Supplemental information

A	**Additional details about the results**	**15**
B	**Additional experiments**	**15**
	B.1 Importance of components of the motion energy model	15
	B.2 Multi-frame optical flow	16
	B.3 Comparison with state-of-the-art motion segmentation	16
C	**Additional details about the human subject study**	**18**
	C.1 Comparison of humans and machines by example difficulty	18
	C.2 Screenshots of the experiment	19

A   Additional details about the results

For an additional overview, view visualize the segmentation performances on random dot stimuli as reported in Table 1.
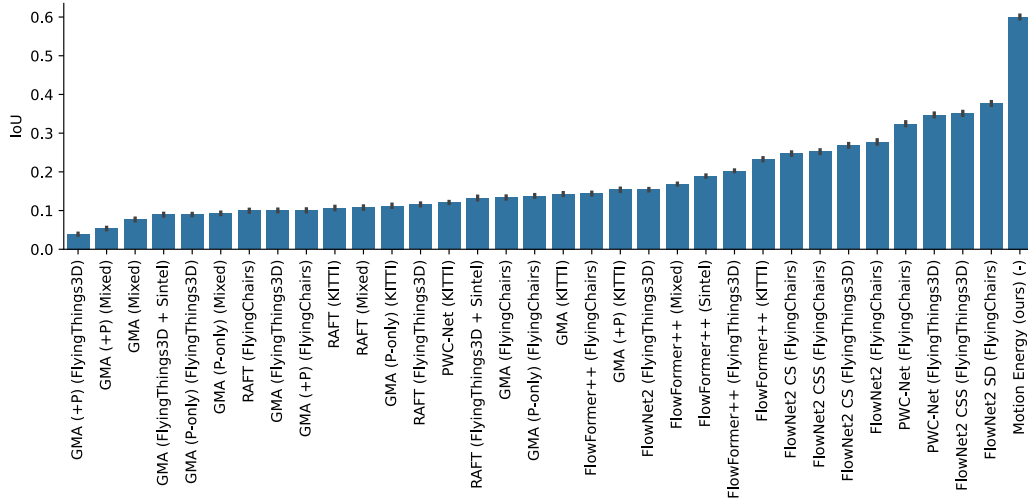
Figure 5: Segmentation performances of the evaluated models on the random dot stimuli. Same data as in Table 1.

B   Additional experiments

B.1   Importance of components of the motion energy model

We conducted an additional ablation study in order to better understand which aspects of the motion energy model are essential for generalization to random dot stimuli. We removed or replaced individual layers as described in Table 3 and trained the ablated models from scratch using in the same way as the baseline model.

The results in Table 2 hint at the normalization and pooling layers being important for generalization. When the Gaussian pooling layers are removed completely, the performance on original videos even slightly improves while the generalization to random dot stimuli is substantially reduced.

15

Replacing the squaring-based nonlinear layers with ReLU layers, however, hardly changes the model's performance.

Condition	Original		Random Dots	
	IoU ↑	F-Score ↑	IoU ↑	F-Score ↑
Baseline	0.759	0.845	0.600	0.718
Replace RectifiedSquare → ReLU (MT)	0.753	0.838	**0.609**	**0.725**
Replace Square → ReLU (V1)	0.770	0.854	0.536	0.663
Remove MT Linear	0.768	0.856	0.481	0.609
Remove MT	0.770	0.854	0.451	0.583
Remove Blur (V1, MT)	**0.801**	**0.872**	0.421	0.540
Replace ChannelNorm → InstanceNorm (V1, MT)	0.592	0.703	0.230	0.340
Remove Normalization (V1, MT)	0.400	0.516	0.018	0.018

Table 3: Ablation study: Performance of the model on original videos and corresponding random dot stimuli with various layers of the motion energy model removed or replaced. Results are ordered by IoU on the random dot stimuli.

B.2 Multi-frame optical flow

The motion energy model uses a window of 9 frames as input, while typical optical flow methods estimate correspondences between only two frames. To rule out the possibility that the results observed in our paper are mainly explained by the different input window lengths, we perform an ablation study in which we apply optical flow methods using the same 9 frame windows. For each window, we compute the optical flow between the central frame, for which the segmentation has to be predicted, to the 8 other frames in the window. The stacked optical flow fields are then used as the input to the segmentation network.

