

We are starting at 18:45 !

# 45th Vienna Deep Learning Meetup

## Intro to Neural Radiance Fields

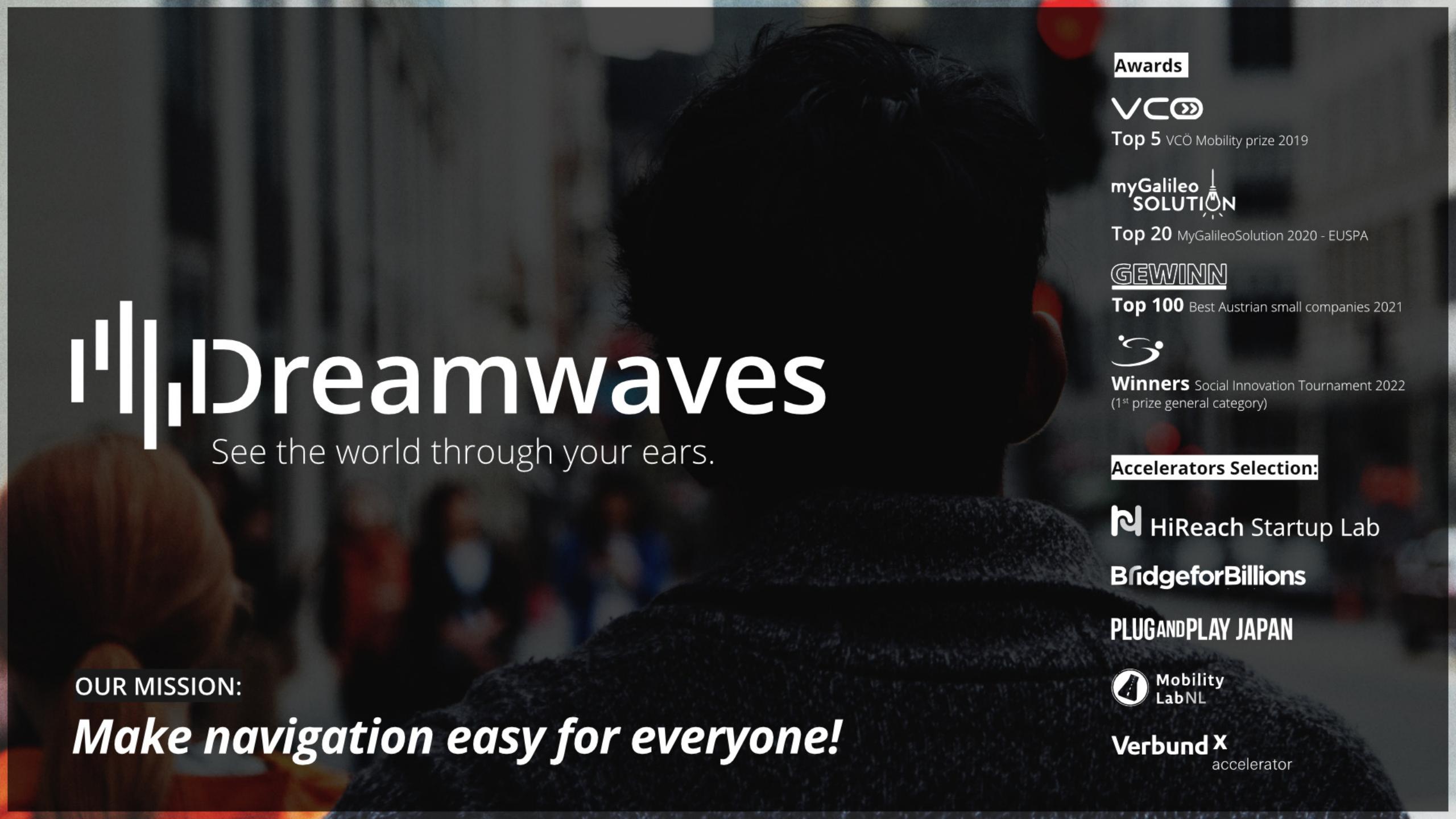
Lukas Zimmermann, dreamwaves.io

## Announcements

Break

## How does Stable Diffusion work?

Michael Pieler, contextflow.com



# Dreamwaves

See the world through your ears.

## OUR MISSION:

***Make navigation easy for everyone!***

## Awards



Top 5 VCÖ Mobility prize 2019



Top 20 MyGalileoSolution 2020 - EUSPA

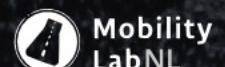


Top 100 Best Austrian small companies 2021



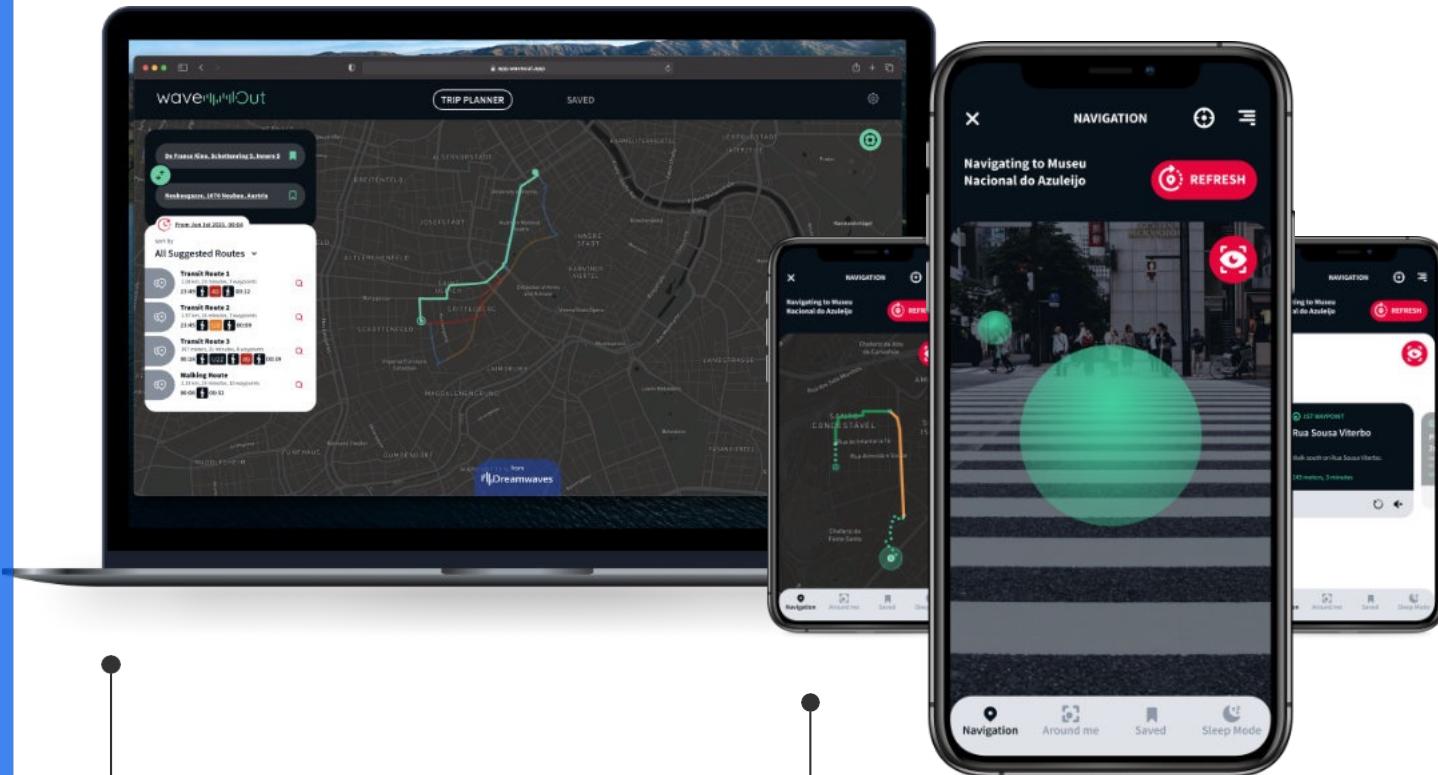
Winners Social Innovation Tournament 2022  
(1<sup>st</sup> prize general category)

