Supplementary Material of "Target Bias Is All You Need: Zero-Shot Debiasing of Vision-Language Models with Bias Corpus"

1. Experimental Details

1.1. Evaluation Datasets

To evaluate the fairness of various debiasing methods, we focus on two key downstream tasks: classification and retrieval. For classification, we evaluate the methods on the following datasets:

- CelebA dataset [15]: The dataset includes 19,868 validation samples with 40 facial attributes including gender, smiling, blond, facial shape, etc. The goal is to predict whether the portrait has blond hair or dark hair, where gender is treated as the sensitive attribute.
- Waterbirds (CUB) dataset [19]: The dataset includes 1,199 samples. The goal is to predict whether the bird in the image is "landbird" or "waterbird". Here, we consider the background ={water background, land background} of the image serving as the spurious attribute.

For retrieval task, we employ:

- FairFace dataset [11]: The dataset consists of 10,954 images of facial portrait annotated with age, gender and ethnicity information. The goal is to make sure the retrieved images given a neutral prompt, such as "A photo of a smart person", follow the demographic ratio of the overall dataset.

1.2. Experimental Setup

All of our experiments are conducted on a system with a single RTX3090 GPU and Ryzen Threadripper 3960X CPU.

1.2.1. ZSDebias

ZSDebias comprises three adaptors, where all adaptors are structured as 2-layer MLP. For the two encoders (bias and neutral adaptors), we incorporate LayerNorm [2] at the last layer followed by Leaky ReLU activation function. For the weight for each loss function, we set $\lambda_{recon} = 10^2$, $\lambda_{CKA}, \lambda_{irrel} = 1$, and $\lambda_{cont} = 10$. ZSDebias is trained for 500 epochs at most, but implemented early stopping base on the saturation of the validation loss. In image modification experiment demonstrated in Sec 5.4, we adopt text-guided latent optimization scheme of StyleCLIP [17], which is pretrained on the FFHQ dataset [12].

1.2.2. ConAdapt

In ConAdapt [23], the Adpter consists of two-layer MLPs where the hidden dimension is 128 while maintaining the input and output dimension as the same as CLIP embeddings. The first hidden layer is followed by a BatchNorm and ReLU activation. During the training, the batch size is set to 128. The SGD optimizer is used with learning rate 10^{-3}, weight decay 5×10^{-5}, momentum 0.9. The number of positive and negative samples is 2048, while the number of nearest neighbors for negative samples is 4096 for the Waterbirds dataset. The maximum epoch for the CelebA dataset is 50, while that of the Waterbirds dataset is 100, following the authors' choice. We adopt the same validation strategy taking the interim result where the worst-group accuracy for the validation set is the highest.

1.2.3. Prompt-debias

The Prompt-debias [3] has a auxiliary adversarial model taking the image-text similarity as input aiming to predict the sensitive attribute. In the debiaisng stage, the part of input token is replaced with the debiasing token to maximizie the adversarial model's loss. In detail, the adversarial model is three hidden layers with 32 dimension, followed by ReLU activation function for each layer. The adversarial model is trained alone for the first two epoch, with Adam optimizer with learning rate 2×10^{-5} by a Binary Cross Entropy loss. Then, both the debiasing prompt and adverarial model are trained alternatively for 10 epochs. The promot is trained with Adam optimizer with learning rate 2×10^{-4}. The text prompt is trained by CLIP's output, to maintain the image-text contrastive similarity (ITC) while maximize the adversarial loss,

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda \mathcal{L}_{itc}$$

where $\lambda = 0.05$. The number of learnable token is set as 2 out of 77 token in the CLIP model.

1.2.4. CLIP-clip

CLIP-clip [22] is a feature-pruning method that measures mutual information between spurious annotations and each column feature in the image embeddings. Based on the measured mutual information for each feature, we retain top

k features having less mutual information with the spurious annotations. In the zero-shot classification task, to find the number of clipped features after being sorted by mutual information, a validation set is used. We grid search the number of clipped features k from 10 to d with 10 intervals, where d is the embedding size of the CLIP model. In the zero-shot retrieval task, we take $k = 400$ following the author's instruction.

