

# Machaverse-XS: A Compact Universe of Masked 3D Objects

Jun-Pu (Yi) Zhang

Project Page: <https://github.com/Jun-Pu/Machaverse-XS>

We propose a small-scale yet fine-grained dataset, namely ***Machaverse-XS***, for photorealistic 3D object generation from a complete (or a defective) single view. We illustrate the details of the proposed dataset (Section 1) and present a preliminary study (Section 2) of our *Machaverse-XS* towards the task of Image-to-3D generation (I23).

## 1. DataSet

We collected 154 clothing accessories from OmniObject3D [3], and masked varying percentages (*i.e.*, 25%, 50% and 75%) of their 3D meshes. Notably, we applied three masking strategies to each of the 3D mesh-based geometries, thus gaining nine 3D meshes corresponding to each 3D asset (Figure 1).

We then generated orbital multi-view sequences corresponding to each of the masked 3D assets. The visualizations of a single key frame of each sequential rendering are shown below. Specifically, there are 22 “Backpack” (Figure 2, 3, 4, 5, 6), 9 “Bumbag” (Figure 7, 8), 48 “Handbag” (Figure 9, 10, 11, 12, 13, 14, 15, 16, 17, 18), 9 “Suitcase” (Figure 19, 20), 29 “Hat” (Figure 21, 22, 23, 24, 25, 26) and 37 “Shoe” (Figure 27, 28, 29, 30, 31, 32, 33, 34) videos.

To facilitate future researches in I23 and 3D object completion [2] from a single defective (or partial) view, we also provide an official split of the proposed *Machaverse-XS* (please refer to our project page for the details).

## 2. A Baseline Model towards I23

To verify the feasibility of the proposed dataset upon the task of I23, we applied SV3D [1] as a baseline model and trained it with a *Machaverse-XS* subset and conducted qualitative experiments with the residual data (Figure 35). Please refer to our project page for more experimental results and a detailed tutorial to conduct I23 with our *Machaverse-XS*.

## 3. Conclusion&Future Work

We propose *Machaverse-XS* for challenging 3D vision tasks such as I23 and 3D object completion from a single partial view. The proposed dataset proved its feasibility towards the task of I23. Future works can take advantage of our provided fine-grained masked 3D meshes to advance 3D object completion from sparse views. The dataset is now made open-sourced to support researches in challenging computer vision tasks, such as object-centric real-world 3D inpainting, and to support XR-based applications towards E-commercial merchandising scenario.

## References

- [1] Vikram Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz, Dmitry Tochilkin, Christian Laforte, Robin Rombach, and Varun Jampani. Sv3d: Novel multi-view synthesis and 3d generation from a single image using latent video diffusion. In *European Conference on Computer Vision (ECCV)*, pages 439–457. Springer, 2024. 1
- [2] Ethan Weber, Aleksander Holynski, Varun Jampani, Saurabh Saxena, Noah Snavely, Abhishek Kar, and Angjoo Kanazawa. Ner-filler: Completing scenes via generative 3d inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20731–20741, 2024. 1
- [3] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Jiawei Ren, Liang Pan, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, et al. Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 803–814, 2023. 1

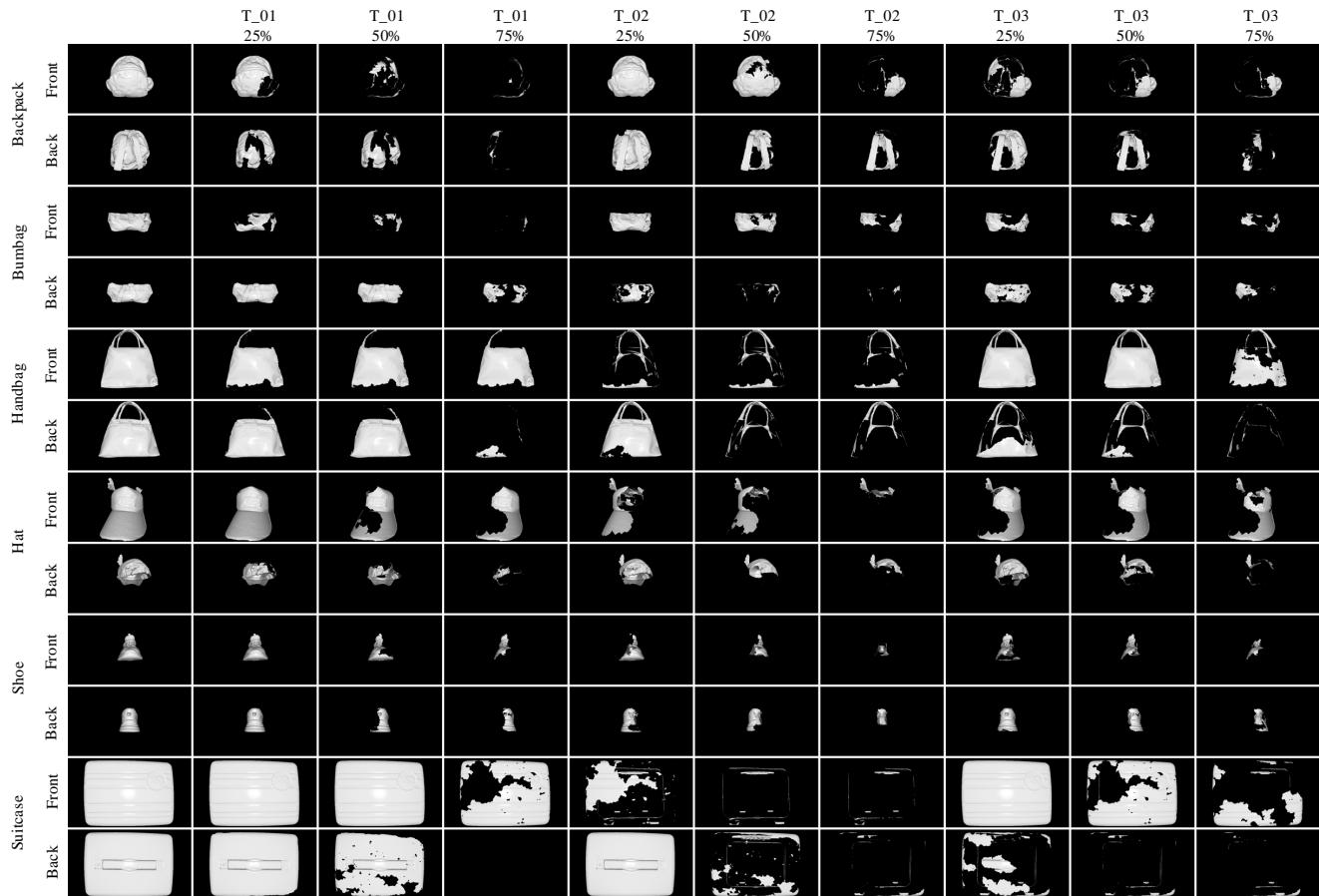


Figure 1. An example that illustrates the three masking strategies.  $T\_0X$  denotes the  $X$  masking strategy.

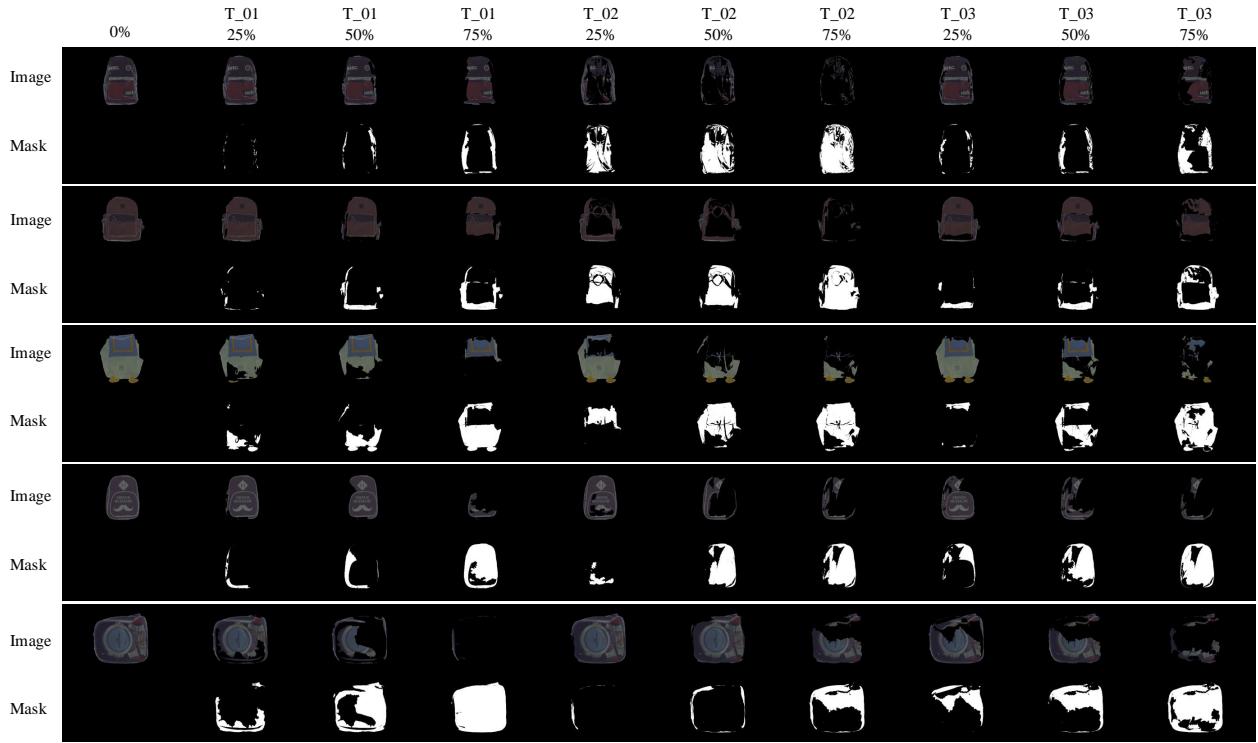


Figure 2. Key frame of each of the **Machaverse-XS** “Backpack” sequences (1/5).  $T_{0X}$  means the  $X$  type of masking strategy.

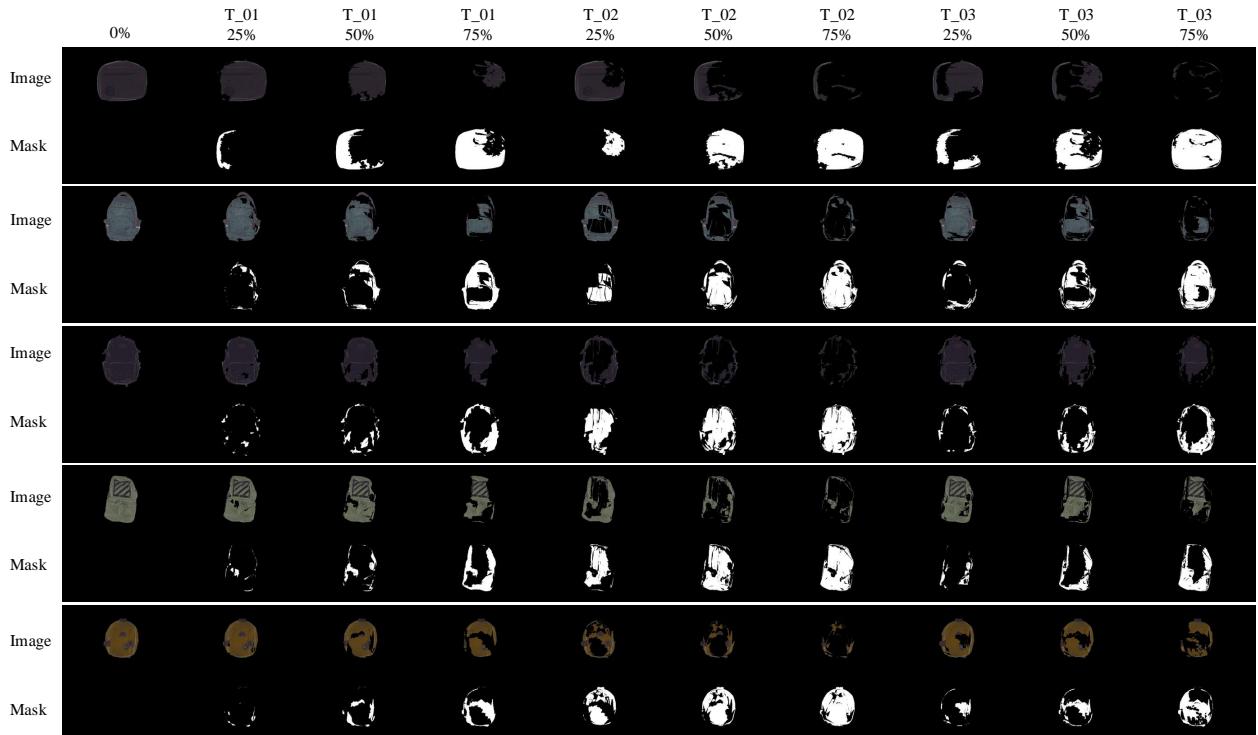


Figure 3. Key frame of each of the **Machaverse-XS** “Backpack” sequences (2/5).  $T_{0X}$  means the  $X$  type of masking strategy.