## Accelerators Selection:



# waveOut - navigation App

Intuitive hands-free audio augmented reality



waveOut  
WEB PLANNER

waveOut  
NAVIGATION SOFTWARE



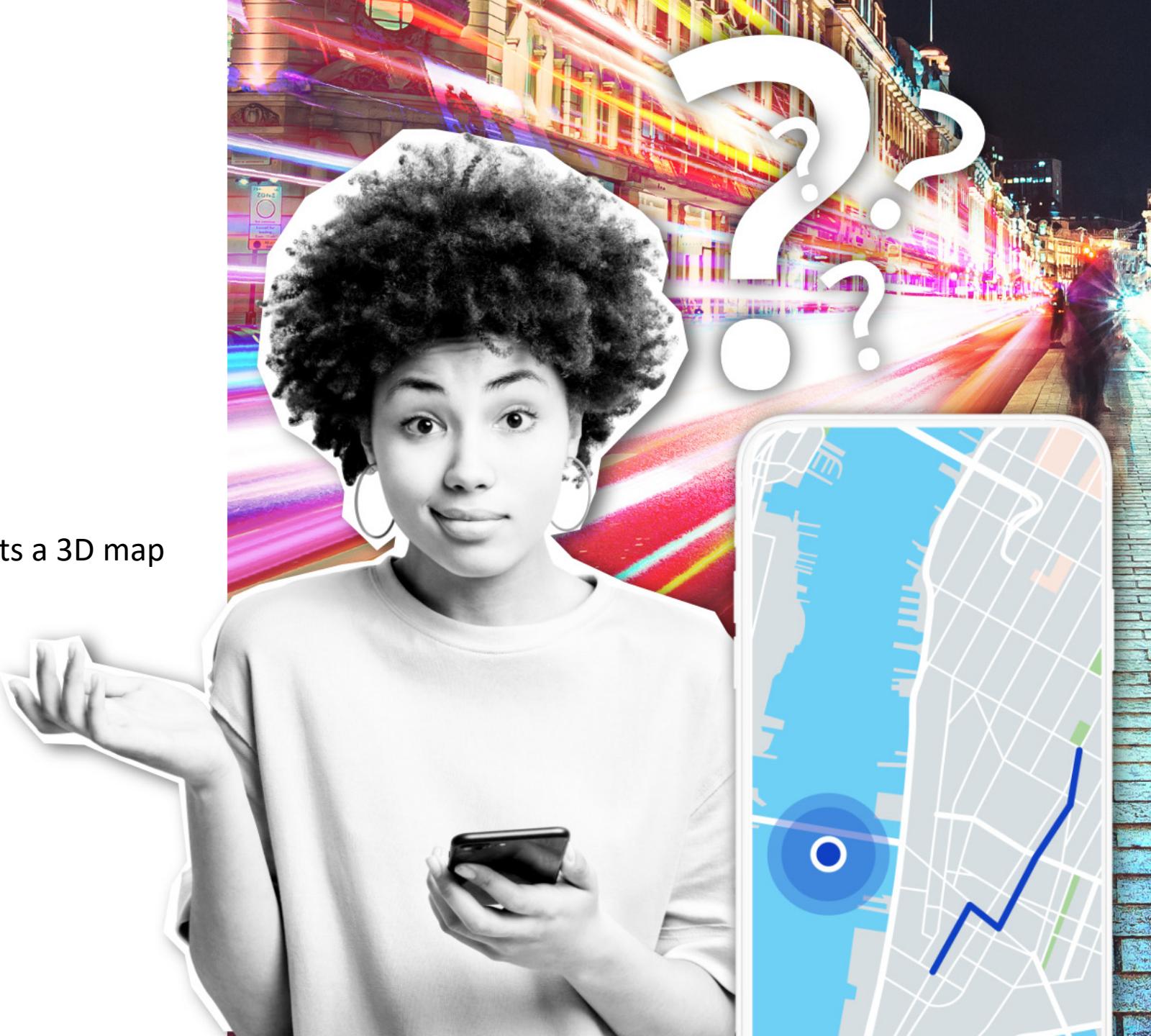
Bose  
AR Frames

Sony  
LinkBuds

**RECOMMENDED HEARABLES**  
MORE IMMERSIVE, EARS ARE NOT BLOCKED

# Motivation

- Accurate navigation requires:
  - Global positioning (e.g. GPS)
  - Local positioning (e.g. VIO)
- Local positioning requires for best results a 3D map
  - Alignment with the map



# Motivation

- Accurate navigation requires:
  - Global positioning (e.g. GPS)
  - Local positioning (e.g. VIO)
- Local positioning requires for best results a 3D map
  - Alignment with the map



# Neural Radiance Fields (NeRF)



# What does NeRF have to do with Augmented Reality?

- Place an object in any virtual scene with mixed reality
  - Good depth estimation
  - Knowledge of the scene representation from different camera positions

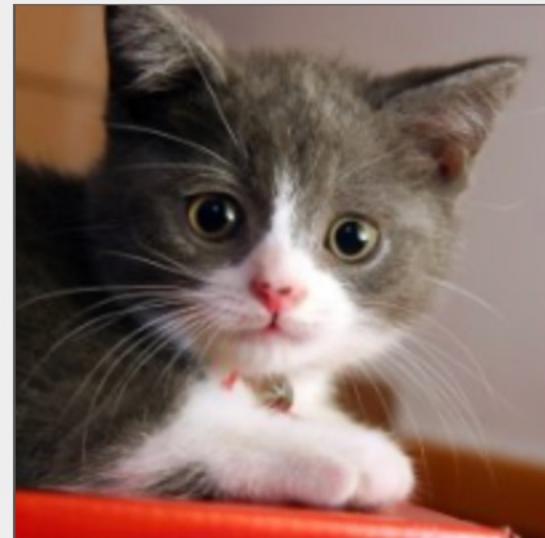


<https://www.matthewtancik.com/nerf>

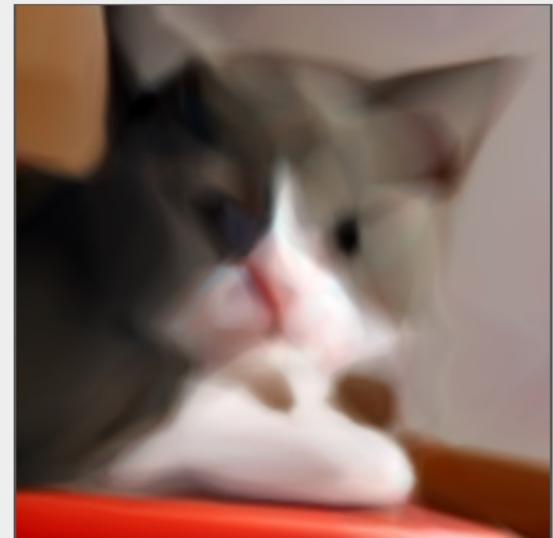
# Modern solution for an old problem

- Andrej Karpathy build back in his PhD days a neural network engine for the browser
- One show example was overfitting a neural network to decode an image
- Input are the coordinates of the single pixels
- Output predictions are the RGB colors of the image

