Text-Driven Stylization of Video Objects



Sebastian Loeschcke[1]

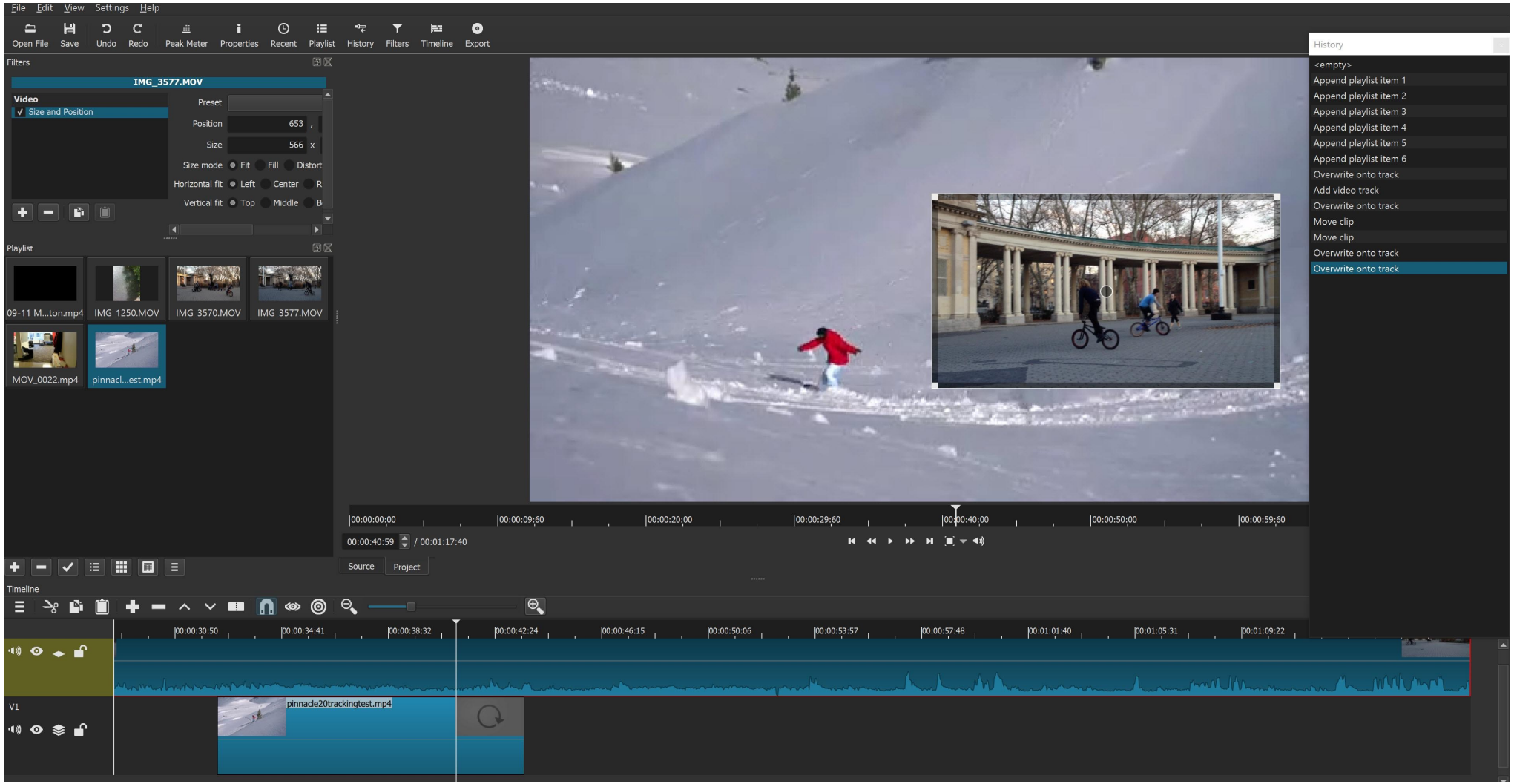
Serge Belongie[2]

Sagie Benaim[2]

