	cow	15/13	body, ears, eyes, head, hooves, horns, legs, mouth, spots, tail, muzzle, neck, nose, pasture, udder, fur_lines, motion_lines, nostrils
	crab	12/9	antennae, body, claws, eyes, legs, mouth, gills, joints, legs bend, mouthparts, shell spikes, shell texture, motion_lines, shell_pattern, tail
	crocodile	13/10	body, eyes, head, legs, nostrils, tail, teeth, feet, hunting pose, jaws, ridges, scales, underbelly, mouth, scale_pattern, snout
	deer	12/12	antlers, body, ears, eyes, head, hooves, legs, mouth, neck, nose, spots, tail
	dog	15/12	body, collar, ears, eyes, head, legs, mouth, nose, paws, tail, fur, fur markings, teeth, tongue, whiskers, fur_lines, motion_lines
	dolphin	12/9	body, head, beak, blowhole, dorsal fin, eye, pectoral fins, skin texture, smile, tail flukes, water splash, wave, dorsal_fin, eyes, motion_lines, mouth, pectoral_fins, skin_pattern, tail_fin
	dragonfly	13/9	abdomen, antennae, head, legs, thorax, wings, compound eyes, flight motion, mouthparts, resting position, tail, wing nodules, wing veins, eyes, motion_lines, wing_veins
	duck	13/11	beak, body, head, neck, breast, eye, feather detail, leg, nostril, tail, water, webbed foot, wing, eyes, feather_pattern, feet, legs, tail_feather_lines, wing_feather_lines, wings
	elephant	12/14	body, ears, head, legs, tail, trunk, tusks, belly, eye, mouth, toes, wrinkles, eyes, feet, foot_pad_detail, inner_ear_folds, skin_ridges, trunk_detail, tusk_detail
	feather	8/10	barbs, tip, base, fluff, fused barbs, quill, separated barbs, vane, barb_pattern, barbules, calamus, feather_curve, feather_fan, motion_lines, root, shaft
	fish	13/9	body, head, mouth, anal fin, dorsal fin, eyes, gills, markings, pectoral fins, pelvic fins, scales, tail fin, water environment, dorsal_fin, eye, pectoral_fins, pelvic_fins, scale_pattern, tail
	flamingo	14/11	beak, body, eye, head, neck, bent neck, feather, foot, knee, leg, standing pose, tail, water, wing, feather_lines, feet, legs, motion_lines, wing_vein_lines, wings
	frog	13/10	body, eyes, head, legs, mouth, nostrils, tongue, belly, body pattern, environment water, jump motion, sitting posture, toes, feet, skin_pattern, webbing
	giraffe	15/11	body, ears, eyes, head, hooves, legs, mouth, neck, spots, tail, horns, mane, nostrils, tongue, trees, nose
	goose	14/13	beak, body, head, legs, neck, tail, wings, eyes, feathers, neck bend, nostrils, water, webbed feet, wing feathers, eye, feather_lines, feet, motion_lines, webbing_lines, wing_folds
	hedgehog	11/10	body, eyes, feet, head, legs, mouth, nose, spines, tail, ears, snout, whiskers
	hippopotamus	12/9	body, ears, eyes, head, legs, mouth, tail, teeth, feet, nostrils, skin, water, snout
	horse	12/14	body, ears, eyes, head, hooves, legs, mane, mouth, tail, muzzle, neck, nostrils, nose, reins, saddle, teeth, whiskers
	jellyfish	6/8	tentacles, bell, markings, oral arms, radial canals, water, body, internal_rings, motion_lines, mouth, radial_lines, symmetry_lines, tentacle_tips
	kangaroo	10/13	body, ears, eyes, head, mouth, nose, pouch, tail, arms, legs, feet, front_legs, fur_lines, hind_legs, motion_lines
	koala	12/11	body, claws, ears, eyes, head, legs, mouth, nose, tail, arms, fur, tree, fur_lines, paws
	lion	15/12	body, eyes, head, legs, mane, mouth, nose, paws, tail, whiskers, claws, ears, ground, jaw, tail tuft, fur_lines, motion_lines
	lobster	12/12	claws, eyes, legs, mouth, tail, antennas, body, environment (water), mouthparts, rostrum, shell segments, shell texture, abdomen, antennae, carapace_lines, head, motion_lines, tail_fan, thorax
	mole	12/10	body, ears, eyes, head, legs, mouth, nose, tail, whiskers, claws, fur texture, tunnel, fur_lines
	monkey	15/13	arms, body, ears, eyes, feet, hands, head, legs, mouth, nose, tail, banana, face, fur, tree branch, fur_lines, whiskers
	moose	13/11	antlers, body, ears, eyes, head, hooves, legs, mouth, nose, tail, mane, muzzle, neck, fur_lines
	mouse	11/9	body, ears, eyes, head, legs, nose, tail, whiskers, feet, fur, teeth, mouth
	octopus	8/8	arms, eyes, head, mouth, siphon, skin texture, suckers, water, ink_lines, mantle, spot_pattern, web
	owl	13/10	beak, body, eyes, head, tail, talons, wings, belly, ear tufts, feathers, feet, flight pose, perch, ear_tufts, feather_lines, legs
	panda	13/13	arms, body, ears, eyes, head, legs, mouth, nose, tail, bamboo, black patches, claws, sitting pose, black_patch_arms, black_patch_ears, black_patch_eyes, black_patch_legs
	parrot	13/11	beak, body, eyes, feathers, feet, head, tail, wings, branch, cheeks, crown, nostrils, throat patch, claws, feather_lines, plumage_pattern
	peacock	12/10	beak, body, crest, eyes, head, neck, wings, feet, legs, tail, tail eyes, tail feathers, motion_lines, tail_eye_spots, tail_fan
	penguin	10/9	beak, body, eyes, feet, head, tail, wings, belly, group, ice, feather_lines, motion_lines
	rabbit	13/11	body, ears, eyes, head, mouth, nose, paws, tail, whiskers, carrot, fur, hopping motion, legs, fur_lines, teeth
	rooster	12/12	beak, body, comb, eyes, head, legs, tail, wings, crowing pose, feather patterns, feet, wattle, claws, feather_lines, motion_lines, wattles
	scorpion	12/9	body, eyes, head, legs, stinger, tail, claw joints, exoskeleton, leg joints, mouthparts, pincers, segmented tail, claws, exoskeleton_pattern, mouth
	seahorse	10/9	body, head, tail, belly, dorsal fin, environment: water, eyes, pectoral fins, ridges, snout, dorsal_fin, eye, mouth, pectoral_fins, spine_lines, tail_fin
	shark	13/10	body, eyes, head, mouth, teeth, anal fin, dorsal fin, environmental water, gills, pectoral fins, pelvic fins, second dorsal fin, tail fin, caudal_fin, dorsal_fin, gill_slits, pectoral_fins, pelvic_fins
	sheep	13/10	body, eyes, head, legs, mouth, tail, ears, eyebrows, grass, hooves, nose, snout, wool texture, fur_lines, horns, motion_lines, nostrils
	sloth	13/12	arms, body, claws, eyes, head, legs, mouth, nose, tail, branch, face, fur, hanging position, ears, fur_lines, motion_lines
	snail	7/9	eyes, mouth, shell, tentacles, body, environment ground, shell spiral, aperture, foot, head, spiral_lines, tail
	snake	9/8	body, eyes, fangs, head, tail, coiling, rattle, scales, tongue, motion_lines, mouth, scale_lines
	spider	10/8	abdomen, cephalothorax, eyes, fangs, legs, body, hairs, joints, pedipalps, web, body_pattern, leg_ends, spinnerets
	squid	10/8	arms, eyes, head, mantle, tentacles, body, fins, mouth, ocean, suckers, beak, motion_lines, tentacle_clubs
	squirrel	13/12	body, ears, eyes, head, legs, nose, tail, whiskers, cheeks, claws, feet, fur, nut, arms, fur_lines, mouth, paws
	swan	11/10	beak, body, eyes, head, legs, neck, tail, wings, feathers, water, webbed feet, feather_lines, feet
	tiger	15/12	body, ears, eyes, head, legs, mouth, nose, paws, tail, whiskers, claws, fur, jaw, stripes, teeth, fur_pattern, stripe_pattern
	turtle	11/10	eyes, head, legs, mouth, shell, tail, claws, flippers, neck, shell patterns, water, motion_lines, shell_crack, shell_pattern, shell_veins
	whale	13/10	blowhole, body, eye, head, mouth, dorsal fin, pectoral fins, skin texture, tail fluke, teeth, ventral pleats, water spray, water surface, dorsal_fin, flippers, motion_lines, spout, tail

