EPFL IVRL Largo.ai BMVC 2023

VETIM: Expanding the Vocabulary of Text-to-Image Models only with Text

Martin Nicolas Everaert, Marco Bocchio, Sami Arpa, Sabine Süsstrunk, Radhakrishna Achanta

The problem we solve

Current methods [1, 2] cannot be used when **sample images are not available**. Instead, our method VETIM **solely** uses **textual descriptions**.

VETIM learns to **represent a complex concept as a single token** S_*.

Optimisation with VETIM is **faster**.

VETIM **does not mimic** visual features from **existing images**.

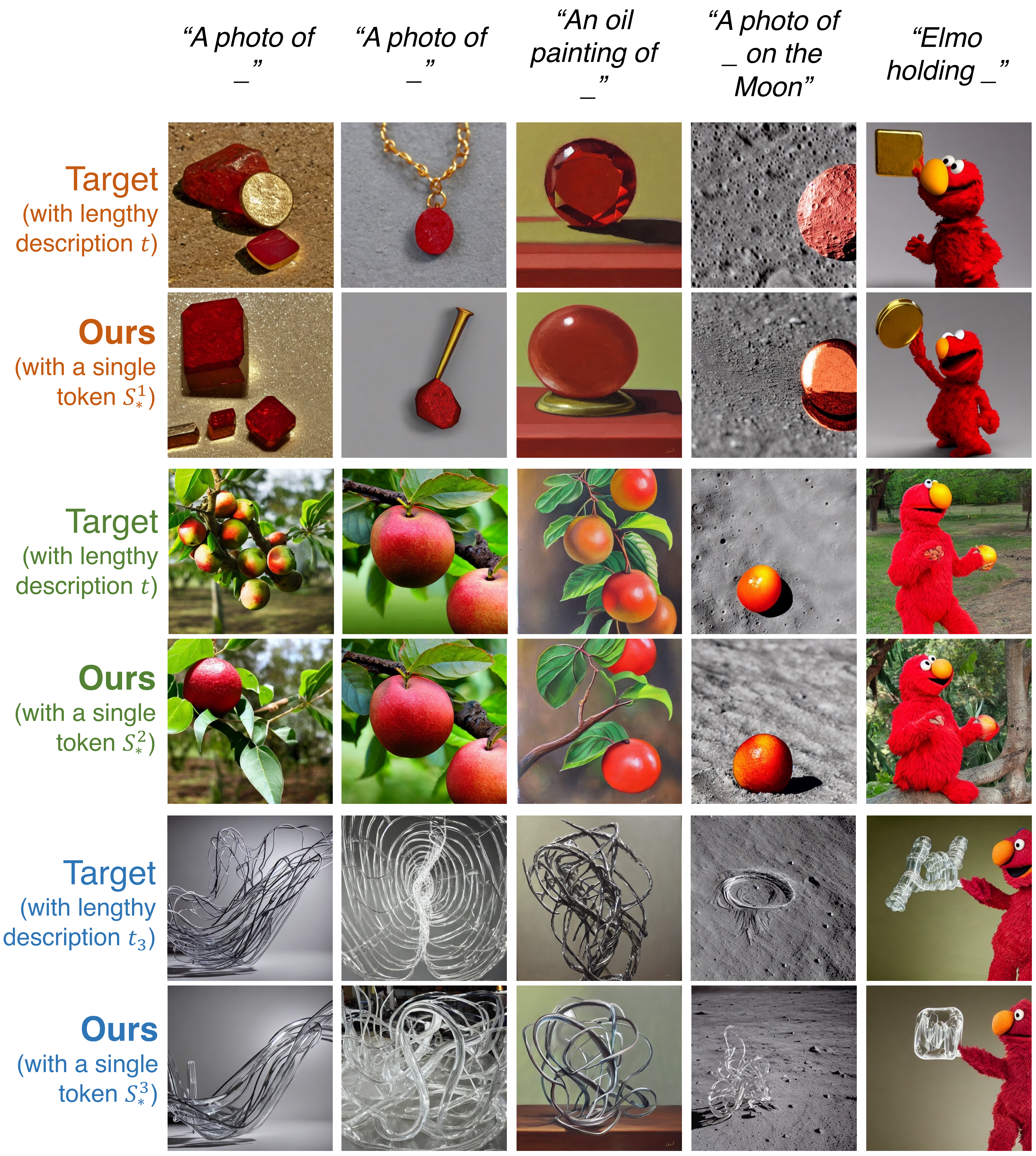
Results

Our newly learned token S_* **can be used instead of a lengthy text description** t:

t_1 = a small, brilliant red stone that can produce the Elixir of Life and turn base metals into gold [3]

t_2 = a common, round fruit produced by the tree Malus domestica, cultivated in temperate climates [4]

t_3 = a twisted, abstract sculpture made of delicate, interlocking tendrils of glass [5]



"A photo of _" *"A photo of _"* *"An oil painting of _"* *"A photo of _ on the Moon"* *"Elmo holding _"*

Target (with lengthy description t)