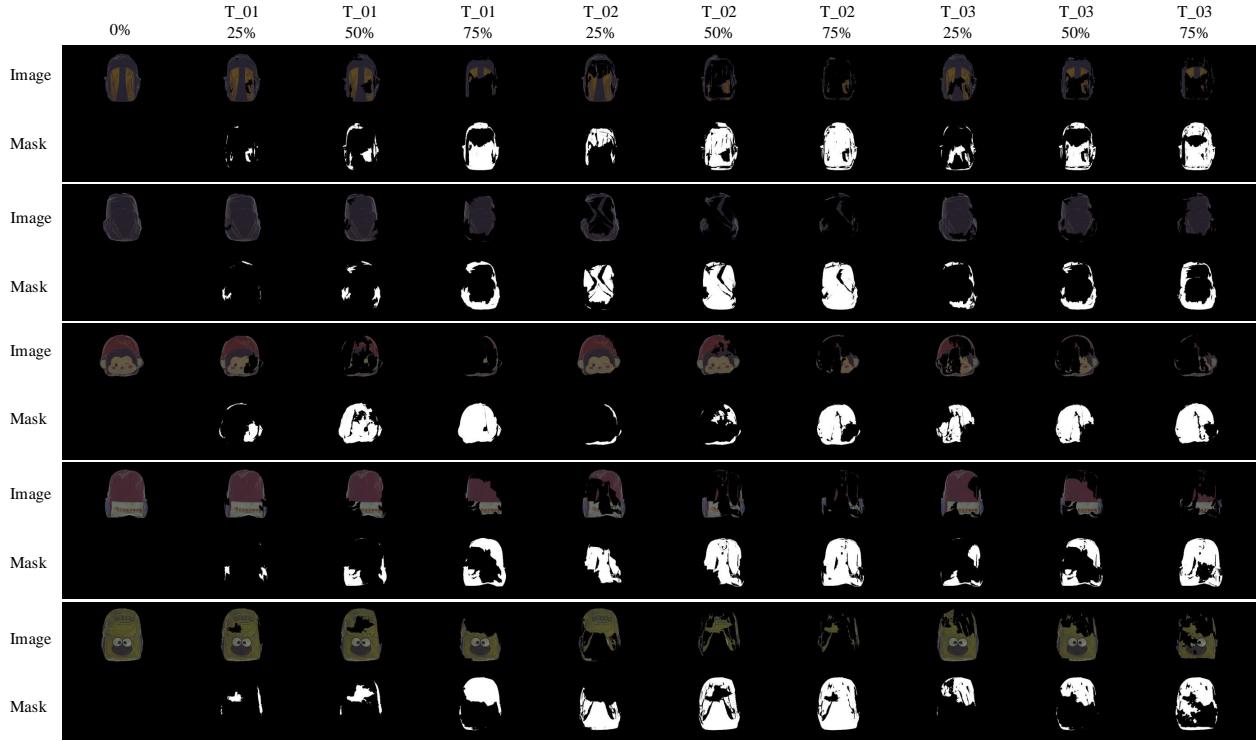


Figure 4. Key frame of each of the **Machaverse-XS** “Backpack” sequences (3/5).  $T_{0X}$  means the  $X$  type of masking strategy.

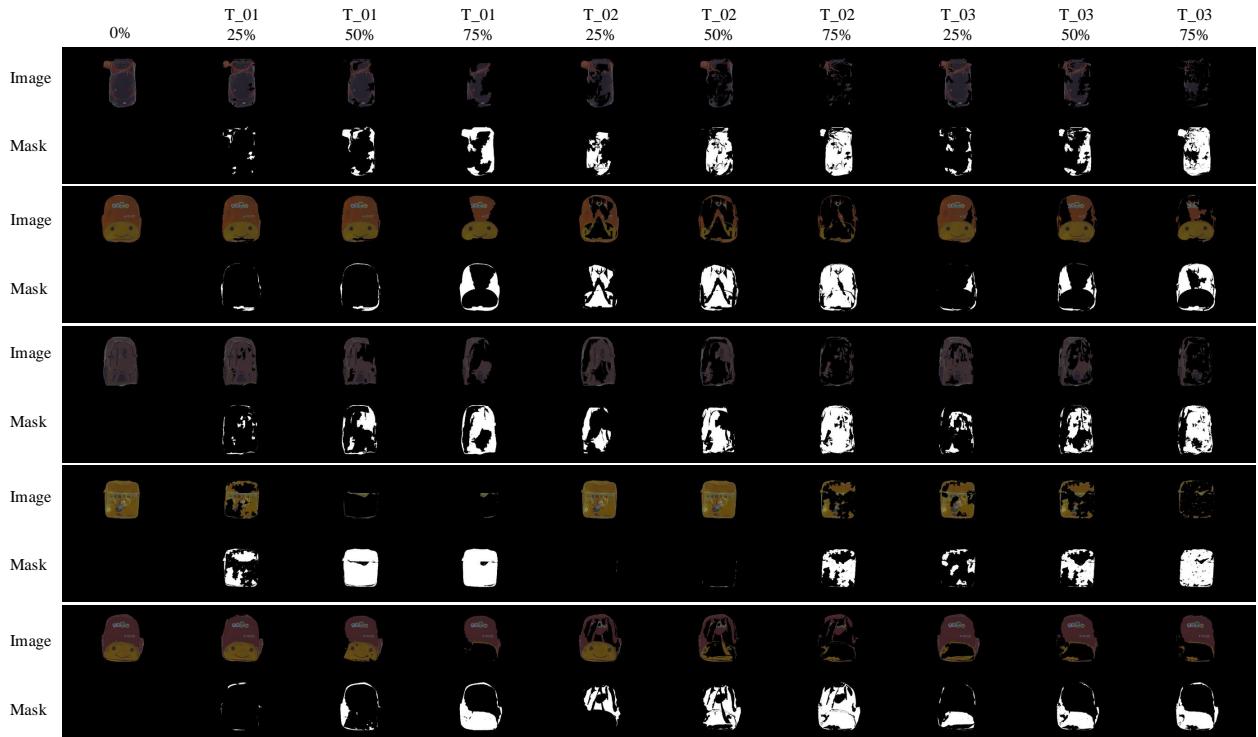


Figure 5. Key frame of each of the **Machaverse-XS** “Backpack” sequences (4/5).  $T_{0X}$  means the  $X$  type of masking strategy.

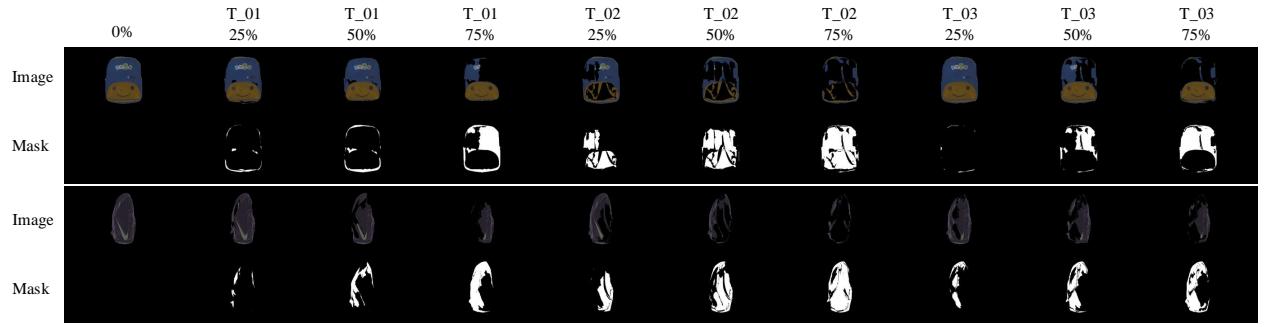


Figure 6. Key frame of each of the **Machaverse-XS** “Backpack” sequences (5/5).  $T\_0X$  means the  $X$  type of masking strategy.

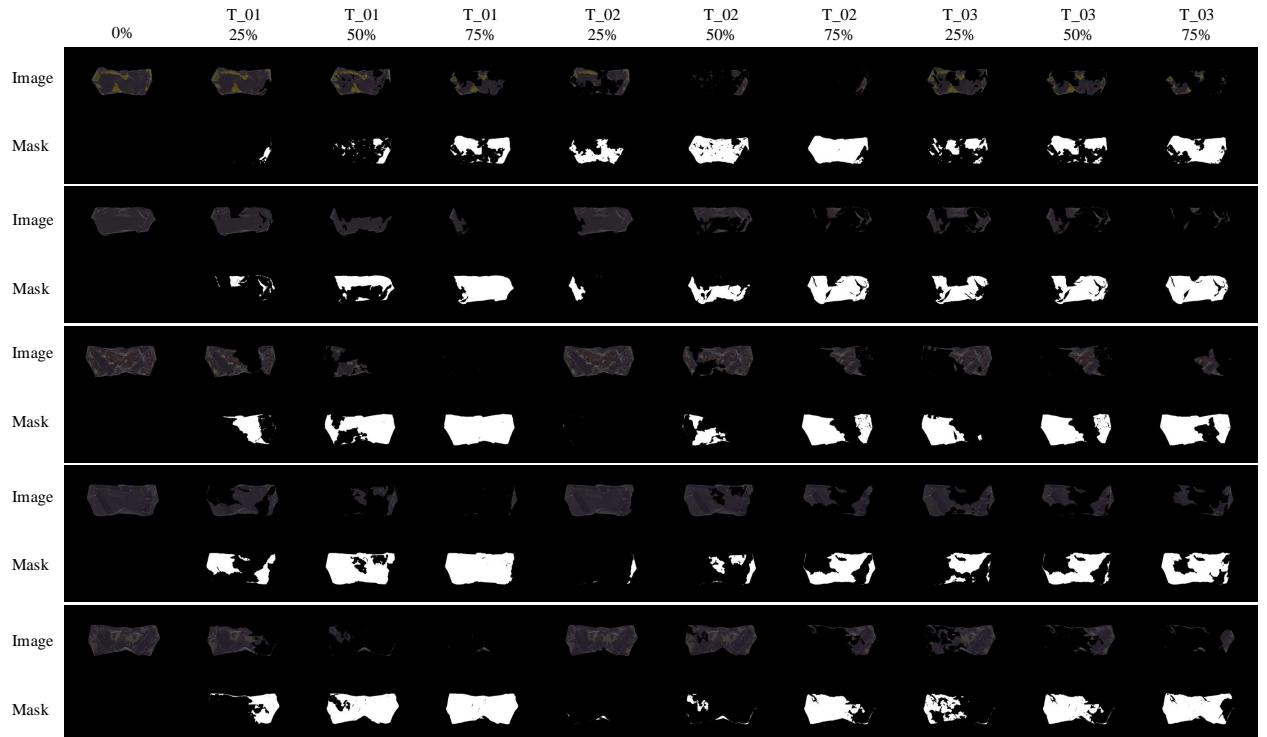


Figure 7. Key frame of each of the **Machaverse-XS** “Bumbag” sequences (1/2).  $T\_0X$  means the  $X$  type of masking strategy.