[1]Aarhus University [2]University of Copenhagen

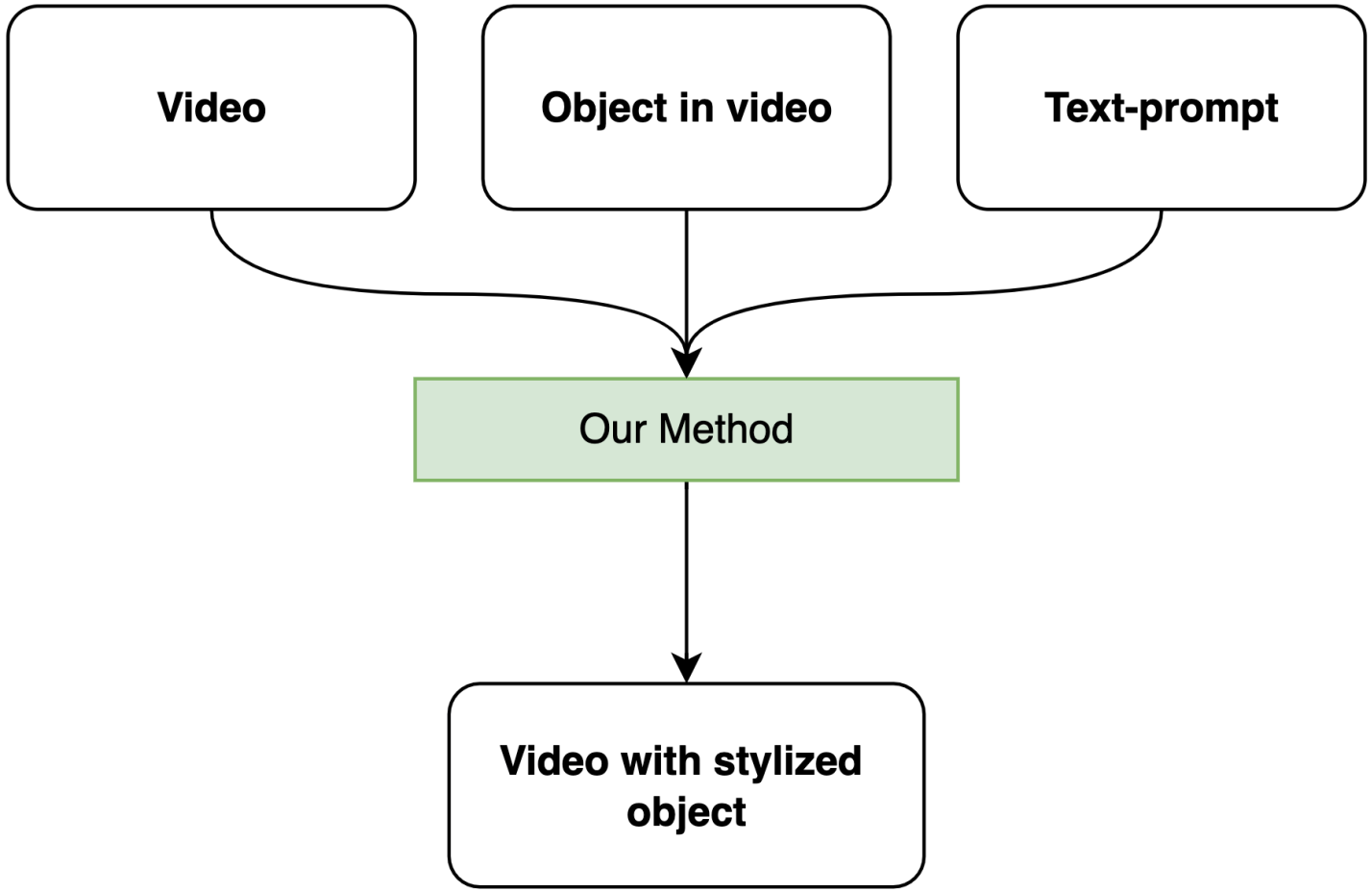
ECCV2022 Workshop on AI for Creative Video Editing and Understanding

Video Object Editing: Unintuitive and Complicated



Source: *https://uk.pcmag.com/video-editing*

Specifying Stylization Through Text



Intuitive and Consistent Video Editing via Text

Input Video

“Swan with a medieval iron amor“

“Swan made out of cactus”

Diverse Set of Videos and Target Texts

Input Video

“Dog with zebra fur“

“Golden dog”