1.2.5. RoboShot

For RoboShot [1], we follow the implementation from the authors. As Roboshot adopts LLM model in the design process, ChatGPT [25] is chosen as the LLM for generating prompts due to its superior generalized performance compared to other LLMs such as Flan-T5 [6], GPT-2 [18], and LLaMA [21]. The LLM is utilized to generate prompts to remove harmful insight and amplify helpful insight. Other than the prompts generation, we follow the released code in RoboShot since no hyperparameter is needed.

1.2.6. B2T

B2T [13], is designed to extract bias keywords from the pre-trained model and aims to generate text prompts for the inference stage by including the bias keywords in the prompts. As the pre-trained weight is not available, we utilize the generated text prompts for each class in the Waterbirds and CelebA datasets provided by the authors. Without any modification, the generated text prompts are used in the test.

1.2.7. DBP

DBP [5] only requires a list of spurious prompts containing the sensitive information only, and a list of candidate prompts containing the joints prompts for each target and sensitive attribute while making a pair of candidates from different attributes and the same target. In zero-shot classification, the spurious prompts and candidate prompts for Waterbirds and CelebA datasets are available from the authors' repository. In the case of the FairFace dataset, the spurious prompts follow that of CelebA. For the candidate prompts, we manually generate text prompt pairs by setting a pair ("A photo of a {target} man.", "A photo of a {target} woman."). Other than spurious and candidate prompts, we follow the implementation provided by the authors.

1.2.8. DEAR

DEAR [20] comprises two networks: the Protected Attribute Classifier (PAC) and the Additive Residual Learner (ARL). The PAC is a simple network with a single linear layer employing a ReLU activation function. This layer is followed by three classification heads dedicated to race, age, and gender, each consisting of two linear layers. The dimensions of these layers follow a progression of $256 \rightarrow 128 \rightarrow m$, where m stands for the number of classes: 7 for race, 4 for age, and 2 for gender provided in FairFace dataset. The PAC is trained using a cross-entropy loss function for each output, with a batch size of 512 and an Adam optimizer with a learning rate of 5×10^{-3} over 10 epochs. Once trained, the PAC module is frozen during the training of the ARL. The ARL involves a single linear transformation that preserves the embedding size, and the final output is obtained by adding the ARL output to the original image embedding. The ARL's objective function aims to maximize the PAC's cross-entropy loss while minimizing both the softmax output of the PAC and the difference between the CLIP output and the final output. The coefficient for the adversarial cross-entropy loss is set to 10^{-4}. The ARL is trained with a learning rate of 5×10^{-4}, a weight decay of 2×10^{-3}, and a batch size of 512 over 30 epochs.

2. Details on Bias Text Corpus and Generic Vision Dataset

One key merit of the proposed approach is that the predefined bias text corpora are easy to obtain and update by leveraging generative power of LLM. Here, we demonstrate how we accumulated the corpora by providing examples on specific target biases. Then we discuss the compiled generic vision datasets for specific target bias.

2.1. Establishing bias text corpus

We generated both prompt and corpus with GPT-3.5 [4]. Example query of how to generate the adequate prompt is as follows:

> "Create a prompt designed to produce a comprehensive corpus representing the diverse aspects of {target}. This corpus should encompass a wide range of specific terms relevant to the chosen category. {example}. It is crucial that the corpus is carefully constructed to avoid terms that may ambiguously apply to multiple subcategories within the chosen target. The goal is to ensure clear and distinct representation of each facet of the target, enabling nuanced understanding and analysis"

Here, examples of {target} includes *age, gender, education, etc*. And we also input example for respective target. For instance {example} for *age* is {*example*}: *"For example, if the target is 'age', the corpus should include terms like 'young', 'old', 'middle-aged', etc."*

Then, we queried with (generated) prompt to generate binary gender corpus as:

> "Generate a corpus of individuals with binary gender information, emphasizing a variety of nuanced expressions. Include both 'male' and 'female' as well as their synonyms, equivalent terms, and other related terms, while avoiding nouns that can be either gender. The expected

samples of the corpus should exhibit a range of gendered descriptions, such as 'man,' 'woman,' 'boy,' 'girl,' 'gentleman,' 'lady,' 'dude,' 'chick,' 'father,' 'mother,' 'son,' 'daughter,' 'brother,' 'sister,' 'husband,' 'wife' and so on, in a back-and-forth fashion, maintaining a diverse and balanced representation of binary gender information while excluding nouns that could apply to both genders. Make it into json file at the end."