Original Image



Neural Network output

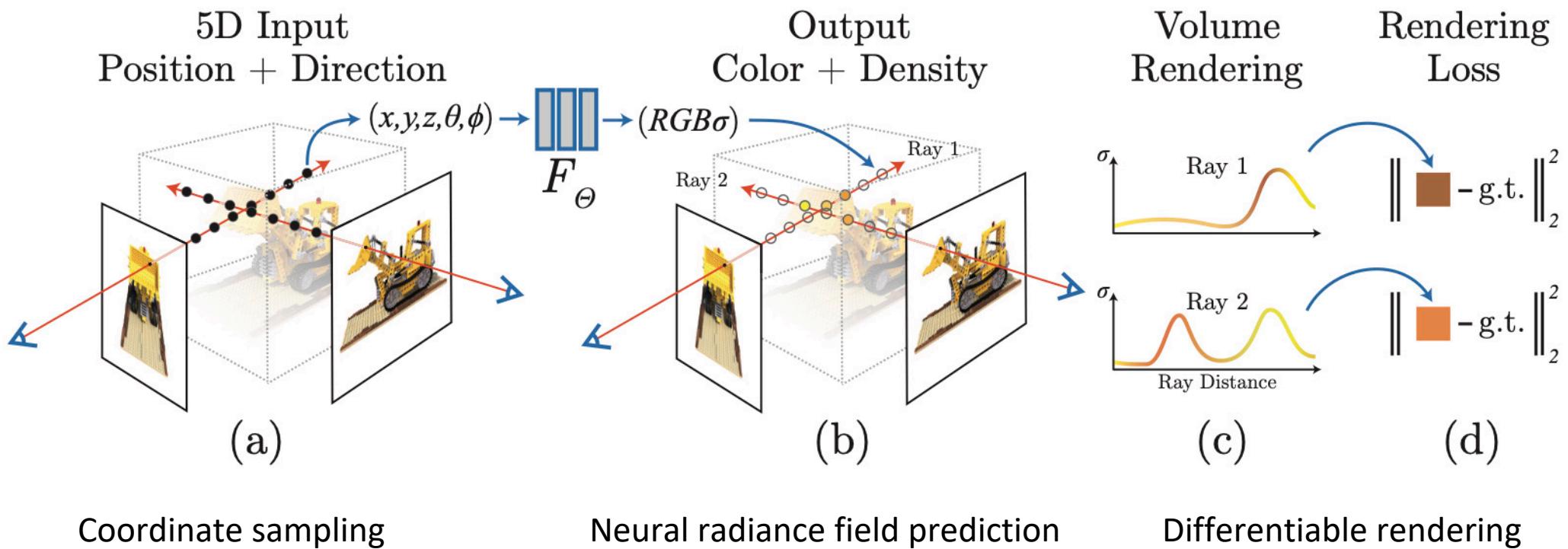


# Another unsolved problem...

- Considering a video taken or multiple images from a scene
  - No view from different angles possible
- Possible to reconstruct sparse scenes via
  - Structure from motion (SfM)
  - Simultaneous localization and mapping (SLAM)



# Now how is it done?

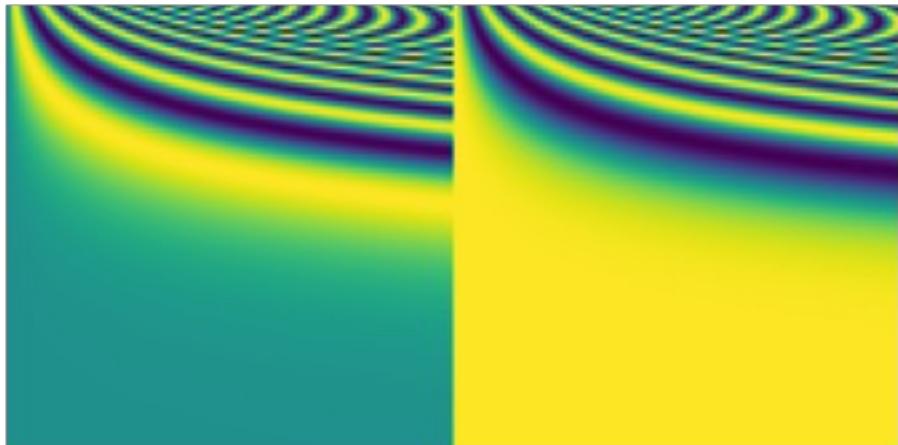


# Let's check the parts together!

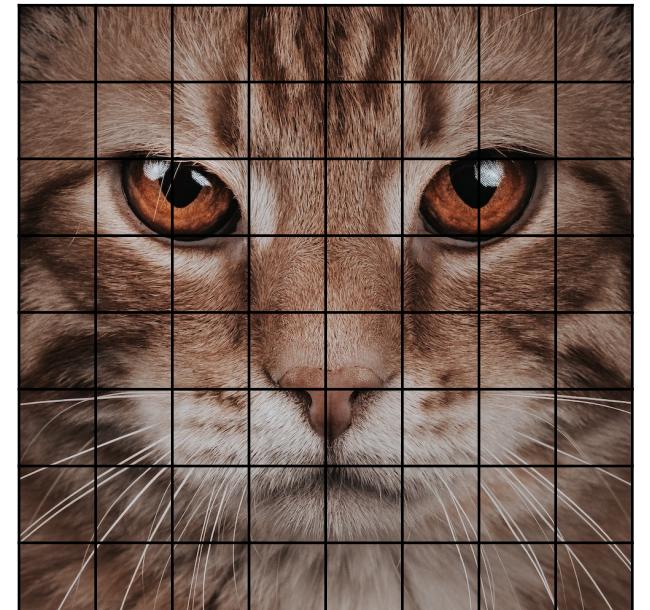
- Coordinate sampling
  - Get camera position
  - Sampling along the rays of a camera
  - Decoding input
- Neural network design
  - Size and activation functions
- Rendering function
  - Dependent on the defined problem

# Coordinate sampling

- Cartesian coordinate system ( $x, y$ )
  - Description is not informative
- Enhance information content
  - Feature transforms (e.g. Fourier transform)
  - $\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \dots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p))$
  - Increase input dimension dependent of amount of fourier features



$p(0,0)$



$p(x, y)$

# Differentiable rendering

- Starting with the general definition of classical volume rendering with  $\mathbf{r}(t) = \mathbf{o} + \mathbf{d}t$

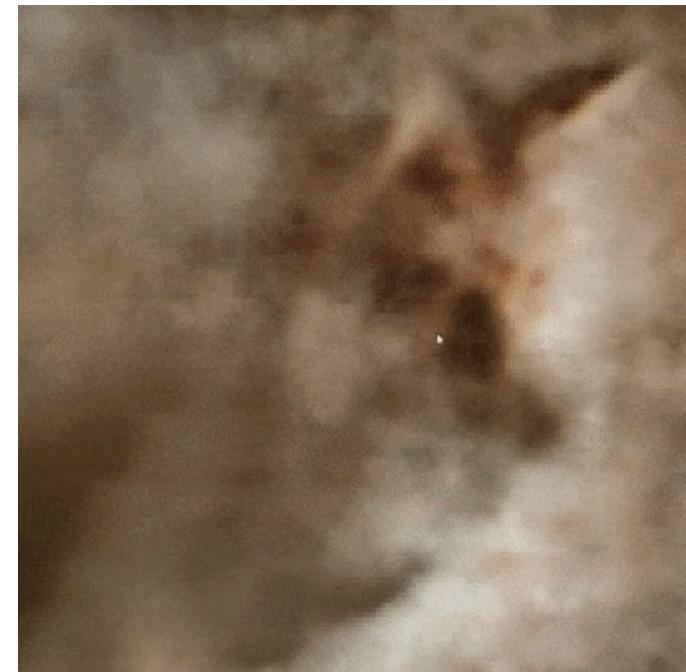
$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt$$
$$T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds\right)$$

- Where the **density** and the **color** are predicted in our case by a neural network
- The integral is discretized by summing  $N$  samples along the ray to a respective position:

$$\hat{C}(r) = \sum T_i (1 - \exp(-\sigma_i \delta_i) \mathbf{c}_i); T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

