StyleBabel: Artistic Style Tagging and Captioning

Supplementary material

Fig. 1. StyleBabel generated caption focuses on style (eg '.. the image has dark tones and high contrast..'), rather than semantics, eg '..a wooded landscape with horse drawn carriage..' from SemArt ground truth

1 StyleBabel Grounded Annotation: Further Information

Screen captures of the annotation tasks (Stages 1a, 1b, 2 and 3) are presented with representative data fields completed in Figures. 2-5 respectively. Please refer to the main paper for a full account of the Grounded Annotation method used to create StyleBabel .
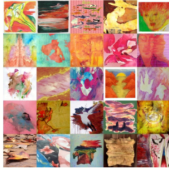
Briefly, moodboards are annotated by in a pair-wise (contrastive) manner the initial Stage 1a (Fig. 2), where an initial tag set (vocabularity) is gathered from individual participants. Each moodboard is annotated by several participants, and the annotation passed into Stage 1b (Fig. 3), where the tags are refined as a participant group activity. Once finished, the tags are verified by the ESP-style game in Stage 2 (Fig.4), where any invalid tags are also filtered out. Descriptive sentences are gathered at this stage. Finally, to refine tag annotation to individual images, these are passed to crowd Stage 3 (Fig.5), where tags invalid for individual images are filtered out through majority consensus.

**Please tag (words/*short* phrases) the common artistic style in each of the moodboards (as a whole, not individual images) with as many tags/details as possible. Focus on artistic STYLE, NOT the semantic content (EG stroke/colouring type, lighting, shading, patterns, shapes, composition, medium, layout, theme, mood etc.; NOT cat, dog, people etc.).
An artist should be able to re-create the style from your descriptions**
Avoid comparative language - each moodboard's descriptions should make sense stand-alone, without both moodboards present.

Fig. 2. Stage 1a: Pair-wise annotation of moodboards as an individual task. Participants annotate a pair of moodboards with style attributes that are shared by, or distinguish the moodboards.

Fig. 3. Stage 1b: Group annotation task to refine annotation and co-create a shared vocabulary among the participant group. Several moodboards are shown to a group of workers with pre-populated tags from Stage 1a. Workers spend roughly one minute per moodboard, making changes to the *stickers* containing tags, after an initial discussion phase.
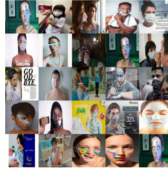
Please identify which moodboard is being described by the tags, and check its checkbox. Then, include **ALL** the given words/phrases to form descriptive captions (sentences) about the moodboard STYLE.
You can click images to view them larger.

☐   ☐   ☐   ☐   ☐

If you absolutely don't agree with a word/phrase, click it to indicate. These are tags that in groups you decided were correct, so there should rarely be any that are incorrect. You can ignore duplicates.

| close up shot | detailed | movement | warm tones | intricate | red tones | high saturation | close up shot | experimentation angles | bold compositions | strong composition | photography |

| angular shots | natural lighting | metallic elements | professional photography | energetic | commercial photography | vivid |

Please use good, formal English with good grammar/punctuation.

Write moodboard caption sentences, describing the moodboard style, making use of the tags above. Write full stop separated sentences using the provided vocabulary. Aim to write at least a couple of sentences. Enter a single dash '-' if the moodboard can't be identified

Please select here if none of the moodboards match the descriptions☐

Fig. 4. Stage 2: individual evaluation task to verify the correctness of the tags, as well as generating natural language sentences from the tags.

Have a look at the art image below, and some keywords that describe its style.
Click tags which DO NOT match the artistic style of the image
Refer only to the ARTISTIC STYLE rather than the content of the image. For example, things like: stroke/colouring type, lighting, shading, patterns, shapes, composition, medium, layout, theme, mood, etc; NOT literal things like cat, dog, person, etc
Please make sure you take the time to read ALL the tags
Responses are manually checked for correctness before approval, with sanity check tags / HITs randomly included.


| line | idea | scamp | sketch | original |

Please select here if absolutely all the tags are 100% correct ☐

Fig. 5. Stage 3: crowd annotation task where non-experts are shown individual images from the moodboards along with the moodboard tags. They are asked to click which tags do not match the style of the image, to filter out any tag which apply to the moodboard generally, but not to a particular image inside it.