Continued on next page

Category	Class	4o/OSS	Elements
	zebra	14/12	body, ears, eyes, head, legs, mane, mouth, neck, tail, hooves, horizon line, muzzle, nostrils, stripes, motion_lines, nose, stripe_lines
body part	ear	11/9	Antihelix, Antitragus, Concha, Ear Canal, Ear Lobe, Earring, Head, Helix, Outer Ear, Sound line, Tragus, ear_antihelix, ear_antitragus, ear_canal, ear_cartilage_lines, ear_helix, ear_lobe, ear_shape, ear_tragus, earring
	eye	7/9	eyelashes, eyelid, iris, pupil, sclera, eyebrow, tear duct, eye, highlight, iris_lines, tear_drop
	foot	9/9	arch, heel, toes, ankle, ankle bone, ball, foot wrinkles, footprint, toenails, ball_of_foot, foot_back, foot_side, motion_lines, sole, toe_nails
	hand	13/11	knuckles, palm, thumb, wrist, finger joints, finger print, index finger, lines, little finger, middle finger, nails, ring finger, veins, finger_lines, finger_tips, fingers, knuckle_lines, nail_bases, palm_lines, thumb_tip
	mouth	7/9	Corners, Dimple, Lower Lip, Teeth, Tongue, Upper Lip, Uvula, frown_line, gums, lower_lip, mouth_corners, mouth_gap, smile_line, teeth, tongue, upper_lip
	nose	6/9	Ala, Bridge, Dorsum, Nostrils, Septum, Tip, nose_bridge, nose_hair, nose_shadow, nose_shape, nose_tip, nostril_lines, nostril_openings, nostrils, skin_dots
	tooth	4/9	crown, root, grooves, gums, canine_point, enamel, enamel_pattern, fissure_lines, incisal_edge, occlusal_surface, root_tip
clothing	belt	8/9	Buckle, Decorative Elements, End Tip, Holes, Loop, Prong, Stitching, Strip, belt_end, belt_holes, buckle, buckle_frame, buckle_prong, buckle_shank, leather_grain, stitching_lines, strap
	bowtie	7/9	Central Knot, Fold Lines, Left Loop, Left Tail, Neck Band, Right Loop, Right Tail, bowtie_shape, center_button, center_knot, fabric_edge_lines, fabric_fold_lines, knot_detail, knot_strut, loops, pattern_lines
	crown	7/9	Arches, Base Band, Cross, Embellishments, Fleur-de-lis, Jewels, Spikes, base_ring, central_jewel, central_spike, decorative_lines, jewel_details, jewels, pattern_lines, sparkle_lines, spikes
	flip flops	6/10	cushion layer, decoration, side straps, sole, strap attachment points, toe strap, heel_area, heel_connector, heel_pattern, motion_lines, sole_outline, strap, strap_pattern, thong, toe_area, toe_pattern
	hat	8/10	band, brim, crown, feather, chin strap, ear flaps, peak, pom-pom, decorative_stone, logo, pattern_dots, pattern_logo, pattern_stripes, strap
	shoe	12/12	eyelets, heel, laces, logo, sole, tongue, upper, ankle collar, insole, strap, toe box, tread, brand_name, heel_strap, sole_tread, stitching_lines, toe
	sock	7/8	body, cuff, heel, toe, ankle, pattern, seam, heel_fold, knit_pattern, sag_lines, stretch_lines
	t-shirt	8/10	body, Detail, Printing, hem, neckline, shoulder seam, side seam, sleeve, buttons, collar, pocket_lines, pockets, print, seam_lines, sleeves, stitch_lines, stripe_lines
container	backpack	11/11	Back Panel, Buckles, Front Pocket, Laptop Compartment, Logo Patch, Main Compartment, Shoulder Straps, Side Pockets, Top Handle, Zippers, chest strap, back_pocket, body, front_pocket, handle, patch, shoulder_straps, side_pockets, stitching_lines, strap_harness, top_flap, zipper_line
	basket	7/12	handle, rim, body, decoration, lid, supports, weave pattern, base, base_pattern, handle_bend, handle_grip, side_pattern, sides, strap, top, top_pattern, weave_pattern
	bucket	6/8	body, handle, rim, contents, grip, pour spout, bottom, crack_lines, lid, splash_lines, water_lines
	envelope	4/8	Address Area, Flap, Rectangular Body, Seal Line, address_area, body, corner_edges, flap, fold_lines, paper_pattern, seal, stamp
	mailbox	9/7	body, door, flag, handle, label, post, base, letter, mail slot, mail_slot
	present	6/9	bow, ribbon, box, gift tag, lid, ribbon tails, bow_clasp, bow_loop, box_body, box_edge_lines, fold_lines, paper_pattern, ribbon_stripe
	purse	10/10	body, clasp, handle, strap, zipper, brand logo, chain strap, flap, lining, stitching, chain, decorative_lines, front_flap, logo, pocket
	suitcase	11/10	handle, wheels, zipper, Sticker, body, corner protectors, feet, locks, luggage tag, retractable handle, side handle, body_texture, case_body, folding_lids, handle_stripe, lock, logo, motion_lines
	wine bottle	8/9	cork, label, body, foil, neck, opening, punt, shoulder, bottle_base, bottle_body, bottle_neck, cork_strand, label_frame, label_pattern, label_text
electronic device	alarm clock	8/10	body, alarm bell, alarm switch, clock face, clock hands, legs, numbers, second hand, alarm_bell, alarm_button, dial, dial_numbers, display_screen, hour_hand, hour_markers, minute_hand, minute_markers
	calculator	10/10	body, button labels, buttons, clear button, display, equal button, function buttons, memory buttons, power button, solar panel, clear_button, display_area, display_borders, equals_button, keypad_grid_lines, memory_buttons, numeric_buttons, operator_buttons, power_button
	camera	10/9	body, flash, lens, grip, hot shoe, lens cap, mode dial, shutter button, strap hooks, zoom ring, buttons, dials, memory_card_slot, mirror, strap, viewfinder
	cell phone	8/10	body, screen, antenna lines, front camera, home button, logo, side buttons, speaker, battery_indicator, camera_lens, charging_port, fingerprint_sensor, home_button, notch, speaker_grille, volume_buttons
	charger	5/8	cable, cable connector, indicator light, main body, prongs, cable_knot, cable_strands, connector, motion_lines, plug, plug_pins, power_symbol
	computer	7/12	mouse, CD drive, USB port, body, keyboard, power button, screen, cable, headphone_jack, keyboard_body, keys, logo, monitor_body, monitor_screen, power_button, speaker, stand, usb_port
	fan	8/8	base, blades, environmental airflow, guard, motor housing, pole, power cord, speed control, cable, hub, motion_lines, rim, shade, shaft
	headphones	7/9	headband, adjustment slider, cushion, ear cup, microphone, speaker, wire, cable, cable_twists, earcup_ear_pad_pattern, earcup_padding, earcups, headband_suspension_lines, headphone_logo, plug
	ipod	8/10	body, center button, click wheel, dock connector, headphone jack, hold switch, logo, screen, apple_logo, bottom_button, charging_port, click_wheel, display, headphone_jack, side_buttons, speaker_grill, top_button
	keyboard	9/12	body, keys, arrow keys, cable, enter key, escape key, indicator lights, numpad, space bar, arrow_keys, case_pattern, function_keys, key_labels, keyboard_backlight, keycap_grooves, num_pad, side_buttons, space_bar, usb_connector
	laptop	5/12	body, keyboard, screen, touchpad, webcam, brand_logo, camera, headphone_jack, hinge, keycap_letters, power_button, screen_border, trackpad, usb_ports
	megaphone	7/9	handle, mouthpiece, cone, microphone, speaker grille, trigger, volume control, back_opening, back_opening_curve, back_opening_edge, body, handle_curve, handle_grip, mouthpiece_edge
	microphone	7/9	body, cable, clip, grille, head, switch, windscreen, grille_pattern, handle, motion_lines, speaker_circles, stand
	microwave	7/12	body, display, door, buttons, control panel, handle, turntable, back_panel, control_panel, door_handle, door_hinge, door_latch, grill, logo, power_knob, side_panel