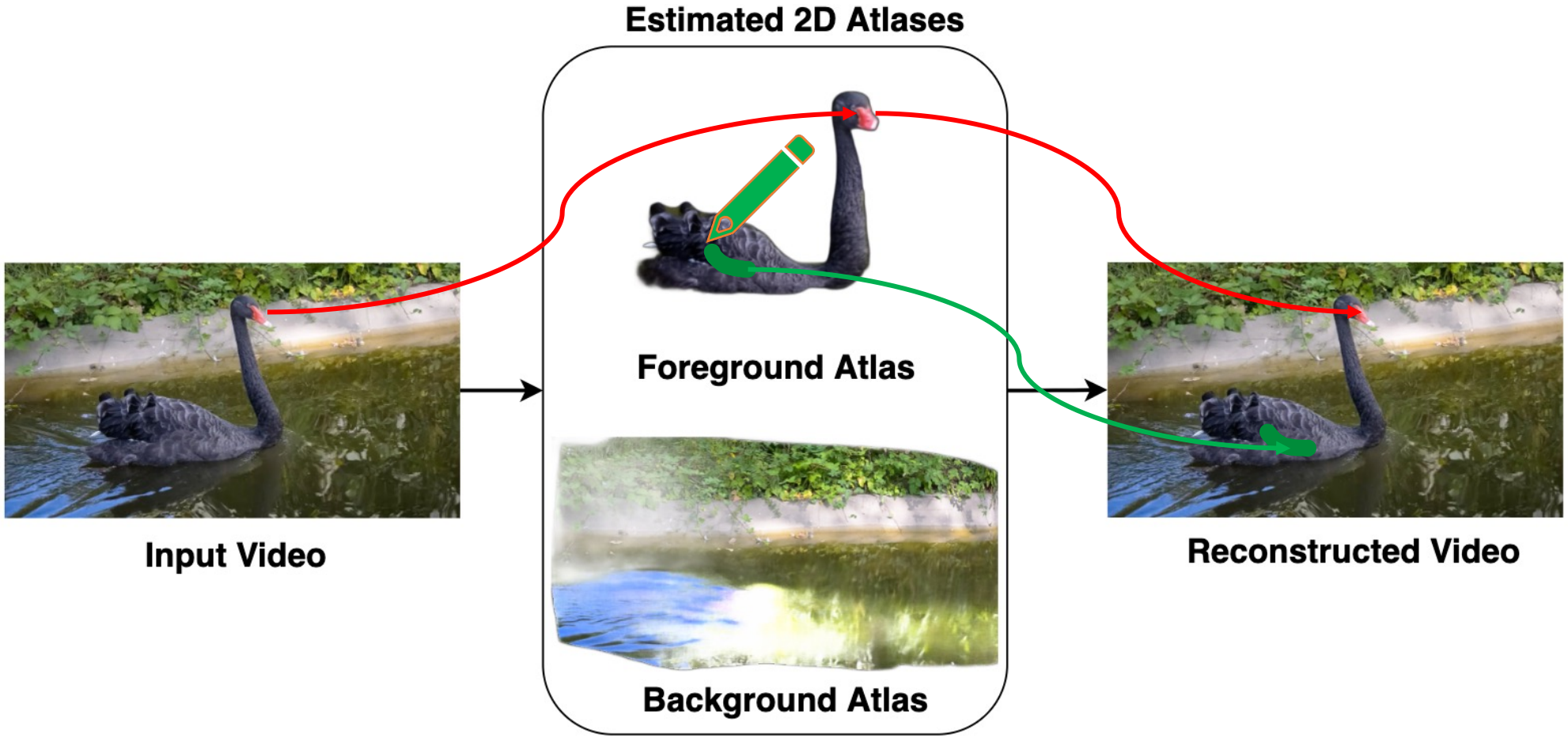
No Pretrained GAN or Video dataset

Input Video

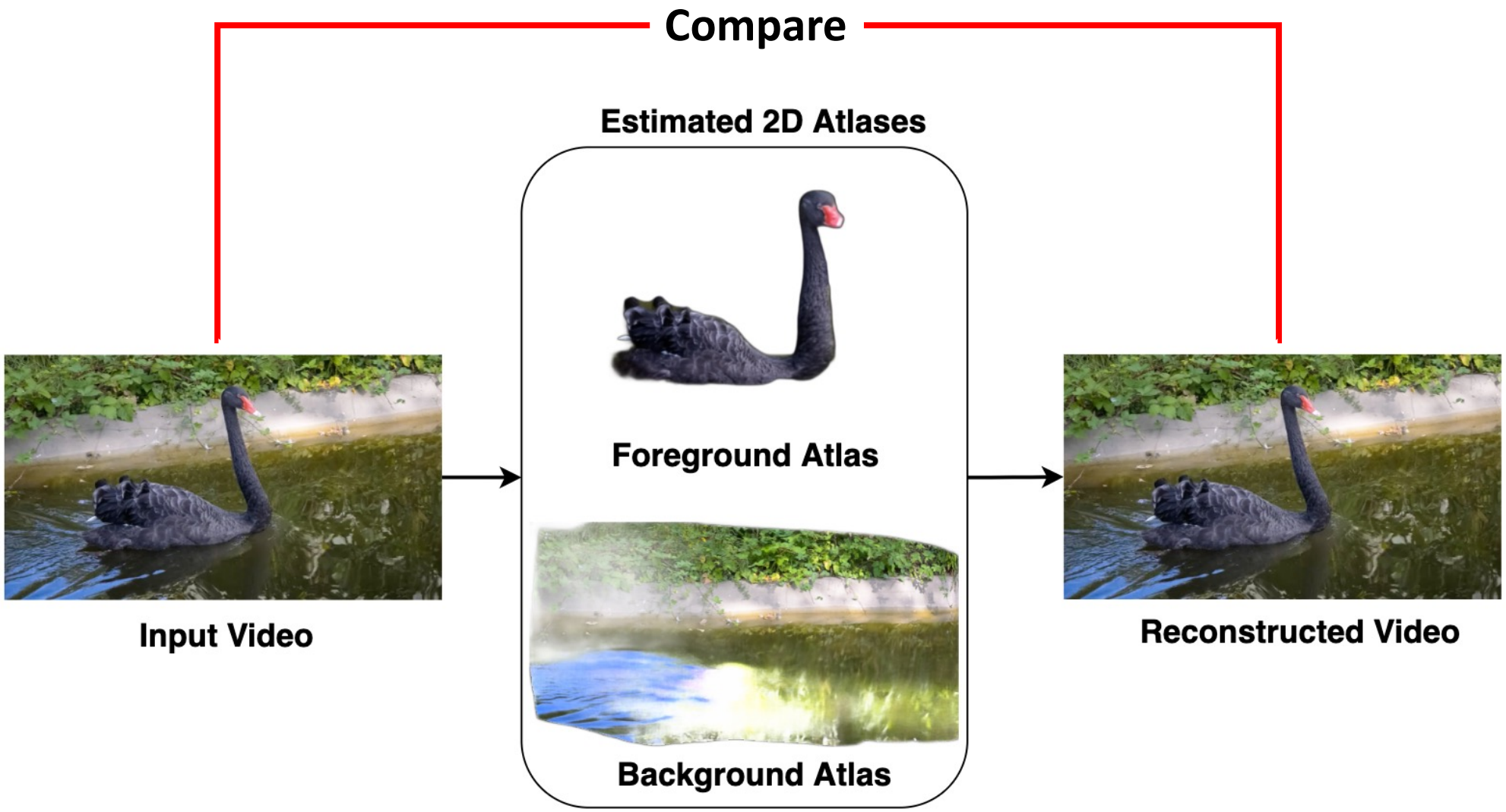