An example of gender corpus derived from the above prompt is enumerated in Table 7. In addition, we generate other potential bias properties of demographics as in Table 8-14. Similarly, we could obtain the variation of background descriptions as in Table 15. Then, to fully articulate and complete the prompts that represents the bias information from a sample in the target bias corpus, we define prefix corpus as in Table. 6. By combining the prefix corpus with a specific bias corpus, *e.g.*, gender corpus, we can establish complete target bias corpus.

- A photo of a male
- A photo of a female
- Portrait of a man
- Portrait of a woman
- Image of a boy
- Image of a girl
- Snapshot of a gentleman

The example corpora is also included in a json file in the Github repository.

2.2. Establishing generic vision dataset

To disentangle bias representation with ZSDebias, we compute CKA with text embedding of target bias corpus. To this end, we require image embeddings to proxy downstream image distribution regarding target bias, *e.g.*, gender bias. For this purpose, we combine existing open source benchmarks to represent diversity of image distribution of interest. Specifically, for background bias, we combine the following two datasets:

- MS-COCO (animal) dataset [14]: The dataset consist of 23,552 samples with various annotations for segmentation, caption, object detection tasks. This is subset of MS-COCO dataset that contain *animal* superclass category in the image.
- ImageNet-100 dataset [7]: The dataset consist of 126,689 samples for image classification task. This is subset of ImageNet-1K dataset.

Whereas, we consider the following datasets when considering gender bias:

- UTKFace dataset [24]: The dataset consist of 10,137 samples of facial portraits. The dataset contains demographic labels including age, gender, and race.
- LFW dataset [10]: The dataset consist of 10,000 samples of facial portraits. It provides identity information of each sample.
- FACET dataset [9]: The dataset has various view of images including human. We selected 5858 samples with `visible_torso=True`.

Note that the unified generic vision dataset is employed to address various downstream tasks regarding the target bias. The overview of generic datasets are presented in Figure. 1 of the main paper.

3. Additional Experiments

In addition to Figure. 4 of the main paper, we present more examples of image modifications, targeting prompts associated with career biases, as discussed in previous studies [8, 16]. For instance, nurse and housekeeper are known to be linked with female, while engineer and firefighter are associated more with male. As in the figure below, ZSDebias consistently exhibit gender invariance in the modified images.

Moreover, we illustrate the image editing process (from left to right) until convergence in Figure 5. For the gender-neutral prompts, the modified images should retain the original gender of the source image (left). However, when we prompt career (engineer) that potentially reflects association bias [8], StyleCLIP equipped with Vanilla CLIP-RN50 model tends to change gender through the optimization process. In contrast, StyleCLIP utilizing CLIP-RN50 debiased by ZSDebias presents gender consistency during the editing procedure.

3.1. Additional Zero-shot Classification Results on Background Bias

We conducted further experiments to evaluate ZSDebias on background bias using the Waterbirds dataset [19]. In this task, the objective is to correctly classify whether a bird in the image is a waterbird or a landbird—a task known to be influenced by background cues (e.g., water or forest scenes).

Table 5 presents the worst-group accuracy, average accuracy, and the gap between them for both the CLIP ResNet-50 and CLIP ViT-L/14 backbones. Here, superior performance is indicated by a higher average accuracy and better fairness by a lower gap. Notably, ZSDebias achieves competitive performance without requiring task-specific sensitive or target label annotations during training. These results underscore the robustness and versatility of our approach in handling diverse zero-shot classification challenges

4. Ablation Studies

We conducted a series of ablation experiments to understand the influence of key hyperparameters on the fairness and overall performance of our method. Below, we detail