Continued on next page

Category	Class	4o/OSS	Elements
	oven	9/9	body, door, handle, control panel, digital display, feet, knobs, light, rack, control_knob, display_panel, door_frame, door_latch, glass_window, vent_grill
	radio	9/10	antenna, body, dial display, handle, knob markers, power button, speaker grill, tuning knob, volume knob, casing, headphone_jack, main_dial, power-button, power_light, sound_waves, speaker_grill, tuning_dial, usb_port
	robot	12/12	arms, body, head, legs, mouth, wheels, antennas, eyes, feet, hands, joints, treads, antenna, hinge_lines, motion_lines, panel_lines, power_core, sensor_eyes
	satellite	6/8	antenna, body, dish, sensor, solar panel, support structure, exhaust_lines, logo, panel_mast, solar_panels, support_arms, thrusters
	telephone	11/9	handset, microphone, speaker, base, cord, dial pad, display, hook switch, receiver hook, ringer, volume control, body, call_button, cradle, keypad, logo, speaker_grill
	television	6/10	screen, stand, control buttons, frame, remote sensor, speakers, antenna, body, buttons, logo, panel_lines, ports, screen_grid_lines, speaker_grills
	toaster	9/10	body, button, lever, brand label, control knob, foot, power cord, slot, toast, crumb_tray, indicator_light, lid, metal_grill_lines, metal_rails, power_cord, slots
	walkie talkie	8/10	antenna, body, display, microphone, button, channel knob, speaker, volume dial, battery_compartment, channel_buttons, power-button, speaker_grill, strap, volume_buttons
food	apple	11/7	body, leaf, stem, bite mark, bug hole, calyx, core, seeds, skin texture, slice, surface_spots, crack_lines, cut_line, motion_lines, seed_cluster
	asparagus	5/9	stalk, tip, base, bundle_tie, scaled buds, leaf, leaf_cluster, leaf_pattern, root, root_cap, root_pattern, stalk_pattern
	banana	6/9	Curved Body, Hands, Peel, Peel Segments, Stem, Tip, body, fruit_head, fruit_tail, leaf, leaf_base, leaf_tip, peel_lines, stem, stem_tip
	bread	9/9	Crumbs, Crust, Ends, Holes, Loaf Shape, Slash Marks, Slice Cross Section, Slices, Surface Features, crumb, crumb_pattern, crust, crust_pattern, knife, loaf, slice, slice_cut_line, steam_lines
	broccoli	6/8	Branch, Cut End, Floret, Floret Cluster, Stalk, Texture Hint, floret_base_lines, floret_leaflets, floret_pattern_lines, floret_tip_lines, florets, leaves, stem, stem_veins
	cake	9/9	base layer, candle, frosting, fruit topping, icing decoration, layer divider, plate, slice, upper layer, base, cake_body, cake_top, candles, frosting_lines, icing, layers, ribbon, sprinkles
	carrot	6/9	body, greens, ridges, root hairs, tip, top, furrows, leaf_cluster, leaves, root_tip, seed_dots, slice_lines, stem, surface_ridges
	cookie	6/8	base shape, bite mark, cracks, icing, texture, toppings, bite_mark, body, chips, crack_line, crumbs, glaze_lines, hole, texture_lines
	cupcake	9/8	frosting, sprinkles, base, topping (cherry, strawberry, heart cookie, etc.), wrapper, wrapper_pleats, cake_body, eyes, frosting_crown, frosting_swirl, mouth, paper_cup
	donut	6/9	Bite mark, Glaze, Inner Hole, Outer Ring, Powdered Sugar, Sprinkles, chocolate_drizzle, chocolate_drizzle_lines, donut_body, glaze_lines, hole, icing, icing_lines, sprinkle_dots, sprinkles
	garlic	8/10	root, stem, bulb, clove, ridges, segment line, shadow, skin, body, bulb_shape_lines, clove_inner_surface_lines, clove_separating_lines, cloves, papery_skin_lines, root_tip, stem_twig
	grapes	5/8	cluster, leaf, stem, grape, pedicle, grapes, seed_dots, vine, vine_branch, vine_twig
	hamburger	9/10	Bottom Bun, Cheese, Lettuce, Onion, Patty, Pickles, Sesame Seeds, Tomato, Top Bun, bun_bottom, bun_top, cheese, ketchup, lettuce, onion, patty, pickles, seed_dots, tomato
	hot_dog	6/10	sausage, bun, grill marks, relish, sauce, sesame seeds, bun_bottom, bun_crust_lines, bun_top, ketchup, lettuce, mustard, onion, sausage_veins, seed_dots
	ice cream cone	9/10	cone, Ice Cream, drip, flake, fruit, multiple scoops, napkin, sprinkles, waffle texture, cone_base, cone_edge, cone_tip, scoop_drip_lines, scoop_slice_lines, scoop_swirl_lines, scoops, sugar_dots, waffle_grid
	lollipop	5/7	stick, candy head, stick base, swirl pattern, wrapper, bite_mark, candy, motion_lines, seed_dots, stripe_lines, swirl_lines
	mushroom	7/8	cap, gills, stem, ring, scales, soil, spots, annulus, cap_base, cap_surface_pattern, stem_base, stem_surface_pattern
	noodle	5/8	bowl, bundle, chopsticks, strand, topping, body, crack_lines, fold_lines, motion_lines, split_lines, steam_lines, swirl_lines, texture_lines
	onion	10/10	stem, bulb, cut surface, dry skin, layers, outer layer, root end, slice cross-section, sprout, stem end, body, concentric_layers, cut_lines, inner_core, root, seed, skin_pattern, sliced_layers, texture_lines
	peanut	6/10	ridges, shell, nut, seam, skin, split shell, body, center_gap, crack_lines, edge_lines, halves, middle_line, outer_surface, seed_dots
	pear	8/8	body, leaf, stem, bottom dimple, core, seeds, slice, surface texture, bottom_curve, motion_lines, seed_dots, stem_bulb, surface_pattern
	pineapple	7/9	body, crown, eyes, base, leaves, skin texture, slice, face, leaf_pattern, mouth, nose, spine_pattern, stem
	pizza	9/9	base, cheese, crust, green peppers, mushrooms, olives, onions, pepperoni, slices, cheese_texture, crust_pattern, sauce, slice, steam_lines, toppings
	pretzel	4/8	knot, ends, loop, surface texture, body, crease_lines, crust_lines, face, glaze_lines, human_mouth, salt_dots
	pumpkin	5/8	body, stem, leaf, ridges, vine, crack_lines, eyes, mouth, nose, pumpkin_pattern, ridge_lines
	sandwich	10/12	lettuce, bread slice, cheese slice, crust, cut line, filling, meat slice, pickle, tomato slice, toothpick, bottom_bun, cheese, ketchup_lines, mayo_lines, mustard_lines, patty, seed_dots, slice_lines, steam_lines, tomato, top_bun
	strawberry	5/9	body, stem, calyx, leaf, seeds, crown, leaf_veins, leaves, seed_dots, seed_pattern, stem_base, stem_tip
	watermelon	9/10	stem, bite mark, cross-section, inner flesh, outer shell, rind, seeds, slice, stripes, body, cut_lines, inner_flesh, leaf, rind_pattern, rind_seam, seed_dots, seed_line, slice_lines
furniture	bath tub	7/9	body, faucet, legs, rim, shower head, taps, water, drain, handle, motion_lines, side_handle, water_flow_lines, water_splash, water_surface
	bed	7/9	blanket, footboard, frame, headboard, legs, mattress, pillows, bed_blanket, bed_box_spring, bed_drawers, bed_footboard, bed_frame, bed_headboard, bed_legs, bed_mattress, bed_pillow
	book	8/9	bookmark, pages, spine, cover, dust jacket, illustrations, page edges, title text, back_cover, cover_corners, front_cover, page_lines, ribbon, title_text
	calendar	7/11	header, binding, body, date grid, date numbers, hanger, notes section, binder_coil, binder_holes, corner_stamp, date_numbers, decorative_element, grid_lines, month_name, page, page_border, year_text
	candle	7/9	body, flame, wick, base, holder, melted wax pool, wax drip, bottom_flat, burn_mark, crackle_lines, label, surface_pattern, top_flat
	ceiling fan	9/9	blades, airflow lines, blade arms, ceiling attachment, central hub, downrod, motor housing, mounting bracket, pull chain, blade_pattern, bolt, central_hub, decorative_dome, handle, motion_lines, mounting_bracket, screw_nut

