

A Appendix

A.1 Limitations and Future Work

In our work, as in most previous approaches, we cannot guarantee recovery of every ground-truth factor of variation. Some subtle or rare attributes may simply fall outside the axes we discover. In fact, perfectly capturing all underlying factors in a complex, real-world dataset is generally intractable. Nevertheless, our method still identifies diverse, meaningful concepts, and extending coverage to additional or more fine-grained factors remains an important direction for our future work. Moreover, our framework depends on the quality and scope of the pretrained vision language model (VLM), so it can only discover concepts the VLM recognizes. Fortunately, as VLMs are improving rapidly and our method is not restricted by a specific VLM, we can adopt stronger models as they become available.

A.2 Broader Impact

Our approach can extract diverse visual concepts from images and reuse them to synthesize new content, which could pose privacy issues such as deepfake generation or unauthorized duplication of digital content.

A.3 Additional Implementation Details

Table 7 summarizes hyper-parameters for model architectures and training used in our experiments. For baselines, we follow the default hyper-parameters recommended by the official codes. All baselines used DDIM inversion with guidance of 7.5 and 50 inference steps.

Table 7: Hyperparameters used in our experiments.

General	Batch Size	32
	Training Steps	100k
	Learning Rate	0.00003
Concept Encoder	Layers	4
	Hidden Dim	768
	Number of Heads	8
Regression Network	Layers	768
	Input Dimension	768
	Hidden Dimensio	768
	Activation Function	ReLU

A.4 Prompt for Concept Axes Extraction

We provide the complete prompt and examples of the discovered concept axes per image in Figure 6 and Figure 7, respectively. As shown in Figure 7, our prompt successfully steers the VLM to identify diverse concept axes across different datasets, even when using only a single output exemplar of a human face.

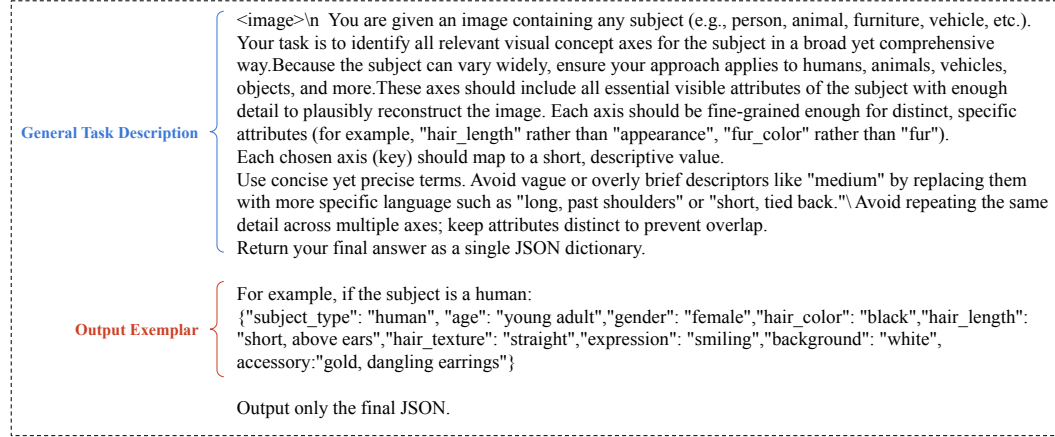


Figure 6: Our complete prompt consists of a general task description and output exemplar.



Figure 7: Examples of outputs from the VLM. Concept axes colored in red are unseen from the given exemplar.

A.5 Human Evaluation

For human evaluation, we randomly select 10 pairs of images for each attribute. Then, we replace an attribute of one image with another one in each pair using each of the methods. We ensure that randomly selected attributes in each pair are different from each other so that the edited image is always recognizable. We collect 10 participants for each dataset (a total of 30) on Prolific [28] and provide a general guideline as in Figure 8 for the task. Our questionnaire (Figure 9) asks participants to rank the images that most closely adhere to the criteria provided in our guideline. Following [16], we used Borda score metrics [32] to differentiate the scores according to each ranking, and final scores are normalized to a 0-1 scale.

Concept Swapping Visual Inspection

B *I* U

In this survey, you will be presented with pairs of images alongside a target concept that describes what the edited image should modify. Your task is to evaluate and rank the edited images that most closely adheres to the following criteria.

✔ Concept Adherence:
Given a target concept, the edited image should accurately reflect the same concept details of the source image. For example, if the target concept is "hair color", the hair color of the edited image should be same as that of the source image.

✔ Preservation of Other Attributes:
Only the details of the given target concept should change, while other attributes must remain unchanged. The other concepts to remain unchanged are specified for each task.

You will be given 50 questions in total, 10 questions for each of the 5 concepts.

Additional Guidelines:

- 1 is the best, while 6 is the worst.
- Take your time to inspect each image carefully before making your selection.
- If none of the answers seem accurate, answer with your best guess.
- No ties.
- Please ZOOM IN your images for better inspection.

