From StyleGAN3 to Diffusion Model

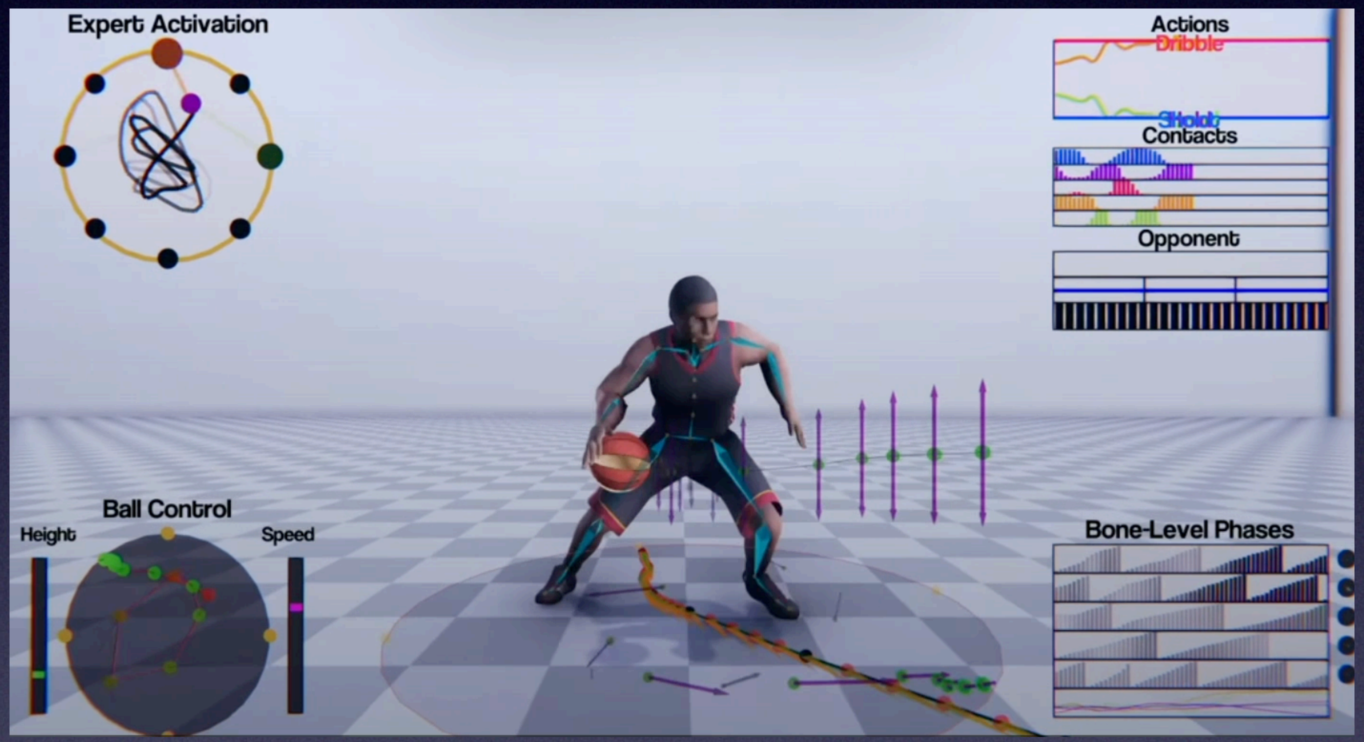
Reviewing two methods in Image-to-image translation task

Weihao Qiu | MAT 255 Fall 2022 | 2022.12.1

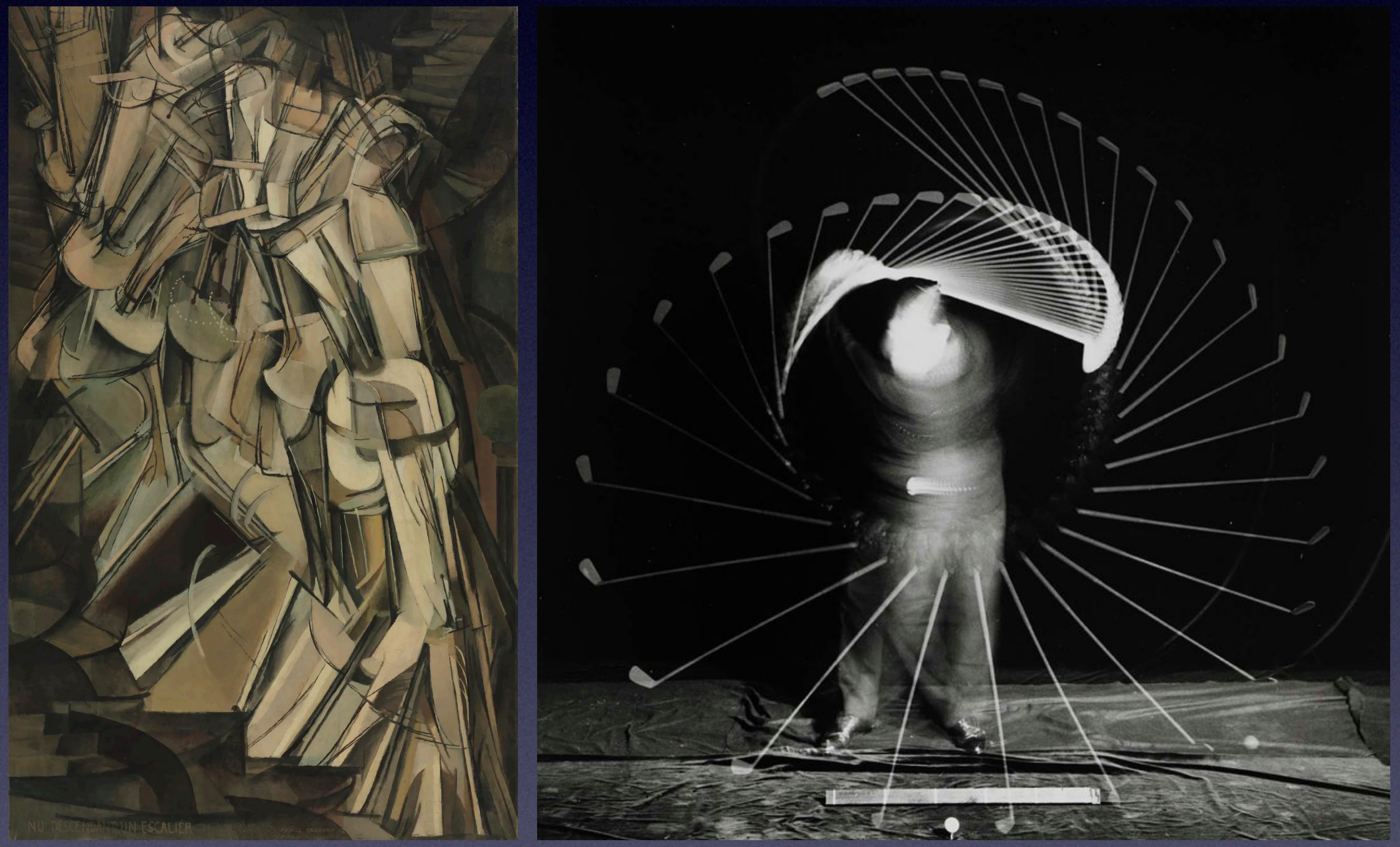
Contents

- Background
- Problem definition
- Solution with StyleGAN3
- Solution with Diffusion Models
- Comparisons
- Other experiments
- Future improvements