Continued on next page

Category	Class	4o/OSS	Elements
	chandelier	10/9	arms, bobeches, candles, ceiling mount, central stem, chain, decorative rings, drops, light bulbs, sockets, arm, base, bulb, bulb_shape, crystal_facet, hanging_chain, light_glow, main_rod, ornamental_pattern
	couch	7/10	backrest, cushion, legs, seat, armrest, throw pillow, tufting, armrest_style, armrests, cushion_pattern, cushion_seams, leg_style, upholstery_pattern
	crayon	6/9	body, base, label, multiple, tip, wrapper, body_pattern, cap, colored_tip, colored_tip_nub, crack_lines, paper_label, paper_sticker, wax_lines
	door	8/8	frame, handle, hinge, knocker, lock, panel, threshold, window, body, door_top, hinges, keyhole, lock_mechanism, wood_grain_lines
	drawer	5/8	body, handle, front panel, legs, sides, corner_lines, drawer_back, drawer_side, drawer_teeth, front_face, handle_loop
	fireplace	9/9	chimney, firebox, hearth, Fire Guard, andirons, ash pit, fire, logs, mantel, body, crack_lines, door, flames, mantle, smoke_lines
	floor lamp	11/10	base, bulb, switch, adjustable arm, cord, decorative elements, diffuser, dimmer, lamp shade, pole, weight, arm, light_beam, shade, shade_edge, shade_hinge, shade_pattern, stand
	hourglass	8/8	Base, Bottom Bulb, Frame, Narrow Neck, Sand, Sand Flow, Top Bulb, Top Cap, base, bottom_bulb, crack_lines, glass_edges, neck, sand, sand_flow, top_bulb
	lantern	9/10	handle, base, decorative elements, frame, glass panels, hanging hook, light source, top cover, ventilation holes, body, bottom, decorative_pattern, flame, frame_bars, glass_pane, light_source, panels, top
	light bulb	6/9	Base, Contact Point, Filament, Glass Envelope, Screw Thread, Support Wires, base_hole, cap_lug, cap_ring, filament, glass_envelope, glass_pattern, light_beam, metal_base, screw_threads
	map	6/12	Borders, Compass Rose, Landmarks, Rectangular Shape, Rivers, Roads, border, city_markers, coastline, coordinates_grid, country_borders, grid_lines, inland_water_bodies, legend, north_arrow, scale_bar, terrain_lines, title
	marker	7/9	body, cap, clip, tip, branding, end plug, grip section, cap_lip, ink_reservoir, label, nozzle, stripe
	paintbrush	6/7	bristles, ferrule, handle, hanging hole, paint residue, tip, bristle_pattern, brush_head, handle_cap, motion_lines
	paper clip	5/9	End Point, Inner Curve, Loop, Outer Curve, Twist, arms, bend_lines, contact_points, ends, gap, loop, metal_edges, small_holes, wire_thickness_lines
	pencil	7/8	body, eraser, tip, ferrule, graphite core, sharpened edges, wooden shell, eraser_cap, metal_ferrule, painted_surface, shading_lines, wood_grain_lines
	stairs	6/8	handrail, baluster, newel post, riser, step, stringer, balusters, landing, risers, stair_stringers, step_bottom_edges, step_edges, treads
	table	11/8	drawer, legs, apron, chair, extension leaf, shelf, stretcher, stuff, tablecloth, top surface, wheels, drawer_handle, frame, leg_cap, surface_pattern, tabletop, top_edge
	toilet	8/9	base, bowl, lid, seat, tank, flush handle, trapway, water surface, drain, flush_handle, splash, water_line
	vase	7/9	body, handle, neck, opening, rim, decorative pattern, water, base, decorative_lines, glaze_lines, pattern_lines
icon	angel	12/13	arms, body, hair, halo, head, legs, robe, wings, bow, face, rod, trumpet, eyebrows, eyes, feather_lines, mouth, sword
	diamond	5/9	Crown, Girdle, Pavilion, Side Facets, Top Facet, culet, cut_lines, edges, facet_lines, facets, reflection_lines, shape, sparkle, vertices
	dragon	15/13	body, claws, eyes, head, horns, legs, scales, tail, teeth, wings, belly, fire, mouth, nostrils, spikes, fire_breath, smoke, tail_spike
	jack o lantern	11/9	stem, base, candle, carved face, cut top, eye holes, light glow, mouth hole, nose hole, pumpkin body, ribbing, carving_lines, eyes, face, light_source, mouth, nose, outer_shell, shell_pattern
	mermaid	11/13	Arms, Earrings, Fins, Fish Tail, Hair, Human Head, Human Torso, Necklace, Scales, Shell Bra, hairpin, arms, body, crown, eyes, fins, hair, head, motion_lines, mouth, necklace, scales_pattern, shell_pattern, tail
	mona lisa	10/12	Background, Dress, Eyebrows, Eyes, Face, Forehead, Hair, Hands, Nose, Smile, arms, dress_pattern, eyebrows, eyes, hair, hands, head, mouth, necklace, nose, torso, veil
	patrick star	9/9	arms, eyes, mouth, belly button, eyebrows, head, legs, sand pants, spots, body_pattern, frown_lines, shorts, skin_texture, smile_lines, star_body
	santa claus	14/16	beard, belt, boots, gloves, hat, buttons, coat, eyeglasses, gift box, pants, rosy cheeks, sack, sleigh, white trim, arms, bag, belt_buckle, body, eyes, hat_fur_trim, hat_pom_pom, head, legs, mouth, nose
	skull	7/10	Cranium, Eye Sockets, Jaw, Mandibular Condyle, Nasal Cavity, Teeth, Zygomatic Arches, cheekbones, crack_lines, cranial_vault, ear_holes, eye_sockets, jaw, mouth_opening, nasal_cavity, skull_body, teeth
	snowman	11/9	Arms, Base Sphere, Buttons, Eyes, Hat, Head Sphere, Middle Sphere, Mouth, Nose, Pipe, Scarf, arms, base, body, eyes, hat, head, mouth, nose, scarf
	sponge bob	15/13	arms, belt, body, eyes, legs, mouth, nose, pants, teeth, tie, eyelashes, hands, holes, pupils, shoes, grid_pattern, hair_tuft, head
	stop sign	5/8	base, border, letters, octagon, post, octagon_corners, octagon_edges, octagon_outline, red_fill, stop_text, text_fill, text_stroke, white_border
	teddy bear	14/11	arms, body, ears, eyes, head, legs, mouth, nose, paws, belly patch, bow tie, seams, stitching, tail, bow_tie, fur_lines
musical instrument	bell	7/8	clapper, handle, rim, decoration lines, dome, mounted bracket, sound wave lines, base, bell_chain, bell_rim_pattern, body, motion_lines
	cello	12/10	body, bridge, endpin, neck, scroll, strings, tailpiece, bow, bridge feet, f-holes, fingerboard, pegs, f_holes, soundhole, tuning_pegs
	clarinet	10/9	bell, keys, mouthpiece, barrel, ligature, lower joint, register key, thumb rest, tone holes, upper joint, body, key_cover, key_holes, key_levers, reed, tailpiece
	drums	9/9	bass drum, cymbals, drum head, drum shell, drum sticks, hi-hat, hoops, tension rods, tom-toms, body, body_side_pattern, head_seam, heads, motion_lines, rim, snare_wire_ring, snare_wires, stand
	guitar	10/13	body, bridge, fretboard, headstock, neck, pickguard, strings, frets, sound hole, tuning_pegs, pickups, soundhole, strap, tone_knob, tuning_pegs, volume_knob
	harp	10/10	Base, Column, Contextual Stand, Decorative Finial, Frame, Neck, Pedals, Soundboard, Strings, Tuning Pegs, base, base_pattern, body, bridge, crossbar, decorative_crown, decorative_scroll, motion_lines, neck, strings
	piano	9/8	keys, legs, lid, bench, body, fallboard, music stand, pedals, wheels, black_keys, key_lines, key_separators, soundboard_pattern, white_keys
	saxophone	8/9	bell, body, keys, mouthpiece, neck, key guard, ligature, thumb rest, key_dial_lines, key_lever_lines, key_levers, mouthpiece_hole
	trombone	9/9	bell, mouthpiece, slide, bell brace, bell rim, brace, counterweight, hand grip, tuning slide, bell_flare, bell_pattern, body, metal_lines, mouthpiece_handle, tuning_slide