Figure 8: General guidelines used in our human evaluation.

1. Which image best represents (C), reflecting the 'fur_color' from (B) while keeping all other details from (A)?

Rank the answers: 1st (Best) - 6th (Worst).

(A) Target image

(B) Source image

(C)

Attributes of (C):
fur_color: (B)
fur_pattern: (A)
eye_color: (A)
expression: (A)
background: (A)

(a)

(b)

(c)

(d)

(e)

(f)

	1	2	3	4	5	6
(a)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(b)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(c)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(d)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(e)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(f)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 9: Questionnaires used in human evaluation.

A.6 Additional Qualitative Results

A.6.1 Additional Qualitative Comparisons on Visual Concept Editing

Figures 10–19 present additional qualitative results along diverse concept axes discovered in ImageNet-S20, CelebA-HQ, and AFHQ datasets. Across all axes, our method consistently outperforms the baselines. Whereas the baselines often fail to accurately capture and transfer the specified visual attributes, our approach reliably extracts the visual concept from the source and transfers it to the target image. Since LIVCL trains a set of separate encoders only for the top–10 frequent axes, it was unable to evaluate “lip color” in Figure 15 and “collar” in Figure 18, which are not among the top–10 most frequent concepts in the dataset, and we therefore mark those entries as N/A.



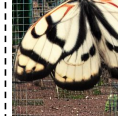






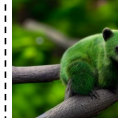





























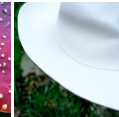




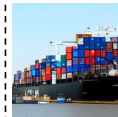
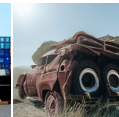


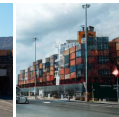



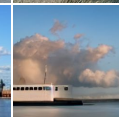
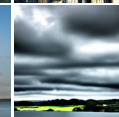
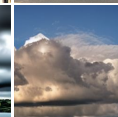
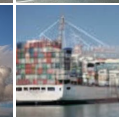
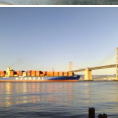


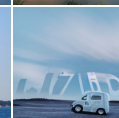


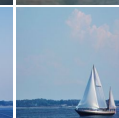
Target Concept	Source	Target	Ours	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Species							
							
							
Cap Color							
							
							
Sky Condition (e.g., clear, cloudy)							
							
							

Figure 10: Additional qualitative comparison to baselines in ImageNet-S20









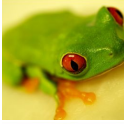
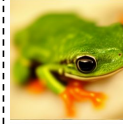
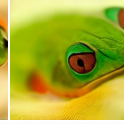
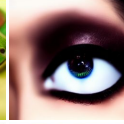
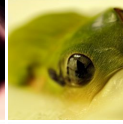
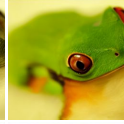
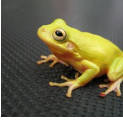


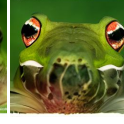








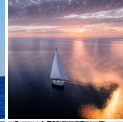
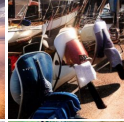






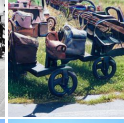


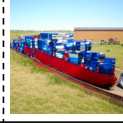




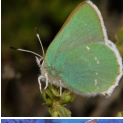


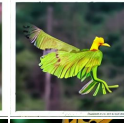
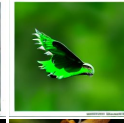
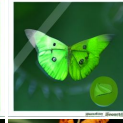


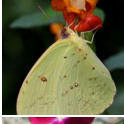
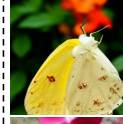

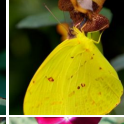
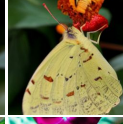
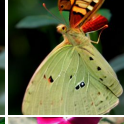

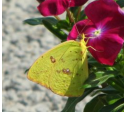
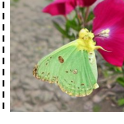
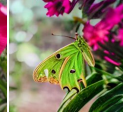
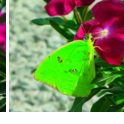
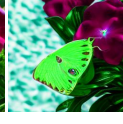
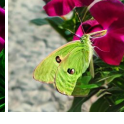
Target Concept	Source	Target	Ours	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Eye Color							
							
							
Vehicle Type							
							
							
Wing Color							
							
							

Figure 11: Additional qualitative comparison to baselines in ImageNet-S20

Target Concept	Source	Target	Ours	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Stem Color							
							
							
Flower Color							
							
							
Age							
							
							

Figure 12: Additional qualitative comparison to baselines in ImageNet-S20














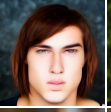
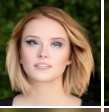






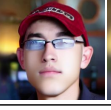
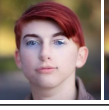




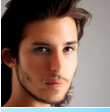
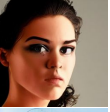
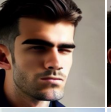
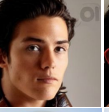