Background



Advancements in
Motion Synthesis



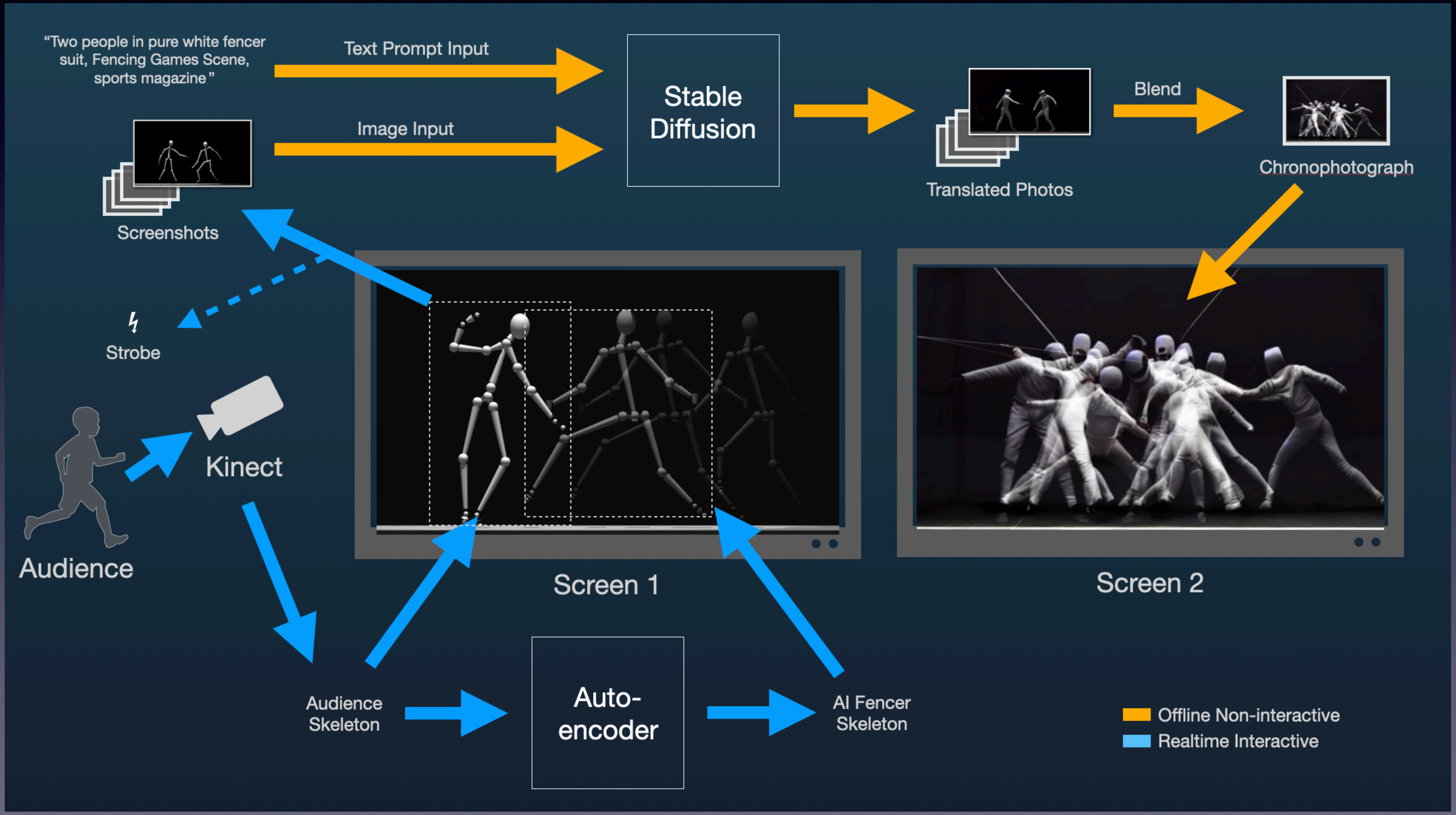
Motion is a popular subject matter for **art creation**

Pushing the limits of the **neural image synthesis**

Background

My Practice

Interactive fencing motion synthesis to generate chronophotograph



Problem

Definition: given a skeleton consisting of the body joints coordinates, synthesize a photo of the fencer in that position.

Interaction

(x1,y1,
x2, y2,
x3, y3, …
Xn,Yn)

Body Joints Coordinates

Synthetic Photo

Chronophotograph

Challenges

Image Quality

image has to
look natural

Limited Data

limited
resources

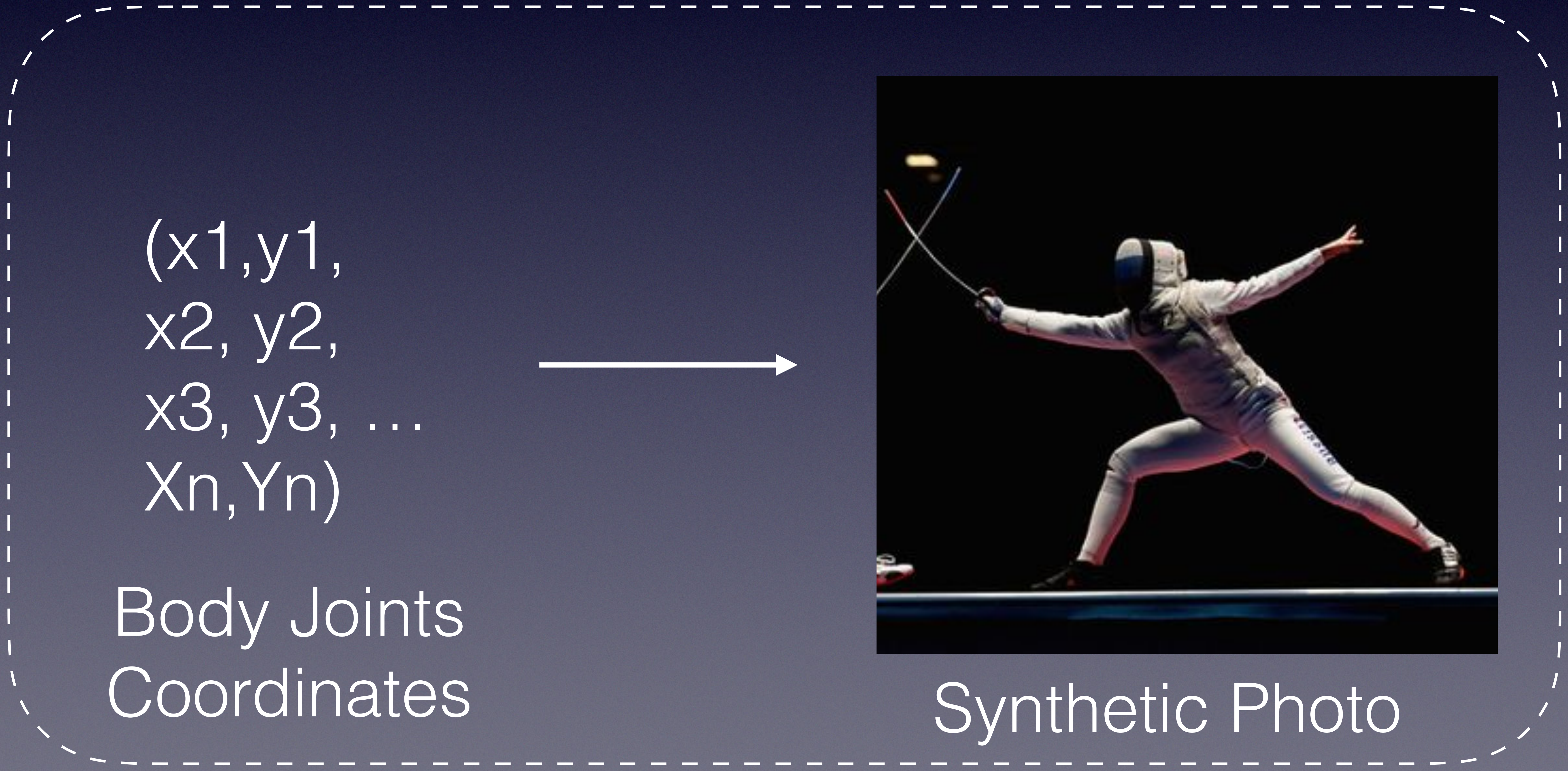
Consistency

generated
images should
contain the
same character

Style Control

e.g., change
the background
of the the
generated
images

Previous Solution: StyleGAN3



(x1,y1,
x2, y2,
x3, y3, …
Xn,Yn)

Body Joints Coordinates

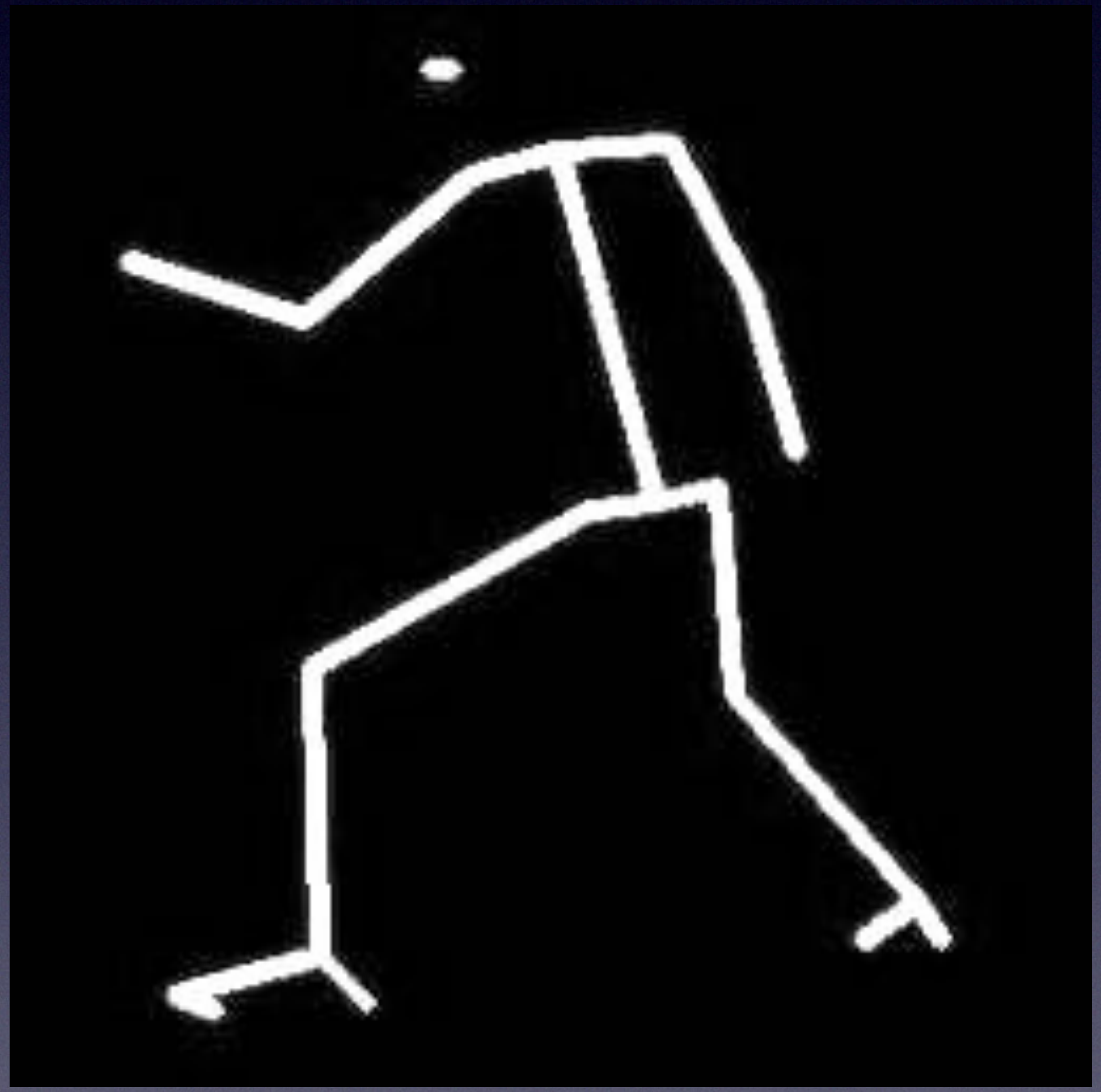
Synthetic Photo

Previous Solution: StyleGAN3

(x1,y1,
x2, y2,
x3, y3, …
Xn,Yn)