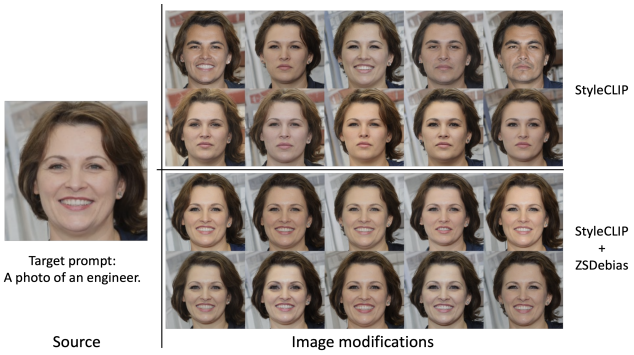
Method	CLIP ResNet-50 (Background)			CLIP ViT-L/14 (Background)		
	WG (↑)	Avg (↑)	Gap (↓)	WG (↑)	Avg (↑)	Gap (↓)
Zero-shot (ZS)	32.08	**92.25**	60.18	45.45	90.76	45.30
Group Prompt ZS	30.19	<u>90.67</u>	60.48	57.78	89.26	31.48
ERM Linear Probe	28.66	82.22	53.56	40.50	88.16	47.66
ERM Adapter	46.42	86.52	40.10	<u>70.87</u>	**92.63**	21.76
CLIP-clip [22]	58.49	62.69	<u>4.19</u>	15.73	71.14	55.41
RoboShot [1]	50.47	79.86	29.39	42.21	85.50	43.29
B2T [13]	55.76	79.60	23.84	43.77	86.00	42.23
ConAdapt [23]	**79.65**	81.69	**2.04**	**85.36**	<u>92.53</u>	**7.17**
DBP [5]	<u>69.78</u>	77.86	8.08	64.17	86.85	22.67
ZSDebias (ours)	50.00	72.45	22.45	66.63	84.56	<u>17.93</u>

Table 5. Performance on the Background group (from the Waterbirds dataset) for various methods using two CLIP models. WG: Worst Group, Avg: Average, Gap: Difference between average and worst group accuracies. For WG and Avg, higher is better; for Gap, lower is better. Best performance is highlighted in bold and second best is underlined.

(a) StyleCLIP (Vanilla CLIP-RN50)

(b) StyleCLIP (ZSDebias CLIP-RN50)

Figure 5. Interpolation of image editing procedure given the prompt "A smiling engineer with curly hair" until convergence. The fair modification should be invariant to gender until it converges. Vanilla CLIP-based StyleCLIP exhibits vulnerability given "engineer", which is known to be gender biased.



two main studies: one varying the weight for the CKA loss and another examining the effect of different top-k zero-shot sample ensembles.

4.1. Effect of CKA Loss

To investigate the role of the CKA loss in enhancing fairness, we varied the weight λ_{CKA} across three values: 10, 1,

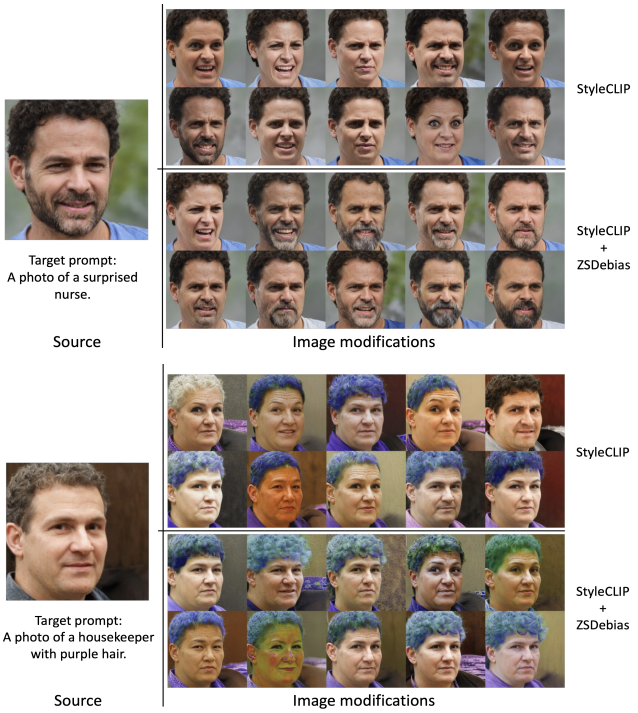


Figure 6. Additional results of a variety of image modifications (right) of the source images (left) given the target prompts. Fair image editing should be invariant to the gender of the source image.

and 0.1. Figure 7 illustrates the impact on worst group accuracy and average accuracy. Our results indicate that a higher CKA loss weight significantly improves the worst group accuracy, which is critical for fairness. In particular, when $\lambda_{\text{CKA}} = 10$, the worst group accuracy reaches 78.475%, compared to 71.748% for $\lambda_{\text{CKA}} = 1$ and 70.237% for $\lambda_{\text{CKA}} = 0.1$. Although the average accuracy remains high (82.79%, 82.53%, and 79.848% for $\lambda_{\text{CKA}} = 10, 1, 0.1$, respectively), the improvement in worst group performance demonstrates that emphasizing the CKA loss aids in mitigating bias.