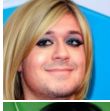

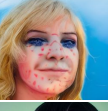


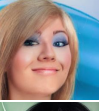





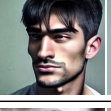
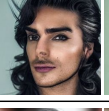
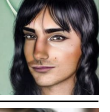




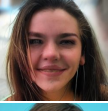
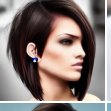




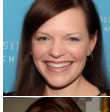

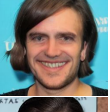


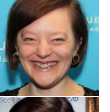
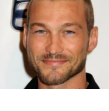

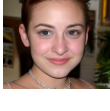
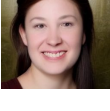

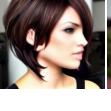
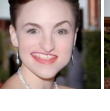
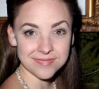
Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Age								
								
								
Gender								
								
								
Hair length								
								
								

Figure 13: Additional qualitative comparison to baselines in CelebA-HQ







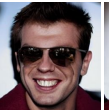

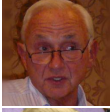

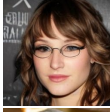
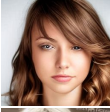
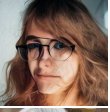







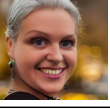
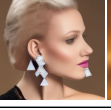
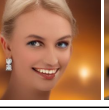


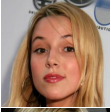
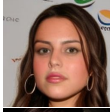
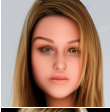
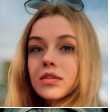
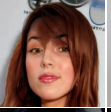
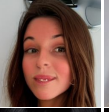
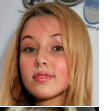
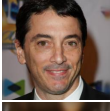



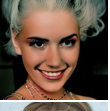


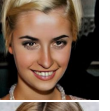
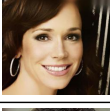

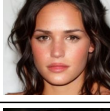
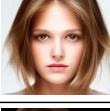
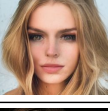

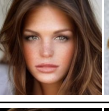
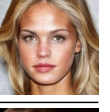


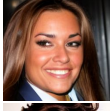

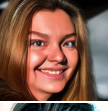











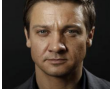


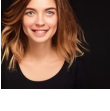
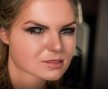
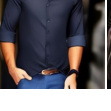
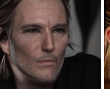

Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Accessory								
								
								
Hair color								
								
								
Clothing								
								
								

Figure 14: Additional qualitative comparison to baselines in CelebA-HQ

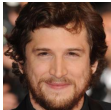
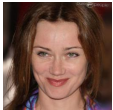




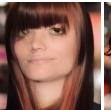

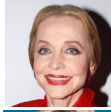
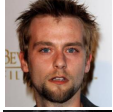
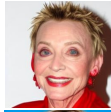

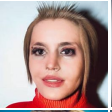
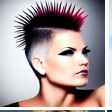
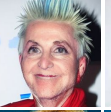

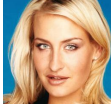


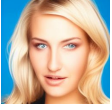
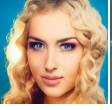
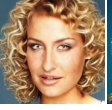
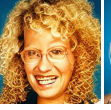
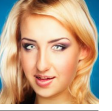
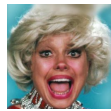







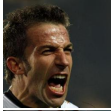





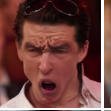




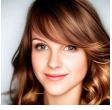
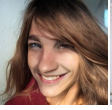
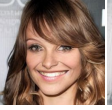
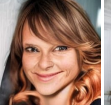

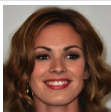
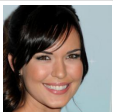
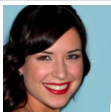
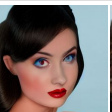
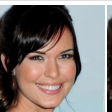
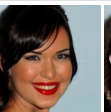









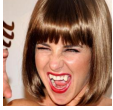
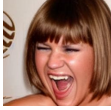
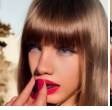

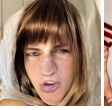
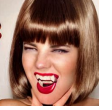
Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Hair Texture								
								
								
Expression								
								
								
Lip Color				N/A				
				N/A				
				N/A				

Figure 15: Additional qualitative comparison to baselines in CelebA-HQ






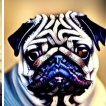



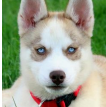
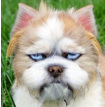














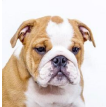


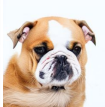
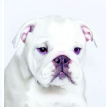
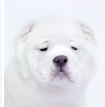
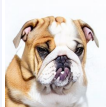

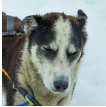


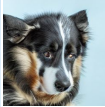


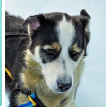




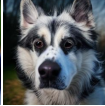
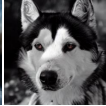