Body Joints Coordinates

→



Stick-figure image

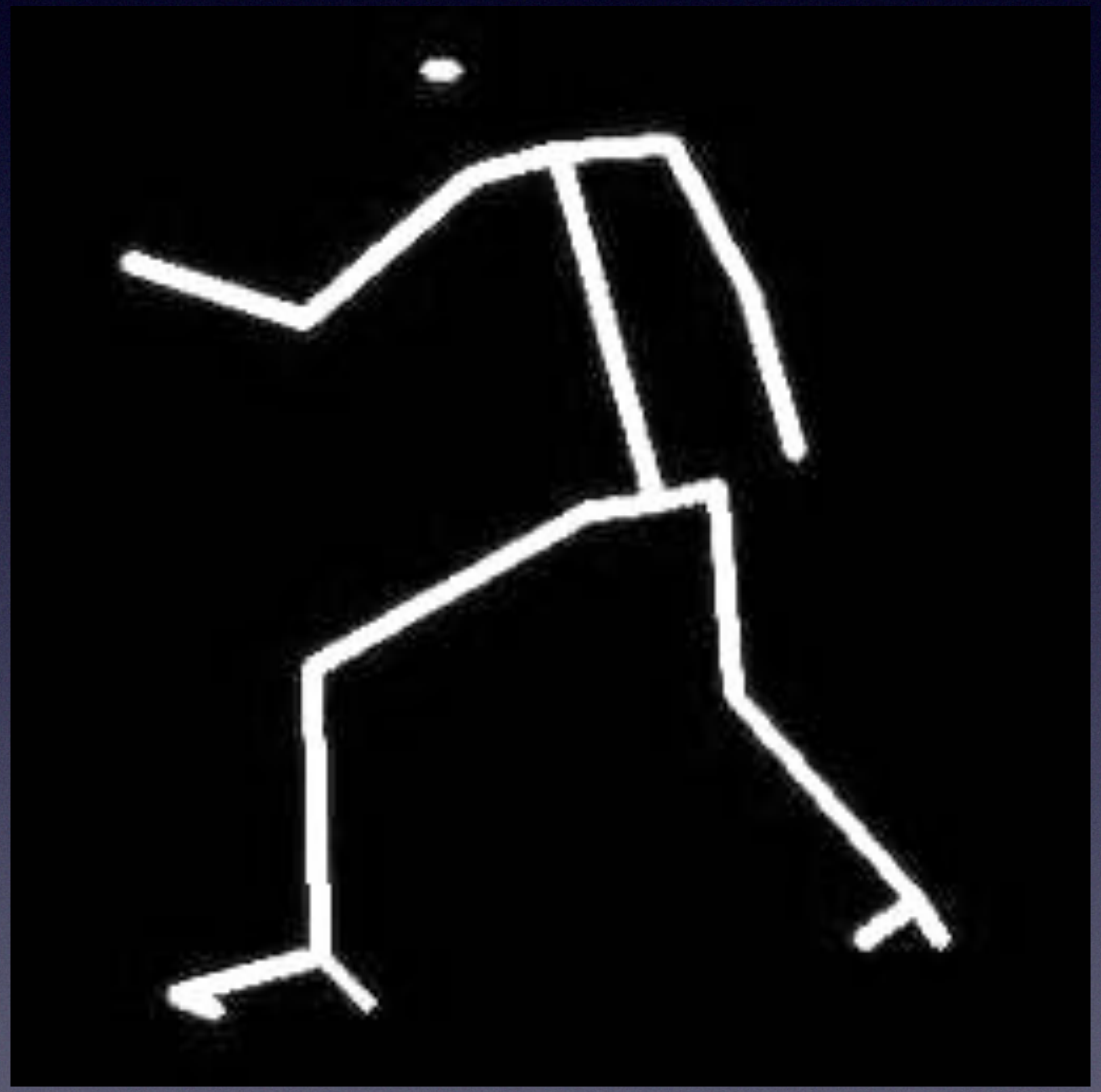
StyleGAN3
+
p2s2p

→

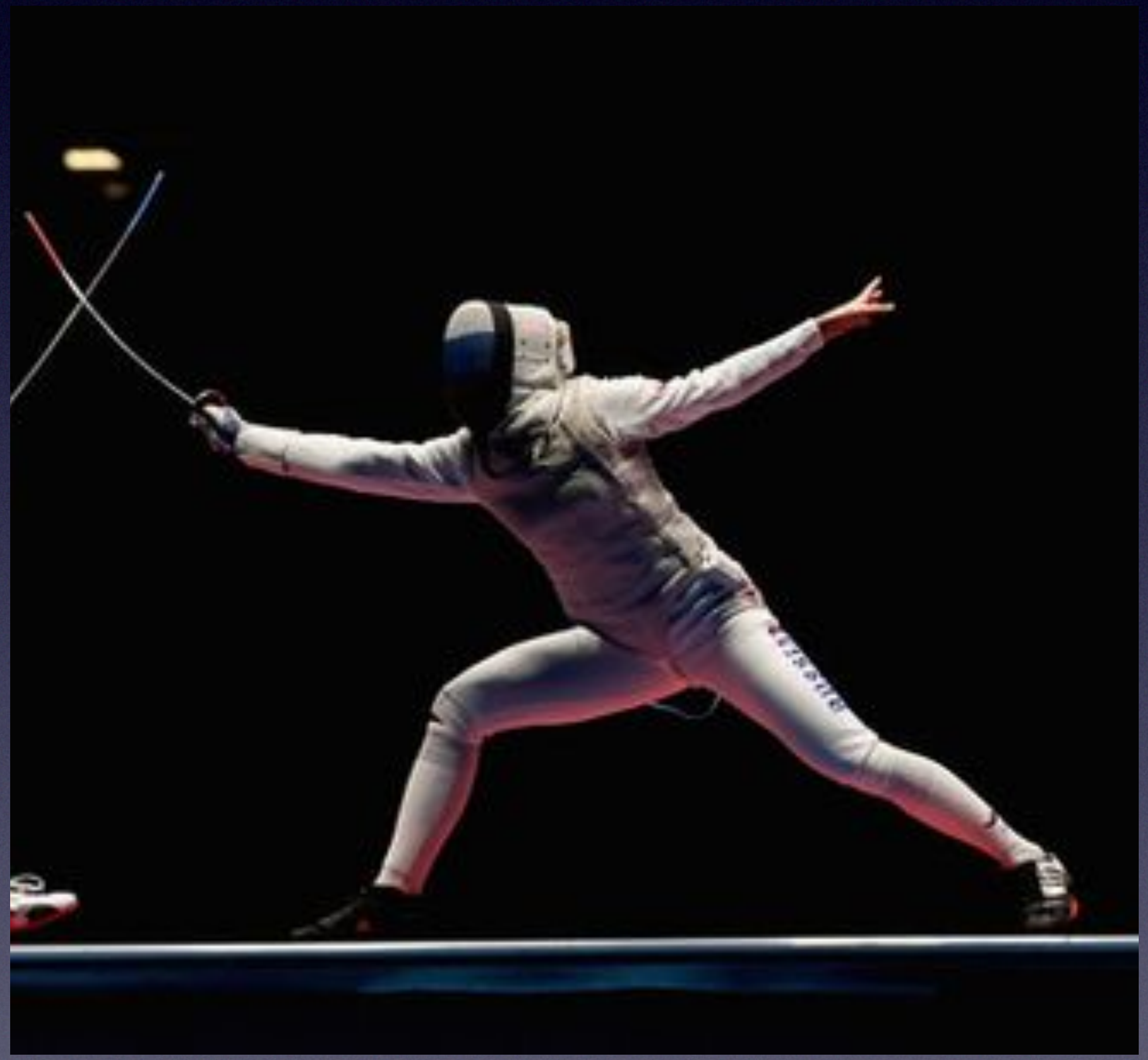
Synthetic Photo

Previous Solution: StyleGAN3

StyleGAN3
+
p2s2p



Stick-figure image

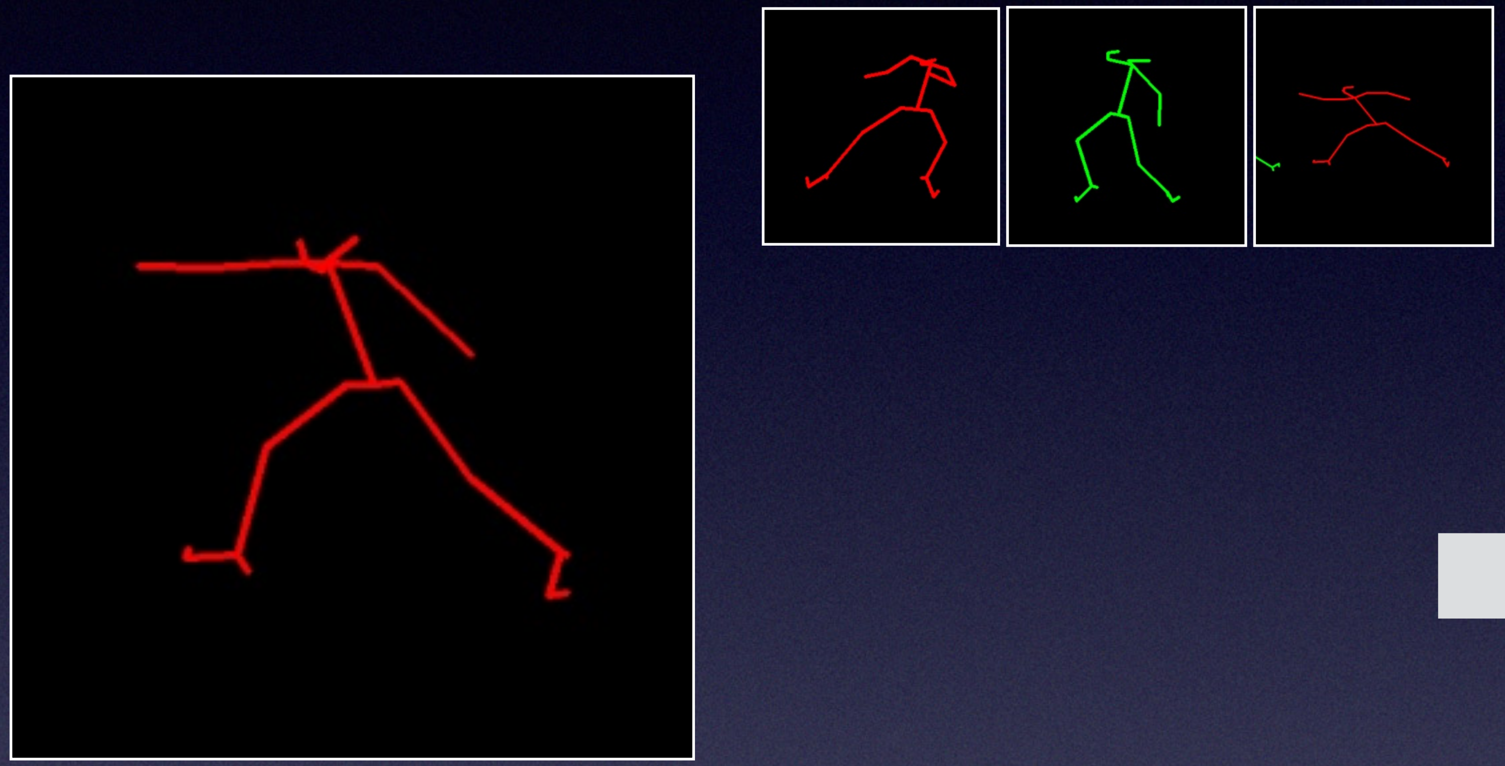


Synthetic Photo

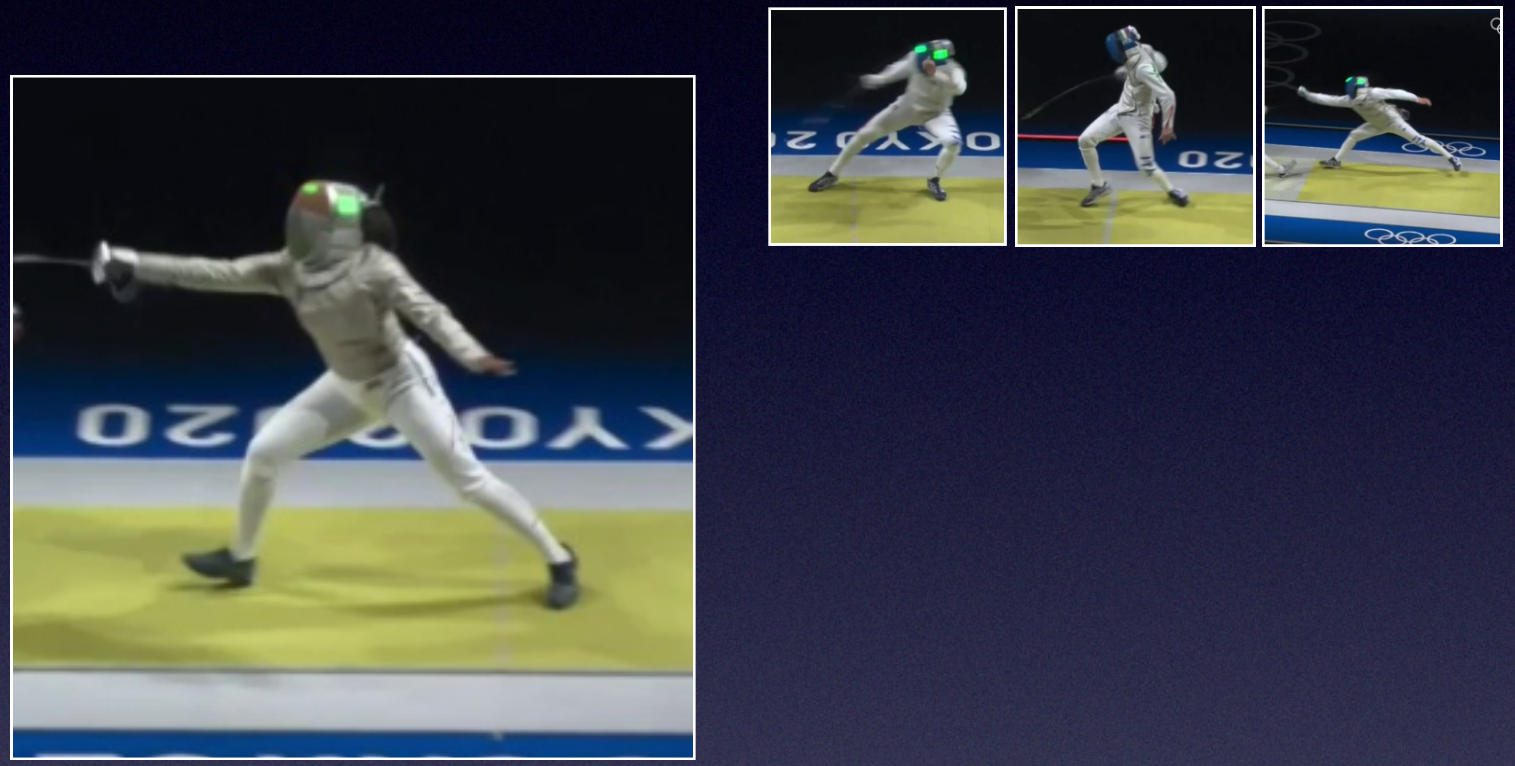
Image-to-image translation problem

Previous Solution: StyleGAN3

Training