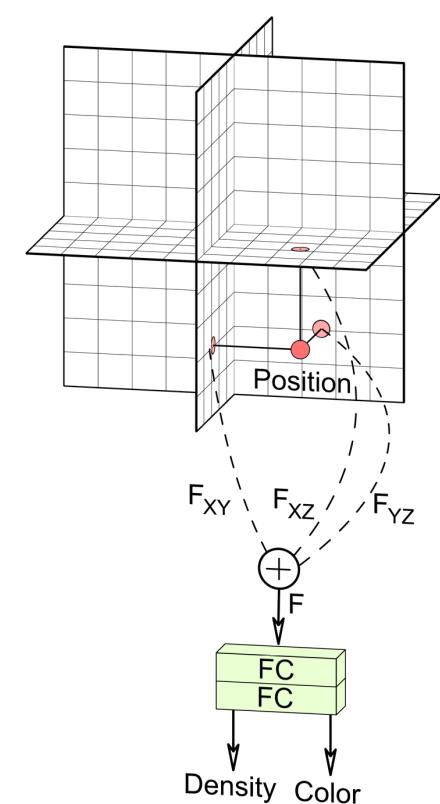
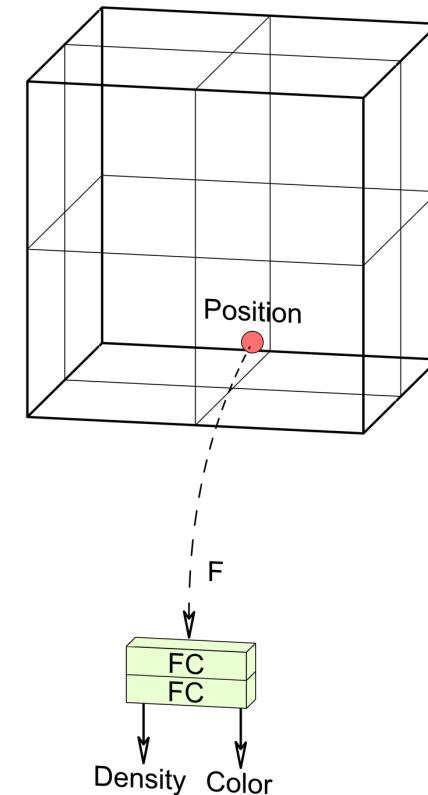
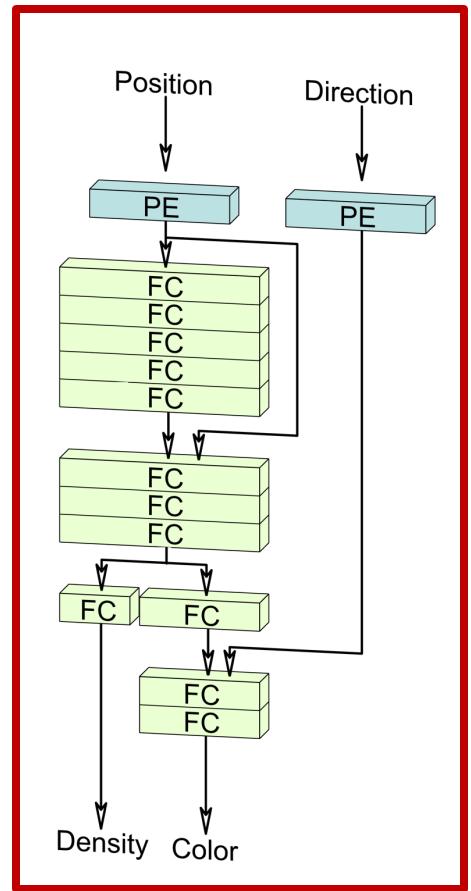
# Differentiable rendering

- Drawbacks:
  - Super slow training with implicit representation (>1d)
  - Super slow image sampling strategy (>30secs/frame)
- How to overcome this challenge?
  - Image dimension (1024 x 512)
  - Video frame rate (60fps)
  - Ray samples (128points)
  - **~4 billion points** for a single second video!



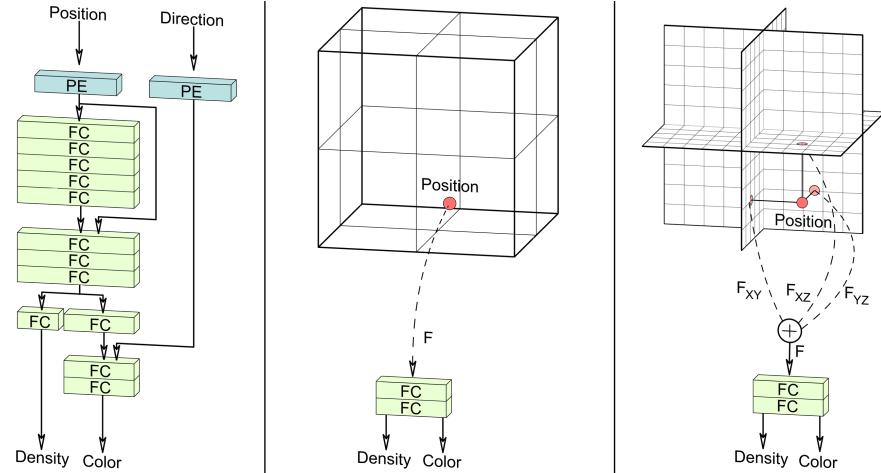
# Secret ingredient

- Biggest difference is how to encode the input signal (position)
- Different methods:
  - Encoding position solely with model
  - Grid based (Penoxels)
  - Orthogonal planes (EG3D)
  - Hash encoding (Instant NGP)
- Reduce memory cost required by the network
  - Latent vector from learned map



# The neural network

- Nothing to add here sorry... :(
- In general it is a simple Multi Layer Perceptron
  - Originally: view dependency
- Most important finding:
  - Instead of using one neural network use two
  - Periodic activation functions can help!  
(SIREN)



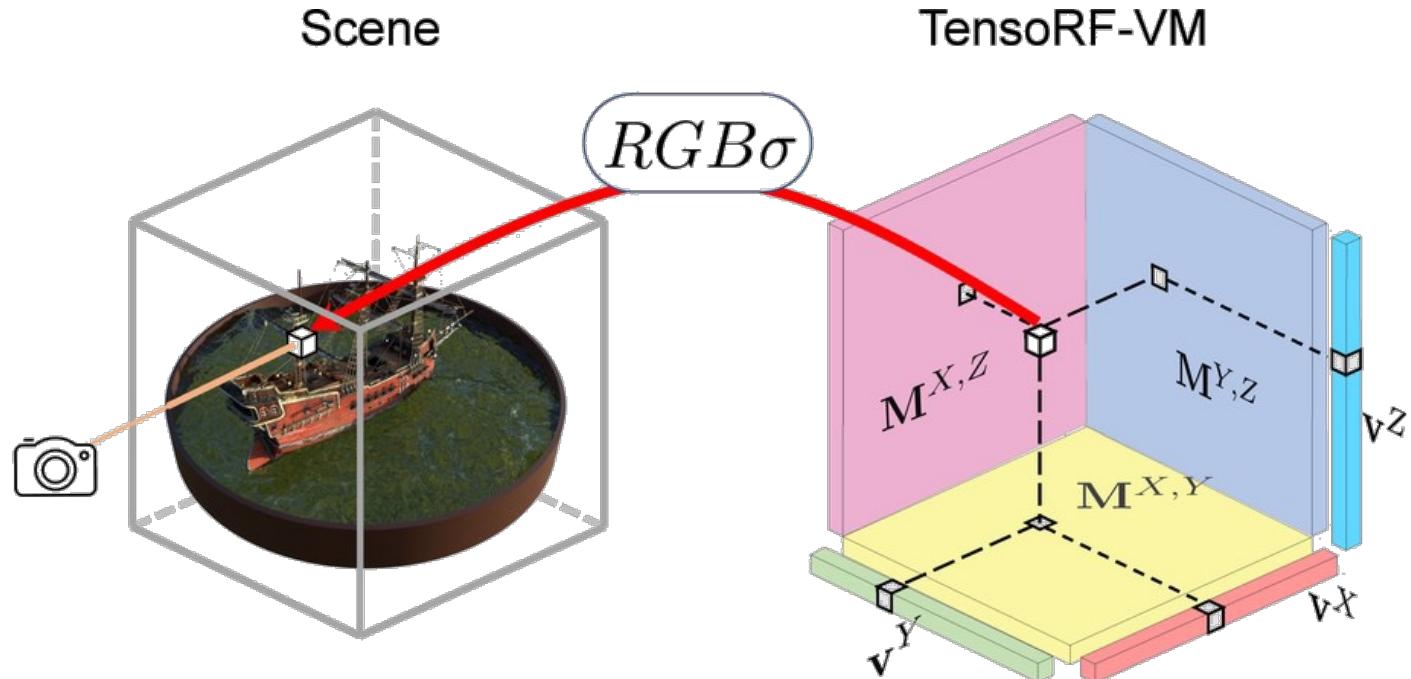
# Current research

- Increasing speed for rendering (Instant NGP)
  - <https://nvlabs.github.io/instant-ngp>



# Current research

- Decreasing memory requirement (TensoRF)
  - <https://apchenstu.github.io/TensoRF/>