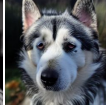

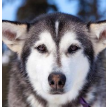
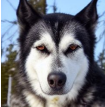


















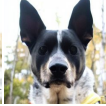


Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Breed								
								
								
Fur color								
								
								
Ear shape								
								
								

Figure 16: Additional qualitative comparison to baselines in AFHQ-Dog

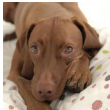

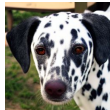
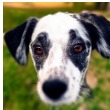
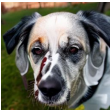
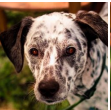


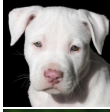

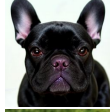



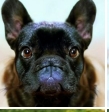
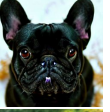

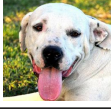



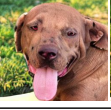
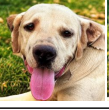

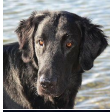
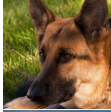
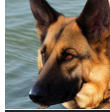
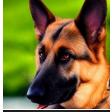
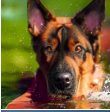
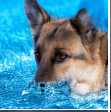
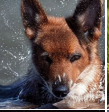
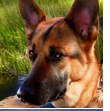
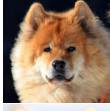
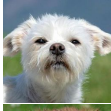
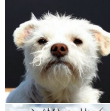
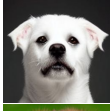
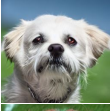

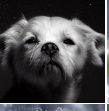
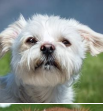
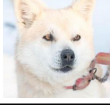
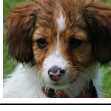

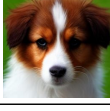


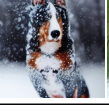

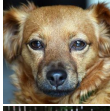
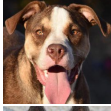
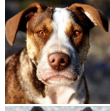

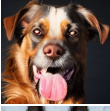
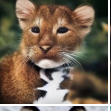

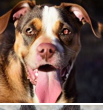

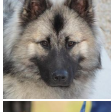
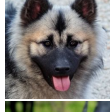
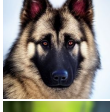
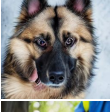
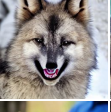
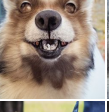
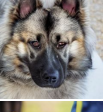
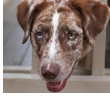
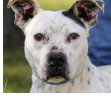
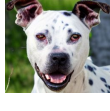
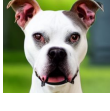
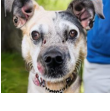
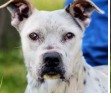
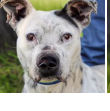
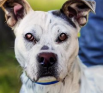
Target Concept	Target	Source	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Nose color								
								
								
Background								
								
								
Expression								
								
								

Figure 17: Additional qualitative comparison to baselines in AFHQ-Dog


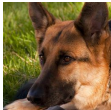
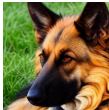
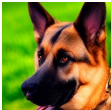
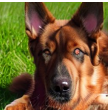
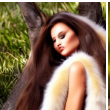
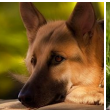
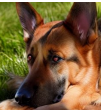
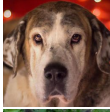




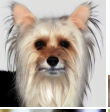







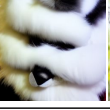
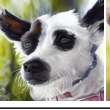

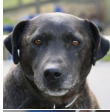
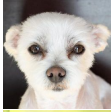
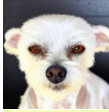
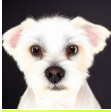

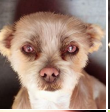
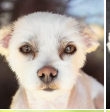
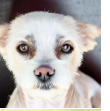

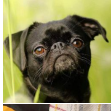
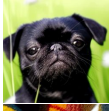
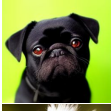



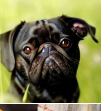

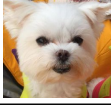
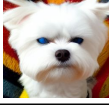
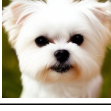
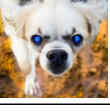
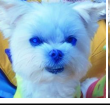
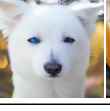
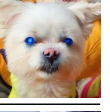
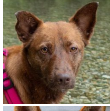
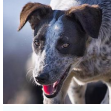
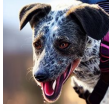
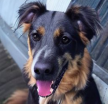
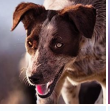
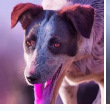
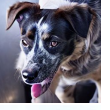
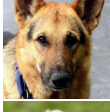





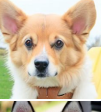
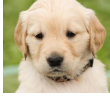
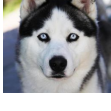
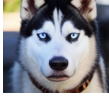

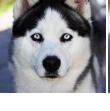
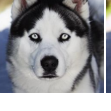
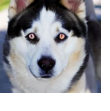
Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Fur Texture								
								