Continued on next page

Category	Class	4o/OSS	Elements
	trumpet	9/9	bell, body, mouthpiece, brace, finger hooks, leadpipe, slides, tuning slide, valves, crown, stem, valve_cables, valve_groove_lines, valve_housing, valve_levers
	violin	11/9	body, bridge, neck, pegs, scroll, strings, tailpiece, bow, chin rest, f-holes, fingerboard, chin_rest, f_holes
plant	bamboo	5/9	branch, clump, leaf, node, stem, culm, flower, leaf_margin_lines, leaf_tip, leaf_vein_lines, leaves, nodes, root
	beach	9/8	sand, cloud, dune, palm tree, sun, towel, umbrella, water, wave, sand_dunes, sand_pattern, sea_surface, sea_swell, shoreline, wave_lines, waves
	bush	7/9	Base, Berries, Branches, Flowers, Leaves, Main Body, Thorns, branch, foliage, leaf_dots, leaf_edges, leaf_lines, seed_cluster, sprout, stems, trunk
	cactus	8/8	Arms, Base, Flower, Fruit, Main Body, Pot, Ridges, Spines, arms, body, cactus_eyes, cactus_tip, flower, root, spines, stem_ribs
	cloud	4/8	flat base, main body, puffs, rain streaks, bottom, center, cloud_body, fluff_lines, middle, outline, soft_edges, top
	clover	6/8	stem, fourth leaf, leaf, petal outline, textural detail, vein, flowers, leaf_base, leaf_margin, leaflet_bud, leaflet_tip, leaflets, veins
	dandelion	9/8	leaves, stem, buds, florets, flower head, ground, pappus, seed heads, seeds, flower_head, leaf_base, leaf_shape, seed_dots, stem_bend, stem_curve
	flower	8/11	stamen, stem, bud, center, ground, leaf, petal, sepal, leaf_edge_lines, leaf_vein_lines, leaves, petal_edge_lines, petal_vein_lines, petals, pistil, seed_dots, thorn
	leaf	7/9	Base, Edge, Main Blade, Midrib, Petiole, Tip, Veins, base, blade, edge, main_vein, petiole, secondary_veins, serrated_edge, tip, vein_pattern
	moon	4/9	craters, circle, crescent, glow, disk, illuminated_area, maria, moon_highlight, phase_lines, rim_lines, shadow_boundary, surface_pattern
	palm tree	7/7	trunk, bark texture, coconuts, flower cluster, leaves, root base, shadow, crown, frond_tip, fronds, leaf_sheath, root, trunk_lines
	rainbow	5/8	arc, bands, clouds, glow, sky context, color_band_blue, color_band_green, color_band_indigo, color_band_orange, color_band_red, color_band_violet, color_band_yellow
	sun	5/9	rays, core circle, curved rays, face, halo, center, core, human_eyes, human_mouth, sun_burst_lines, sun_face, sun_glow, sun_spots
	tree	9/8	branches, roots, trunk, bark texture, crown, flowers, fruits, hollow, leaves, bark_pattern, canopy, fruit, leaf_cluster, leaf_vein_lines
sports equipment	barbell	6/9	bar, double weight plates, floor line, support stand, weight markings, weight plate, bar_end_caps, bar_holes, knurling_lines, plate_label, plate_ring_lines, plates, sleeve, weight_markings
	baseball	5/8	bat, core, motion lines, seams, stitches, ball, brand, logo, motion_lines, number, seam, shadow, stitching
	baseball bat	5/9	barrel, handle, grip, knob, taper, barrel_shape, grip_lines, handle_curve, handle_shape, motion_lines, tip, wood_grain_lines
	basketball	5/8	hoop, net, panel lines, sphere, valve, body, center_dot, highlight_lines, motion_lines, panel_pattern, panel_seams, shading_lines, texture_lines
	dumbbell	4/9	End Cap, Handle, Weight Plate, Weight Plate Hole, bar, bar_surface_lines, grip_lines, head_edge_lines, head_engraving, head_numbers, head_ring, head_surface_lines, heads
	golf club	8/9	face, grip, head, hosel, shaft, heel, sole, toe, clubhead_surface_lines, face_pattern, grip_texture, swing_motion_lines
	helmet	10/8	brim, shell, visor, chin strap, ear cover, face shield, neck guard, padding, retention system, vent, crack_lines, dents, ear_covers, logo, strap
	parachute	5/9	canopy, harness, person, sky, suspension lines, anchor_ring, canopy_seams, harness_cup, harness_straps, lines, steering_lines, stitch_lines
	roller skate	7/10	boot, ankle support, buckle, laces, toe stop, truck, wheel, brand_logo, deck, heel_cup, motion_lines, sole, straps, toe_stop, tread_lines, wheels
	skateboard	6/9	deck, trucks, wheels, bolts, nose, tail, deck_edges, deck_pattern, grip_tape, sticker, truck_arms, wheel_treads
	snorkel	6/9	mouthpiece, tube, clip, flexible joint, purge valve, top opening, air_flow_lines, mouthpiece_air_port, mouthpiece_cup, mouthpiece_handle, valve, valve_handle, water_line
	soccer ball	6/9	hexagonal patches, high contrast patches, motion lines, pentagonal patches, seams, spherical shape, ball_body, hexagon_center_lines, hexagon_pattern_lines, highlight_lines, motion_lines, pentagon_center_lines, pentagon_corners, pentagon_pattern_lines, shadow_lines
	table tennis	9/9	Table Legs, ball, net, net supports, racket, racket handle, racket surface, table, table lines, center_line, legs, service_area_lines, table_corners, table_edges, table_leg_bases, table_surface, table_surface_pattern, table_surface_shading
	tennis racquet	7/9	grip, handle, strings, ball, butt cap, frame, throat, handle_taper, head, head_ring, logo, motion_lines, strap
structure	arch of triumph	9/9	Attic, Bases, Central Arch, Cornices, Decorative Sculptures, Entablature, Keystone, Pillars, Side Arches, arch_body, arches, balcony, base, columns, inscription, pediment, relief_sculpture, statues
	barn	10/8	Doors, Fence, Foundation, Hay Bales, Loft, Main Structure, Roof, Silo, Weather Vane, Windows, body, chimney, door, eaves, loft_hatch, roof, shutters, windows
	bench	8/8	armrests, backrest, legs, seat, crossbar, frame, ground contact, slats, decorative_slats, railings, support_braces, wood_grain_lines
	big ben	8/11	Base, Belfry, Clock Face, Clock Hands, Pinnacles, Spire, Tower Body, Windows, bell, clock_face, clock_face_glass, clock_face_markings, clock_hands, clock_numbers, motion_lines, tower_base, tower_columns, tower_spire, tower_surface_pattern
	bridge	10/8	cables, deck, supports, abutments, arch, piers, railings, road lines, towers, water, arches, guardrail, rail_tracks, roadway_markings, traffic_lights
	campfire	4/9	flames, logs, spark, stones, fire_ember, firewood, flame, flame_outline, smoke, smoke_lines, spark_lines, wood_crack_lines, wood_grain
	castle	10/10	flag, gate, roof, archway, battlement, keep, moat, tower, turret, wall, arch, arrow_slits, bridge, courtyard, crenellations, towers, walls
	church	10/9	Arch, Bell, Cross, Door, Main Building, Platform, Roof, Spire, Tower, Window, buttress, cross, door, porch, roof, spire, structure, tower, windows
	eiffel tower	9/8	antenna, base, spire, arches, cross bracing, pillars, second level, third level, upper pillars, balcony, flag, lattice, lattice_pattern, legs