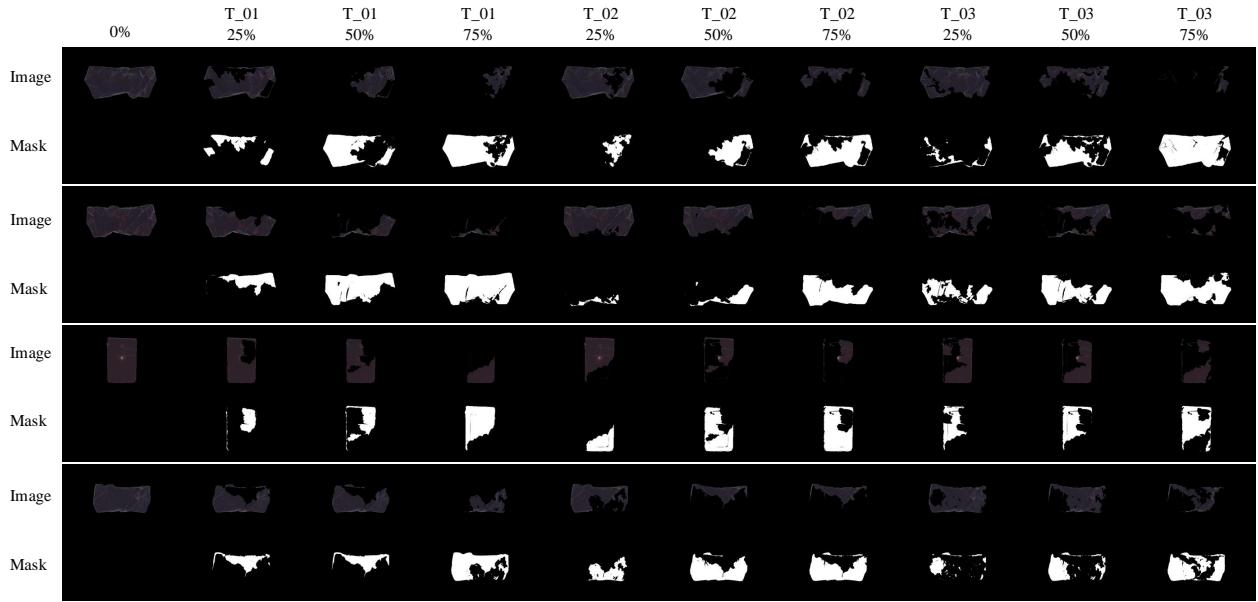


Figure 8. Key frame of each of the **Machaverse-XS** “Bumbag” sequences (2/2).  $T_{.0}X$  means the  $X$  type of masking strategy.

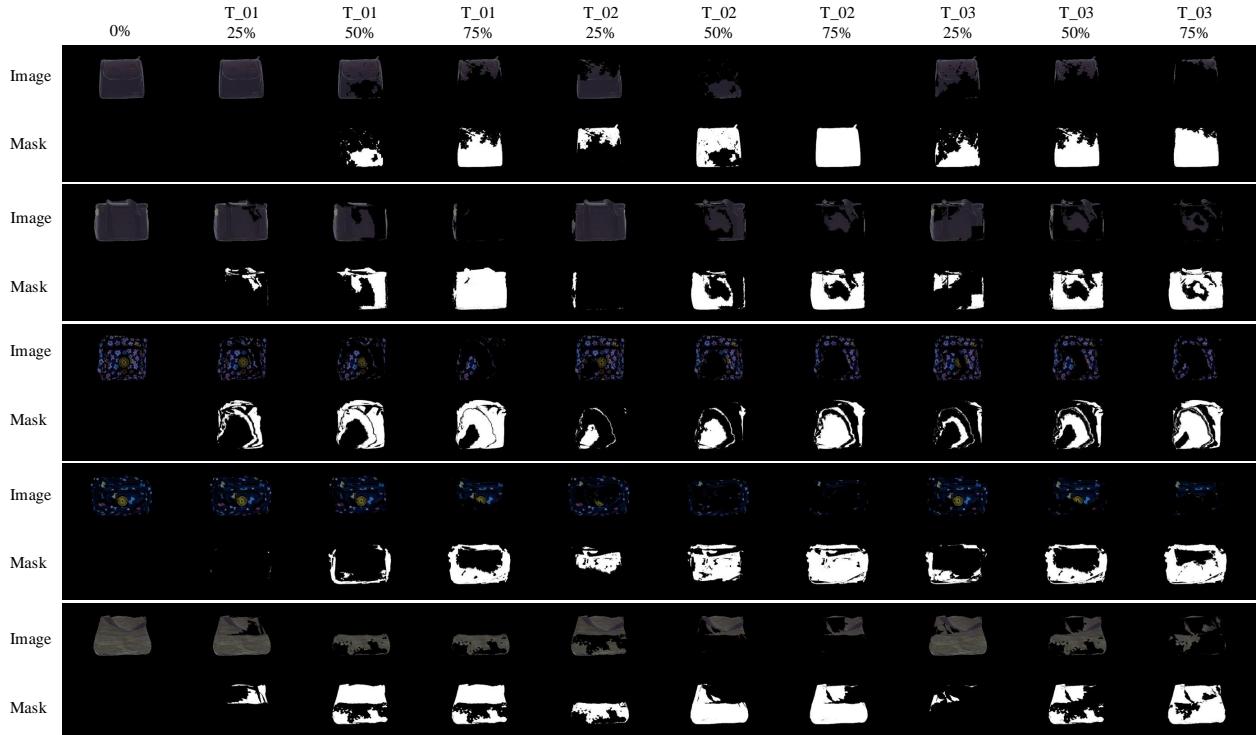


Figure 9. Key frame of each of the **Machaverse-XS** “Handbag” sequences (1/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

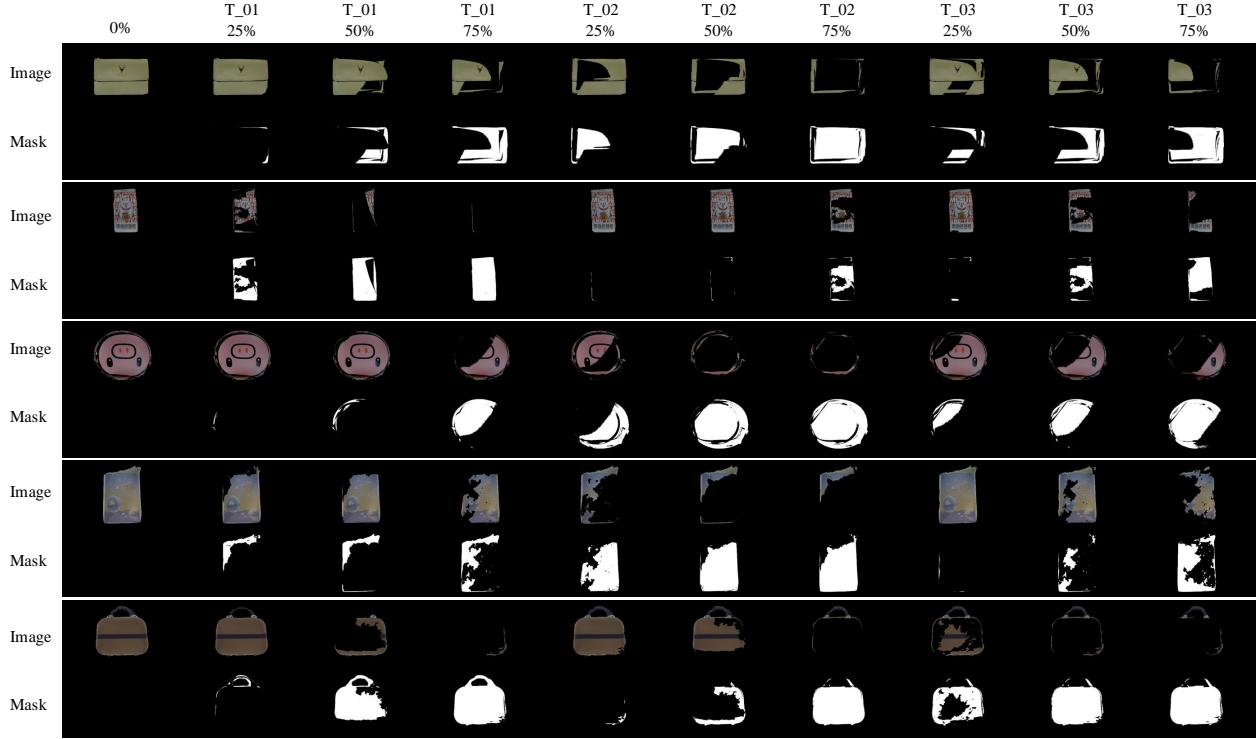


Figure 10. Key frame of each of the **Machaverse-XS** “Handbag” sequences (2/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

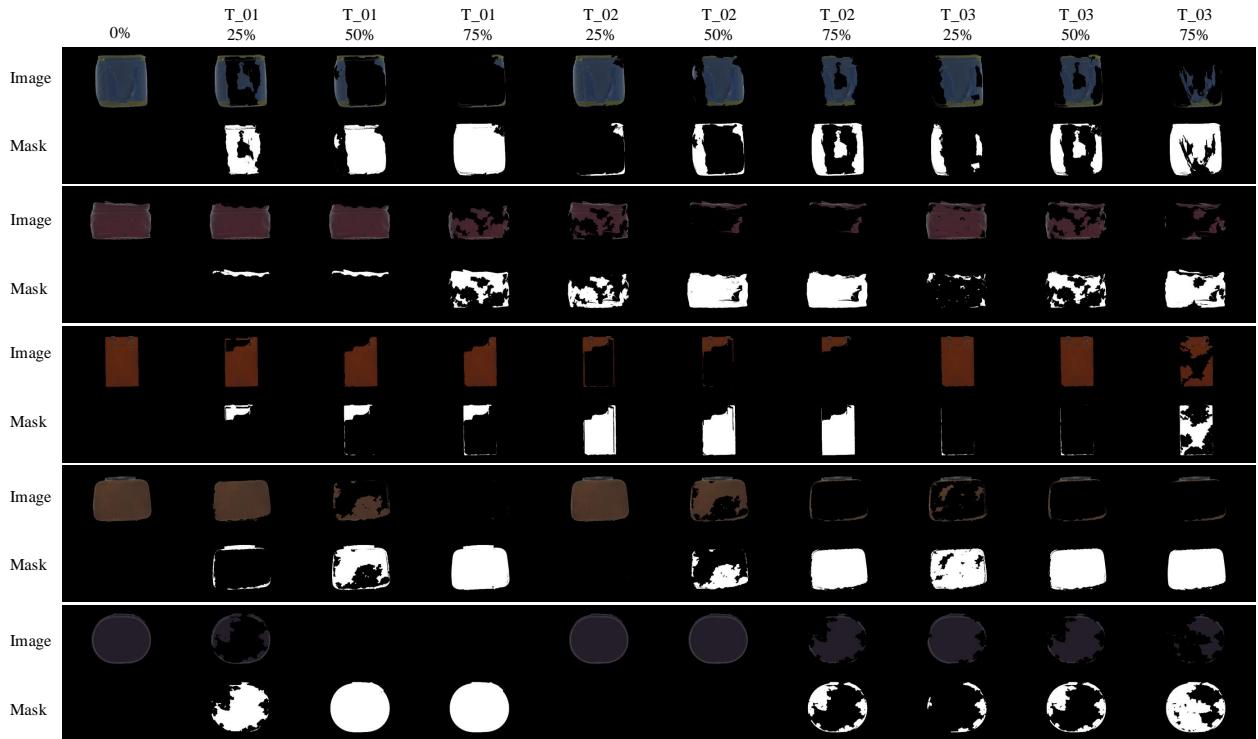


Figure 11. Key frame of each of the **Machaverse-XS** “Handbag” sequences (3/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