								
Eye Color								
								
								
Collar				N/A				
				N/A				
				N/A				

Figure 18: Additional qualitative comparison to baselines in AFHQ-Dog












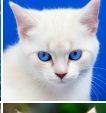
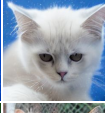

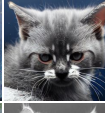
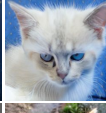



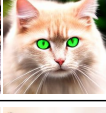
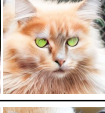
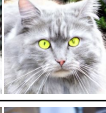

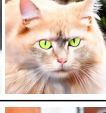






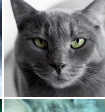








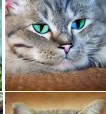



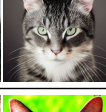


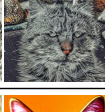
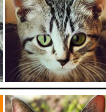




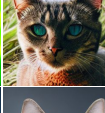

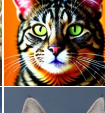












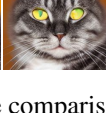
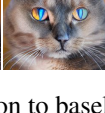


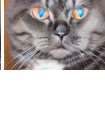
Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Fur color								
								
								
Background								
								
								
Eye color								
								
								

Figure 19: Additional qualitative comparison to baselines in AFHQ-Cat

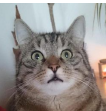




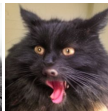





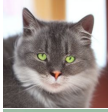


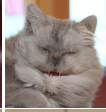
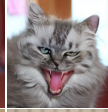
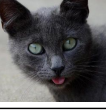
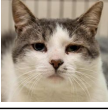
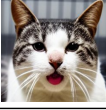

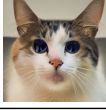
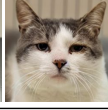
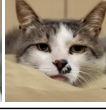
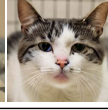
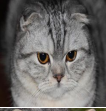


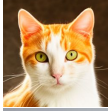
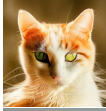
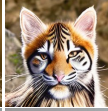
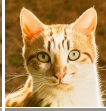
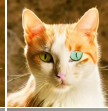

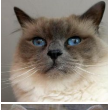


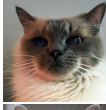
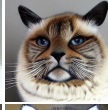
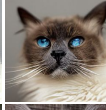
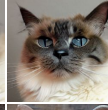

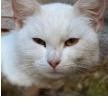
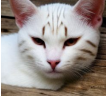
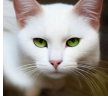
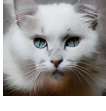
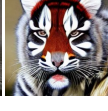
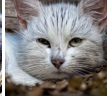
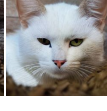
Target Concept	Source	Target	Ours	LIVCL	SDEdit	Instruct Pix2Pix	NullText Inv.	DDPM Inv.
Expression								
								
								
Fur Pattern								
								
								

Figure 20: Additional qualitative comparison to baselines in AFHQ-Cat

A.6.2 More Qualitative Results on Compositions from Multiple Images

We provide more qualitative results on the composition of visual concepts from multiple images in Figure 23-27. We extract N distinct visual concepts from N different images and replace the corresponding visual concepts of the target images with them. Our method successfully transfers multiple visual concepts to target images, which implies that each visual concept extracted from source images is disentangled along other axes.