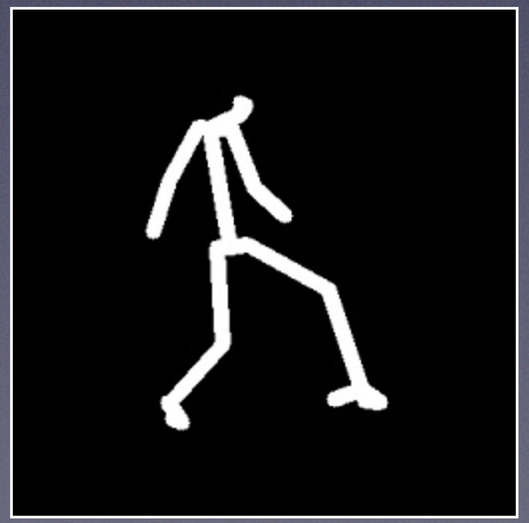


Input: pose sample

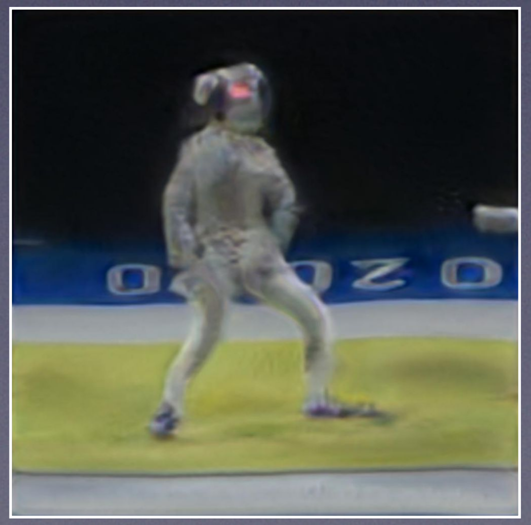


Output: ground truth

Prediction

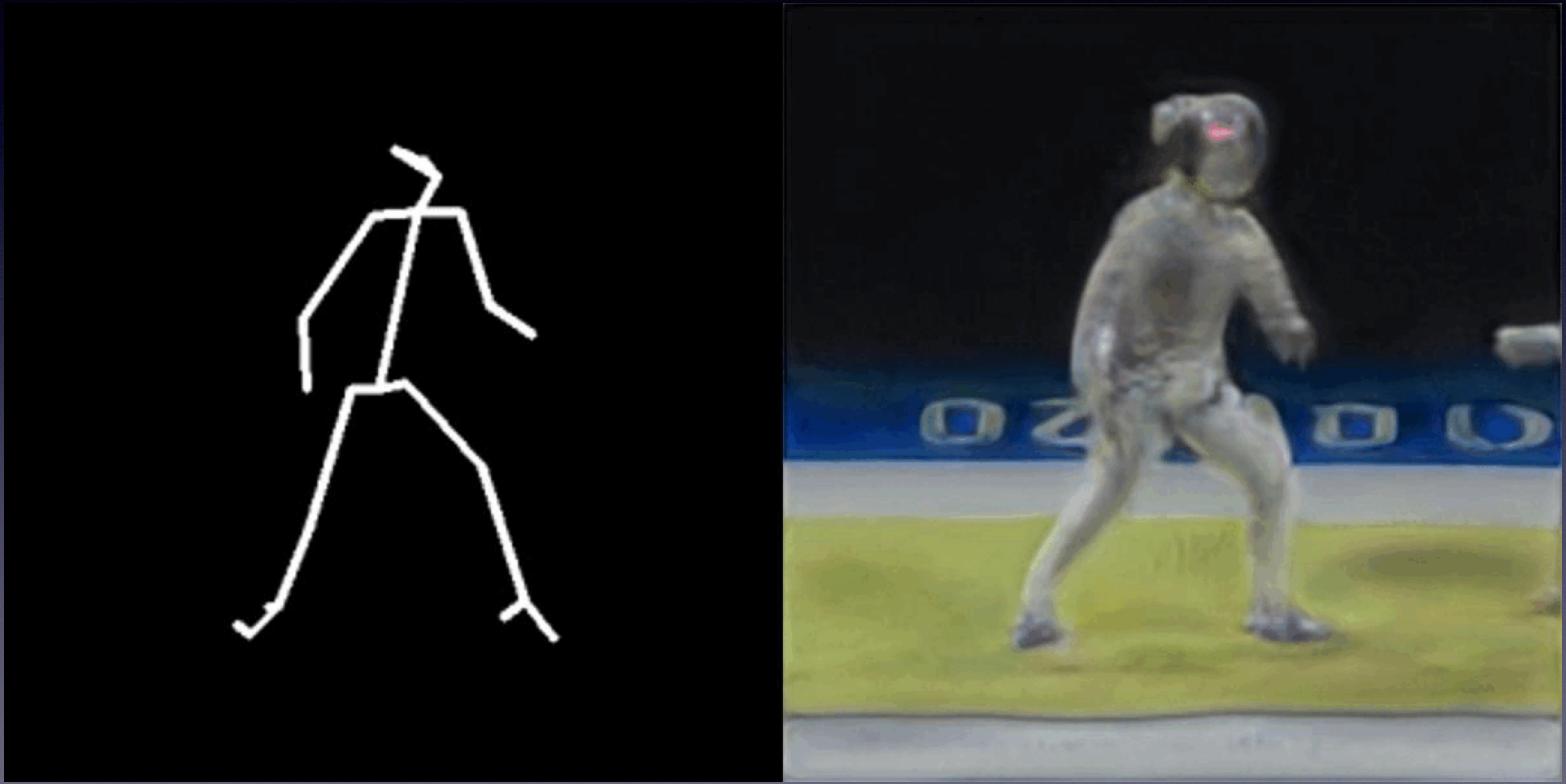


Input: arbitrary pose

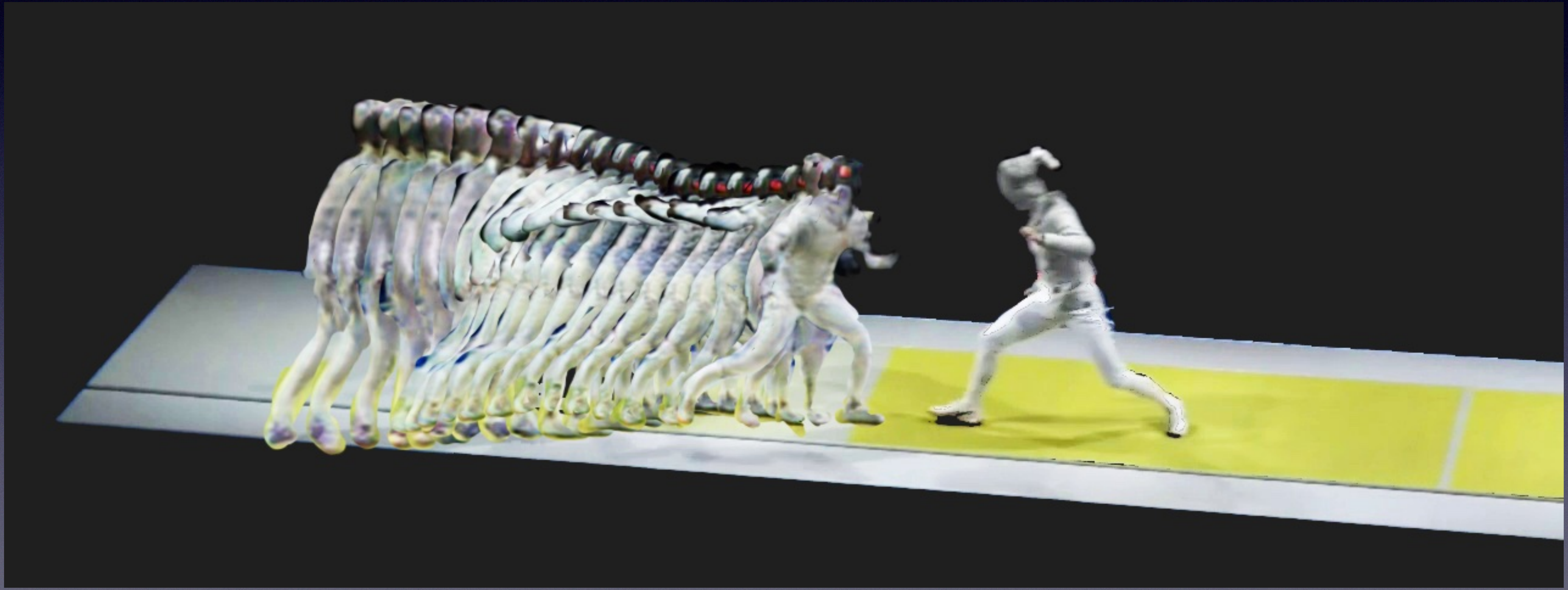


Prediction: pose image

Previous Solution: StyleGAN3 - Results



Previous Solution: StyleGAN3 - Results



Previous Solution: StyleGAN3 - Limitations

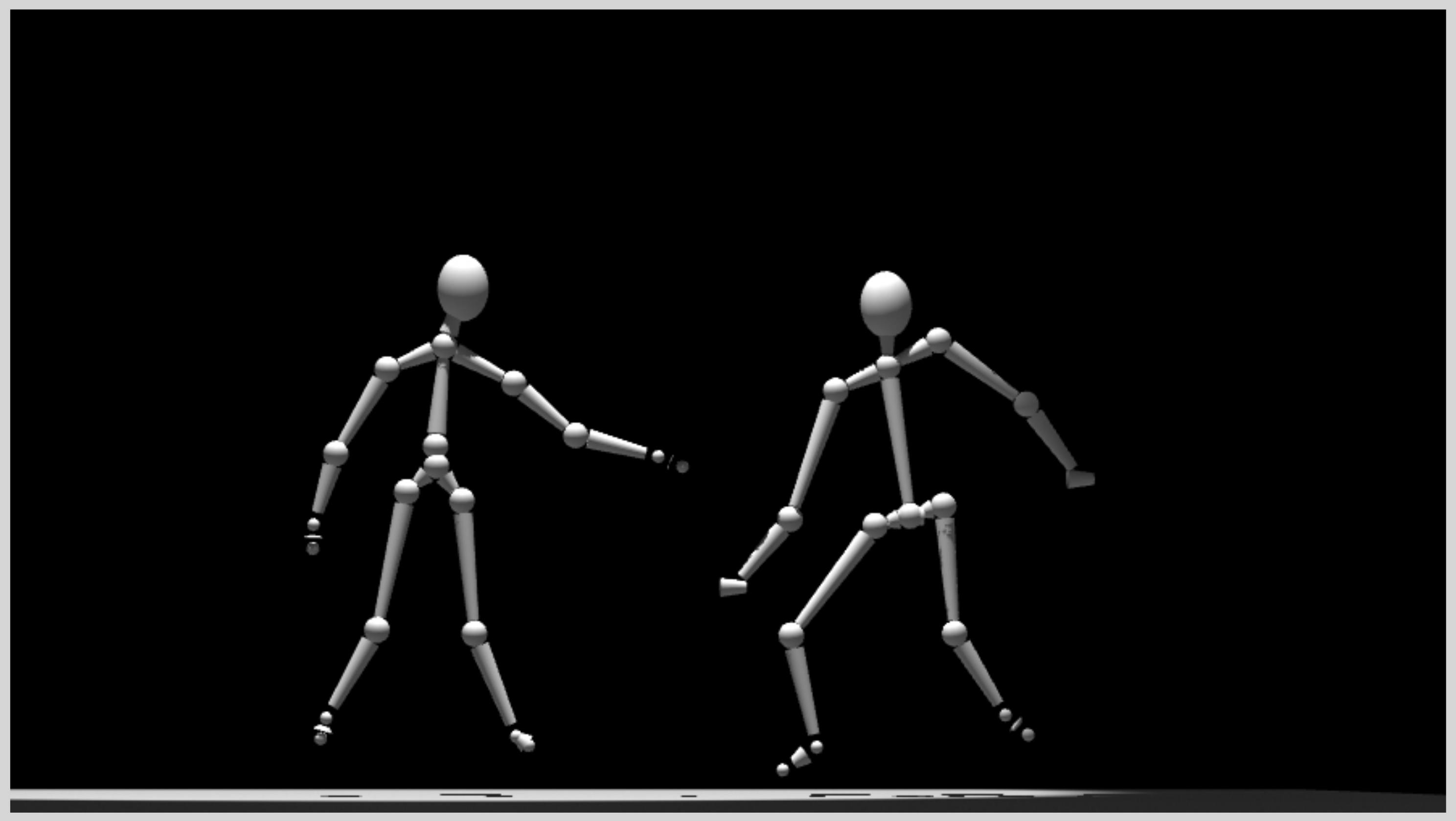
- **Difficulty to train:** 3 days of training time; data has to be calibrated
- **Limited Generalization**: unfamiliar pose could results in artifacts in produced images
- **Complicated workflow:** this process generates a squared photo for every pose individually. On top of that, I need to 1) project every photo to the correct area in the final canvas, 2) remove its background, and finally 3) blend all photos as a multi-exposure.