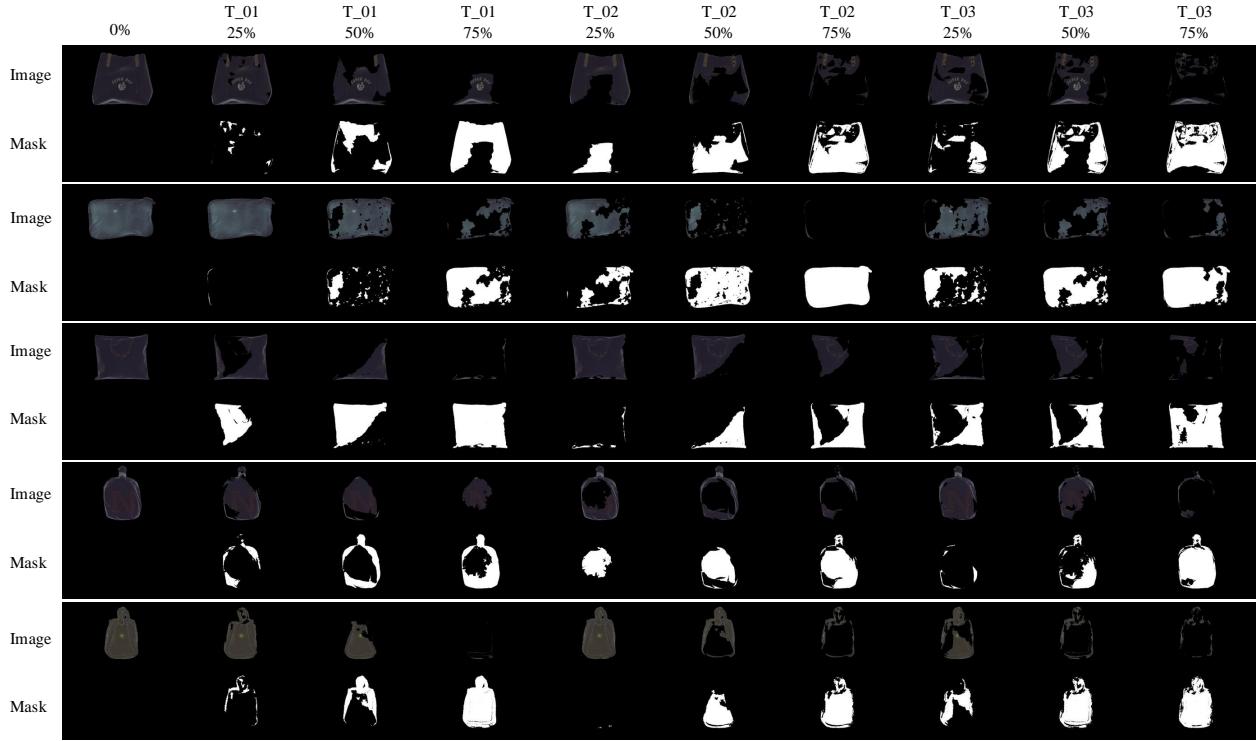


Figure 12. Key frame of each of the **Machaverse-XS** “Handbag” sequences (4/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

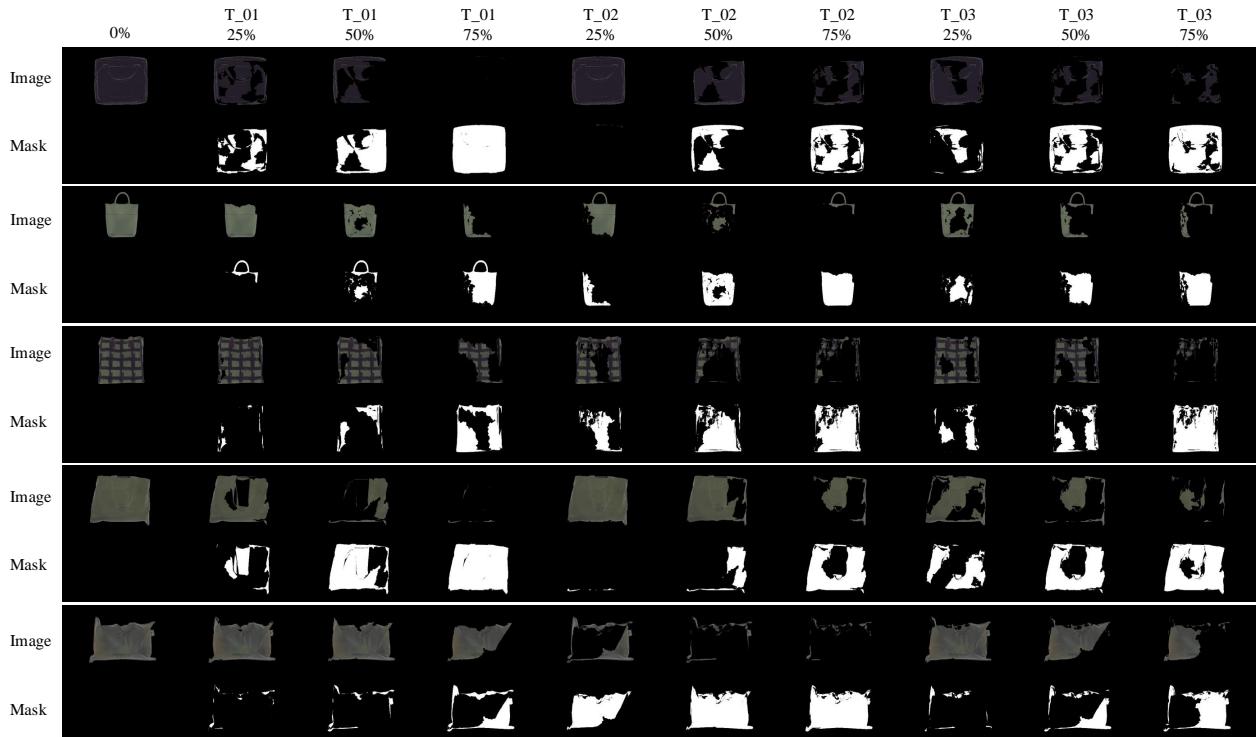


Figure 13. Key frame of each of the **Machaverse-XS** “Handbag” sequences (5/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

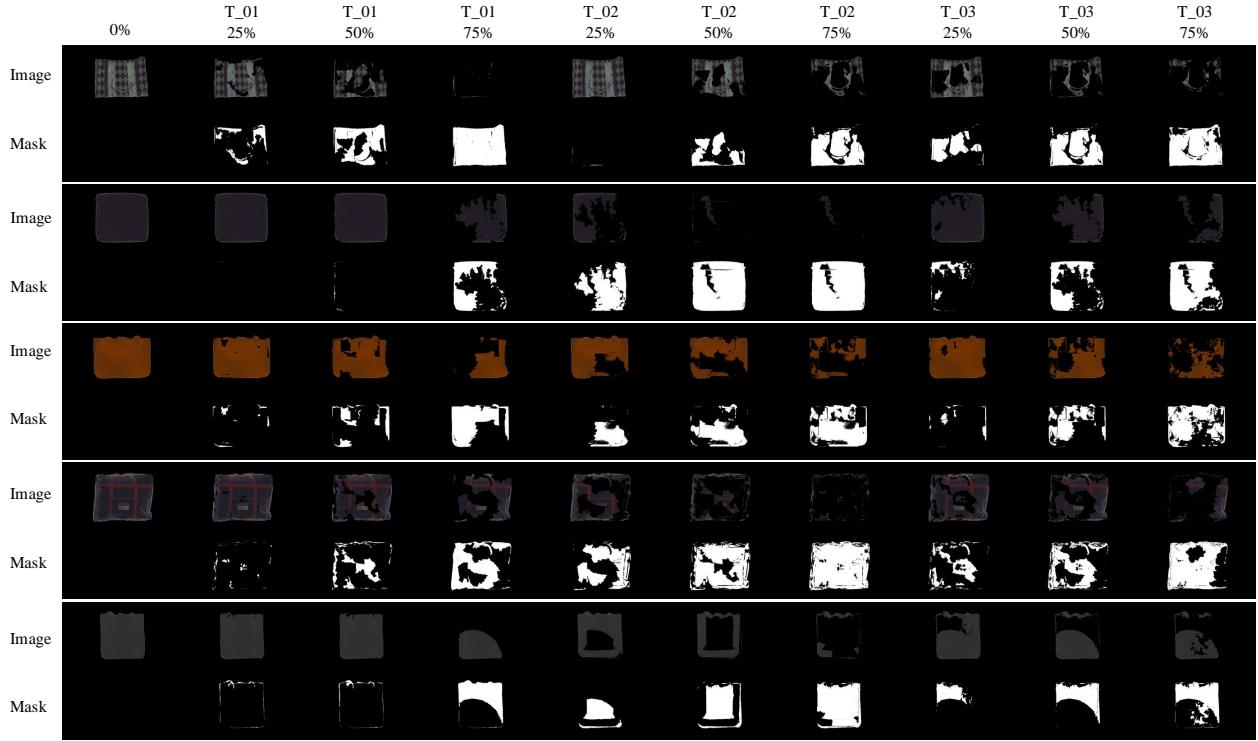


Figure 14. Key frame of each of the **Machaverse-XS** “Handbag” sequences (6/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

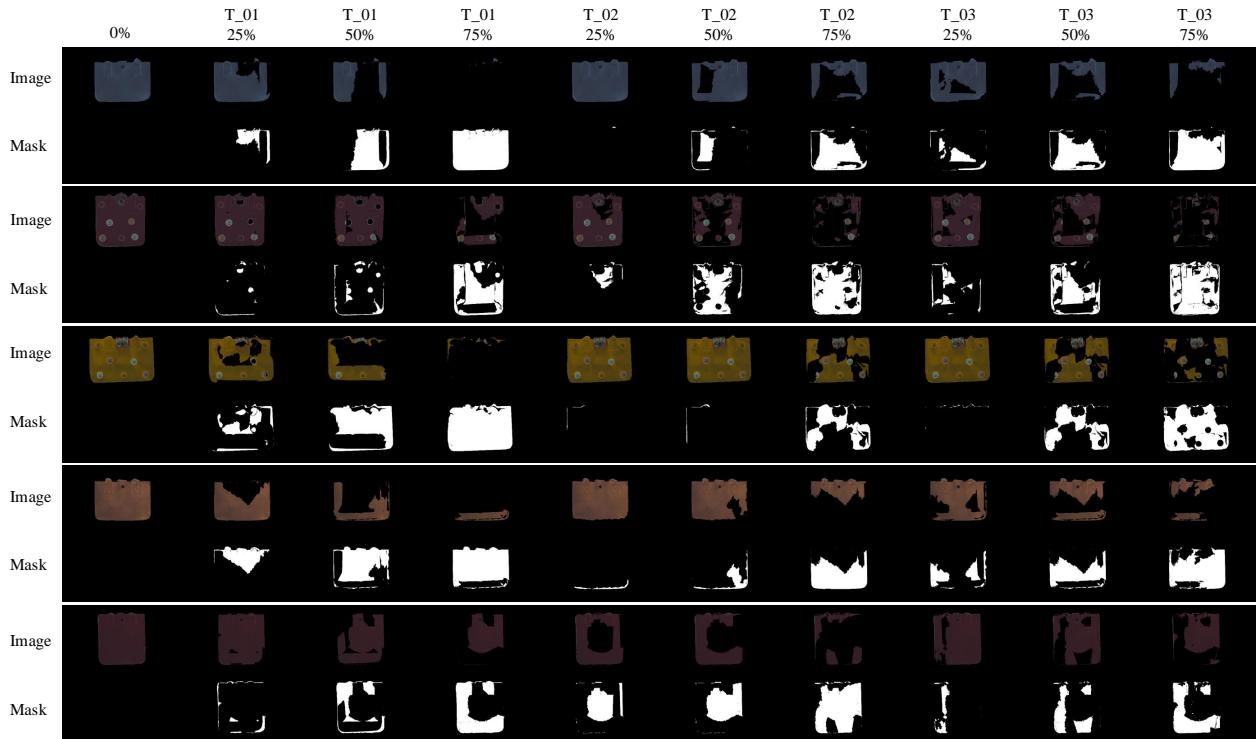


Figure 15. Key frame of each of the **Machaverse-XS** “Handbag” sequences (7/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