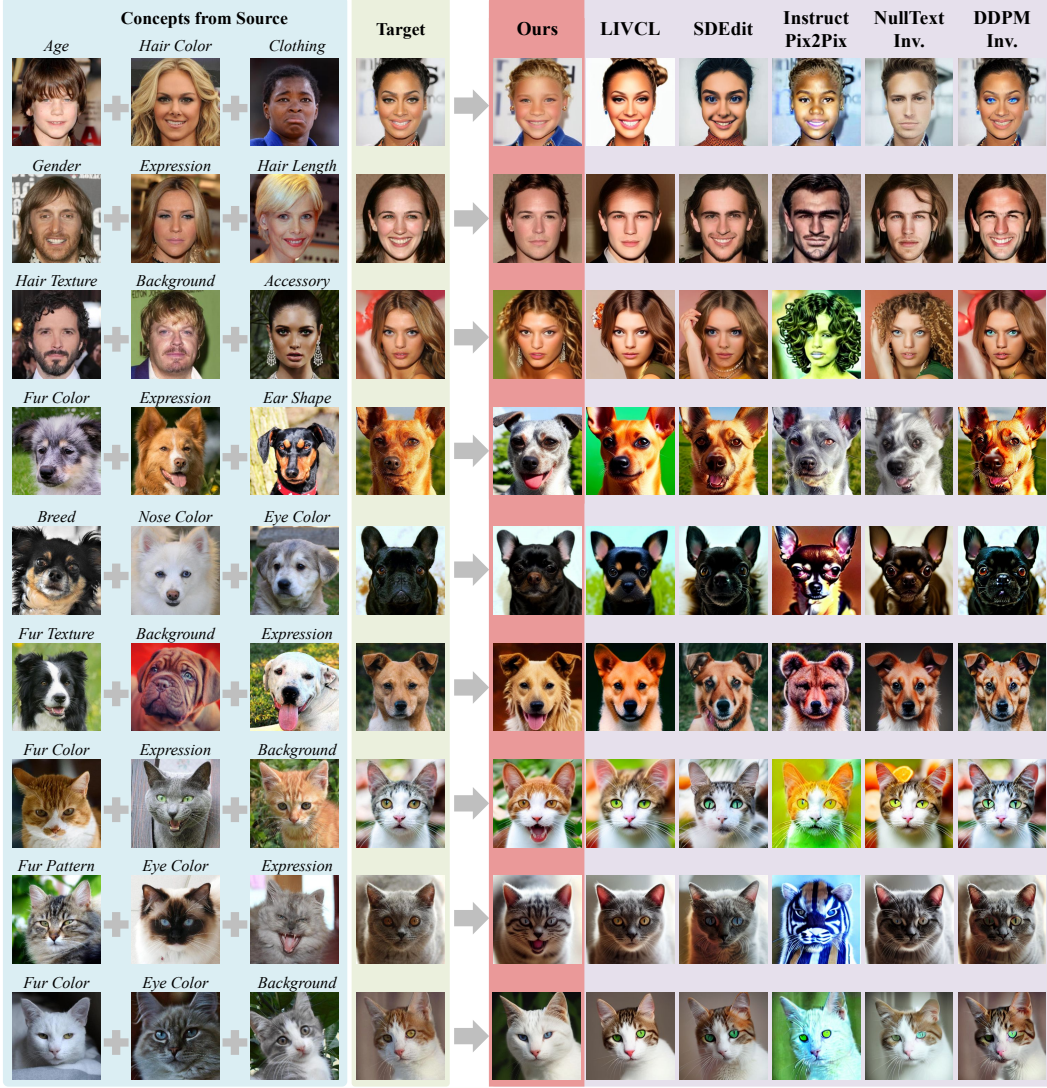


Figure 21: Compositions of visual concepts from multiple images ($N = 3$).

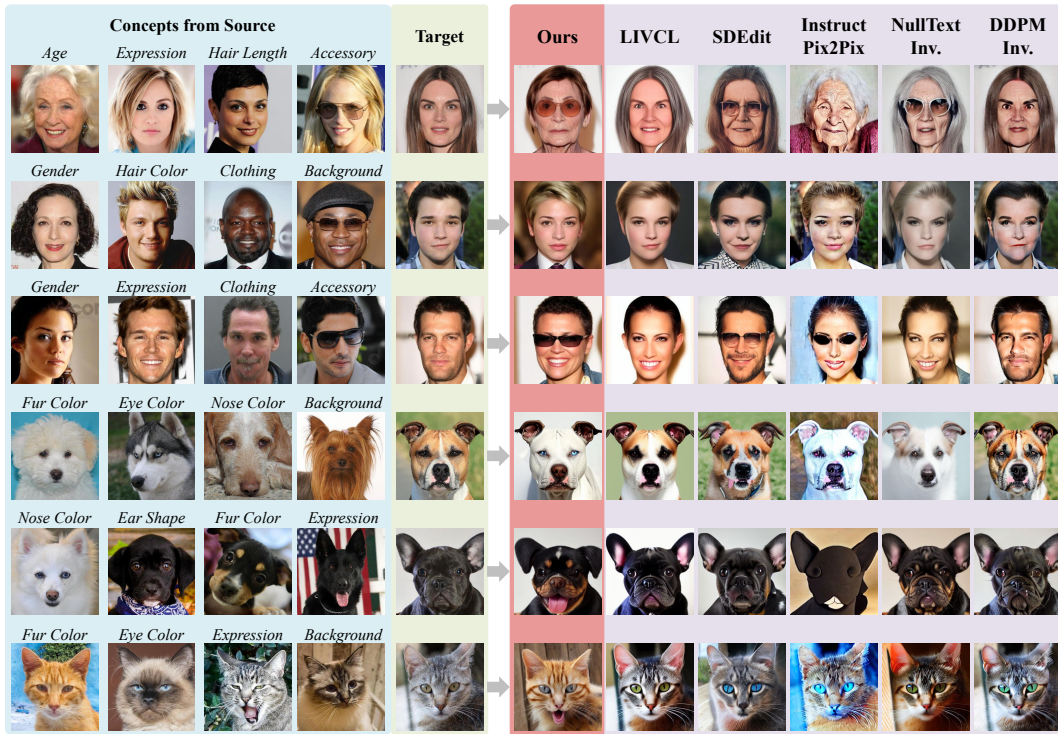


Figure 22: Compositions of visual concepts from multiple images ($N = 4$).

A.6.3 More Qualitative Results on Visual Nuance Transfer

We provide more qualitative results on transferring visual nuance from source to target images in Figure 23-27.