2   Data examples

Tbls. 1 and 2 show some additional samples from the StyleBabel dataset. We also include a short sample of the dataset in the companion .csv file, containing image, tags, and caption data.
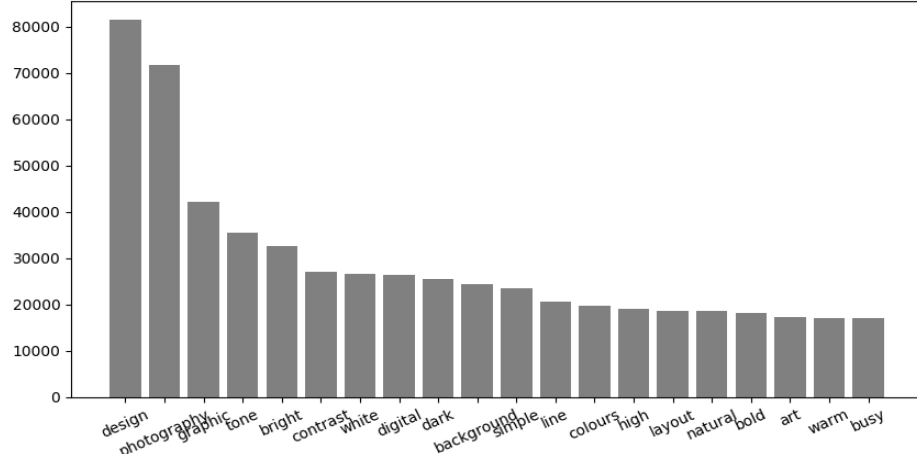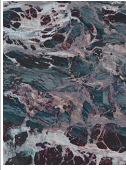

Fig. 6. Visualizing occurence of the top tags in StyleBabel.

3   Model generated results

In Fig.7 we show captions generated using our VirTex model trained with MS-COCO (top) and StyleBabel (bottom). We present some further examples (in Figs. 8 and 9) of tags generated using our ALADIN-ViT style embeddings and our CLIP model. Fig 11 shows word type distribution.

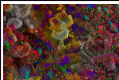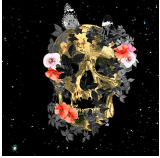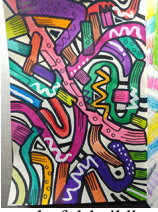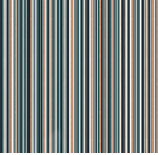|					
Image					
Tags	abstract, portrait, digital, vibrant, texture, painting, up, warm, close, illustration, colorful, bold	abstract, chaotic, complex, texture, colors, pastel, sparse, scale, large, colorful, expressionism	dated, random, promotional, advertising, poster, typography, ux, graphic	line, fantasy, cartoon, bold, black, sketchbook, drawing, story, white, fan, monotone, mystical, art, character, illustration	old, traditional, worn, western, digital, paper, advertisement, stained, cowboy, warm, child, poster, typography, wild, character, graphic
Caption	Portrait illustration featuring vibrant colors against a plain background. The digitally rendered painting gives a bold feeling.	Abstract oil painting with textures and splashes is composed of paintbrush strokes. The expressionism artwork is complex and chaotic.	Digital poster illustration featuring bold typography and vector images. The graphic designs in vibrant colors give a dated feel to the promotional material.	Sketch art of line drawing featuring decorative patterns and curved lines against a plain background. The illustration use thin linear and repetitive patterns.	Digital advertisement poster layout using retro style with linear line art graphics against a stained background with border. The content consists of digital graphic designs with cluttered informative typography using stylized fonts.
Image					
Tags	complementary, pattern, bright, crowded, rounded, fluid, motif, doodle, monochrome, sporadic, repetitive, graphic	abandoned, dark, derelict, dull, ghostly, spooky	logo, identity, clean, outline, thick, animal, black, visual, branding, white, iconography, type, monochrome, vector, icon, typography	360, industrial, photography, aerial, natural, conceptual, architectural, drone, render, dull, landscape	abstract, chaotic, complex, texture, colors, pastel, sparse, scale, large, colorful, expressionism
Caption	Layout design featuring a graphical design with a crowded and repetitive style. The design contains small motifs and doodles in different shapes with a vibrant and bright style against the white background.	Wide-angle photography using natural warm lighting and shadow effect of an abandoned architectural structure with cylindrical shapes and grids. The dramatic photograph captures the golden hue and foamy wave pattern in nature.	Digital typographic logo design using graphic elements and stylized font against a dark background. The color scheme is contrasting.	3d rendered mockup of architectural design from an aerial shot. The conceptual layout shows industrial structures with symmetrical shapes and designs against a landscape background.	Abstract oil painting featuring chaotic and complex linear patterns. This expressionism also features sparse textures using pastel and muted colors.