# Current research

- Scaling (BlockNeRF)
  - <https://waymo.com/research/block-nerf/>



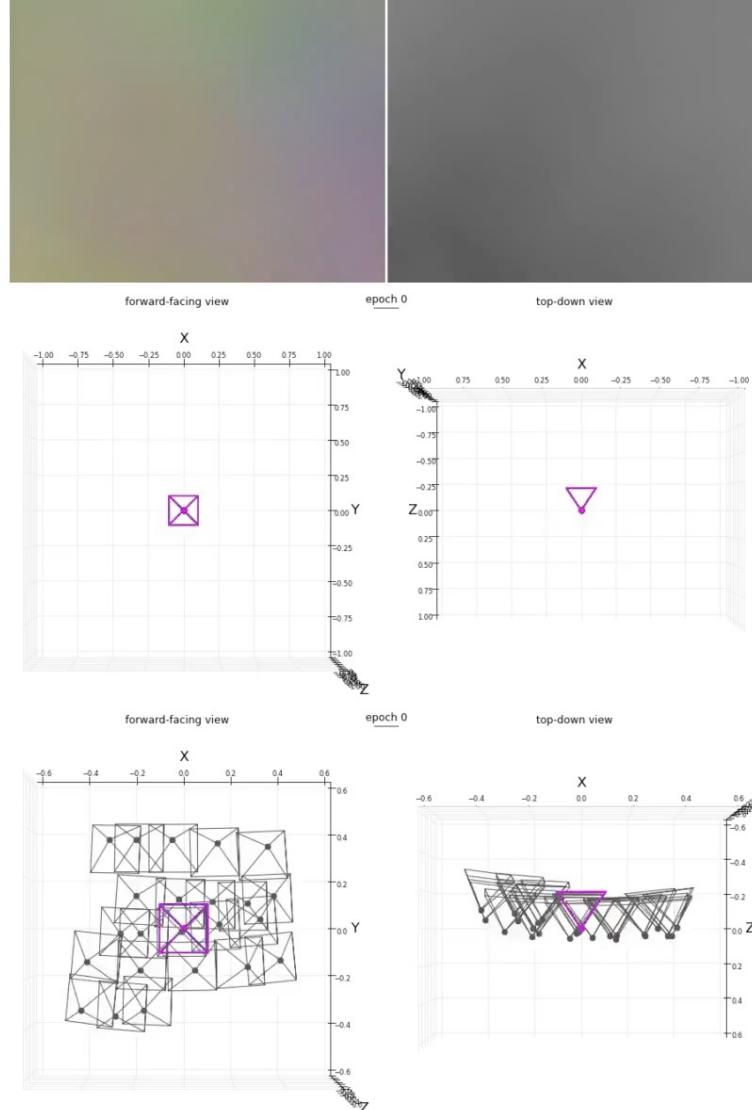
# Current research

- Position aware image generation (EG3D)
  - <https://nvlabs.github.io/eg3d/>



# Current research

- Camera pose optimization (BARF)
  - <https://chenhsuanlin.bitbucket.io/bundle-adjusting-NeRF>



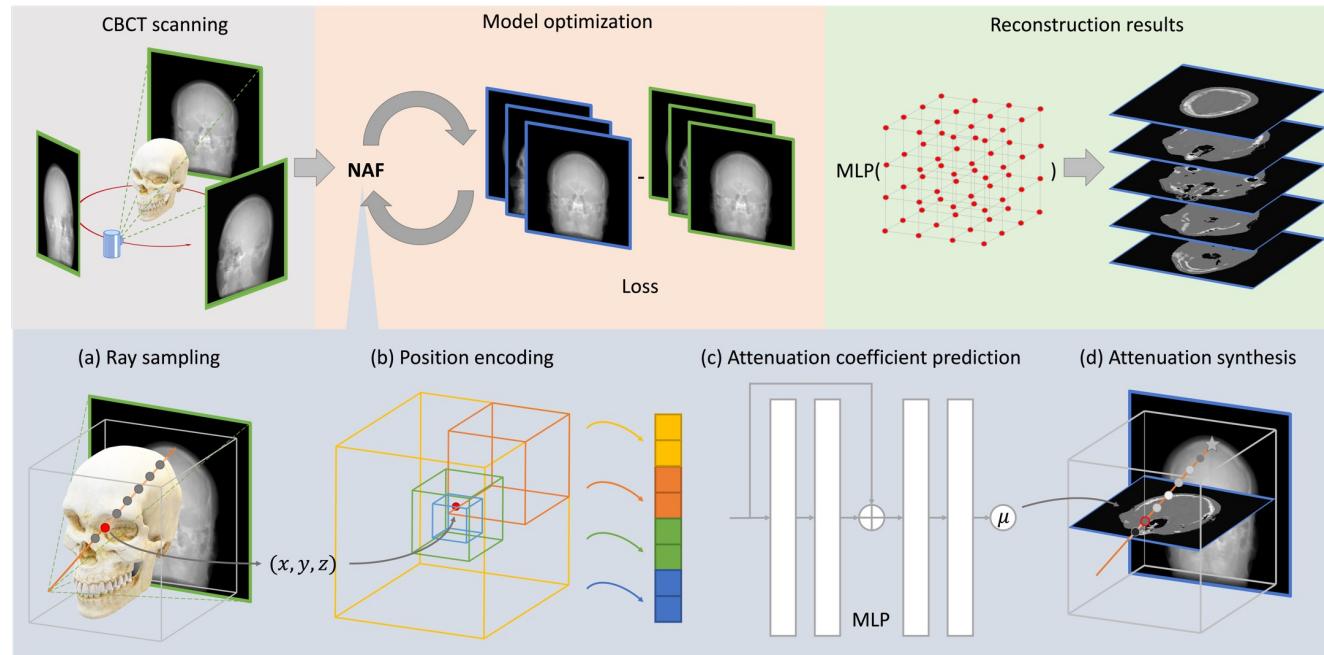
# Current research

- SLAM approach with neural rendering
  - <https://pengsongyou.github.io/nice-slam>
  - Similar to BARF but with sequential image inputs
  - Define camera pose from RGB-D images

# Application in medicine

- The differentiable rendering can be nearly directly applied for CT reconstruction

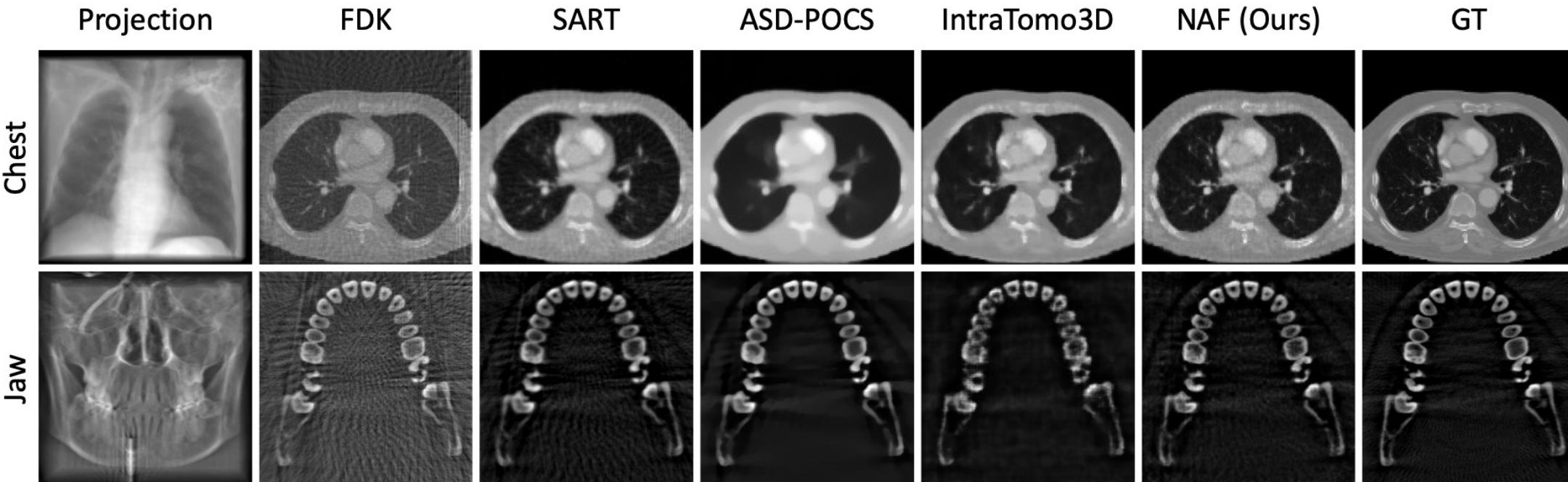
$$-\ln\left(\frac{I}{I_0}\right) = \sum \mu_i d_i$$



# Application in medicine

- The differentiable rendering can be nearly directly applied for CT reconstruction

$$-\ln\left(\frac{I}{I_0}\right) = \sum \mu_i d_i$$



# Conclusion and future outlook

- NeRF representation has been a massive research topic the past 2 years
- Novel view synthesis and the improved understanding of 3D geometry by a neural network will help to build more accurate “ground truth” data
- For all interested to start:
  - <https://github.com/NVIDIAGameWorks/kaolin-wisp>



# Meet the Team

Intuitive & advanced technology can make a difference in people's lives.