Continued on next page

Category	Class	4o/OSS	Elements
	fence	7/10	Base, Horizontal Rails, Nails, Pickets, Post Caps, Vertical Posts, Wire, bottom_rail, gate, horizontal_rails, panel, post_top_ends, posts, rail_end_lines, top_rail, vertical_rails, wood_grain_lines
	ferris wheel	8/9	hub, spokes, wheel, axle, cabins, queue area, rotation direction, support structure, base, counterweight, gondola_roof, gondola_side, gondolas, support_towers
	fire hydrant	8/9	body, base flange, bolts, chain, outlet caps, ridges, side nozzles, top cap, drip_lines, handle, label, paint_stains, pipe, rust_lines, top_valve, water_flow
	fountain	8/9	base, basin, central column, decorative elements, nozzles, upper tier, water pool, water streams, sculpture, spouts, steps, stone_pattern_lines, tiers, water_flow_lines, water_pool
	hospital	5/10	Cross Symbol, Entrance, Main Building, Signage, Windows, base, building_structure, door_handle, doors, facade, pillars, roof, sign, window_grilles, windows
	house	8/13	chimney, door, roof, walls, attic window, door knob, garage, window, chimney_flue, door_handle, eaves, porch, porch_railing, roof_vent, shingle_lines, window_glass, windows
	igloo	5/10	blocks, dome, entrance tunnel, flag, snow pile, body, crack_lines, door_handle, entrance, entrance_wall, ice_layer_lines, ice_pattern, snow_layer_lines, snow_pattern, step
	leaning tower of pisa	7/9	arches, base, bell chamber, columns, flag, lean, tiers, buttresses, foundation, middle, roof, roof_spire, stone_pattern_lines, top
	lighthouse	7/9	base, tower, entrance, lantern room, light, railing, windows, beam_lines, door, flag, lantern_room, paint_strips, stone_pattern, window
	moai stone	9/9	eyes, head, mouth, nose, brow ridge, ears, face, ground, torso, arms, body, crack_lines, neck, stone_surface_lines
	pyramids of giza	8/9	Blocks, Desert Sand, Entrance, Main Pyramid, Secondary Pyramids, Shadow, Sky, Sphinx, apex, base, crack_lines, facets, shadow_lines, small_temple, steps, stone_lines, terraces
	roller coaster	9/9	cars, loops, drops, passengers, rails, station platform, supports, tracks, turns, chain, motion_lines, seats, support_columns, track, track_rail_lines, wheels
	skyscraper	8/11	base, spire, windows, cloud, entrance, ground, skyline context, tower, antenna, billboard, door, floor_bands, roof, sign, skylight, walls
	sphinx	9/10	body, ears, head, tail, base, environment, face, headdress, paws, eyes, hair, legs, mane, mouth, nose
	statue of liberty	11/10	Base, Crown, Face, Head, Left Arm, Right Arm, Robes, Tablet, Torch, Torch Flame, Torso, arms, body, crown, crown_points, head, necklace, robe, tablet, torch, torch_flame
	stonehenge	8/8	Altar Stone, Bluestones, Earthworks, Heel Stone, Lintels, Sarsen Circle, Standing Stones, Trilithons, bluestone, capstone, crack_lines, ditch, inner_circle, outer_circle, standing_stones, stone_base
	streetlight	9/9	base, pole, crossbar, ground surface, lamp cover, light bulb, light expression, light fixture, street, decorative_ornament, lamp_bulb, lamp_housing, light_emission_lines, light_shade, ornamental_pattern, power_switch
	traffic light	6/9	Crosswalk Signal, Light Housing, Lights, Mounting Bracket, Pole, Street, base, body, green_lamp, lamp_frame, mounting_bracket, pole, red_lamp, reflective_glass, yellow_lamp
	windmill	7/10	blades, tower, door, ground, moving_line, wind, windows, base, foundation, gear, nacelle, rotor_hub, windmill_door, windmill_house, windmill_roof
tool	axe	8/8	blade, handle, eye, ferrule, head, heel, poll, toe, blade_edge, blade_tip, guard, handle_base, handle_pattern, motion_lines
	bandage	4/9	adhesive ends, pad, perforations, strip, adhesive_line, body, bottom_edge, corner_folds, fold_lines, left_edge, pattern_lines, right_edge, top_edge
	binoculars	10/10	Barrels, Bridge, Diopter Adjustment, Eyepieces, Focus Wheel, Lens Caps, Neck Strap Attachments, Objective Lenses, Pivot Joint, Rubber Eyecups, barrels, body, collar, eye_adjustment_knob, eye_lens, eyepieces, focus_ring, handle, strap_loop, zoom_ring
	boomerang	6/10	aerodynamic features, angle, arms, edge, hand grip, main body, back_edge, body, center_groove, front_edge, grain_pattern, grip_markers, motion_lines, scratch_lines, weight_mark, wings
	bottlecap	6/9	Fluted Sides, Logo or Text, Material Texture, Rim, Top Surface, liner / seal, body, bottom_seam, cap_print, handle_ring, label, logo, seal_mark, thread_lines, top_rim
	broom	8/9	bristles, handle, brush head, brush head connector, brush head edge, brush head support, dust_particles, hanging hole, bristle_pattern, brush_base, dust_bag, handle_knob, head, motion_lines, sweep_line
	cannon	7/9	Barrel, Breech, Carriage, Muzzle, Touch Hole, Trunnions, Wheels, barrel, base, breech, carriage_wheels, gun_shield, loading_port, muzzle, sight, smoke_lines
	comb	5/9	handle, teeth, decorative elements, frame, spacing, handle_curve, handle_hole, handle_screw, handle_texture, teeth_pattern, tooth_edges, tooth_spacing
	compass	6/9	Center, Decorative Elements, Direction Labels, Needle, Outer Circle, Tick Marks, bearing_markers, body, card, decorative_pattern, grid_lines, needle, north_marker, ring, scale_numbers
	drill	9/8	body, chuck, handle, battery, clutch setting, cord, drill bit, trigger, vent, drill_bit, grip_lines, key, keyhole, motion_lines
	fork	7/9	handle, base, decorative elements, neck, tine curve, tine gap, tines, fork_tip, handle_curve, handle_end, handle_pattern_lines, prong_bend, prong_end, prong_spacing_lines, prongs
	frying pan	6/10	base, contents, handle, lid, rivet, sides, burn_lines, handle_loop, handle_straight, pan_base, pan_handle, pan_logo, pan_rim, pan_surface, rim_ridge, steam_lines
	grenade	6/8	body, fuse, handle, pin, ring, segments, explosion_lines, explosion_smoke, groove_lines, motion_lines, safety_lever, safety_pin, trigger
	hammer	7/9	claw, handle, head, face, grip, neck, peen, handle_sharp_edge, handle_tip, handle_wood_grain_lines, head_flat_side, motion_lines, swing_direction_lines
	key	7/9	teeth, bit, bow, keyring hole, ridges, shank, shoulder, cut_lines, head, head_knob, head_pattern, inscriptions, key_base, ridge_pattern, shaft
	knife	8/8	blade, handle, bolster, choil, edge, rivet, spine, tip, blade_handle_joint, blade_tip, guard, handle_pattern, motion_lines, serrated_edge
	ladder	7/11	rungs, feet, platform, rear rails, side rails, spreaders, top cap, base, base_connector, handrail, rung_dots, rung_labels, rung_lines, rung_pattern, side_rails, top, top_connector
	lighter	8/10	body, flame, button, cap, flame guard, fuel tank, hinge, wheel, brand_label, flame_spark, flaming_point, flint_ring, handle, ignition_button, metal_cover, smoke_lines
	matches	7/9	flame, box, bundle, head, stick, striking surface, sulfur coating, ash, burnt_end, match_head, match_head_color, match_head Flake, match_stick, smoke, spark
	mug	5/8	body, handle, lip, interior, rim, bottom, logo, paint_pattern, rim_line, steam_lines
	pipe	5/9	bend, joint, main body, opening, smoke, ash_line, bowl_cork, mouthpiece_handle, pipe_body, pipe_bowl, pipe_mouthpiece, pipe_stem, smoke_lines, stem_handle