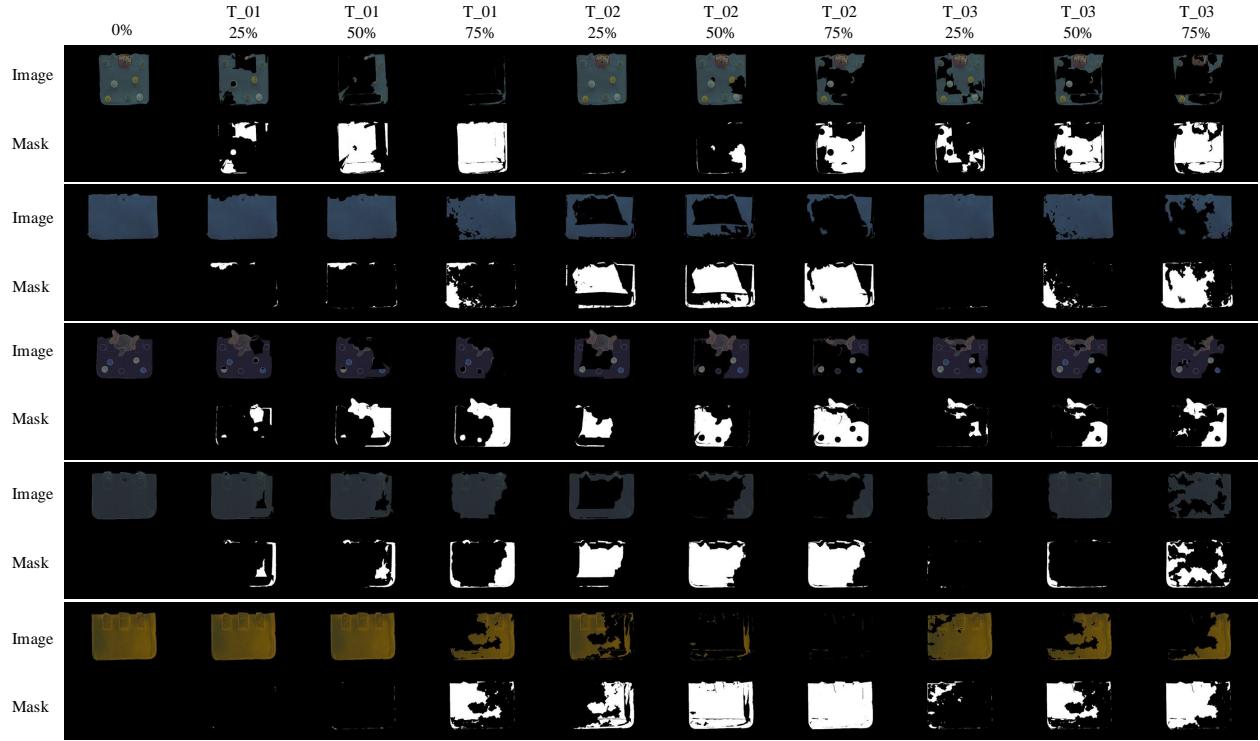


Figure 16. Key frame of each of the **Machaverse-XS** “Handbag” sequences (8/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

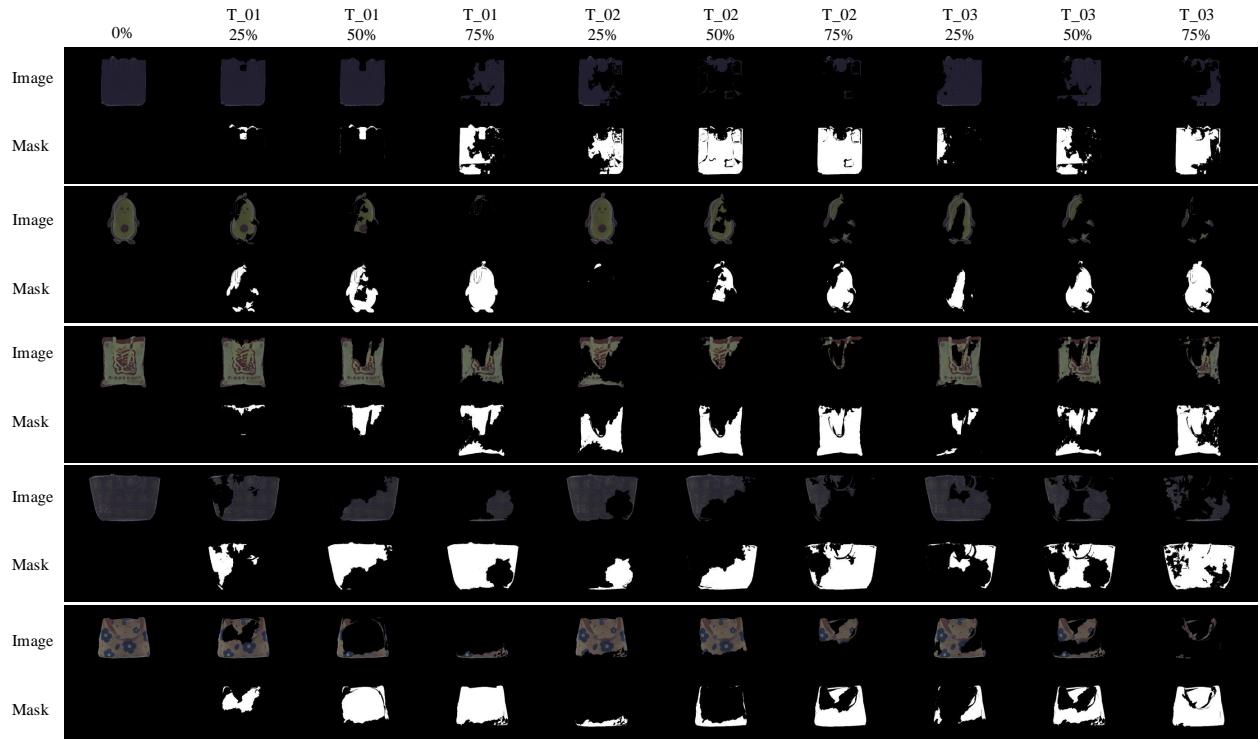


Figure 17. Key frame of each of the **Machaverse-XS** “Handbag” sequences (9/10).  $T_{.0}X$  means the  $X$  type of masking strategy.

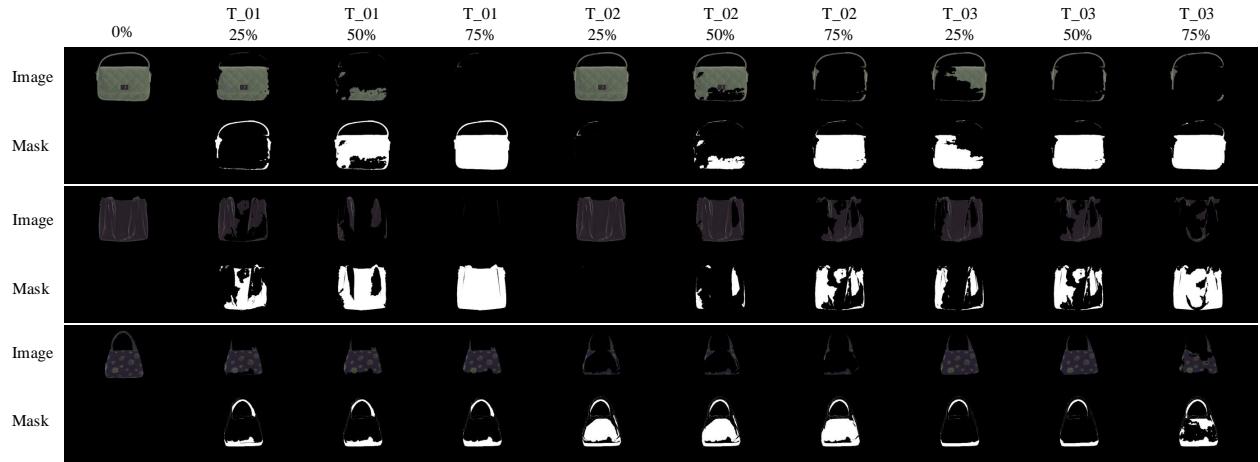


Figure 18. Key frame of each of the **Machaverse-XS** “Handbag” sequences (10/10).  $T_{0X}$  means the  $X$  type of masking strategy.

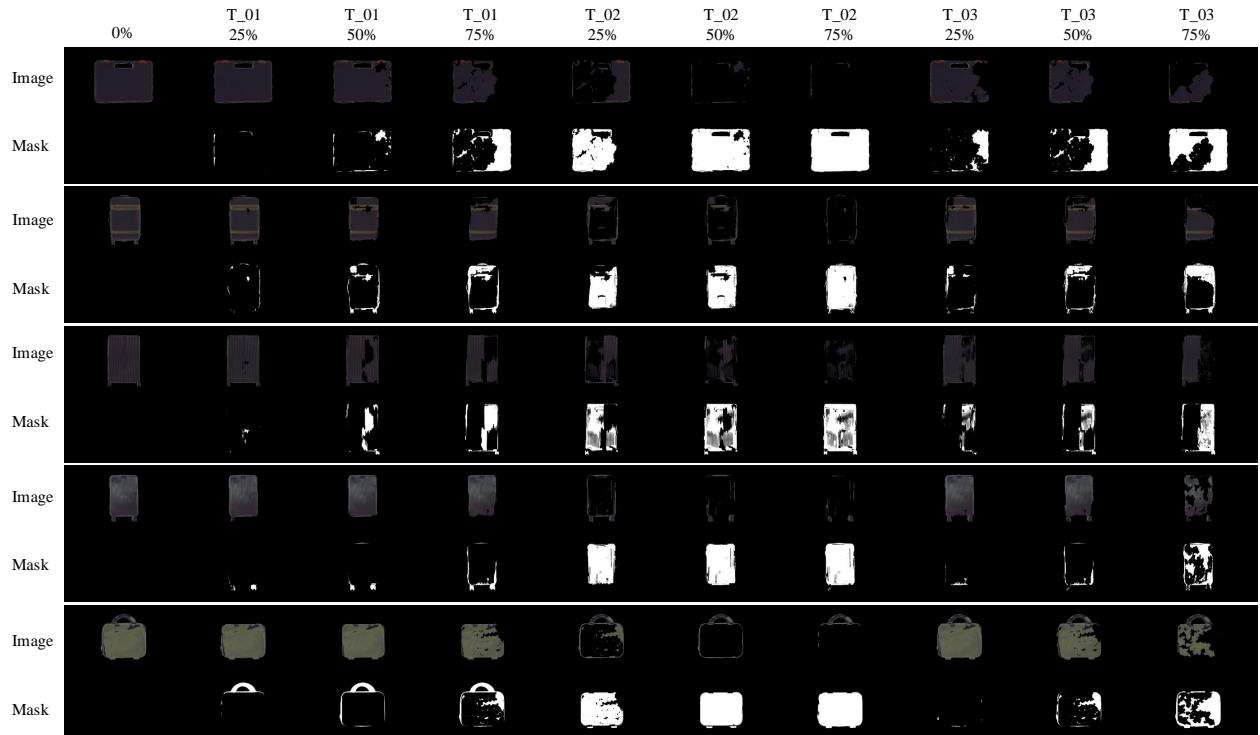


Figure 19. Key frame of each of the **Machaverse-XS** “Suitcase” sequences (1/2).  $T_{0X}$  means the  $X$  type of masking strategy.

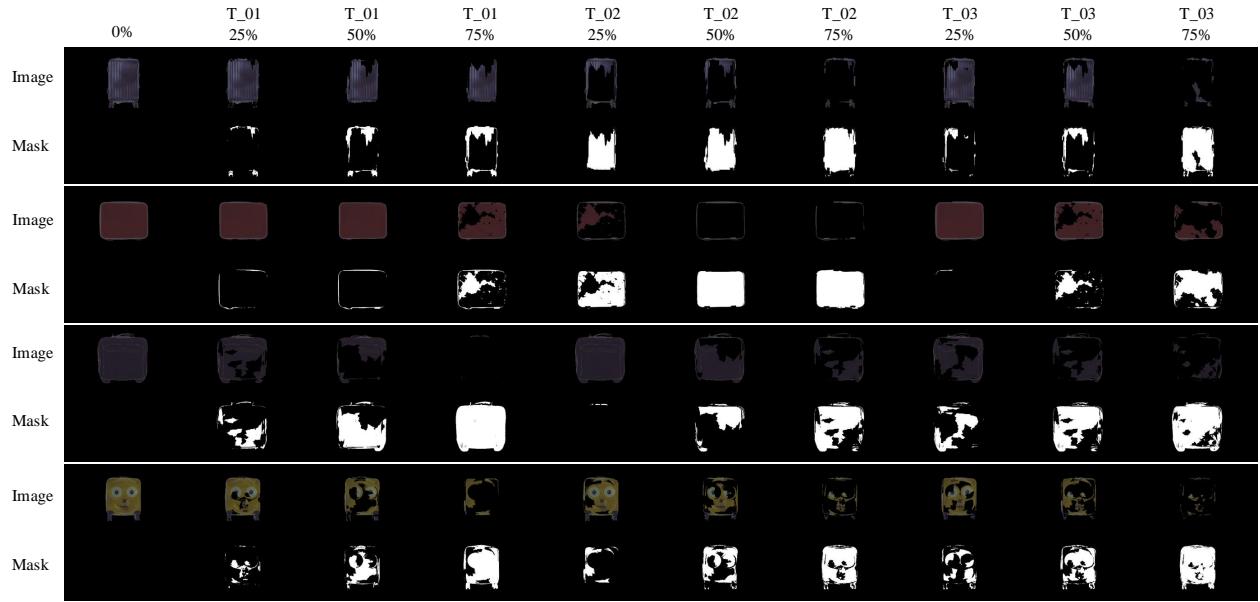


Figure 20. Key frame of each of the **Machaverse-XS** “Suitcase” sequences (2/2).  $T\_0X$  means the  $X$  type of masking strategy.