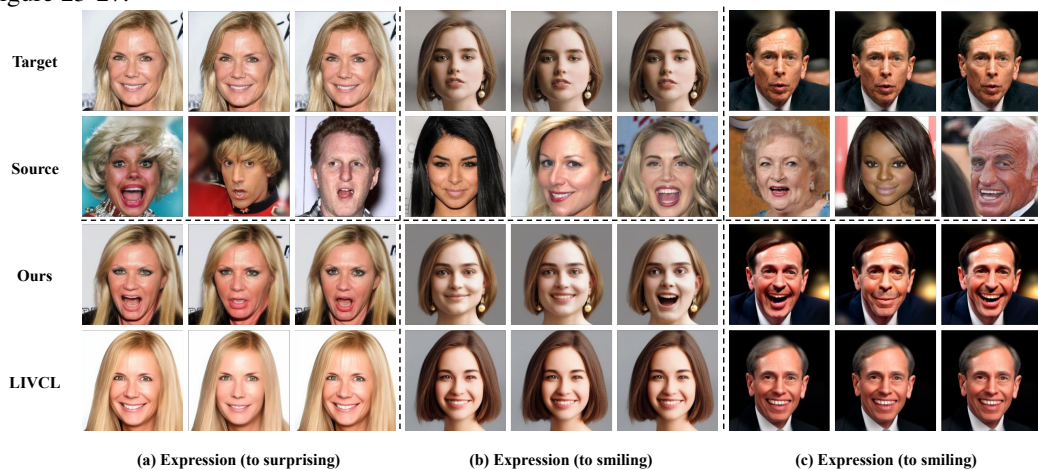


Figure 23: Transferring Visual Nuances from source to target images

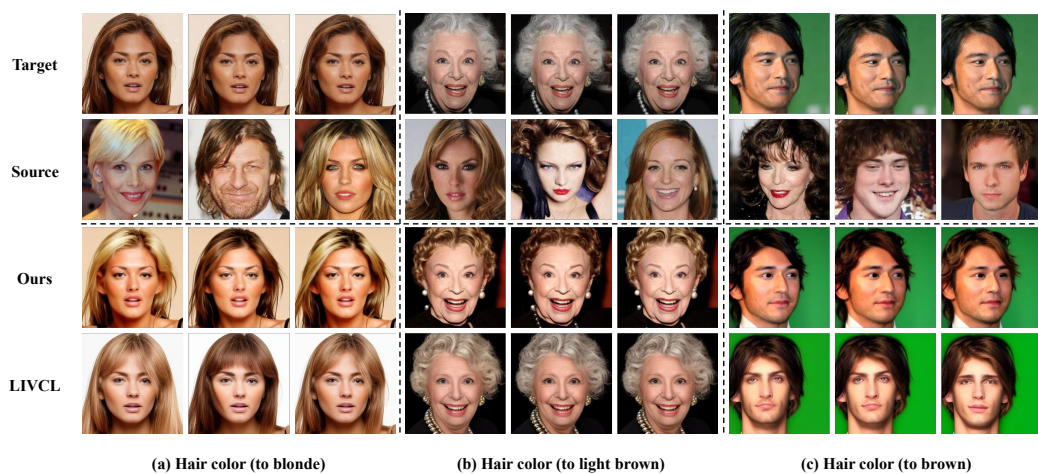


Figure 24: Transferring Visual Nuances from source to target images.

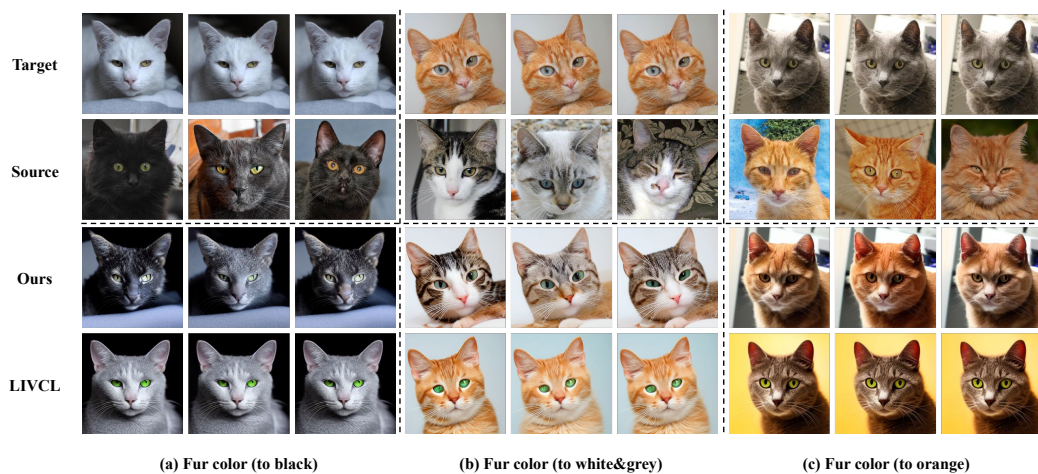


Figure 25: Transferring Visual Nuances from source to target images.