**HUGO FURTADO**  
CEO  
**PhD in medical Augmented Reality**  
10+ years of experience in R&D and innovation management to translate scientific results into products  
Software Developer @ **Altitude Software** (PT)  
Researcher @ **CERN** (CH)  
**Marie Curie** PhD Research Fellow (SLO)  
Post-Doc @ **Medical University of Vienna** (AT)



**ALESSANDRO MULLONI**  
CTO  
**PhD on AR navigation systems**  
15+ years developing interactive mobile experiences  
PhD @ **TUGraz** (AT)  
Senior Engineer @ **Qualcomm Research** (AT)  
Freelancer Mobile AR apps



**LILIANA CALAPEZ**  
CXO & Product Leader  
Degree in Communication & multimedia design  
UI and UX designer for 10+ years  
Communication design, Uii&UX @ **Bookpaper Design** (PT)  
**addsolutions** (PT)  
**studio novo** (AT)  
**E2 Communications GmbH** (AT)



**ELKE MATHEISS**  
COO & Data Analytics  
Degree in psychology, BsC Software Design  
10+ years in project management and quantitative research  
Usability Engineer @ **CURE** (AT)  
Usability Scientist @ **AIT** (AT)  
Freelancer in Data Science



**LUKAS ZIMMERMANN**  
CSO & Machine Learning  
**PhD in Deep-Learning for medical imaging**  
5+ years experience in computer vision  
Software Developer @ **Medical University of Vienna** (AT)  
Software Developer @ **AIT** (AT)



**GIL CAMPOS**  
Software Developer  
Master's Degree, Computer Science  
2+ years experience in web applications development  
Software Developer @ **Collab** (PT)  
Mobile Applications Developer @ **Build Up Labs** (PT)



**SAM PARKE-WOLF**  
Software Developer  
BSc Music Computing  
Software Developer @ **Miniclip** (PT)  
Junior Software Developer @ **Vochlea Music** (UK)



**MICHAEL BUSBOOM**  
**Marketing Advisor**  
**Marketing evaluation of product and technology (AT)**  
Marketing and evaluation of our product and technology 20+ years of experience in assistive technology for the blind. Member of the Technical Commission of the European Blind Union



**MIKEL ECHEGOYEN**  
**Business Development Advisor**  
VP Business Development & Strategy at **Tieto** (FI)  
Business development, investments and strategic alliances for our company.VP for business development for Tieto Product Development Services. Focus on R&D services and system integration for IoT

## PARTNERS

Research and innovation:



Distributors and customers:  
 **Hilfsgemeinschaft**  
der Blinden und Sehbehinderten Österreichs

Commercial partner:  
 **YUNEX TRAFFIC**  
A Siemens Business

# Dreamwaves

*“Let's implement innovative use cases together!”*



Hugo Furtado CEO

[hugo@dreamwaves.io](mailto:hugo@dreamwaves.io)

+43 660 401 5739

[dreamwaves.io](http://dreamwaves.io)

Aug. 2022

# Job Openings

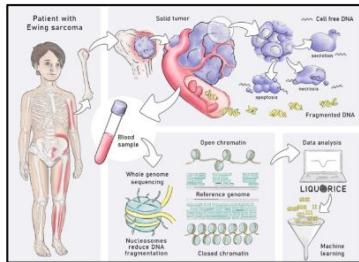
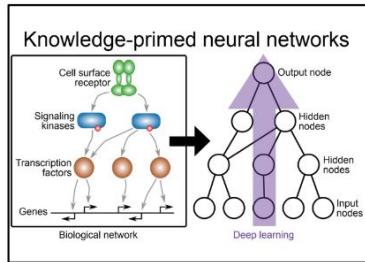
# PhD Position: Machine Learning in [Bio]Medicine

with applications in precision medicine, high-throughput biology, or synthetic biology.



[tinyurl.com/MedUniPhD](https://tinyurl.com/MedUniPhD)

## RESEARCH EXCERPT



## SOCIETAL IMPACT & FUNDING



## Computational Method Development

- Interpretable Deep Learning<sup>(1)</sup>
- Causal & Time Series Modeling
- Federated Machine Learning

## Proof-Of-Concept Applications

- Personalized Medicine<sup>(2)</sup>
- Systems Biology<sup>(2)</sup>
- Decision Support

## Dissemination and Impact

- Clinical Applications
- Creation of Startups
- Research-Centric Teaching

## Institute of Artificial Intelligence

- Director Christoph Bock<sup>(3)</sup>
- PhD funded for 4 years
- International Research stay

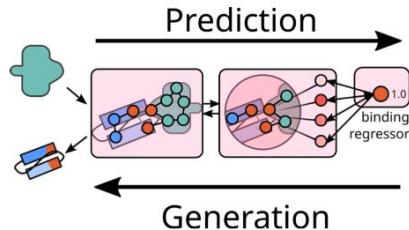
(1) Fortelny, 2020, Genome Biol. (2) Peneder, 2021, Nat. Comm. (3)<https://tinyurl.com/chrbock>

# Junior Group Leader: Machine Learning in [Bio]Medicine

## RESEARCH EXAMPLE



### Graph Deep Learning in Structural Biology



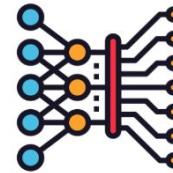
## RESEARCH OPPORTUNITIES



patient data  
high-throughput data  
knowledge-base



MEDICAL UNIVERSITY  
OF VIENNA  
Institute of Artificial Intelligence



NLP  
Geometric DL  
Bioinformatics

- **Relevance:** Shape the future of data-driven precision medicine
- **Freedom:** Drive your research in an entrepreneurial environment
- **Biology:** Work on the most complex and exciting system ever created

Required: Research expertise in STEM/Machine Learning (PhD)

Recommended: Experience in Biology/[Bio]Medicine



[tinyurl.com/MedUniFuture](http://tinyurl.com/MedUniFuture)



e l l i s  
European Laboratory for Learning and Intelligent Systems



MEDICAL UNIVERSITY  
OF VIENNA

CeMM

# WE ARE HIRING

---

We are looking for motivated talents to join our team in Vienna.