Continued on next page

Category	Class	4o/OSS	Elements
	rake	8/8	handle, tines, angle , attachment point , contextual ground , tine base , tine spacing , tine tips , handle_end, handle_grip, handle_tip, head, head_base, tine_spacing
	rifle	11/14	barrel, bolt, magazine, muzzle, scope, sights, stock, trigger, butt plate , butt sling , trigger guard , ejection_port , grain_lines , handle_grip, safety_switch, scope_reticle, sling
	saw	6/9	blade, handle, teeth, angle , frame , nut and bolt , blade_edge, blade_handle_joint, handle_pattern, hinge, motion_lines, teeth_pattern
	scissors	9/9	blades, handles, pivot, blade edge , blade overlap , blade tip , finger hole , hinge screw , thumb hole , blade_edge, blade_inner_edge, blade_tip, handle_knob, handle_pattern, pivot_pin
	screwdriver	7/10	handle, tip, blade , ferrule , grip texture , hanging hole , shank , connector, handle_pattern, handle_ridges, metal_part, shaft, shaft_marks, tip_notch, tip_sharpness
	shovel	8/9	blade, handle, blade edge , context ground , grip , rivet , shaft , step , blade_base, blade_edge, handle_ends, handle_grip, handle_grip_ridges, head, tip
	spoon	5/8	bowl, handle, tip, branding , neck , bowl edge , bowl surface pattern , handle_curve, handle_ornament, handle_pattern
	stethoscope	9/9	Bell , Chest Piece , Diaphragm , Ear Tips , Spring , Tubing , Tubing Connector , Tubing Split , Yoke , bell, chest_piece, connector, diaphragm, earpieces, strap, tubing, tubing_end, tubing_loop
	sword	9/9	blade, pommel, crossguard , edge , fuller , hilt , quillons , ricasso , tip , blade_edge, blade_pattern, blade_tip, guard, handle, hilt_grip, hilt_pattern
	syringe	7/8	barrel, needle, plunger, barrel tip , flange , plunger handle , scale markings , barrel_cap, gauge_lines, label, motion_lines, needle_guard
	teapot	7/10	body, handle, lid, spout, decorative pattern , lid knob , spout tip , base, handle_lip, pattern, rim, spout_tip, steam_lines
	tent	8/12	pegs, poles, context ground , entrance flap , guy lines , main body , rainfly , windows , canvas, canvas_pattern, entrance, frame, lantern, pole_lines, rope_lines, rope_ties, sleeping_area, stove
	toothbrush	8/8	bristles, handle, brush head , flexible joint , grip , hanging hole , neck , toothpaste , brand_label, brush_head, brush_head_pattern, handle_curve, handle_end_cap, handle_pattern
	toothpaste	7/9	cap, tube, brand label , folded end , nozzle , paste , thread , cap_stripe, label, label_icon, label_text, opening, toothpaste_blob, tube_ridges
	umbrella	10/9	canopy, handle, collar , ferrule , ribs , shaft , spring mechanism , tie , tips , vent , arms, base, decorative_lines, folding_lines, handle_curve, motion_lines, pattern
	wine glass	7/10	base, bowl, stem, bowl curve , reflection , rim , wine surface , bowl_cap, bowl_lip, crackle_lines, reflection_lines, sparkle, stem_bottom, stem_cap
vehicle	airplane	10/12	cockpit, fuselage, nose, tail, windows, wings, engines , horizontal stabilizers , landing gear , vertical stabilizer , engine , flaps, jet_exhaust_lines, landing_gear, motion_lines, propeller
	ambulance	8/12	body, wheels, windows, cross symbol , front cab , light bar , rear doors , siren , doors, emergency_light_bar, emergency_signage, front_bumper, headlights, rear_light, roof, side_mirrors, stripe_pattern
	bicycle	10/12	chain, chainring, fork, frame, handlebars, pedals, seat, spokes, saddle , wheel , brakes , gear_teeth , tire_treads , wheels
	blimp	8/8	gondola, propeller, cables , envelope , fins , logo , tail, windows, hull, motion_lines, nose, paint_stripes, tether, window
	bulldozer	8/9	blade, tracks, cabin , engine hood , exhaust pipe , hydraulic cylinders , ripper , undercarriage , cab, cab_door, cab_window, front_bumper, rear_skid, steering_wheel, windshield
	bus	12/11	body, roof, wheels, windows, bumper, door, exhaust pipe, front grille, headlights, logo, mirrors, rear lights, bus_lights, bus_sign, doors, front_bumper, motion_lines, rear_bumper, windshield
	canoe	8/10	bow, gunwales, hull, stern, paddles , seats , thwarts , water, bottom, deck, deck_lines, hull_lines, sidewall_lines, sidewalls
	car	13/11	body, roof, bumper, door, exhaust pipe, grill, headlight, hood, license plate, side mirror, taillight, wheel, window, doors, headlights, license_plate, motion_lines, skid_lines, taillights, trunk, wheels, windows
	cruise ship	9/12	Bow , Bridge , Deck , Hull , Railings , Smokestack , Stern , Waterline , Windows , anchor, bow, decks, flag, funnel, helicopter_pad, hull, lifeboats, railings, smoke_lines, stern, windows
	flying saucer	6/12	antenna, beam , disk , dome , landing gear , lights , body, center_dome, control_panel, door, glow_lines, light_lines, motion_lines, reflection_lines, rim, shadow_areas, thrusters
	helicopter	8/8	Cockpit , Door , Fuselage , Landing Skids , Main Rotor , Tail Boom , Tail Rotor , Windows , cockpit, engine_cowl, fuselage, landing_skids, main_rotor_blades, motion_lines, tail_boom, tail_rotor
	hot air balloon	4/9	basket, envelope, gores , rigging , basket_handle, basket_sides, burner, envelope_seam_lines, flame, motion_lines, rope
	motorcycle	10/9	frame, handlebars, headlight, seat, wheels, exhaust pipe , fairing , foot pegs , mirrors, suspension, chain, exhaust, tail_light, wheel_spokes
	pickup truck	11/15	Bed , Bumpers , Cab , Door Handles , Exhaust Pipe , Grille , Headlights , Side Mirrors , Wheel Arches , Wheels , Windows , cabin, cargo_area, driver_seat, fender, front_bumper, front_lights, front_windshield, license_plate, logo, rear_bumper, rear_lights, rear_mirror, side_windows, steering_wheel, wheels
	rocket	10/10	body, fins, boosters , engines , exhaust flame , nose cone , payload fairing , smoke trail , stages , window , engine , logo , motion_lines, nose_cone, panel_lines, rocket_flag, separation_device, strap_lines
	sailboat	11/12	cabin, deck, hull, mast, boom, bow, flag, sail, stays, stern, waterline , deck_lines, hull_lines, keel, rigging_lines, rudder, sail_flag, sail_lines, sails
	space shuttle	8/9	Cockpit Windows , Engine Nozzles , External Fuel Tank , Main Body , Nose Cone , Rocket Boosters , Vertical Stabilizer , Wings , body, engines, landing_gear, mission_patch, motion_lines, nasa_patch, nose_cone, payload_bay, tail_fins
	submarine	10/11	hull, periscope, propeller, antennas, conning tower, dive planes, hatch, rudder, sonar dome, water, conning_tower, dorsal_fin, horizontal_fins, hull_lines, motion_lines, rivets, torpedo_tubes, vertical_fins
	tractor	13/10	body, cabin, engine hood , exhaust pipe , front wheels, fuel tank, grill, headlights, mudguards, rear wheels, seat, steering wheel, steps, cab, cab_seat, cab_window, engine_exterior, fender_lines, gear_shift_lever, hood, steering_wheel, wheels
	train	13/9	cab, cars, wheels, windows, bogies , buffer , couplers , doors, engine, headlight, roof, smokestack, tracks, caboose, door, grill_lines, locomotive_body, whistle
	truck	11/10	cabin, door, headlights, wheels, windshield, bumper, cargo area, exhaust pipe, grille, side mirrors, taillights, body, cargo_box, grill, side_mirrors, tail_lights
	van	12/11	body, headlights, wheels, windows, bumper, door, exhaust pipe, grille, roof rack, side mirrors, taillights, windshield, bumpers, doors, exhaust_pipe, license_plate, rear_window, roof, side_mirrors
	wheel	7/8	rim, spokes, axle hole , inner circle , lug nuts , outer circle , tread , hub, hubcap, rim_pattern_lines, tire, tire_side_lines, tire_tread_lines