The results in Table 4 and Figure 6 show some improvement on the original videos but an ever wider gap to the motion energy model in terms of of generalization to random dots. The differences between the motion energy and optical flow models therefore cannot be explained by the different input lengths.
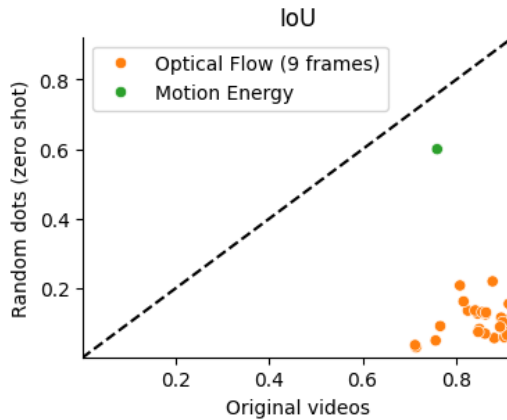
Figure 6: Performance of multi-frame optical flow based models on the original videos and corresponding random dot videos.

B.3 Comparison with state-of-the-art motion segmentation

In our study we used a relatively small segmentation network downstream to the respective motion estimator. State-of-the-art motion segmentation models typically target multi-object segmentation in real world videos and therefore use more complex segmentation networks. In order to verify that the results in our paper are not caused by using a smaller segmentation network, we evaluated the state

16

Motion Estimator	Training Dataset	Original		Random Dots	
		IoU	F-Score	IoU	F-Score
Motion Energy (ours)	-	0.759	0.845	**0.600**	**0.718**
FlowNet2 SD	FlyingChairs	0.878	0.928	0.221	0.325
FlowNet2	FlyingChairs	0.808	0.868	0.209	0.300
	FlyingThings3D	0.881	0.929	0.058	0.100
PWC-Net	FlyingChairs	0.816	0.886	0.163	0.250
	FlyingThings3D	0.825	0.886	0.137	0.221
	KITTI	0.712	0.811	0.038	0.060
RAFT	FlyingThings3D + Sintel	**0.912**	**0.948**	0.156	0.222
	FlyingChairs	0.863	0.914	0.126	0.195
	Mixed	0.896	0.934	0.117	0.164
	FlyingThings3D	0.894	0.934	0.090	0.132
	KITTI	0.714	0.794	0.031	0.053
FlowNet2 CS	FlyingChairs	0.841	0.899	0.137	0.220
	FlyingThings3D	0.847	0.904	0.075	0.129
GMA (+P)	FlyingChairs	0.856	0.912	0.132	0.212
	Mixed	0.900	0.936	0.114	0.179
	FlyingThings3D	0.899	0.936	0.104	0.171
GMA	FlyingChairs	0.864	0.917	0.131	0.212
	Mixed	0.900	0.937	0.090	0.139
	FlyingThings3D + Sintel	0.909	0.943	0.066	0.100
	FlyingThings3D	0.903	0.943	0.060	0.098
	KITTI	0.756	0.834	0.051	0.084
GMA (P-only)	FlyingChairs	0.846	0.901	0.128	0.207
	KITTI	0.766	0.847	0.092	0.155
	FlyingThings3D	0.903	0.940	0.083	0.139
	Mixed	**0.912**	0.947	0.077	0.117
FlowNet2 CSS	FlyingChairs	0.850	0.908	0.084	0.141
	FlyingThings3D	0.862	0.918	0.070	0.121

Table 4: Ablation study: We apply the optical flow estimators to a window of 9 frames by using the central frame as references and computing optical flow to each of the 8 other frames. The stacked optical flow fields are used as inpute for the segmentation network.

of the art OCLR model [51] in our setting. The OCLR model uses optical flow estimated by RAFT [43], which we also included in our experiments. The segmentation network however uses a U-Net architecture with Transformer bottleneck and was trained to segment multiple objects on a synthetic dataset. We use the published weights and do not retrain the model on our data.

The results in Table 5 show that the model performs very well on the original data. OCLR outperforms our motion energy based model and achieves a performance similar to the best optical flow based models considered in this work. At the same time, the model does not generalize to the corresponding random dot stimuli. These results provide further evidence that the low generalization to random dots is not due to the architecture of the segmentation network or the RGB training data, but a property of the motion estimator.