“Shiny aluminium fishing boat“

“Fishing boat made of wood”

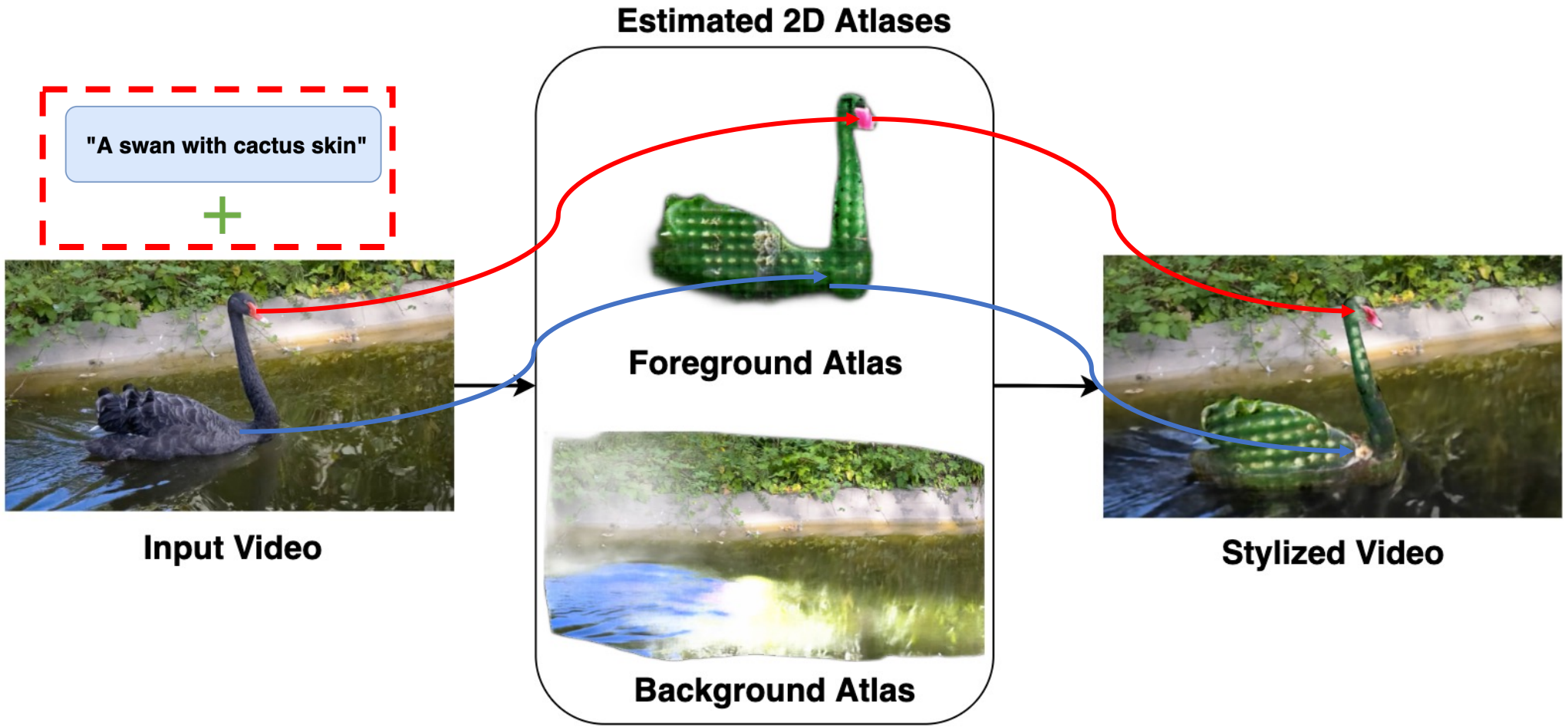
Neural Layered Atlases (NLA) [Kasten et. Al]