4.2. Impact of Top-k Zero-Shot Sample Ensembles

In addition to the CKA loss weight, we evaluated how the number of top-k zero-shot sample ensembles affects performance. Figure 8 shows the performance for different k values: 1, 3, 5, and 10. The results reveal that a lower k is preferable for fairness, as the worst group accuracy decreases from 82.03% with $k = 1$ to 75.192% with $k = 10$. Similarly, the average accuracy declines from 0.8513 to 0.8089 as k increases. These findings suggest that while ensemble methods can provide robustness, using too many zero-shot samples for matching may dilute the alignment quality, adversely affecting performance on the worst-performing groups.

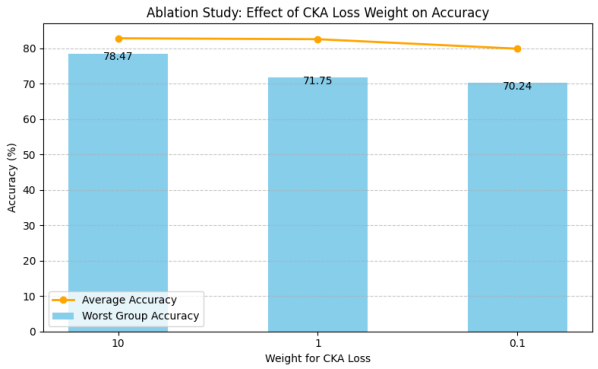


Figure 7. Effect of varying λ_{CKA} on worst group and average accuracy. A higher CKA weight improves worst group accuracy, demonstrating enhanced fairness.

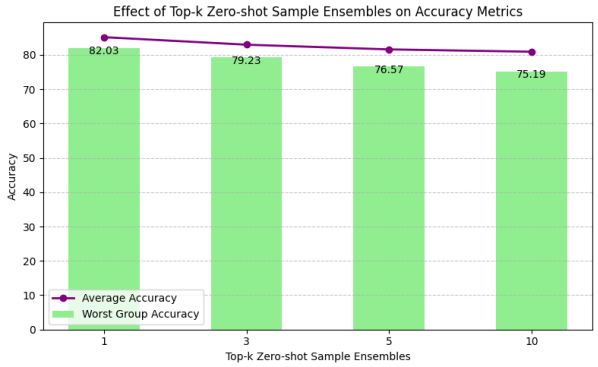


Figure 8. Effect of top-k zero-shot sample ensemble size on worst group and average accuracy. Lower k values yield higher worst group accuracy, indicating a better balance between fairness and overall performance.

These ablation studies validate our design choices: a stronger emphasis on the CKA loss and a careful selection of ensemble size are both critical for achieving a fairer debiasing outcome without significantly compromising overall performance.

References

[1] Adila, D., Shin, C., Cai, L., Sala, F.: Zero-shot robustification of zero-shot models with foundation models. arXiv preprint arXiv:2309.04344 (2023) 2, 4

[2] Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016) 1

[3] Berg, H., Hall, S.M., Bhalgat, Y., Yang, W., Kirk, H.R., Shtedritski, A., Bain, M.: A prompt array keeps the bias away: Debiasing vision-language models with adversarial learning. arXiv preprint arXiv:2203.11933 (2022) 1

[4] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. NeurIPS **33**, 1877–1901 (2020) 2

[5] Chuang, C.Y., Jampani, V., Li, Y., Torralba, A., Jegelka, S.: Debiasing vision-language models via biased prompts. arXiv preprint arXiv:2302.00070 (2023) 2, 4

[6] Chung, H.W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., et al.: Scaling instruction-finetuned language models. arXiv preprint arXiv:2210.11416 (2022) 2