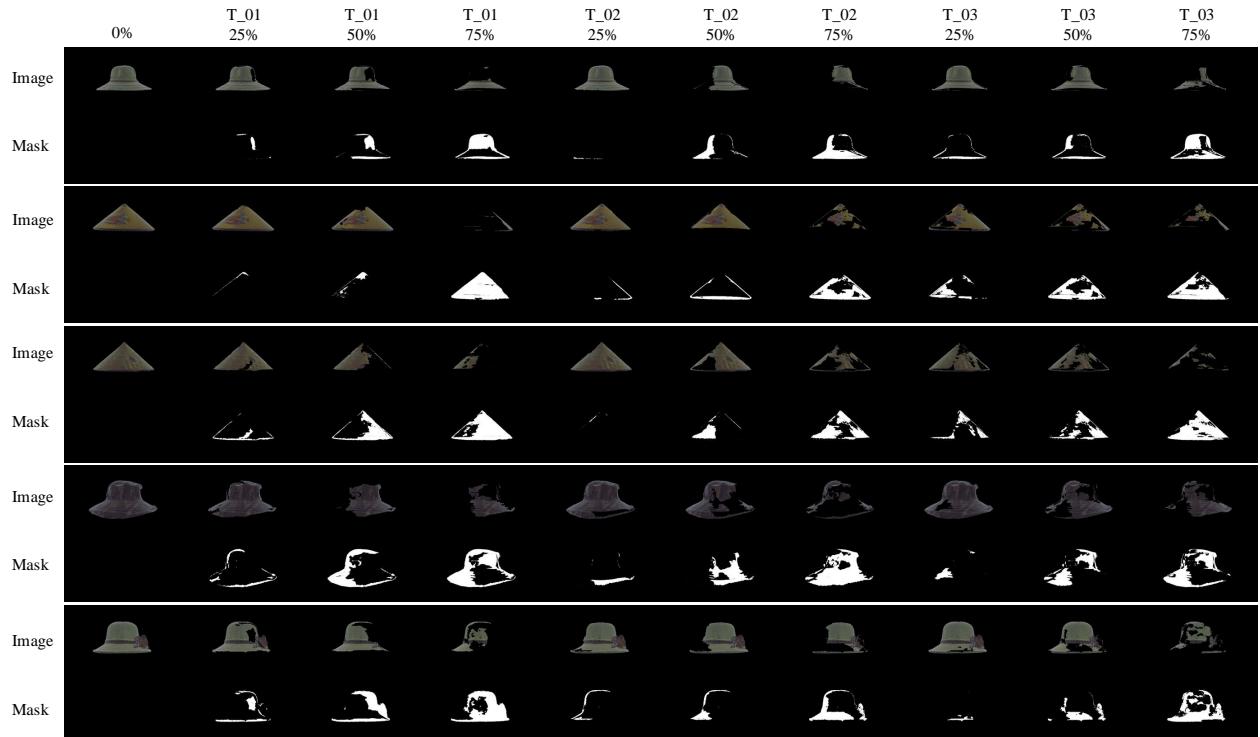


Figure 21. Key frame of each of the **Machaverse-XS** “Hat” sequences (1/6).  $T\_0X$  means the  $X$  type of masking strategy.

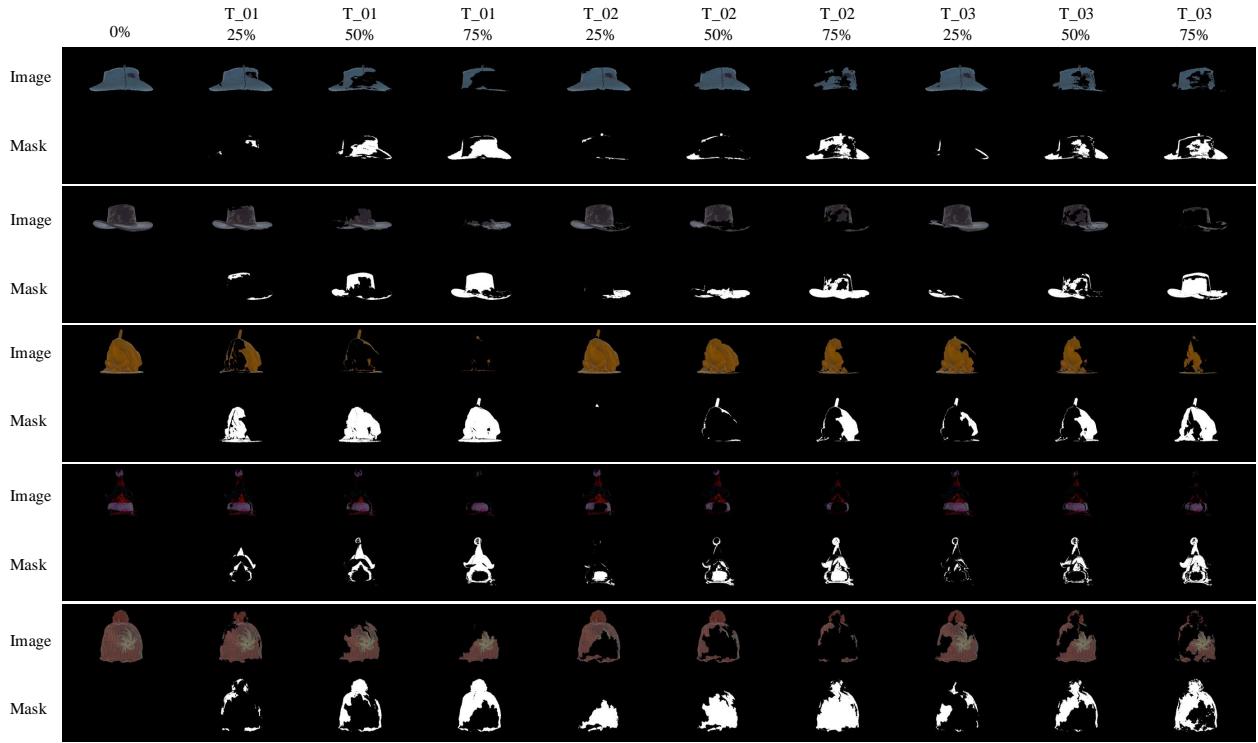


Figure 22. Key frame of each of the **Machaverse-XS** “Hat” sequences (2/6).  $T_{.0}X$  means the  $X$  type of masking strategy.



Figure 23. Key frame of each of the **Machaverse-XS** “Hat” sequences (3/6).  $T_{.0}X$  means the  $X$  type of masking strategy.

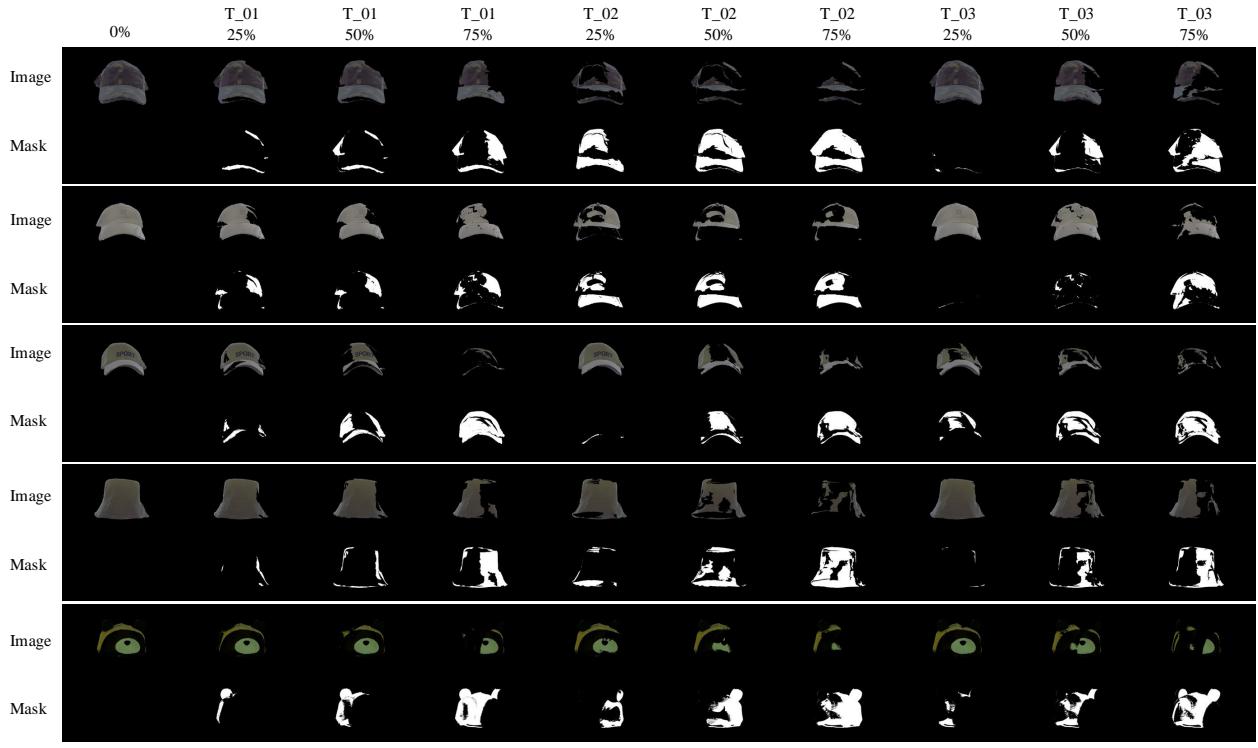


Figure 24. Key frame of each of the **Machaverse-XS** “Hat” sequences (4/6).  $T_{.0}X$  means the  $X$  type of masking strategy.

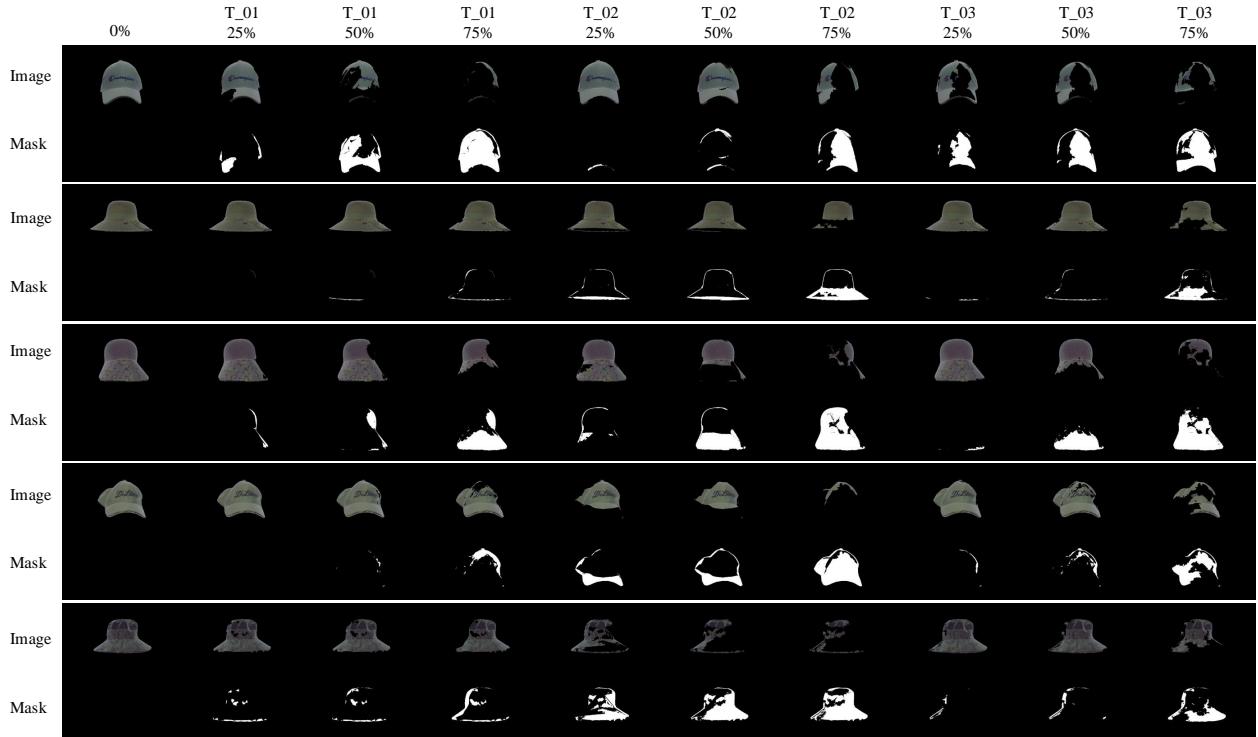


Figure 25. Key frame of each of the **Machaverse-XS** “Hat” sequences (5/6).  $T_{.0}X$  means the  $X$  type of masking strategy.