Solution with Diffusion Model

“Two people in pure white fencer suit, Fencing Games Scene, sports magazine”

Text Prompt

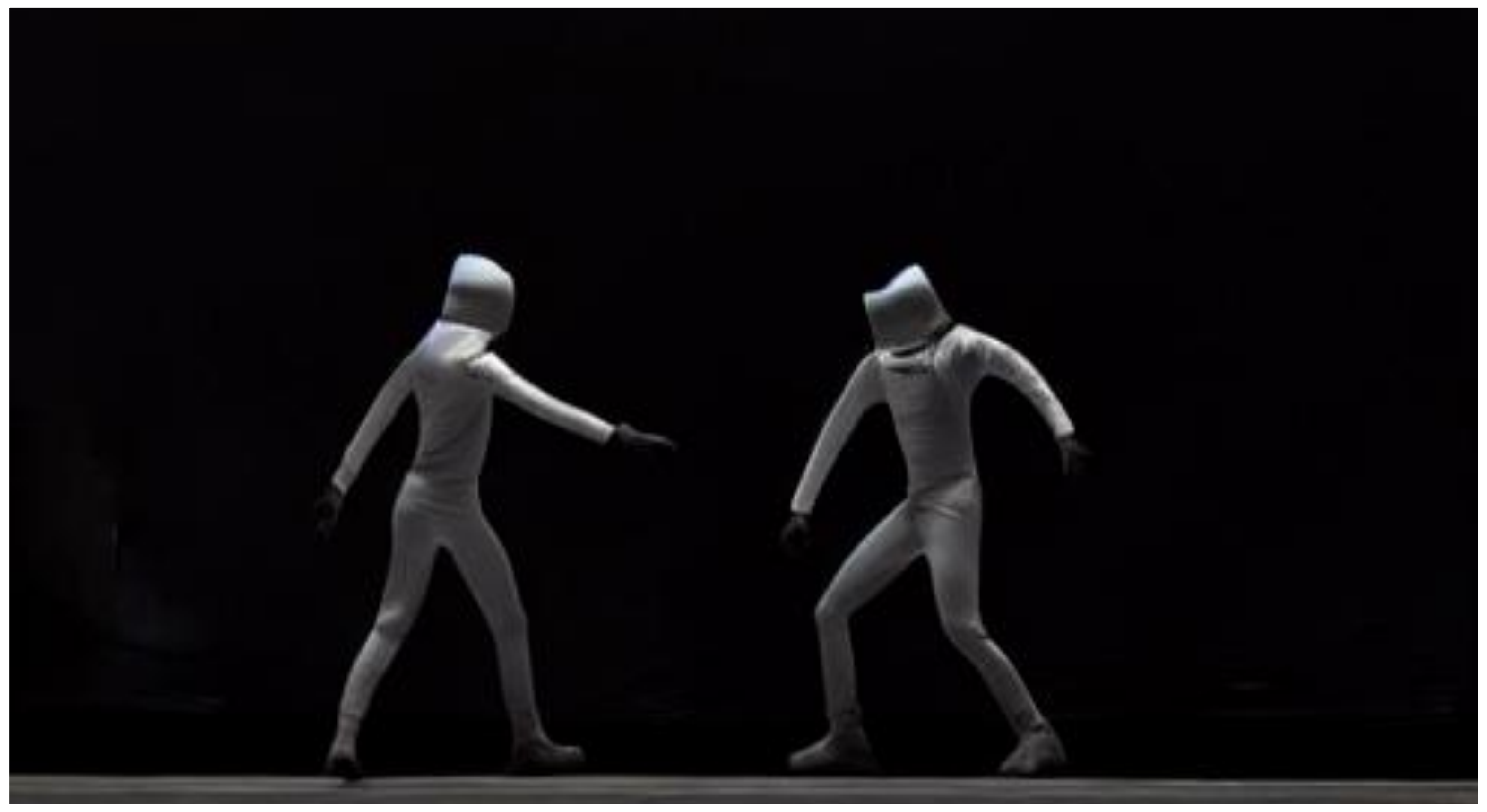
+



Stick-figure image

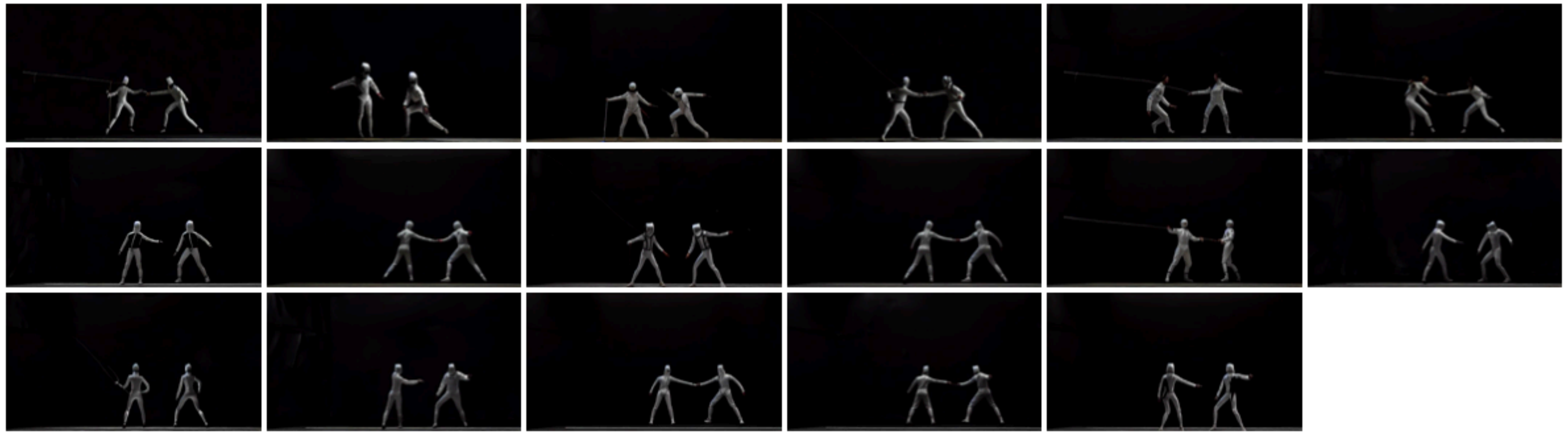
Stable Diffusion

→



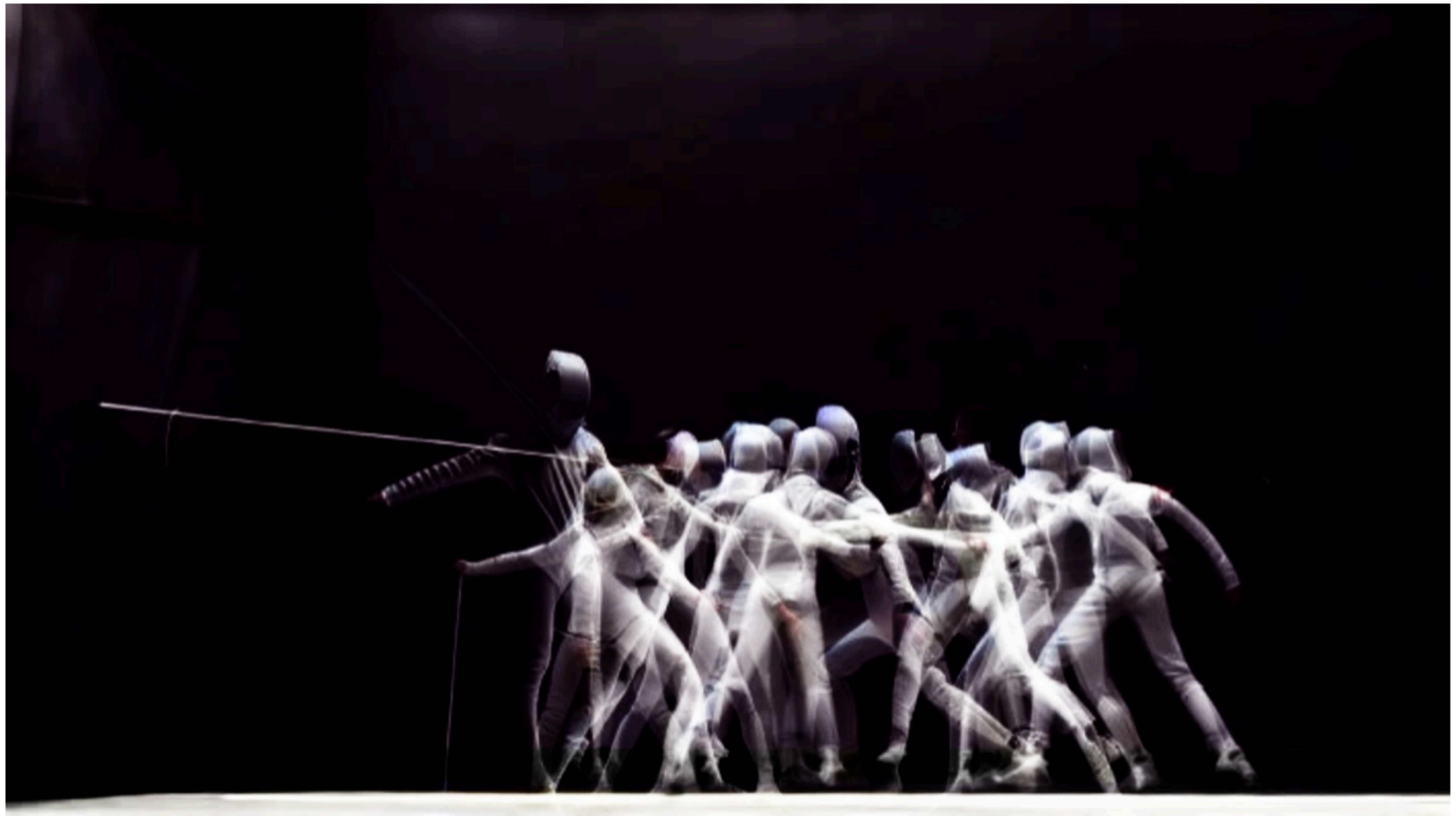
Synthetic Photo

Series of Photos generated by Stable Diffusion from the screenshots and the text prompts



Blend as multi-exposure

Chronophotograph-styled result



Solution with Diffusers - Discussions

Improvements

- **Model training not required** for a given style or a new style; style is entirely controlled by the text prompt