Table 1. Excerpt of StyleBabel dataset. Images and the corresponding tags and sentences collected via our Grounded Annotation.

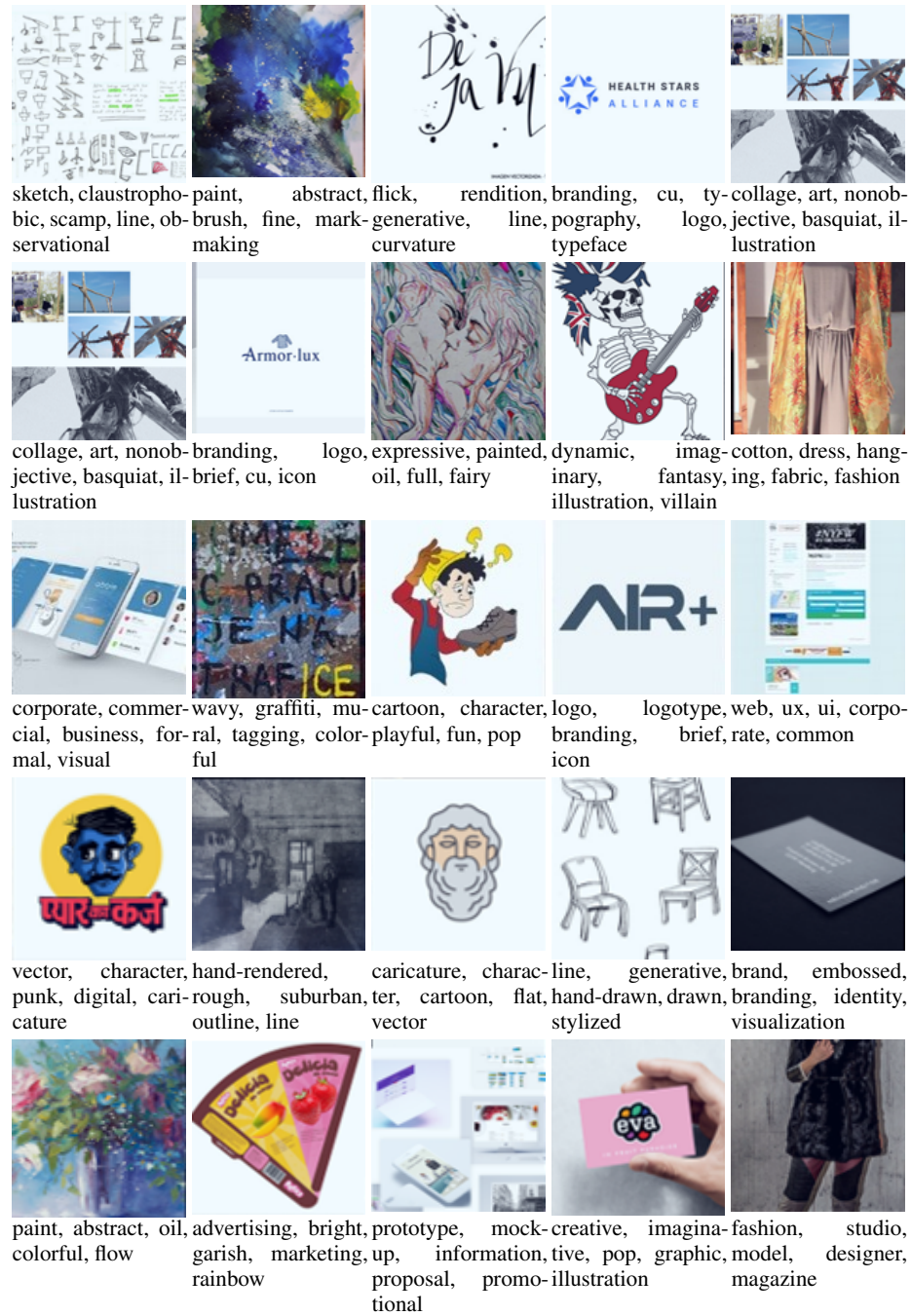| Image					
Tags	business, logotype, identity, personal, serif, mock-up, simple, sleek, visual, branding, minimal, sans, promotional, professional, monochrome, formal, font, card	creative, graphic, paper, animated, flat, development, sketch, drawing, artistic, amateur, professional, illustration, animation, character, storyboard	creative, graphic, paper, flat, animated, development, sketch, drawing, artistic, amateur, professional, illustration, animation, character, storyboard	mock-up, black, simple, branding, white, classy, stylish, book	contrast, kaleidoscope, psychedelic, complex, intricate, complicated, paisley, colored, colorful, repetitive, surreal
Caption	Mock-up of formal business cards with simple and sleek typography using serif font. This visual identity layout also features a logo design with green outlining displayed on a plain background.	Sketch illustration featuring creative graphic design. Flat animated drawing is detailed with artistic lines for character development.	Raw digital sketch representing an artistic storyboard on development stage with structured character design and animated setup formation. The creative drawing has an amateur approach with flat visual presentation.	3D rendered book mock-up for publication. The simple monochromatic cover page looks classy, featuring graphic design, cropped photograph of a female and minimal typography against a black background.	kaleidoscopic illustration featuring repetitive psychedelic pattern with abstract shapes against a plain background. The colorful artwork is complex and intricate.
Image					
Tags	industrial, clean, structured, bright, architectural, masculine, modern, futuristic, linear, monotone, mechanical	fine, texture, pattern, colors, detailed, geometric, secondary, repetitive	interior, concept, retail, model, architectural, mock-up, prototype, commercial, 3d, render, floorplan, magazine, generation, section, editorial	abstract, paint, watercolor, psychedelic, fine, texture, pattern, oil, colors, floral, crowded, warm, vivid, art, overwhelming, colorful, repetitive	dark, contrast, digital, vibrant, bright, colors, dynamic, manipulation, neon, fan, illustration, art, object
Caption	Photography of a luxurious interior design with structured grids and artificial light. The formation of the multifunctional industrial space is symmetrically aligned and looks clean.	Illustration featuring repetitive patterns and fine textures against a plain background. The detailed artwork is composed of geometric shapes and secondary tones which creates an illusionary visual effect.	Digital layout, presenting a centered conceptual 3D architectural interior floor-planning design developed in linear, and textured patterns in warm neutral tones. The prototype is displayed against a plain background.	Abstract watercolor painting featuring repetitive psychedelic patterns and floral design. The colorful crowded artwork is overwhelming with the composition of fine textures and shades.	Digitally manipulated illustration featuring a futuristic theme, a photograph with perspective distortion, geometric shapes and ripple and dot patterns. The dynamic image is eye catching with vibrant neon colors against a dark background.