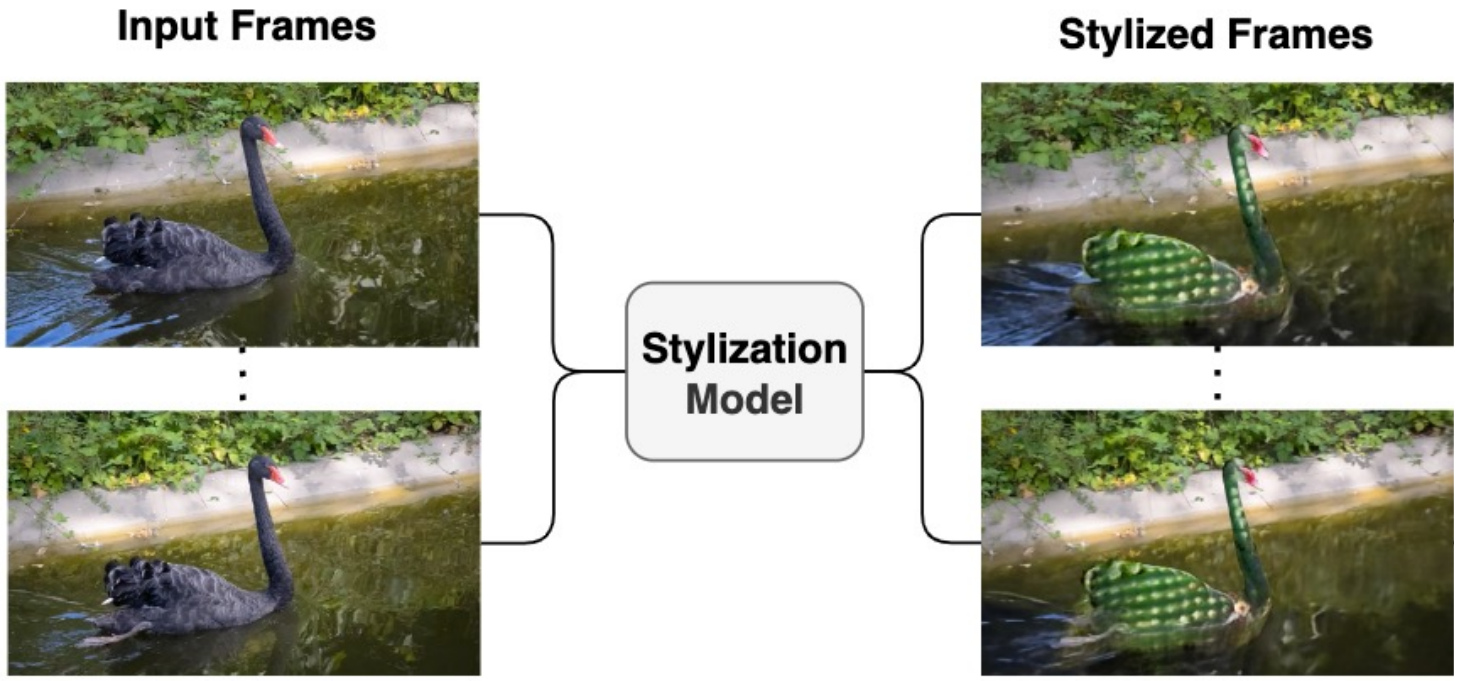
Neural Layered Atlases (NLA) [Kasten et. Al]