Model	IoU (original)	IoU (random dots)
OCLR	0.838	0.026
Motion Energy Segmentation	0.759	0.600

Table 5: Comparison of the state-of-the-art motion segmentation model OCLR, and our segmentation model based on a motion energy model.

17

C Additional details about the human subject study

C.1 Comparison of humans and machines by example difficulty

As a measure of task difficulty, we count the number of *informative dots*. A dot is informative, if it is contained in either the target and distractor shape but not both (see Figure 7, left). Only these dots allow discriminating between the different shapes.

We fitted psychometric curves for human participants and models as a function of the number of informative dots, using the psignifit toolbox [35]. The results in Figure 7 confirm that only the motion energy model is able to match the performance of human subjects, especially for stimuli with a medium number of informative dots.
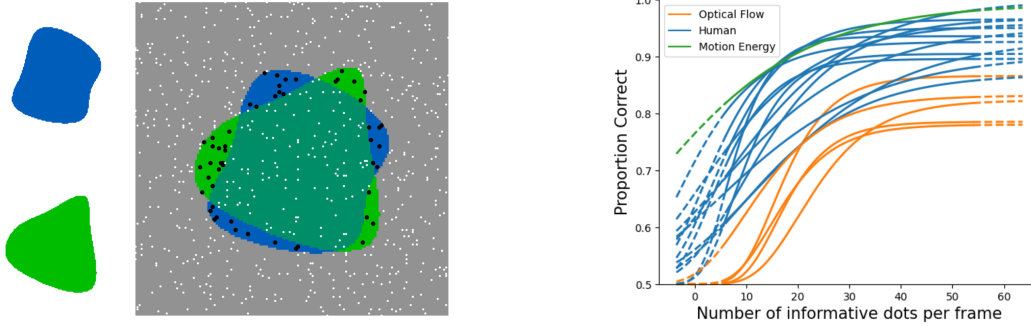
Figure 7: (*left*) As a measure of task difficulty, we count the number of informative dots that allow discriminating between the two shape alternatives. (*right*) Psychometric curves for humans, the motion energy based model and the four best optical flow models for the task as in 8.
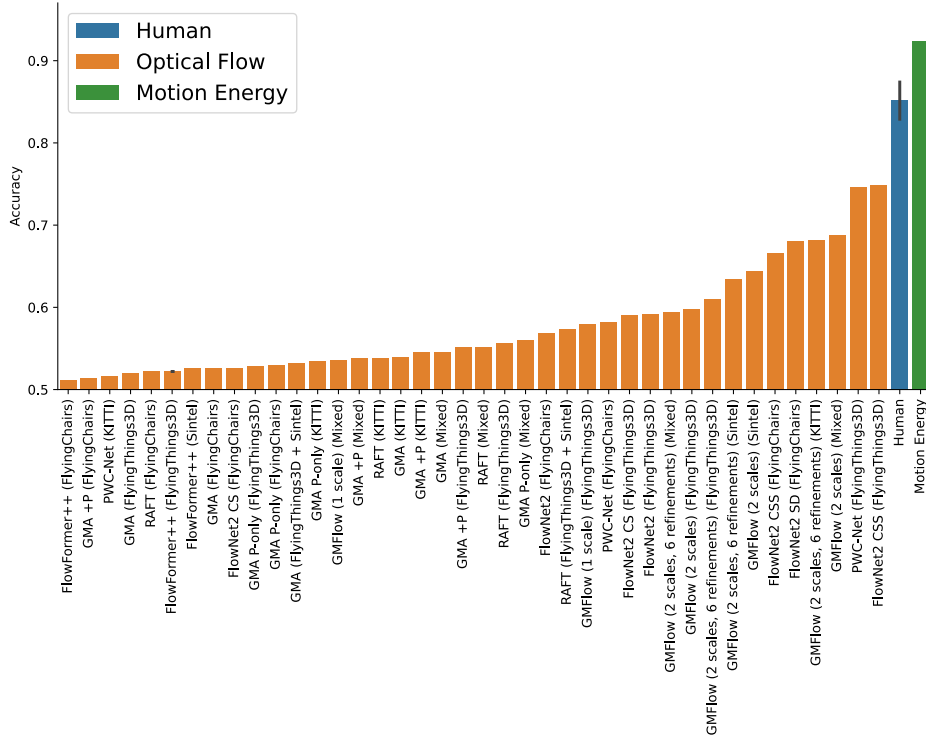
Figure 8: Comparison of the human and model performances for the random dot shape matching task.