Ours (with a single token S_*^1)

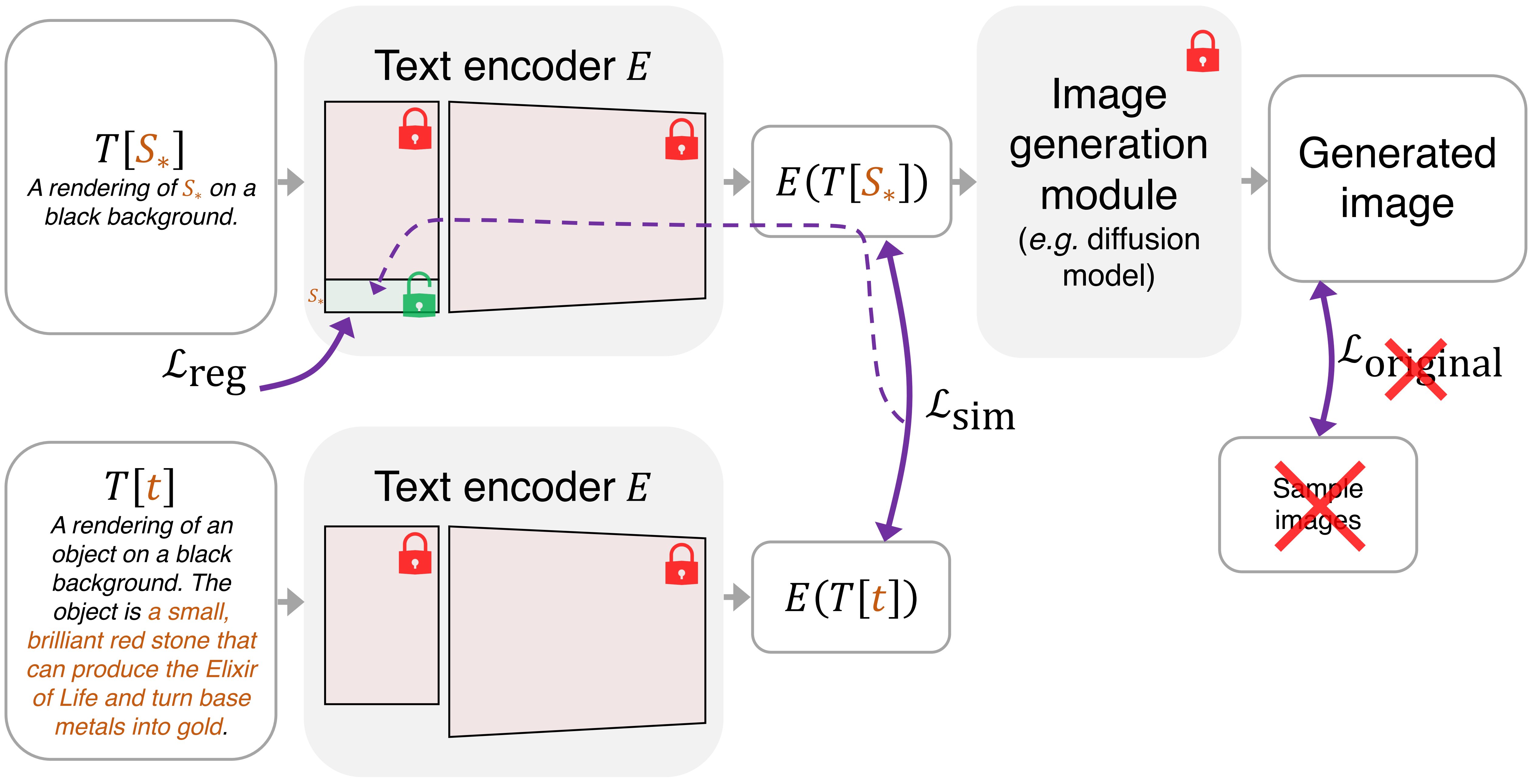
Target (with lengthy description t)

Ours (with a single token S_*^2)

Target (with lengthy description t_3)

Ours (with a single token S_*^3)

Method



$T[S_*]$ *A rendering of S_* on a black background.*

Text encoder E

S_*

$\mathcal{L}_{\text{reg}}$

$E(T[S_*])$

Image generation module (*e.g.* diffusion model)

Generated image

$\mathcal{L}_{\cancel{\text{original}}}$

~~Sample images~~

$\mathcal{L}_{\text{sim}}$

$T[t]$ *A rendering of an object on a black background. The object is a small, brilliant red stone that can produce the Elixir of Life and turn base metals into gold.*

Text encoder E

$E(T[t])$

References

[1] Gal *et al.* An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. ICLR 2022

[2] Kumari *et al.* Multi-Concept Customization of Text-to-Image Diffusion. CVPR 2023

[3] Description generated with ChatGPT

[4] Definition of "apple" on wiktionary.org

[5] Description generated with ChatGPT

This work is supported by Innosuisse grant 48552.1 IP-ICT.

Project website:

https://ivrl.github.io/vetim/