[7] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255. Ieee (2009) 3

[8] Gonen, H., Goldberg, Y.: Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. arXiv preprint arXiv:1903.03862 (2019) 3

[9] Gustafson, L., Rolland, C., Ravi, N., Duval, Q., Adcock, A., Fu, C.Y., Hall, M., Ross, C.: Facet: Fairness in computer vision evaluation benchmark. In: ICCV. pp. 20370–20382 (2023) 3

[10] Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: Workshop on faces in'Real-Life'Images: detection, alignment, and recognition (2008) 3

[11] Karkkainen, K., Joo, J.: Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In: WACV. pp. 1548–1558 (2021) 1

[12] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR. pp. 4401–4410 (2019) 1

[13] Kim, Y., Mo, S., Kim, M., Lee, K., Lee, J., Shin, J.: Bias-to-text: Debiasing unknown visual biases through language interpretation. arXiv preprint arXiv:2301.11104 **2** (2023) 2, 4

[14] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: ECCV. pp. 740–755. Springer (2014) 3

[15] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: ICCV (December 2015) 1

[16] Mandal, A., Little, S., Leavy, S.: Multimodal bias: Assessing gender bias in computer vision models with nlp techniques. In: ICMI. pp. 416–424 (2023) 3

[17] Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., Lischinski, D.: Styleclip: Text-driven manipulation of stylegan imagery. In: ICCV. pp. 2085–2094 (2021) 1

[18] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019) 2

[19] Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. arXiv preprint arXiv:1911.08731 (2019) 1, 3

[20] Seth, A., Hemani, M., Agarwal, C.: Dear: Debiasing vision-language models with additive residuals. In: CVPR. pp. 6820–6829 (2023) 2

[21] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023) 2

[22] Wang, J., Liu, Y., Wang, X.E.: Are gender-neutral queries really gender-neutral? mitigating gender bias in image search. arXiv preprint arXiv:2109.05433 (2021) 1, 4

[23] Zhang, M., Ré, C.: Contrastive adapters for foundation model group robustness. NeurIPS **35**, 21682–21697 (2022) 1, 4

[24] Zhang, Zhifei, S.Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: CVPR. IEEE (2017) 3

[25] Ziegler, D.M., Stiennon, N., Wu, J., Brown, T.B., Radford, A., Amodei, D., Christiano, P., Irving, G.: Fine-tuning language models from human preferences. arXiv preprint arXiv:1909.08593 (2019) 2

snapshot of	portrait of	image of
depiction of	rendering of	illustration of
capture of	representation of	scene of
view of	glimpse of	close-up of
aerial view of	detailed look at	panoramic view of
sketch of	digital creation of	vivid depiction of
artistic interpretation of		snapshot capturing

Table 6. Generated prefix corpus consist of 20 objects.

Male	Female	Man	Woman	Boy	Girl
Guy	Gal	Gentleman	Lady	Dude	Chick
Father	Mother	Son	Daughter	Brother	Sister
Husband	Wife	Uncle	Aunt	Nephew	Niece
Grandfather	Grandmother	Groom	Bride	Father-in-law	Mother-in-law
Son-in-law	Daughter-in-law	Stepfather	Stepmother	Stepson	Stepdaughter
Godfather	Godmother	Godson	Goddaughter	Boyfriend	Girlfriend
Fiancé	Fiancée	Bachelor	Bachelorette	Widower	Widow
King	Queen	Prince	Princess	Actor	Actress
Waiter	Waitress	Host	Hostess	Master	Mistress
Lord	Lady	Duke	Duchess	Emperor	Empress
Count	Countess	Sir	Madam		

Table 7. Generated gender corpus of 72 objects.

Newborn	Infant	Toddler	Preschooler	Preteen	Adolescent
Teenager	Octogenarian	Nonagenarian	Centenarian	Quadragenarian	Quinquagenarian
Sexagenarian	Septuagenarian	Tween	Youngster	Juvenile	Minor
Adult	Senior	Geriatric	Neonate	Weanling	Yearling
Fledgling	Pubescent	Postpubescent	Middlescent	Supercentenarian	Dotard
Venerable	Youth	Elder	Golden-ager	Retiree	Pensioner
Middle-ager	Thirtysomething	Fortysomething	Fiftysomething	Sixtysomething	Septuagenarian
Octogenarian	Nonagenarian	Centenarian	Silver	Codger	Cradle
Fogy	Ancestor	Descendant	Baby	Child	Kid
Teen	Young adult	Middle-aged	Elderly	Ancient	Millennial
Generation Z	Baby boomer	Generation X			

Table 8. Generated age corpus of 59 objects.