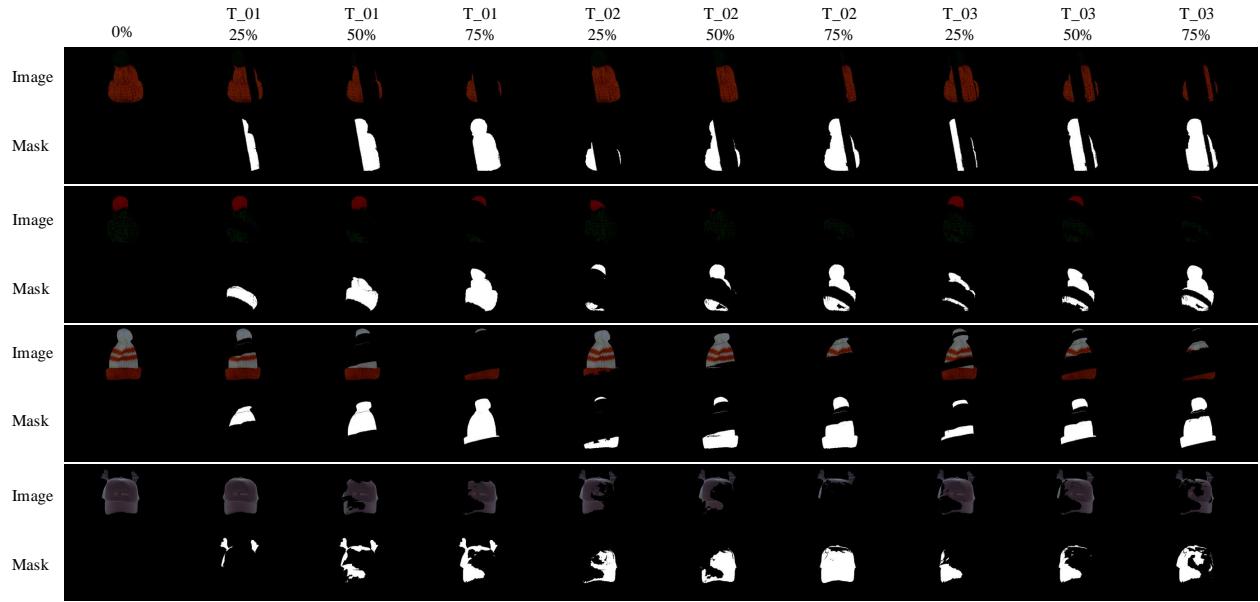


Figure 26. Key frame of each of the **Machaverse-XS** “Hat” sequences (6/6).  $T_{.0}X$  means the  $X$  type of masking strategy.

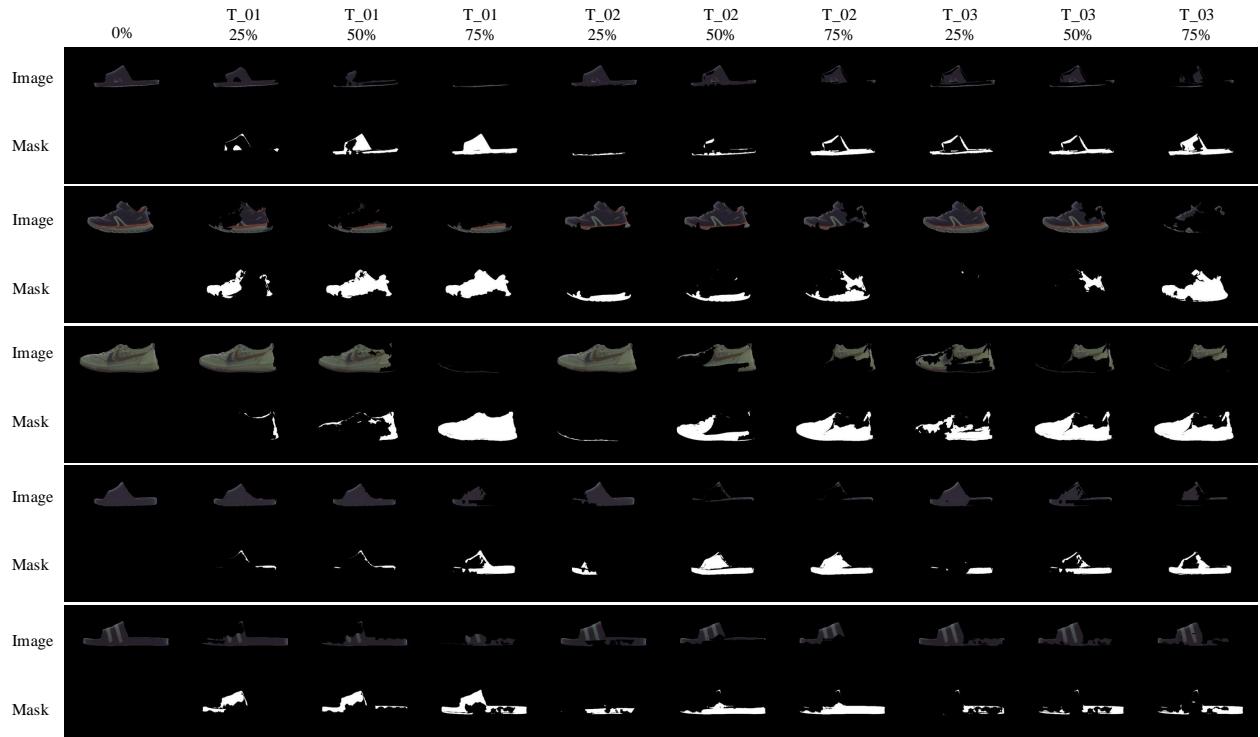


Figure 27. Key frame of each of the **Machaverse-XS** “Shoe” sequences (1/8).  $T_{.0}X$  means the  $X$  type of masking strategy.

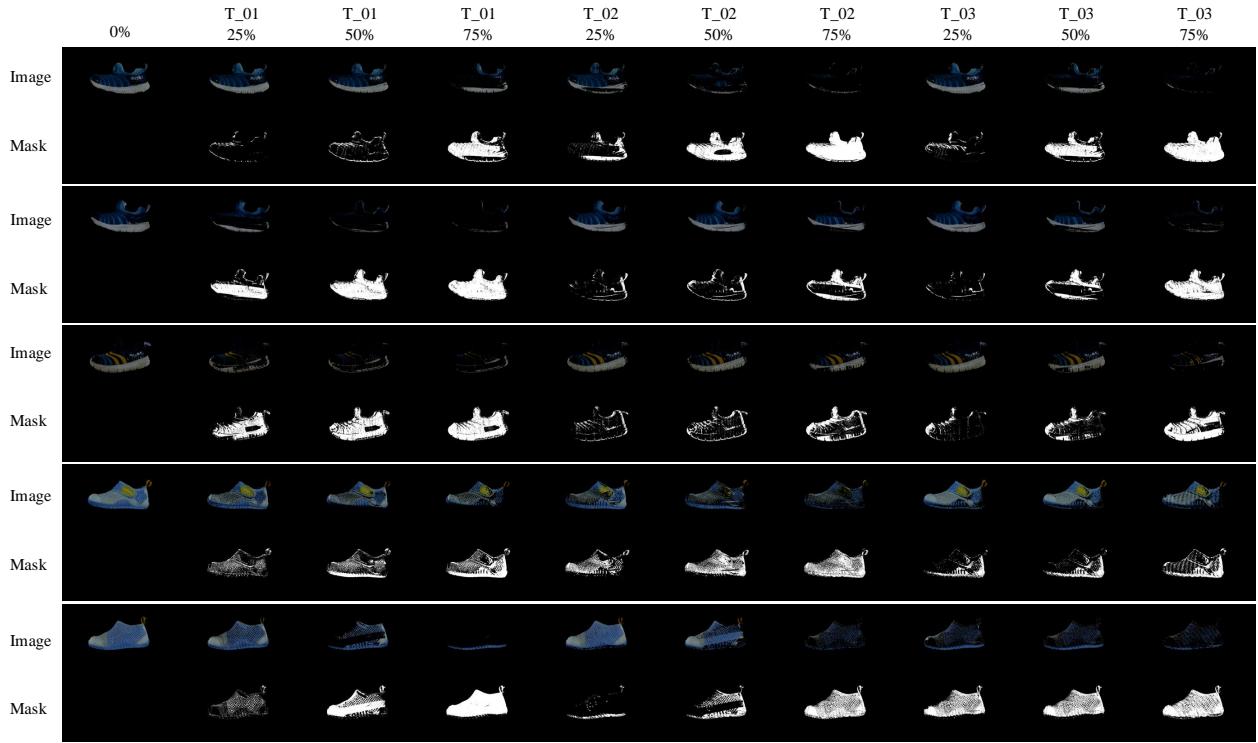


Figure 28. Key frame of each of the **Machaverse-XS** “Shoe” sequences (2/8).  $T\_0X$  means the  $X$  type of masking strategy.

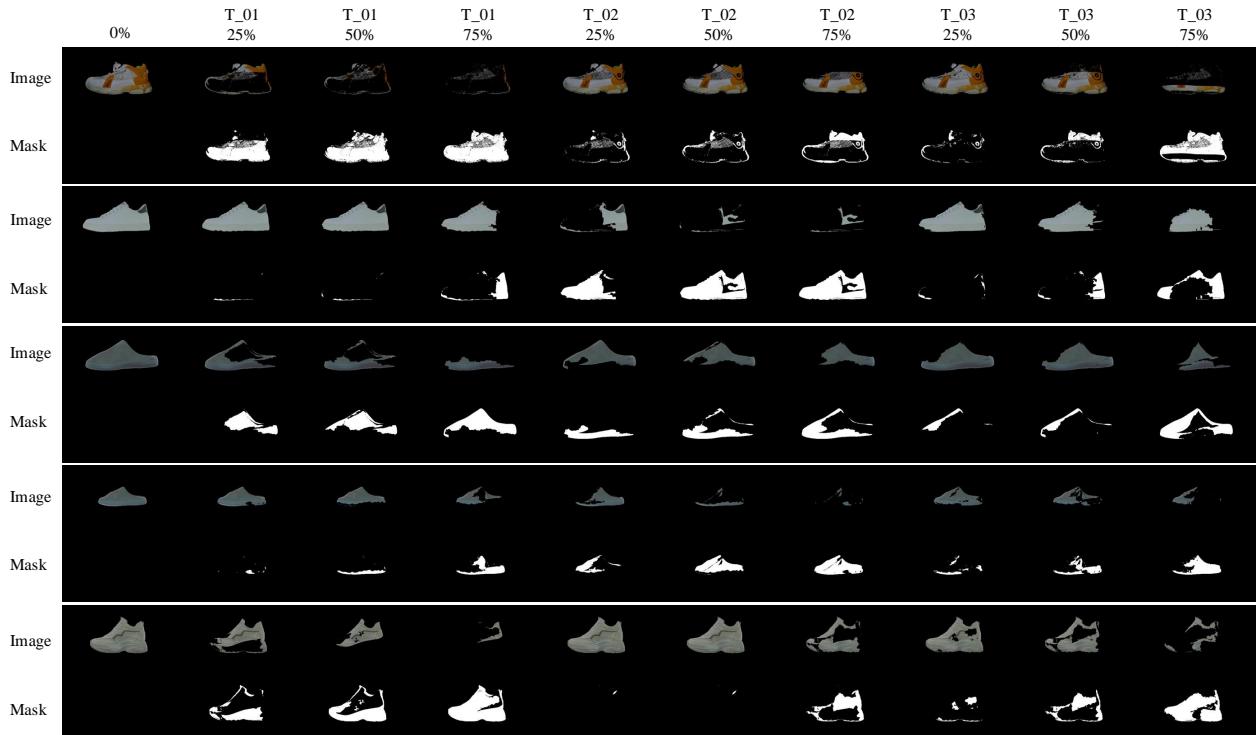


Figure 29. Key frame of each of the **Machaverse-XS** “Shoe” sequences (3/8).  $T\_0X$  means the  $X$  type of masking strategy.



