SUPPLEMENTARY MATERIALS: SELF-SUPERVISED DISENTANGLED REPRESENTATION LEARNING OF ARTISTIC STYLE THROUGH NEURAL STYLE TRANSFER

Dan Ruta[⋆†] *Gemma Canet Tarrés*[⋆] *Alexander Black*[⋆] *Andrew Gilbert*[⋆] *John Collomosse*[⋆†]

[⋆] Centre for Vision Speech and Signal Processing (CVSSP), University of Surrey. UK.
[†] Adobe Research, San Jose. USA.

1. TRAINING DETAILS

We train our model with the NeAT, PAMA, and SANet NST methods for roughly 3 days on a single A100 GPU until convergence. We stylized images for training using 512px resolution, which we downsample to 256x256 for the VGG branch and 224x224 as needed for ViT-B_16, using the same ViT as ALADIN. We disable the prior blurring in NeAT for speed. We use the Adam optimizer, and a target batch size of 1024 via logit accumulation. We decay the learning rate by 0.999875 every 100 iterations.

2. MULTIMODAL RESULTS

Figure 1 contains a visual comparison of multimodal style tagging results using our method, and ALADIN-ViT.

3. MULTIMODAL RESULTS

Figure 2 contains some visual results of image retrieval, compared to methods from literature.

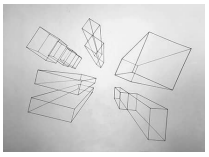
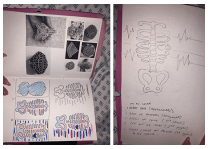
					
ALADIN-ViT	neon-like, warm toned lighting, light trail, strong outline, all cap	narrative driven, manga, Japanese, cartoon drawing, graphic art	regulated layout, posh, promoting, triangle composition, branding package	mixed media, layered composition, watercolor painting, handmade artwork painting, illustrative	color splash, marker drawing, expressionist, word sound effects, subtle colors
Ours (fused)	dark image, chiaroscuro, black and red, red, dark picture	documentary, vigorous, past, documentary shot, dark contrast	trustful, housing, posh, housing architecture render, solarpunk	colorful drawing, abstract art, abstract artwork, nonobjective, cubism	storyboarding, interpretive, panel, story, storyboarding
Ours	flame, spark, dark vibe, explosion, dark space	contrasted, high contrast, suburban, documentary shot, documentary	glass, trustful, architectural landscape, pentagonal, pointy	soft, soft color, soft and bright color, feminine, soft colors	comic book art, black and white art, comic art, comic, doodle art
					
ALADIN-ViT	cold hue, product-focused, product description, blue-based, digital publication	had material, watery, bird eye view, nobody	measuring, technical sketch, design sketch, sketch, typography element	user input, 3d building plan, internet, gathering information, datum collecting	expressive, irregular angle, clean stroke, blue glow, copy
Ours (fused)	mystical, fantasy concept art, fantasy art, aqua, fantasy painting	reflective, commercial shot, rough texture, geometric shape, geometric line	sketch scamp, scamp, sketch work, sketch, sketched line	uxui design, design interface, user interface, ui instructional design, interface	poverty, struggle, slum, evocative, documented
Ours	layered composition, aqua, mystical, fantasy painting, digital print	embossed, small shape, cup, stone image, light grey	idea, blended, pen and pencil, sketch, quick	page layout, image of article, blocky layout, conceptual layout, stationery design	high-contrast, retouched, noir, poverty, slum
					
ALADIN-ViT	cool hue, mockup, magazine booklet layout, blue themed, graphic layout	hand drawn, changing proportion, sketchy, saddening, line drawing	colorful, bright, bright bold, flow, color-heavy	curved line, colorful drawing, mark making, blue ink drawing, thin stroke	fading, sans serif and serif, red highlighting, thin letter, typeface
Ours (fused)	editorial design, editorial, editorial mockup, editorial work, editorial mockup	line, ink work, ink, ink drawing, outline	pastel, pastoral, black tea painting, art appreciation, colorful drawing	abstract line, mark-making, line, fine, fluid line	spontaneous, beginner, sketchbook, paper pattern, light pencil work
Ours	editorial, editorial design, readable, editorial mockup, professional style	animation sketch, fine-line drawing, figurative drawing, line drawing illustration, sketch of cartoon character	pastel, art therapy, fine art, traditional illustration, child illustration	fluid, blue,cut paper, plump, material sample	printmaking, delicate type, various printed paper work, lino, handmade

Fig. 1. Please zoom for more image detail. Zero-shot automatic style tagging comparison, between ALADIN-ViT, our model, and our fused variant, joining our disentangled embeddings with ALADIN-ViT. We show the top 5 tags for each image.