18

C.2 Screenshots of the experiment
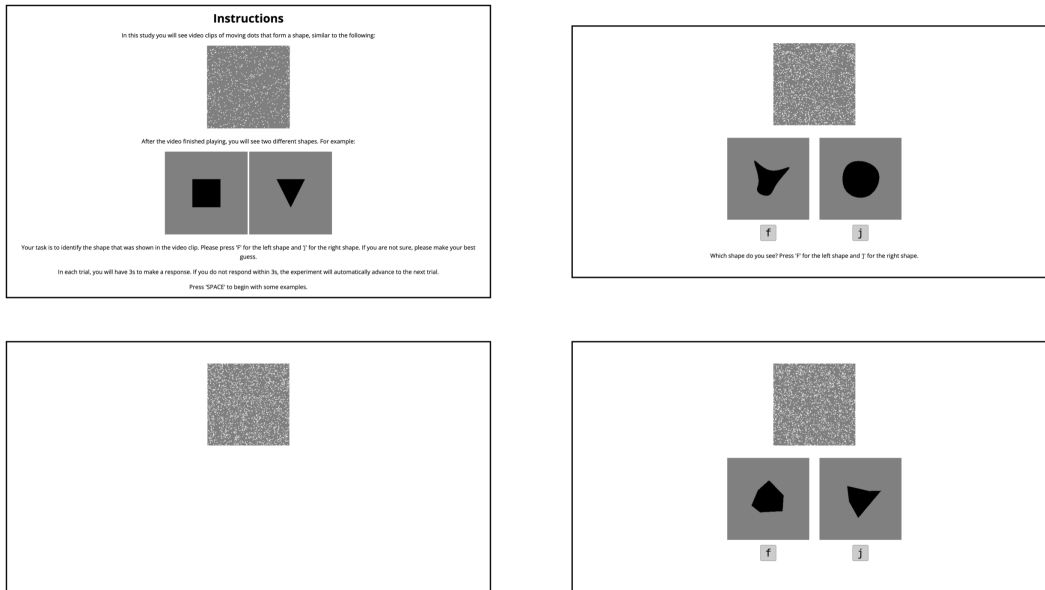
Figure 9: Screenshots from the human subject study on random dot shape identification. (*top left*) Instructions that were shown prior to the experiment. (*top right*) We showed 20 training trials during which subjects could familiarize themselves with the task. (*bottom left*) The training was followed by 500 test trials. A video with the random dot stimuli was shown first. (*bottom right*) Once the video finished playing, the two shape options were shown below.

19

NeurIPS Paper Checklist

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 clearly state the main contributions of our work in both the abstract and the introduction. All mentioned results are supported by the experimental data presented in the paper.

   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: The paper includes a separate section that discusses the limitations of our work in detail, including limitations due to computational constraints.

   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?

20

Answer: [NA]

Justification: The paper does not include 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 describe the models, data and evaluation protocol used in the paper in detail. Additionally, the code, pretrained models and the contributed dataset are publicly released.

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**

21

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: [Yes]

Justification: All code and data for the paper is publicly released.

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, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Hyperparameters and training details are explicitly reported with the description of the models.

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: [Yes]

Justification: Due to space constraints we did not include further statistical information in the main table. However we included a Figure showing the same data with error bars in the supplemental material.

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.

22

- 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 reported the computational resources of our models with the description of the training details.

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 have read the code of ethics and conform to it in every respect.

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 are discussing potential broader impacts of our work in a dedicated section.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

23

- 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: [NA]

Justification: This paper does not pose such risks.

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 used the Kubric generator with the built-in asset library and a range of pretrained models. We cited the sources of all data and implementations that we used.

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.

24

- 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: [Yes]

Justification: The synthetic data we generated for the paper is described in the paper, and published with the code used to generate it.

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: Screenshots of the experiment are included in the supplemental material.

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: [Yes]

Justification: The study in this paper does not pose any particular risk on participants. IRB approval exists.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

25

- 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.

26