Figure 30. Key frame of each of the **Machaverse-XS** “Shoe” sequences (4/8).  $T_{0X}$  means the  $X$  type of masking strategy.

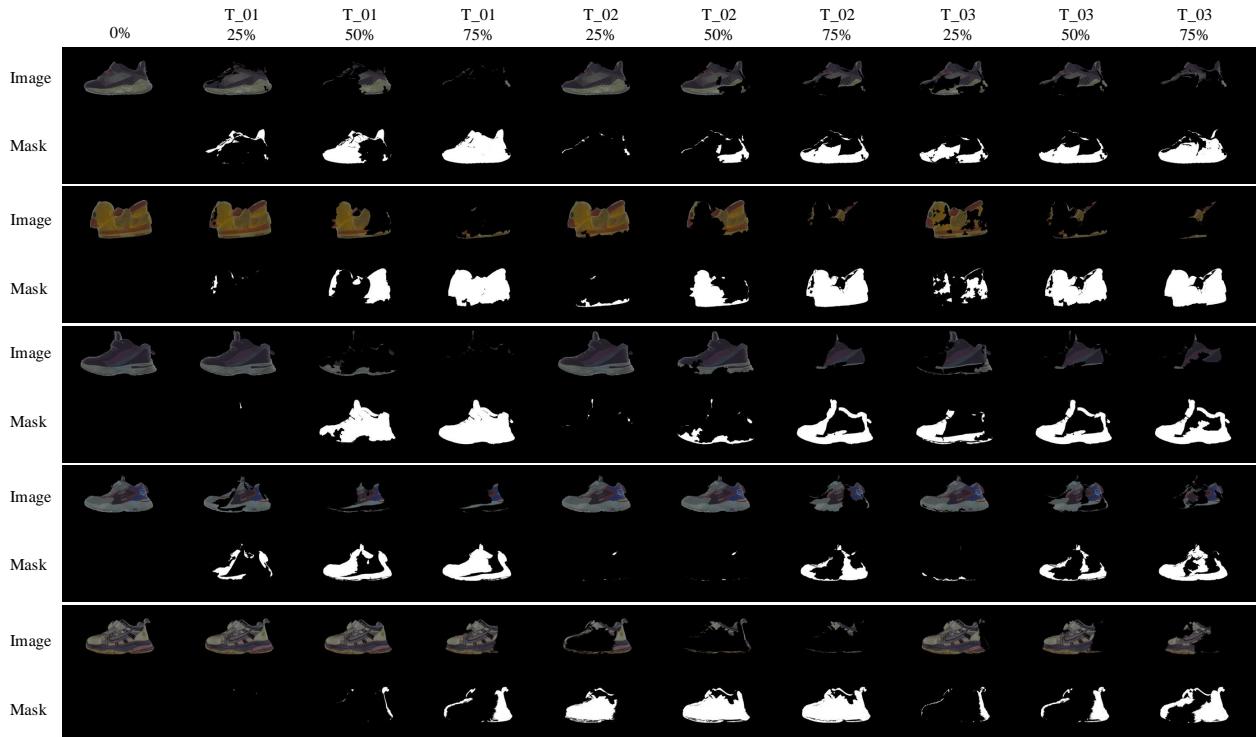


Figure 31. Key frame of each of the **Machaverse-XS** “Shoe” sequences (5/8).  $T_{0X}$  means the  $X$  type of masking strategy.

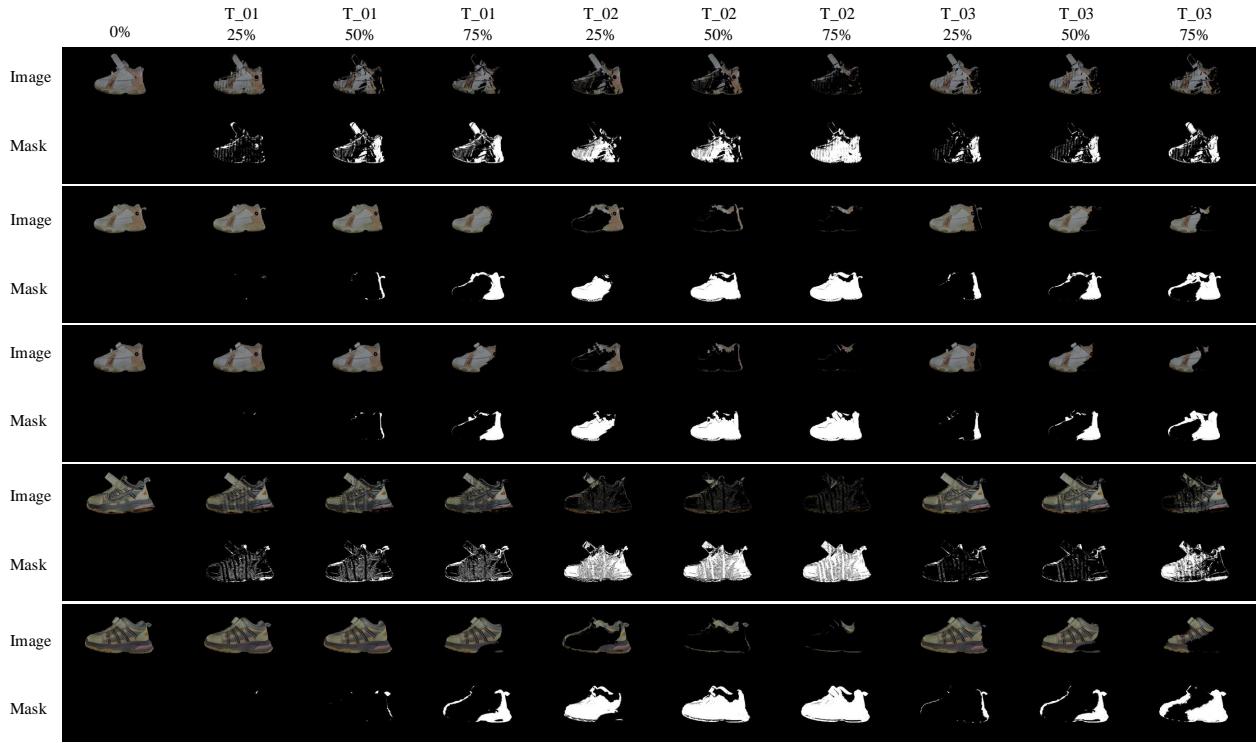


Figure 32. Key frame of each of the **Machaverse-XS** “Shoe” sequences (6/8).  $T_{0X}$  means the  $X$  type of masking strategy.

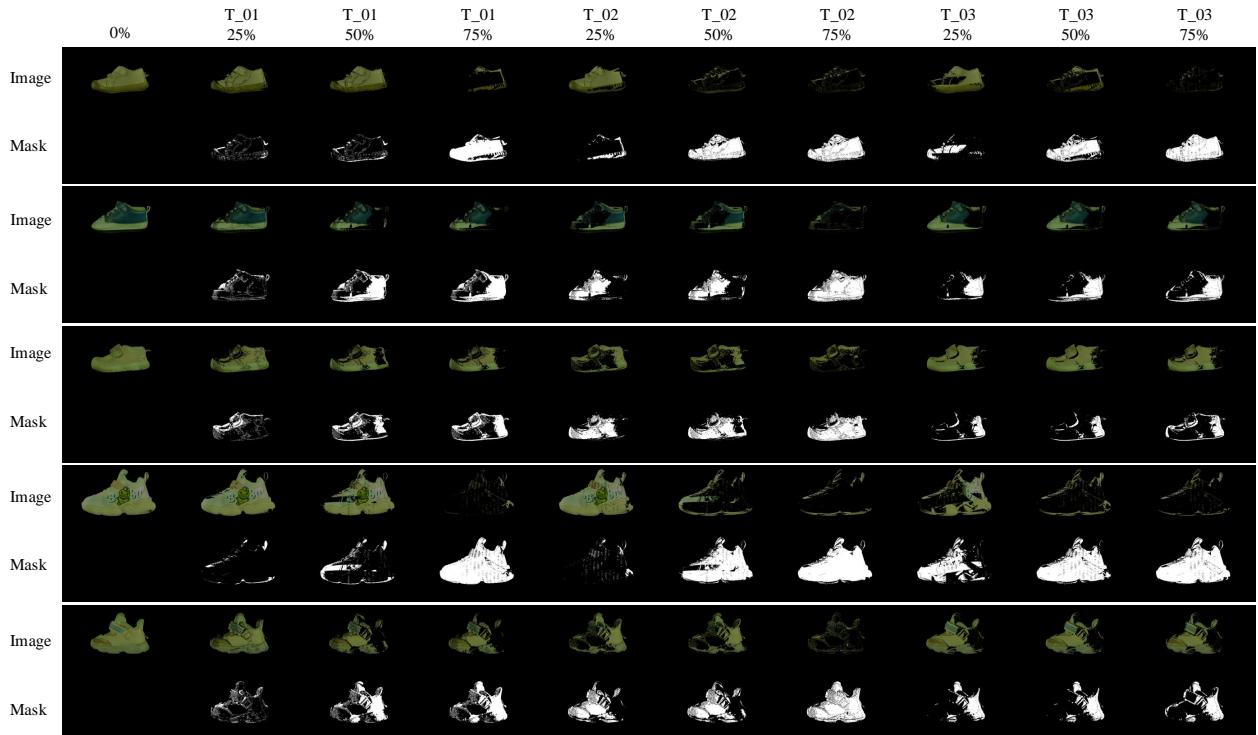


Figure 33. Key frame of each of the **Machaverse-XS** “Shoe” sequences (7/8).  $T_{0X}$  means the  $X$  type of masking strategy.

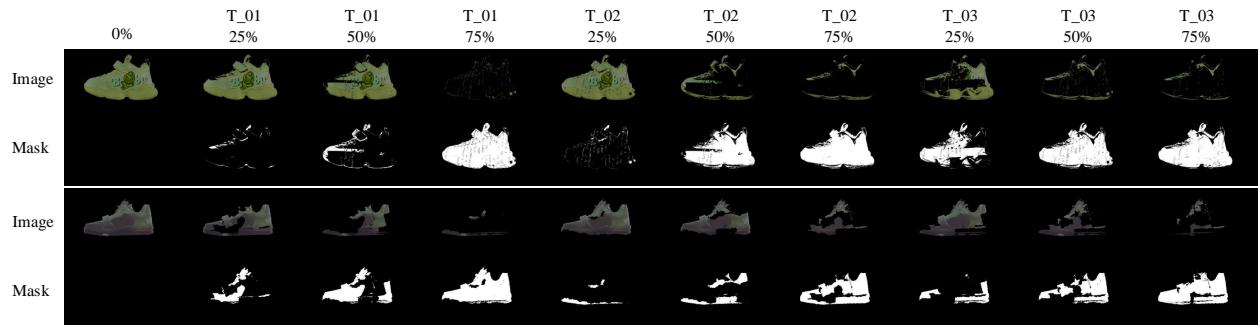


Figure 34. Key frame of each of the **Machaverse-XS** “Shoe” sequences (8/8).  $T\_0X$  means the  $X$  type of masking strategy.



Figure 35. An illustration of our *Machaverse-XS* used for I23 and 3D object completion from a single defective view.