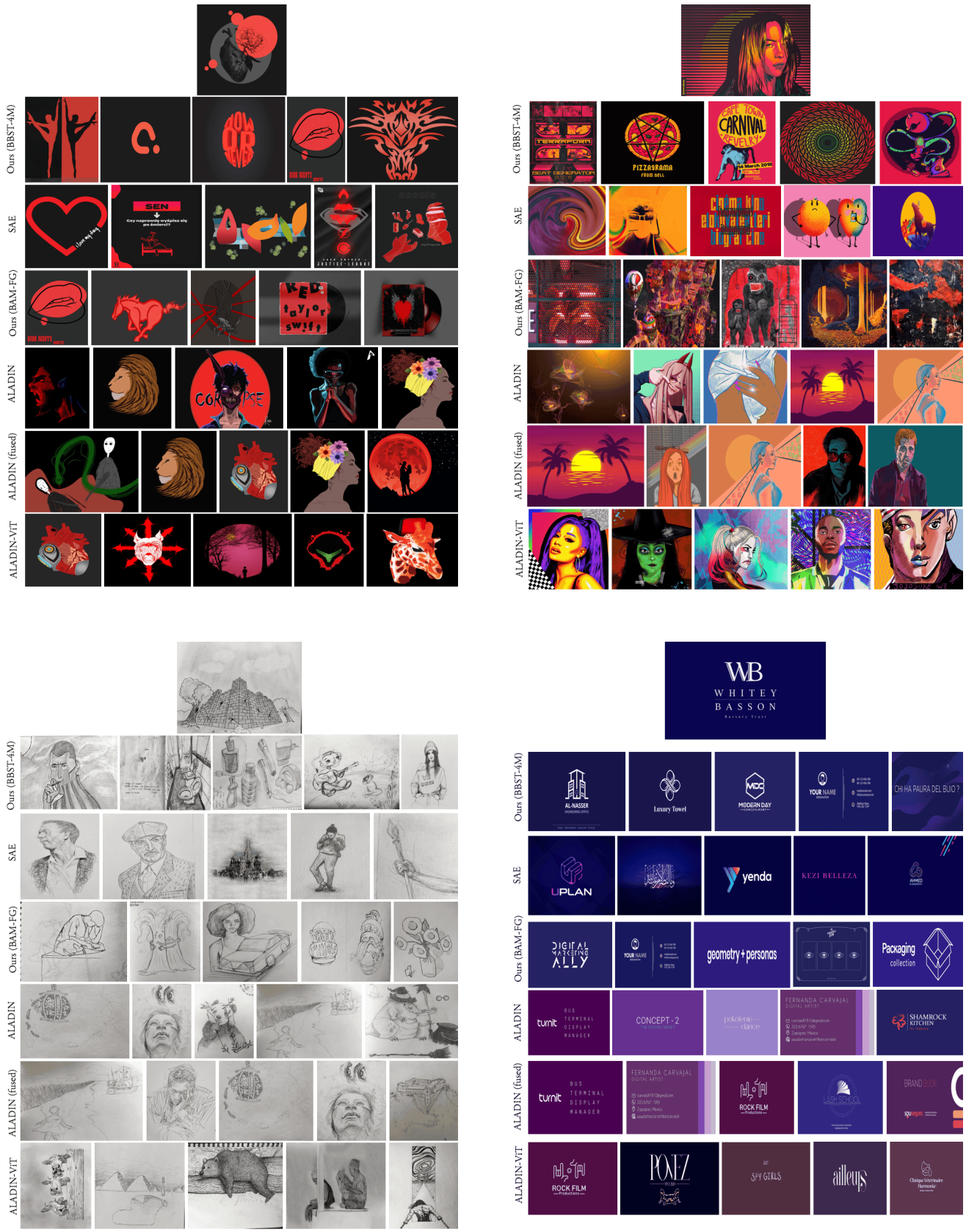


Fig. 2. Style-based image retrieval comparison between our method variants and previous literature.