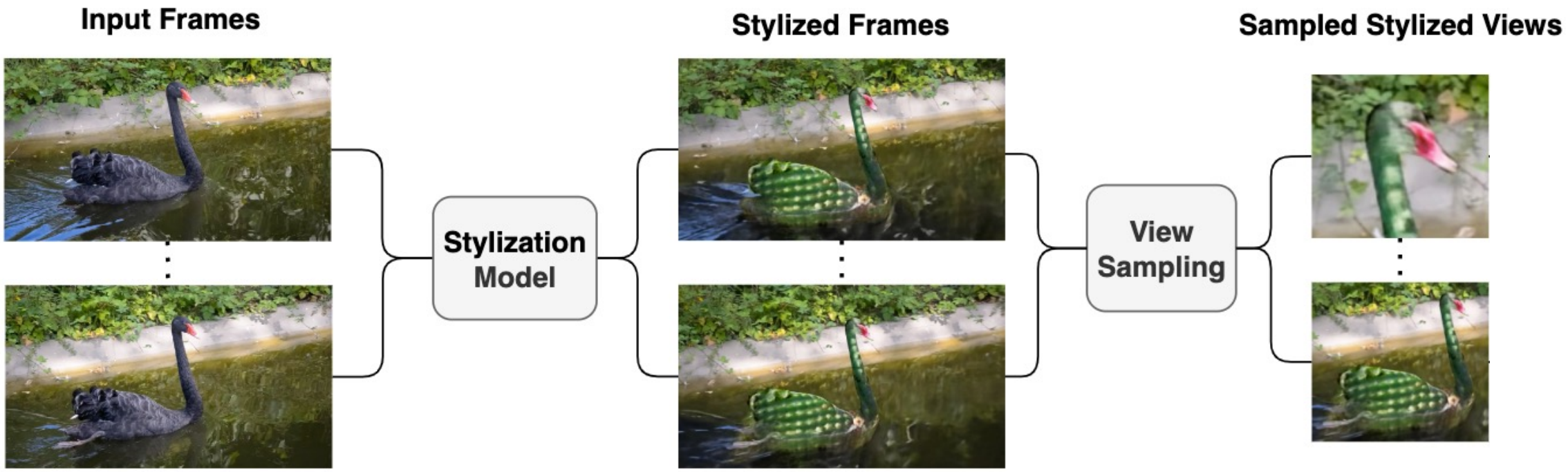
Our Stylization Model



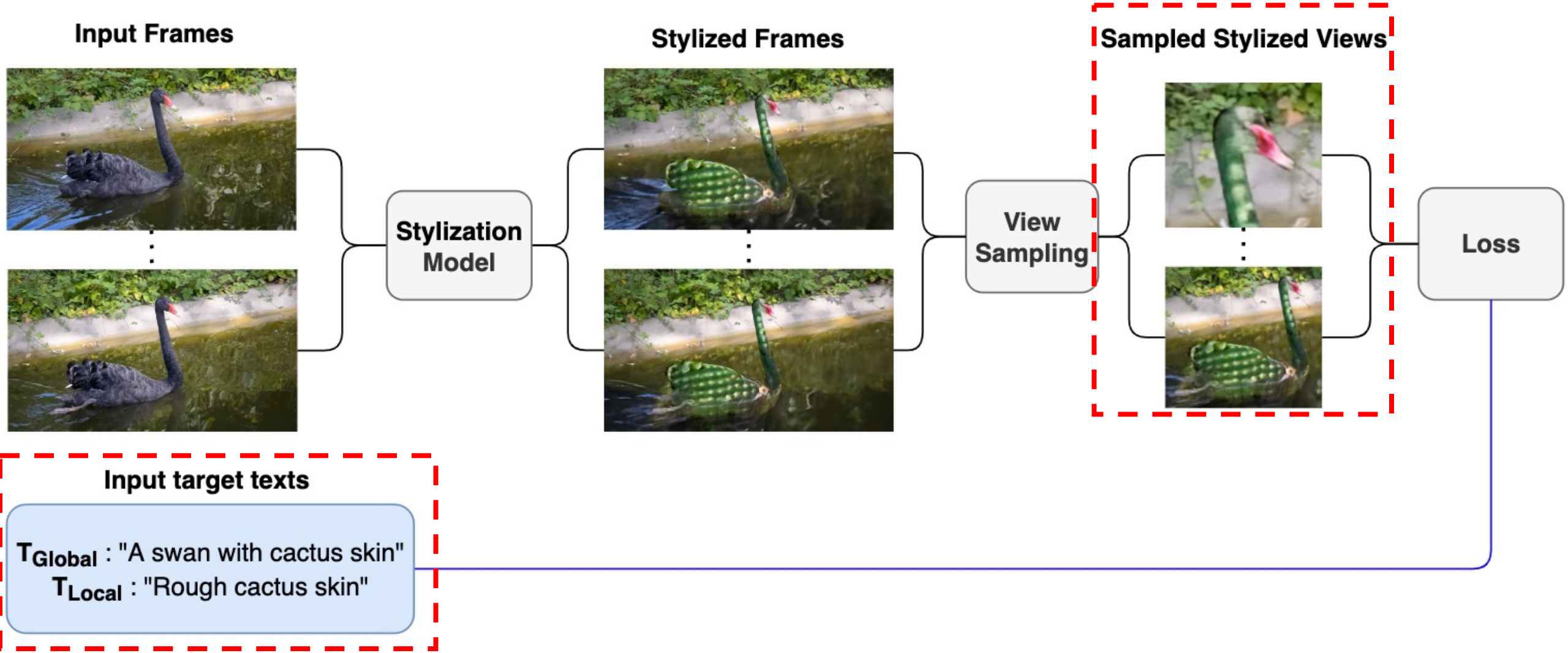
Stylization Pipeline



Stylization Pipeline

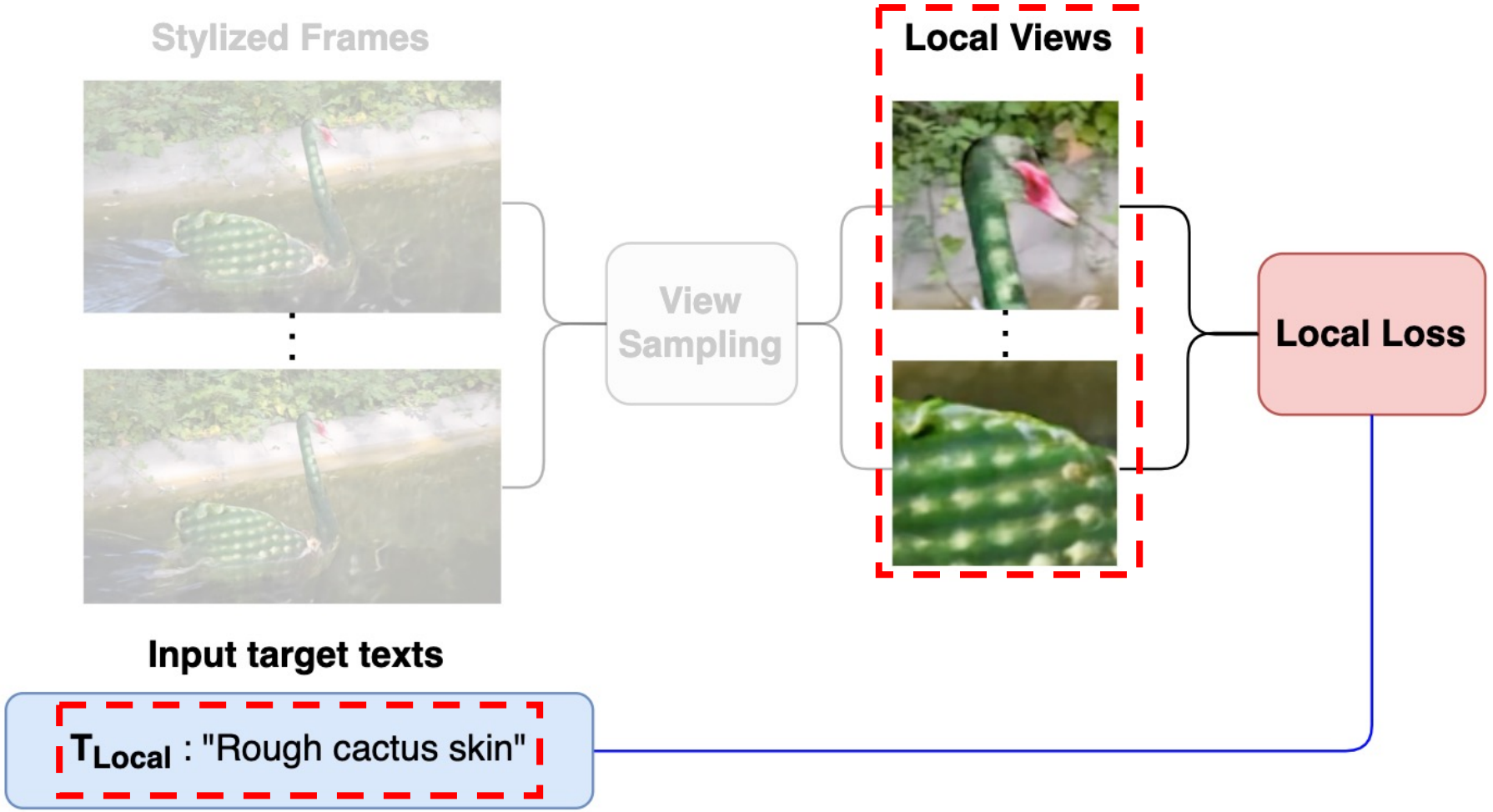


Stylization Pipeline



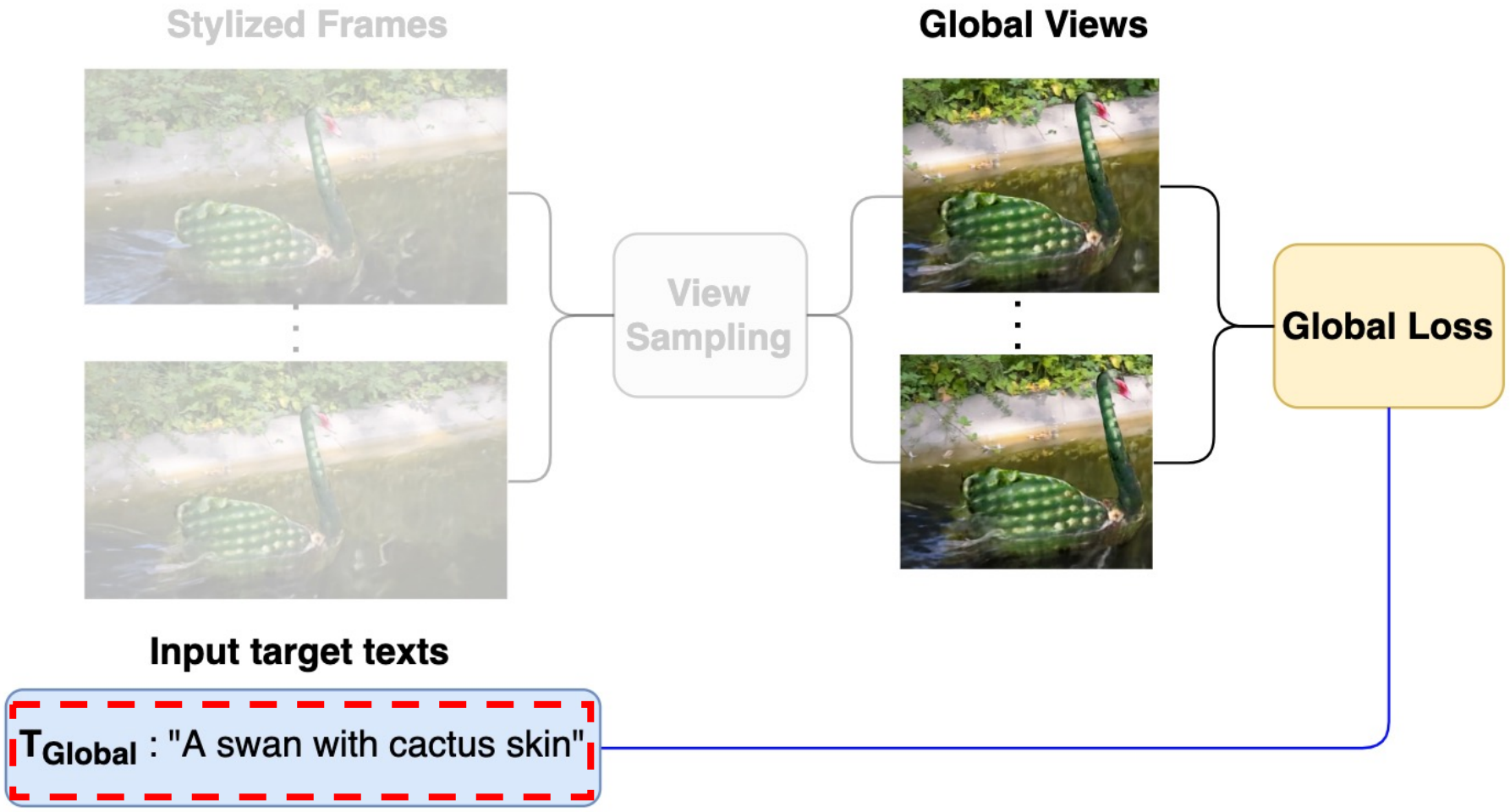
Local Loss

- Semantic Loss Computed with CLIP
- Focus on Fine-grained Details



Global Loss

- Semantic Loss Computed with CLIP
- Preserve Overall Context



Affect of Loss Terms– “Swan with crocodile skin”

All Losses

w/o Local Loss

w/o Global Loss

Affect of Loss Terms– “Swan with crocodile skin”

All Losses

w/o Local Loss

w/o Global Loss

Prompt Specificity

Input

”Boat made of wood”

“Boat made of **dark walnut** wood“

“Fishing boat made of wood **planks”**

Prompt Specificity

Input

”Boat made of wood”

“Boat made of **dark walnut** wood“

“Fishing boat made of wood **planks”**

Prefix Augmentations

- *Target text*: “Dog with Bengal tiger fur”
- E.g. ”A photo of a {*Target text*}”

Input

No prefixes

4 global, 4 local

8 global, 8 local

Prefix Augmentations

- *Target text*: “Dog with Bengal tiger fur”
- E.g. ”A photo of a {*Target text*}”

Input

No prefixes

4 global, 4 local

8 global, 8 local

Conclusion

- Text-Driven Stylization of Video Objects
- Fine-grained control of both local and global semantics
- Specificity and the prefixes of the target texts impact the details
- Shape change and object generation

Thank You

Visit https://sloeschcke.github.io/Text-Driven-Stylization-of-Video-Objects/ for more.

linkedin.com/in/sebastian-loeschcke/

sloeschcke@post.au.dk