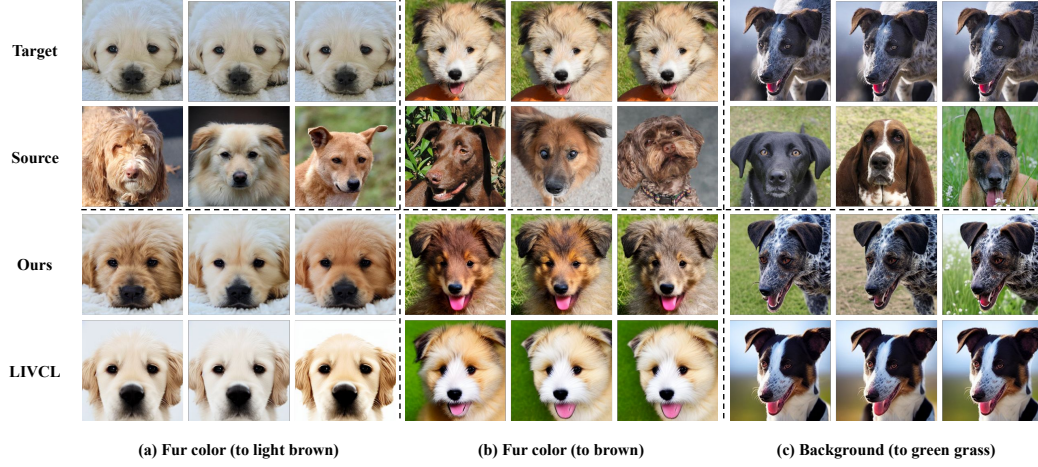


Figure 26: Transferring Visual Nuances from source to target images.

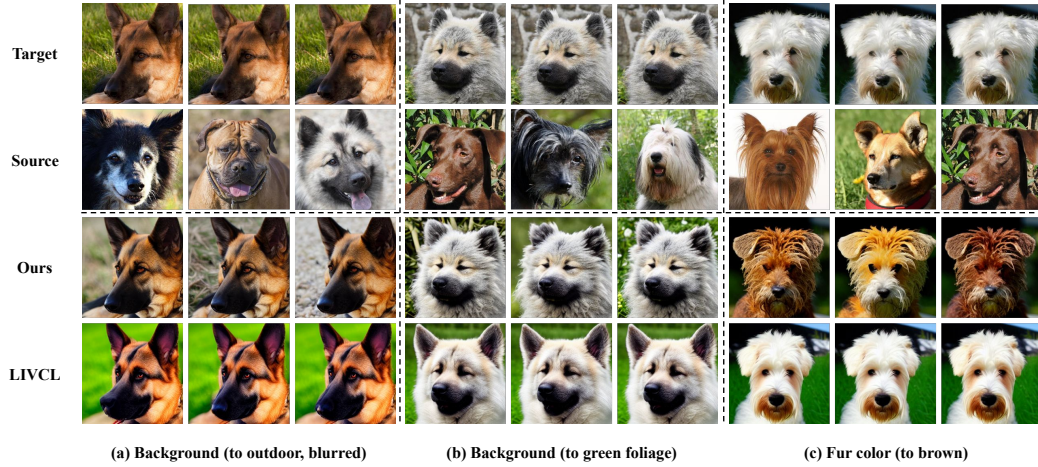


Figure 27: Transferring Visual Nuances from source to target images.

A.7 Computing Resources

All of our experiments are conducted on a GPU Server that consists of an Intel Xeon Gold 6230 CPU, 256GB RAM, and 8 NVIDIA RTX 6000 GPUs (with 48GB VRAM). It takes about 48 GPU hours for each dataset.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [NA].
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes]" is generally preferable to "[No]", it is perfectly acceptable to answer "[No]" provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No]" or "[NA]" is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading “NeurIPS Paper Checklist”,**
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers.**

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [Yes]

Justification: We states our motivation, contributions, scope of our work in abstract and introduction.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We provide limitation of our work in Appendix.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: We do not claim for theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide all the information needed to reproduce the experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No] ,

Justification: Our code is not cleaned and prepared enough for sharing.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [\[Yes\]](#)

Justification: We specify all the details for experimental setting.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [\[No\]](#)

Justification: Since our method requires costly GPU cost and time in training the diffusion model on real images, we were not affordable to conduct and provide repetitive experiments. We will add it in future.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: We provide information of computing resources used for the experiments in Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: We followed the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss it in Appendix.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [No]

Justification: : Our paper possess no risk

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [\[Yes\]](#)

Justification: We cite all the codes, data, paper, and pretrained model in our paper.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [\[NA\]](#)

Justification: We do not release any new assets

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [Yes]

Justification: We provide detailed instructions of our human evaluation and we provide proper rewards to participants through Prolific website.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Our method and human evaluation possess no risk.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: We used VLM for automatic extraction of visual concepts and provide detailed information in the paper.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.