**SOFTWARE ENGINEER: COMPUTER VISION & DEEP LEARNING**

**EMBEDDED SOFTWARE ENGINEER: DEEP LEARNING**

**SOFTWARE ENGINEER: 3D PERCEPTION APPLICATIONS**

**DIPLOMA STUDENTS (VARIOUS TOPICS)**

**EMBEDDED SOFTWARE ARCHITECT**

**QUALITY ENGINEER**

**AUTOMOTIVE REQUIREMENT ENGINEER**

If you'd like to shape the future of automotive safety & driving experience with us, please let us know at [career@emotion3d.ai](mailto:career@emotion3d.ai)



---

We are searching for  
**Senior Machine Learning Researcher**  
full-time (f/m/d)



**contextflow**

See beyond a single case.



contextflow\_rad



contextflow

[contextflow.com](http://contextflow.com)



# OpenBioML

<https://OpenBioML.org>



OpenBioML

*OpenBioML is a grassroots open, collaborative research laboratory  
at the intersection of machine learning and biology  
for the community by the community.*

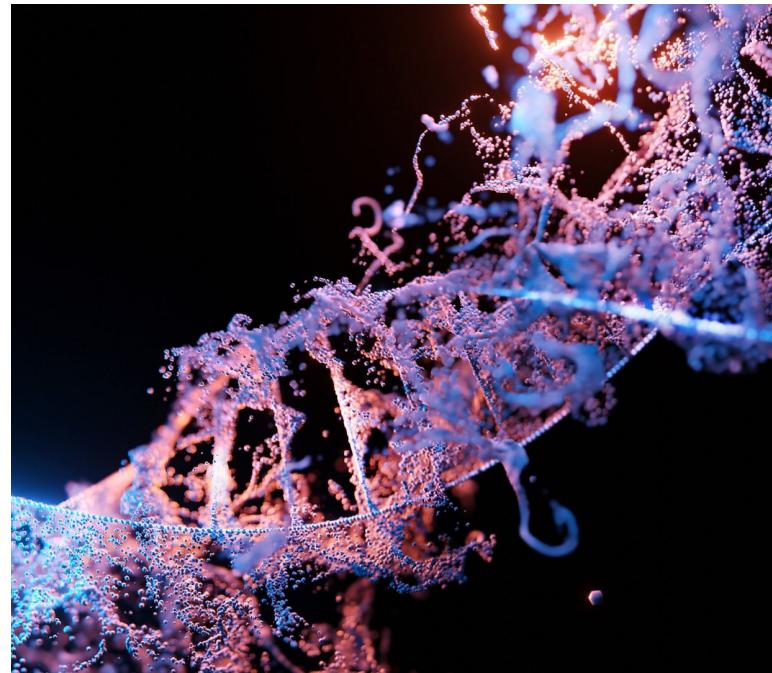
We are generously supported by [Stability AI](#).

# LibreFold

- A “collaborative research workshop” around AlphaFold 2-like protein structure prediction methods
- Set up OpenFold code base comparable to AlphaFold 2
- As open as possible:
  - all trained models will be released under a permissive license
  - all “tacit knowledge” like training logs, negative results, scripts and internal datasets will be made available as well
- Basis for new research!

# DNA diffusion

- Investigate the application and adaptation of recent diffusion models to genomics data.
- A collaborative project initiated by [Luca Pinello's lab](#) (MGH, Harvard Medical School, Broad Institute)





# OpenBioML

- You can find all the information on our website: <https://OpenBioML.org>
- Discord server: <https://discord.gg/GgDBFP8ZEt>
- Twitter: [@openbioml](https://twitter.com/openbioml)
- YouTube channel: [OpenBioML](https://www.youtube.com/c/OpenBioML) (bi-weekly journal club)
- Code: <https://github.com/OpenBioML>
- Project proposal:

<https://github.com/OpenBioML/project-proposal-template>

Join our community! 😊

# Break

# How does Stable Diffusion work?

Michael Pieler

# Stable Diffusion

- How does it work?
- Applications

# What is Stable Diffusion?

<https://www.reddit.com/r/StableDiffusion/>

# What is Stable Diffusion?



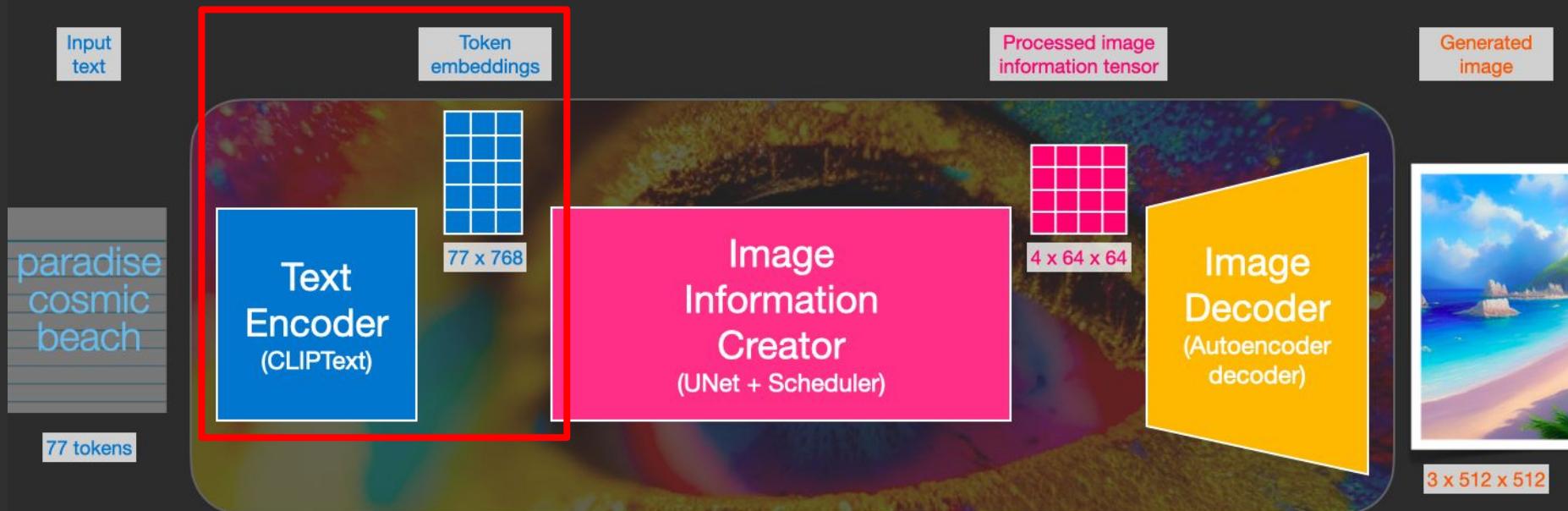
- Stable Diffusion is a latent text-to-image diffusion model.
- Trained on 512x512 images from a subset of the [LAION-5B](#) dataset.
- Compute donation from [Stability AI](#) and support from [LAION](#).

*This slide deck is heavily based on [The Illustrated Stable Diffusion by Jay Alammar](#) so please check out his blog post! :-)*

# Components

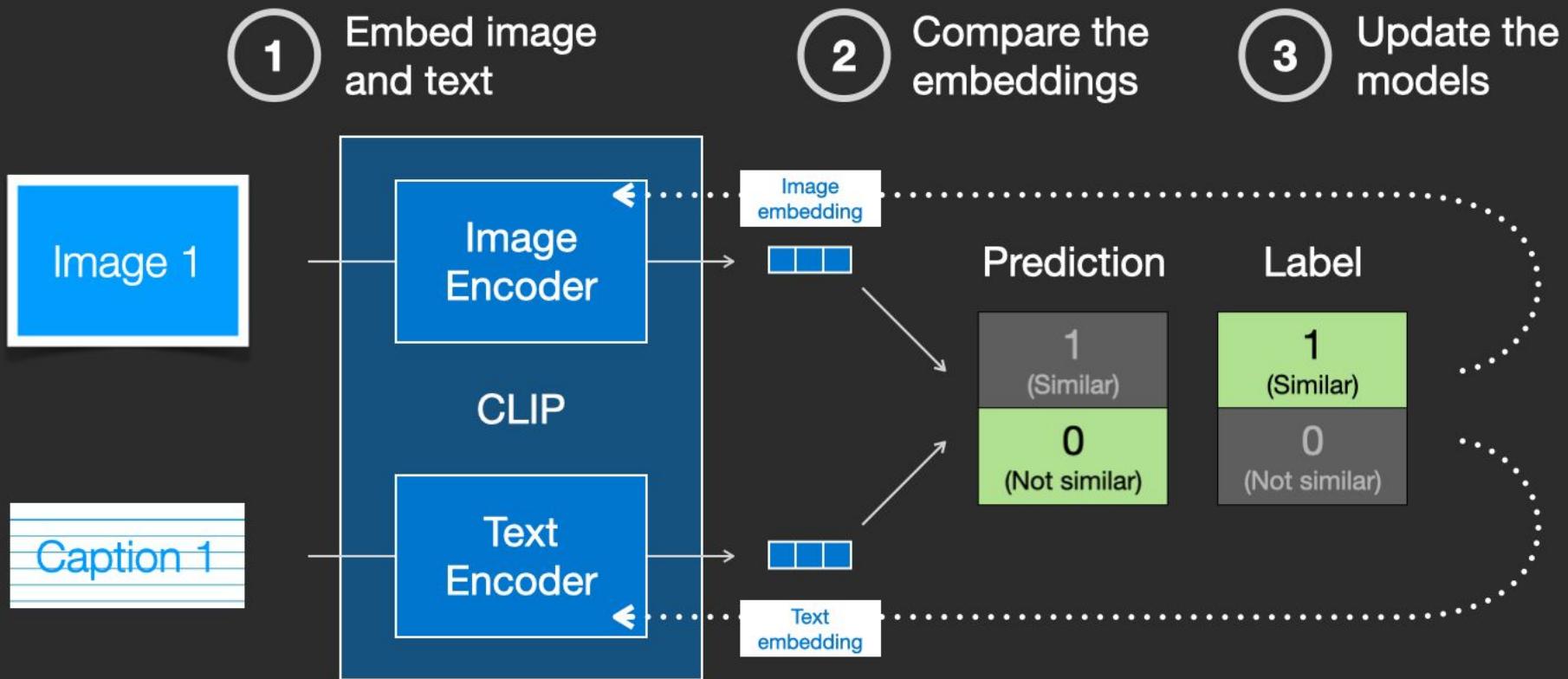
- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