- **Simplified workflow**: entire image of two fencers are generated and blended without extra step of back projection.

Limitations

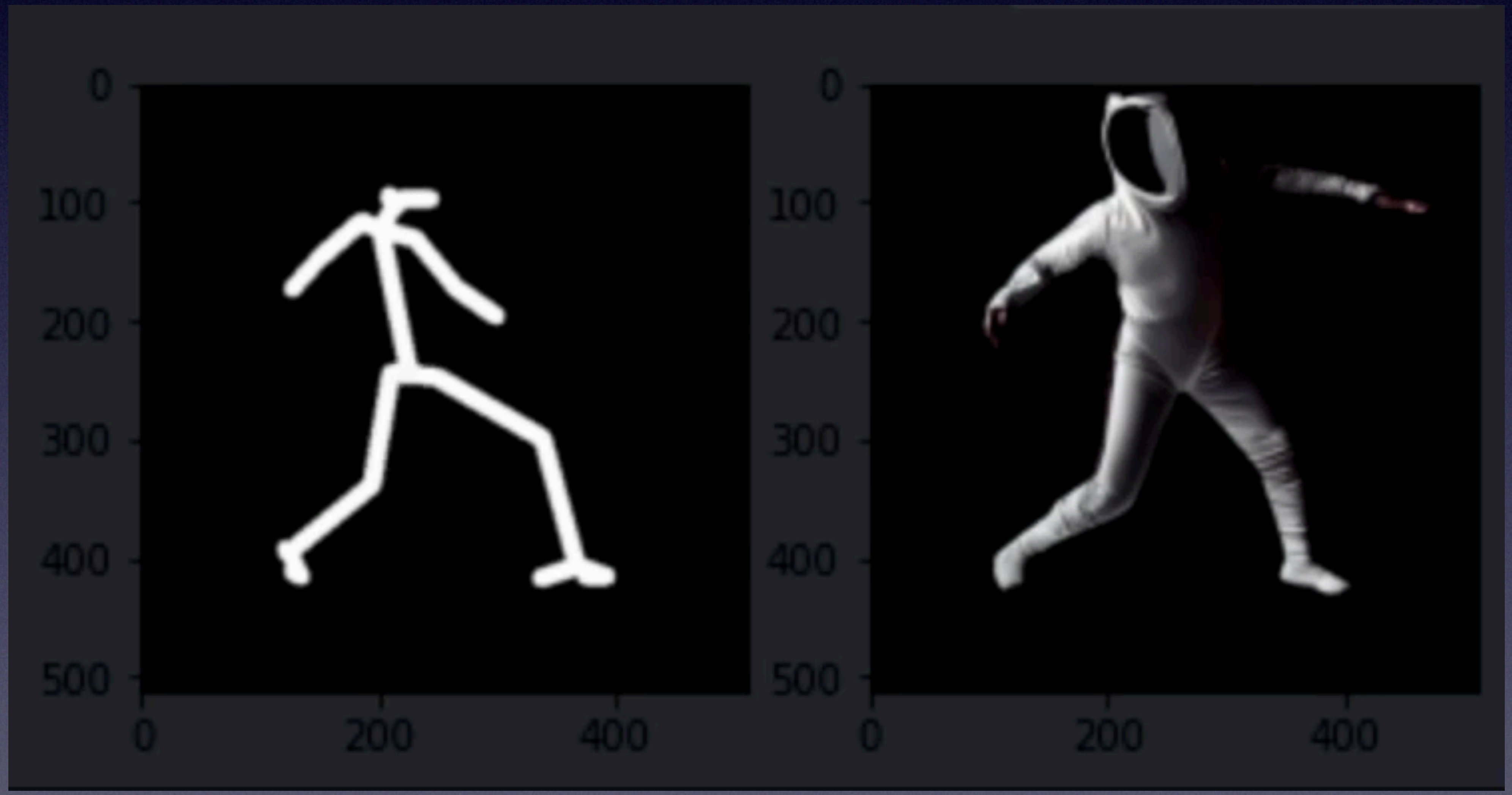
- **Slow**: a 12-frame-blended chronophotograph takes 2 minutes
- **Consistency**: the generated “fencer” is not always wearing the same suits

StyleGAN vs Diffusion Model

	Image Quality	Limited Data	Consistency	Style Control	Ease to use
StyleGAN	Slightly worse	Barely enough	**Better**	No overall style change	**Fast** + longer workflow
Diffusion Model	Slightly better	**No training**	Worse	**Much Better**	Slow + **Simpler workflow**

Other trials with Diffusion Model

Patch translation (for higher resolution)