Table 2. Excerpt of StyleBabel dataset. Images and the corresponding tags and sentences collected via our Grounded Annotation.

MS-COCO	a close up of a bird in a tree	a bunch of flowers are in a vase	a bottle of beer sitting on top of a red	a colorful building with a painted painted on it	a picture of a picture of a picture of a wall
StyleBabel CL+IL	the pencil sketch is a soft linear drawing with the use of shades and strokes against a plain background	digital illustration of a fictional character using a dark color palette	digital illustration featuring repetitive shapes and patterns using a color palette of blue and red	abstract painting with geometric shapes and patterns using bright colors	digital illustration of a fictional character using a color palette of black and white

MS-COCO	a group of birds are in the sky	a display of a display with many clocks	a group of birds that are on a pole	a close up of a picture of the water	a close up of a person with a black background
StyleBabel CL+IL	digital layout of a logo with minimal typography and a white background	digital illustration of cartoon characters using bright colors against a blue background	digital illustration of a sketch with line drawings on a white background	abstract oil painting using paintbrush strokes and scribbles using a cool color palette	digital illustration of repetitive linear pattern using blue and white colors

MS-COCO	a picture of a wall of a wall with a wall	a close up of a woman with a hair	a colorful decorated decorated decorated with colorful decorations	a man in a red jacket is standing in front of a store	a white and white building with a white and white building
StyleBabel CL+IL	digital illustration of a 3d rendered illustration with repetitive shapes and patterns	the sketch is a soft linear drawing with the use of shades and strokes against a white background	digital illustration of an animated character using bright colors	event photography featuring subject in center with colorful lighting against a blue background	3d rendered modern interior design with repetitive straight lines and shapes

Fig. 7. Examples of natural language sentences generated for various art styles; The top row shows captions generated using a VirTex model trained on MS-COCO, and the bottom row shows captions generated using VirTex trained on StyleBabel.