Caucasian	African	Asian	Hispanic	Indigenous	Aboriginal
Maori	Sami	Inuit	Berber	Polynesian	Melanesian
Micronesian	Amerindian	Mestizo	Mulatto	Creole	Romani
Bedouin	Tuareg	Fulani	Yoruba	Zulu	Xhosa
Masai	Han	Bengali	Tamil	Sikh	Malay
Ainu	Tibetan	Uyghur	Kurd	Pashtun	Hazara
Tajik	Baloch	Sinhalese	Karen	Hmong	Maasai
San	Pygmy	Bantu	Aztec	Maya	Inca
Mapuche	Quechua	Aymara	Métis	First_Nations	Aborigine

Table 9. Generated race corpus of 57 objects.

Toxic person	Harmful individual	Abusive person	Destructive person
Malevolent individual	Dangerous person	Pernicious person	Hostile person
Fat	Lazy person	Criminal	Terrorist
Drug addict	Racist	Ugly	Mad
Unattractive	Vindictive person	Deceitful individual	Manipulative person
Negligent person	Corrupt official	Despised individual	Malicious actor
Delinquent	Reckless individual	Ruthless person	Unscrupulous individual
Conniving person	Troublemaker	Vicious criminal	Unethical person
Selfish individual			

Table 10. Generated toxicity corpus of 32 objects.

Person with albinism	Person with freckles	Person with vitiligo	Person with alopecia
Hairy person	Person with epicanthic folds	Person with brachia	Right-handed person
Left-handed person	Person with amputation	Person with visual impairment	Person with hearing impairment
Person with muteness	Able-bodied person	Neurotypical person	Person with amblyopia
Tattooed person	Pierced person	Scarred person	Burn victim
Person using mobility aids	Person using prosthetics	Person using wheelchair	Person with orthopedic condition
Person with dwarfism	Person with gigantism	Pigmented person	Person with melanism
Androgynous person	Intersex person	Person with hemophilia	Person with diabetes
Person with asthma	Person with allergies	Person with autism	Person with dyslexia
Ambidextrous person	Person with albinism	Person with vitiligo	Eunuch
Person with kyphosis	Person with lactose intolerance	Person with celiac disease	Person with epilepsy
Person with paraplegia	Person with quadriplegia	Person with hemiplegia	Person with strabismus
Person with tremors	Person with nystagmus	Person with dysphonia	Person who stutters

Table 11. Generated appearance corpus of 52 objects.

High school dropout	High school graduate	College student	Undergraduate	Graduate student
PhD candidate	Postdoctoral researcher	Professor	Teacher	Principal
Dean	Academic	Scholar	Researcher	Scientist
Student	Alumni	Valedictorian	Salutatorian	Honors student
Exchange student	International student	Teaching assistant	Research assistant	Fellow
Lecturer	Associate professor	Full professor	Emeritus professor	Department head
School board member	Education administrator	Vocational student	Trade school student	MBA student
Law student	Medical student	Engineering student	Art student	Music student
Drama student	ESL student	Adult learner	Distance learner	Homeschooled student
Special education student	Gifted student	Transfer student	First generation student	

Table 12. Generated education corpus of 49 objects.

Conservative	Liberal	Progressive	Moderate	Independent	Centrist
Leftist	Rightist	Socialist	Capitalist	Libertarian	Anarchist
Communist	Fascist	Democrat	Republican	Green party member	Populist
Nationalist	Globalist	Activist	Pacifist	War hawk	Peace dove
Reformist	Radical	Monarchist	Republican	Federalist	States rights advocate
Constitutionalist	Authoritarian	Totalitarian	Democratic socialist	Social democrat	Neo liberal
Neo conservative	Traditionalist	Modernist	Isolationist	Interventionist	Environmentalist
Tea party member	Occupy protester	Alt right member	Alt left member	Swing voter	Single issue voter
Party loyalist	Independent voter				

Table 13. Generated political corpus of 45 objects.