1. Generate Image patch

2. Project patches to the canvas and blend

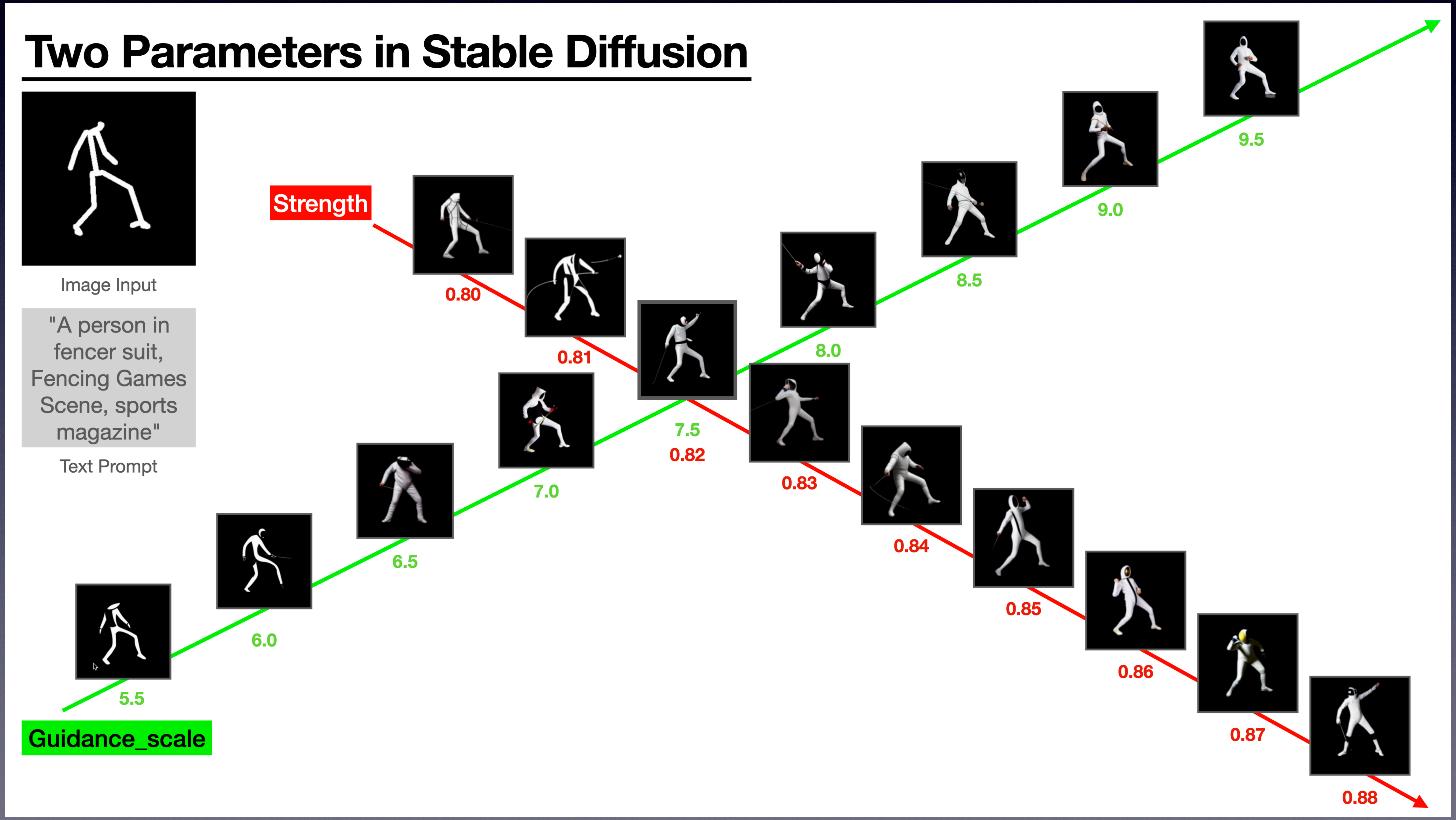
Other trials with Diffusion Model

Patch translation (for higher resolution)

Choosing different parameters:

- Low Strength -> stick to the original pose too much, limited variation, trying to fit a body to the stick-figure;
- High Strength -> drift away from the **original pose**
- Low guidance_scale -> stick to the original **style**
- High guidance_scale -> match the **style described in text prompt** scale better, such as keywords of “fencing”, “sports magazine”, etc.

There has to be a **tradeoff** between the **pose fitness** and the **style fitness**.

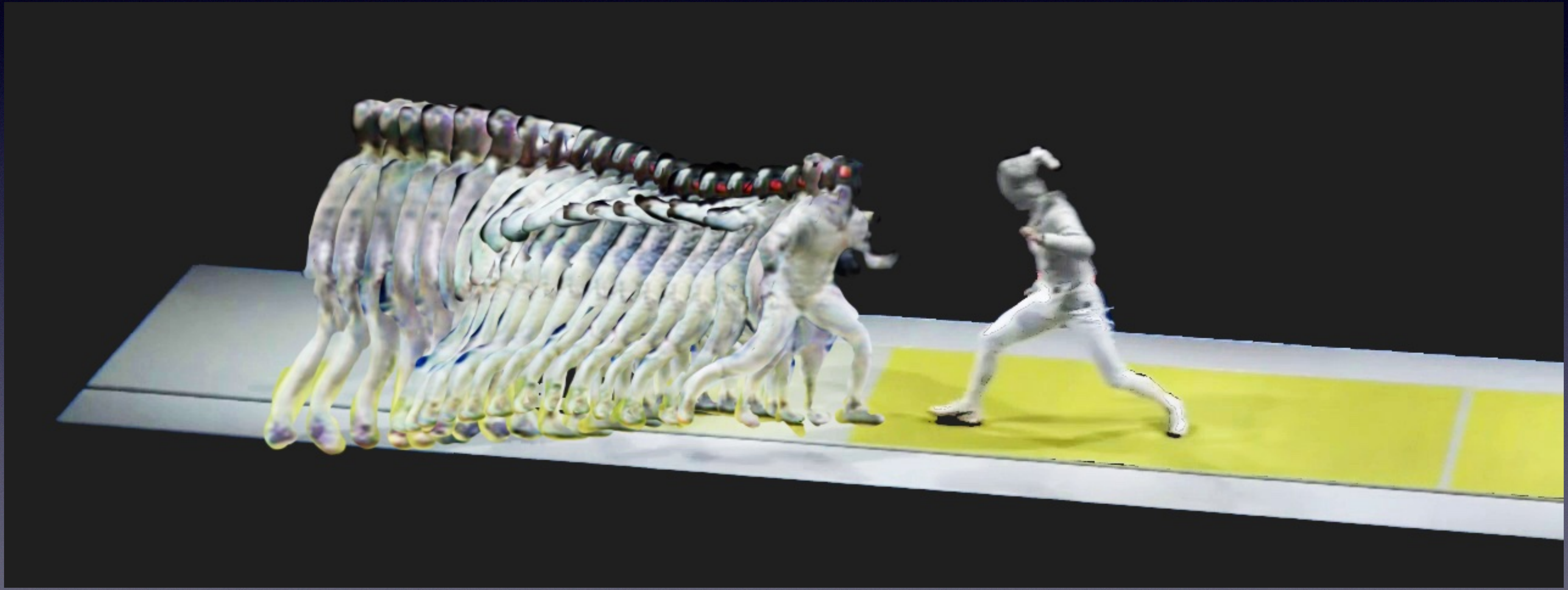


Two Parameters in Stable Diffusion

"A person in fencer suit, Fencing Games Scene, sports magazine"

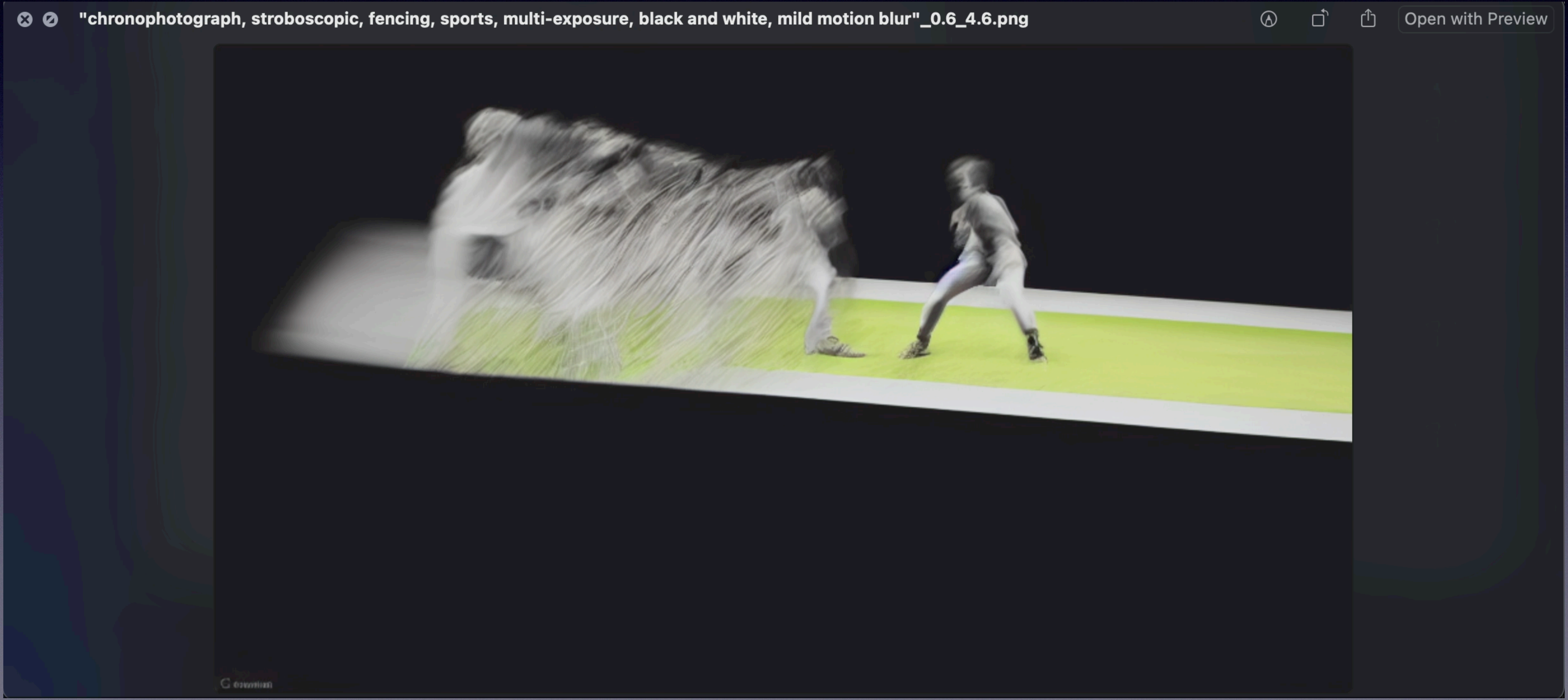
Other trials with Diffusion Model

Skeleton array translation (faster speed)



Other trials with Diffusion Model

Skeleton array translation (faster speed)



"chronophotograph, stroboscopic, fencing, sports, multi-exposure, black and white, mild motion blur"_0.6_4.6.png

Future Improvements

- Consistency: tweaking parameters; using algorithm such as DreamBooth to fine-tuned the model to specialize at generating fencing images with consistent style; create domain specific encoder to better guidance the diffusion model.

- Speed: optimizing the workflow, such as inserting upscaling module; reducing the steps; using multiple computers or cloud GPU to render in parallel.