sketch, claustropho-bic, scamp, line, ob-servational

paint,   abstract, brush, fine, mark-making

flick,   rendition, generative,   line, curvature

branding,   cu,   ty-pography,   logo, typeface

collage, art, nonob-jective, basquiat, il-lustration

collage, art, nonob-jective, basquiat, il-lustration

branding,   logo, brief, cu, icon

expressive, painted, oil, full, fairy

dynamic,   imag-inary,   fantasy, illustration, villain

cotton, dress, hang-ing, fabric, fashion

corporate, commer-cial, business, for-mal, visual

wavy, graffiti, mu-ral, tagging, color-ful

cartoon, character, playful, fun, pop

logo,   logotype, branding,   brief, icon

web, ux, ui, corpo-rate, common

vector,   character, punk, digital, cari-cature

hand-rendered, rough,   suburban, outline, line

caricature, charac-ter, cartoon, flat, vector

line,   generative, hand-drawn, drawn, stylized

brand,   embossed, branding, identity, visualization

paint, abstract, oil, colorful, flow

advertising, bright, garish, marketing, rainbow

prototype,   mock-up,   information, proposal,   promo-tional

creative, imagina-tive, pop, graphic, illustration

fashion,   studio, model,   designer, magazine

Fig. 8. Style2Text tag generation experiment, using CLIP trained with ALADIN-ViT encoder. Top 5 tags shown for each image.

generative, lined, artisan, complex, pattern

brand, sleek, branding, profession, printing

paint, abstract, particle, whimsical, oil

cartoon, illustration, comic, trip, storyboard

nerdy, black, generative, monochrome, dark

lightened, cozy, interior, homely, luxurious

distorted, creative, bold, illustration, nonobjective

mark, icon, vector, monogram, logo

freehand, line, stroke, thin, freeform

cartoon, character, avatar, child, animation

informative, copy, complex, presentation, web

sketchbook, original, sketching, line, quick

graffiti, bold, abstract, tagging, mural

headshot, portrait, selfie, hair, expression

centered, dainty, hand-lettering, calligraphy, informal

typeface, basic, logo, ligature, font

calligraphy, flick, curvature, generative, hand-lettering

interior, fengshui, cozy, cosmopolitan, lightened

fashion, magazine, persona, material, emo

promotion, beverage, upbeat, advertisement, conclusive

novel, abstract, lined, stencil, hand-rendered

informative, publication, page, academic, readable

abstract, paint, oil, fine, mark-making

rasping, typography, red, eye-catchy, ambiguous
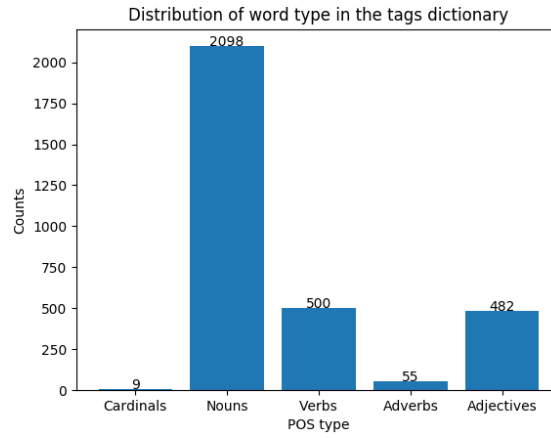
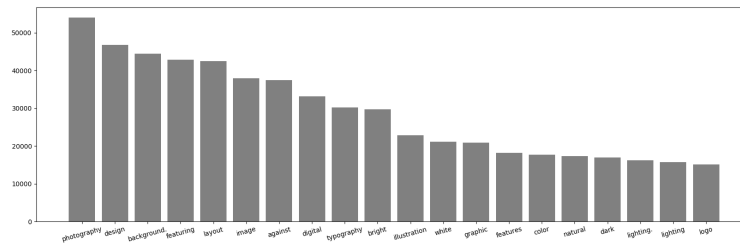pattern, agglomeration, texture, kaleidoscope, 80

Fig. 9. Style2Text tag generation experiment, using CLIP trained with ALADIN-ViT encoder. Top 5 tags shown for each image.

Fig. 10. Distribution of word types for the set of unique style attributes (tags) collected for images within the StyleBabel dataset

Fig. 11. Occurences of the top words in the StyleBabel captions