Christian	Muslim	Hindu	Buddhist	Sikh	Jewish
Bahai	Jain	Shinto	Taoist	Zoroastrian	Pagan
Wiccan	Atheist	Agnostic	Mormon	Catholic	Protestant
Orthodox	Sunni	Shia	Sufi	Salafi	Hasidic
Reform	Zen	Tantric	Vajrayana	Mahayana	Theravada
Vaishnavite	Shaivite	Amish	Mennonite	Quaker	Unitarian
Calvinist	Lutheran	Anglican	Coptic	Druze	Rastafarian
Scientologist	Shaker	Swaminarayan	Zulu	Santeria	Voodoo
Pastafarian	Methodist	Presbyterian	Baptist	Evangelical	Pentecostal

Table 14. Generated religion corpus of 50 objects.

A forest with tall trees.	A beach with clear waters.	A snow-covered mountain.	A lakeside with dense foliage.
A coral reef underwater.	A thunderstorm with lightning.	A pond with lily pads.	A view of the Grand Canyon.
A foggy morning in a forest.	A starry night in the wilderness.	A forest glade with deer.	A snowy landscape with a cabin.
A colorful autumn forest.	A secluded beach with a cove.	A tropical paradise with palms.	A rugged canyon with cliffs.
A serene waterfall in a canyon.	A river winding through a valley.	A green pasture with grazing animals.	A vast savanna with wildlife.
A field ready for harvest.	A city park with greenery.	A bamboo forest with winding paths.	An ocean under a clear sky.
A desert at sunset.	A dense jungle with a waterfall.	A tundra covered in snow and ice.	A clear stream in a quiet glen.
A rocky coastline with waves.	A quiet pond surrounded by trees.	A redwood forest with tall trees.	A sunflower field in sunlight.
An alpine lake reflecting mountains.	A coral reef with marine life.	A meadow with birds.	A vineyard with grapevines.
A pagoda surrounded by blossoms.	A waterfall in a rainforest.	A lavender field in full bloom.	A glacial lake with clear waters.
A city square with historic architecture.	A sun-soaked beach with palm trees.	A pine forest with needles.	A bustling harbor with boats.
A meadow with wildflowers.	A pond with water lilies.	A bayou with cypress trees.	A waterfall with a rainbow.
An olive grove under the sun.	A hillside with vineyards.	A mangrove forest by the coastline.	A koi pond in a garden.
A cottage in a vineyard.	A lavender field with purple blooms.	A forest glen with a brook.	A beach with footprints in the sand.
An urban park with families.	A garden with a stone fountain.	A fern-covered forest floor.	A lake at dawn with mist.
A marsh with tall reeds.	A garden with butterflies.	A garden with a wooden bridge.	A mangrove swamp with waterways.
A cottage in a garden.	A lavender field with fragrant flowers.	A beach with seashells.	A market square with vendors.
A garden with a pond.	A fern-covered forest floor.	A lake at sunset with reflections.	A marshland with tall grasses.
A field of poppies in full bloom.	A desert landscape under a starry sky.	A village nestled in a valley.	A cityscape with skyscrapers.
A forest path with fallen leaves.	A mountain peak at sunrise.	A river cutting through a dense forest.	A tranquil beach at dusk.
A misty valley at dawn.	A desert oasis with palm trees.	A volcanic landscape with steam vents.	A rice terrace in morning light.
A cherry blossom garden in spring.	A medieval castle on a hilltop.	A lighthouse on a rocky shore.	An ancient temple in ruins.
A cobblestone street in old town.	A mountain stream with rapids.	A field of tulips in bloom.	A zen garden with raked sand.
A mountain peak at sunrise.	A river cutting through a dense forest.	A tranquil beach at dusk.	A snowy street in a small town.
An ancient temple in a jungle.	A desert oasis with palm trees.	A flower garden in full bloom.	A frozen lake surrounded by pine trees.
A cave entrance surrounded by vines.	A country road lined with autumn trees.	A traditional village with thatched roofs.	A deep forest trail with sunlight filtering through.
A snowy village with lights.	A castle ruin on a hill.	A serene pond with ducks.	

Table 15. Generated background corpus of 94 objects.