References

- [1] Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, et al. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*, 2025. 1
- [2] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216, 1993. 3
- [3] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibao Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 15
- [4] O Belitskaya. Art pictograms. <https://www.kaggle.com/datasets/olgabbelitskaya/art-pictogram/data>, 2020. Kaggle. 15
- [5] blinixsolutions. Flaticon. https://www.flaticon.com/free-icon/horse-breed_15374551?term=horse&page=1&position=40&origin=search&related_id=15374551. Horse. 15
- [6] Matt Deitke, Christopher Clark, Sangho Lee, Rohun Tripathi, Yue Yang, Jae Sung Park, Mohammadreza Salehi, Niklas Muennighoff, Kyle Lo, Luca Soldaini, et al. Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models. *arXiv e-prints*, pages arXiv–2409, 2024. 15
- [7] DinosoftLabs. Flaticon. https://www.flaticon.com/free-icon/lift_3773839?term=crane&page=1&position=3&origin=search&related_id=3773839. Crane. 15
- [8] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10, 2012. 3
- [9] Freepik. Flaticon. https://www.flaticon.com/free-icon/flowers_2545551?term=flower&page=1&position=58&origin=search&related_id=2545551. Flower. 15
- [10] Freepik. Flaticon. https://www.flaticon.com/free-icon/daisy_6900993?term=flower&page=3&position=30&origin=search&related_id=6900993. Flower.
- [11] Nikita Golubev. Flaticon. https://www.flaticon.com/free-icon/shipping_6414036?term=crane&page=2&position=42&origin=search&related_id=6414036. Crane. 15
- [12] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. 1
- [13] David Ha and Douglas Eck. A neural representation of sketch drawings. In *International Conference on Learning Representations*, 2018. 3
- [14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 9
- [15] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 1
- [16] Flat Icons. Flaticon. https://www.flaticon.com/free-icon/gardenia_5433719?term=flower&page=1&position=79&origin=search&related_id=5433719. Flower. 15
- [17] Gabriel Ilharco, Mitchell Wortsman, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, et al. Openclip. *Zenodo*, 2021. 3, 9
- [18] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *ICLR*, 2017. 9
- [19] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Léo Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023. 1
- [20] Han Jiawei and Kamber Micheline. *Data mining: concepts and techniques*. Morgan kaufmann, 2006. 3
- [21] Park Jisun. Flaticon. https://www.flaticon.com/free-icon/horse_15622258?term=horse&page=1&position=25&origin=search&related_id=15622258. Horse. 15
- [22] jocularart. Flaticon. https://www.flaticon.com/free-icon/plumeria_7091477?term=flower&page=1&position=86&origin=search&related_id=7091477. Flower. 15
- [23] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. BliP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR, 2022. 15
- [24] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023. 15
- [25] Chris Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. In *ICLR*, 2017. 9
- [26] Andrés Marafioti, Orr Zohar, Miquel Farré, Merve Noyan, Elie Bakouch, Pedro Cuenca, Cyril Zakka, Loubna Ben Allal, Anton Lozhkov, Nouamane Tazi, et al. Smolvlm: Redefining small and efficient multimodal models. *arXiv preprint arXiv:2504.05299*, 2025. 15
- [27] Kushin Mukherjee, Holly Huey, Xuanchen Lu, Yael Vinker, Rio Aguina-Kang, Ariel Shamir, and Judith Fan. Seva: Leveraging sketches to evaluate alignment between human and machine visual abstraction. *Advances in Neural Information Processing Systems*, 36, 2024. 9
- [28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry,

- Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 3, 9
- [29] Andreas Steiner, André Susano Pinto, Michael Tschannen, Daniel Keysers, Xiao Wang, Yonatan Bitton, Alexey Gritsenko, Matthias Minderer, Anthony Sherbondy, Shangbang Long, et al. Paligemma 2: A family of versatile vlms for transfer. *arXiv preprint arXiv:2412.03555*, 2024. 15
- [30] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 1992. 9
- [31] An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024. 1
- [32] Jiabo Ye, Haiyang Xu, Haowei Liu, Anwen Hu, Ming Yan, Qi Qian, Ji Zhang, Fei Huang, and Jingren Zhou. mplug-owl3: Towards long image-sequence understanding in multi-modal large language models. *arXiv preprint arXiv:2408.04840*, 2024. 15
- [33] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*, 2022. 3, 9
- [34] zero_wing. Flaticon. https://www.flaticon.com/free-icon/cargo-crane_16133203?term=crane&page=2&position=41&origin=search&related_id=16133203. Crane. 15
- [35] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*, 2025. 15