# The Components of Stable Diffusion



Stable Diffusion

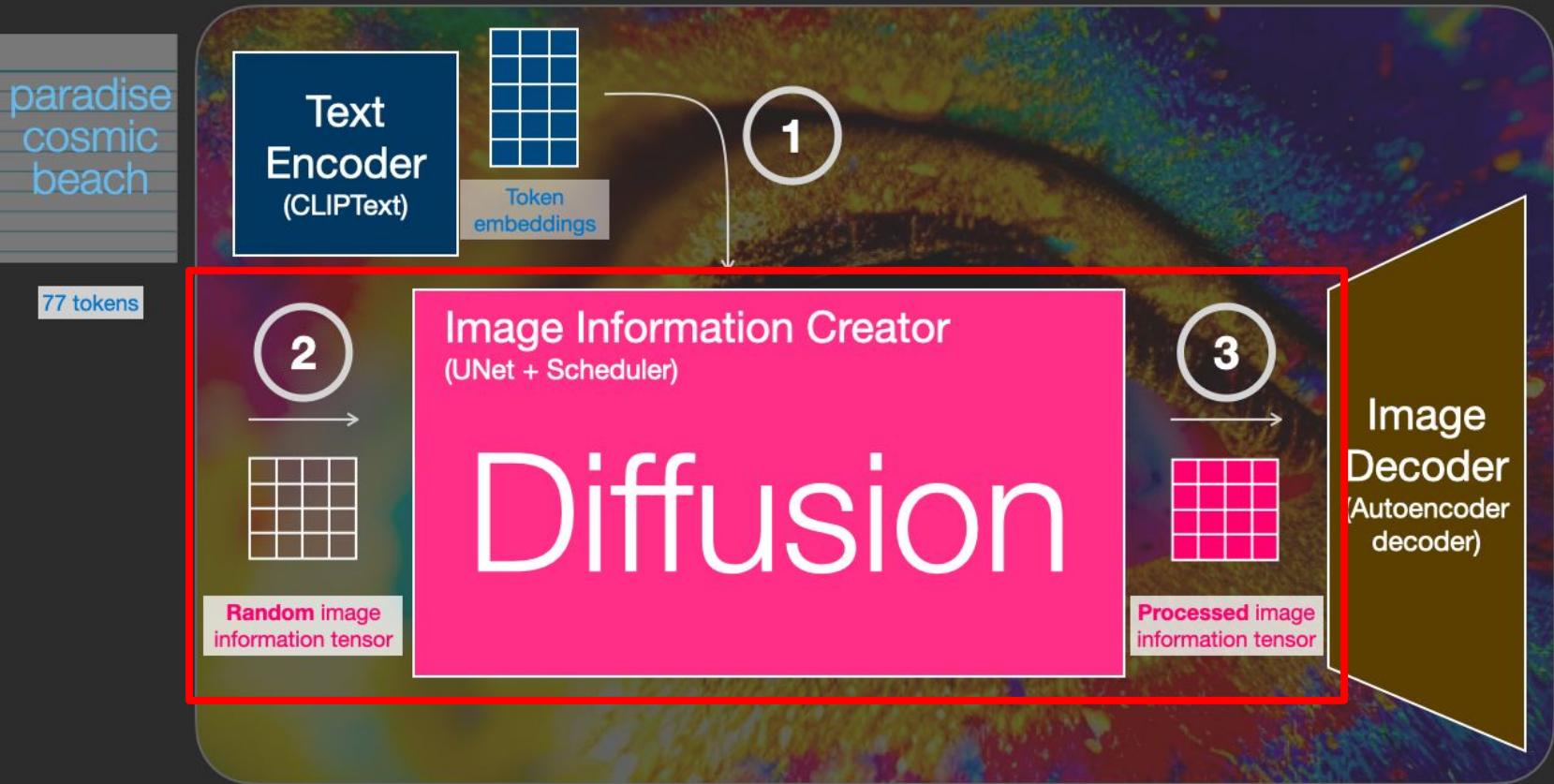
# The text encoder from CLIP



# Components

- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

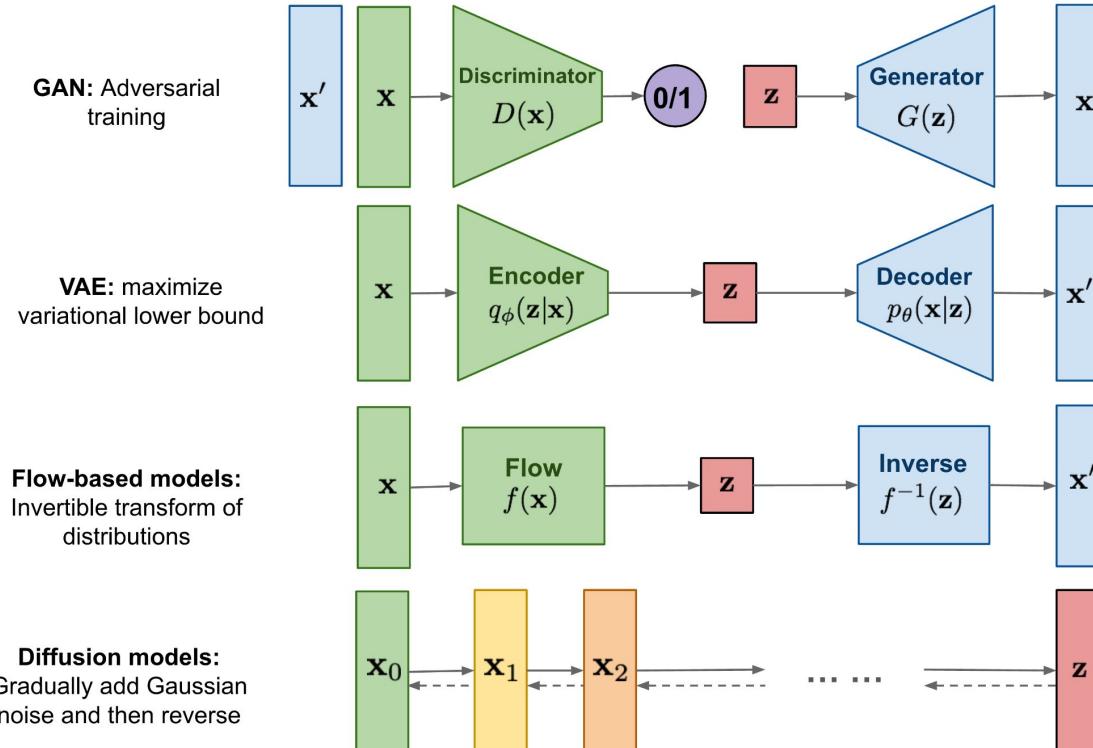
# Stable Diffusion



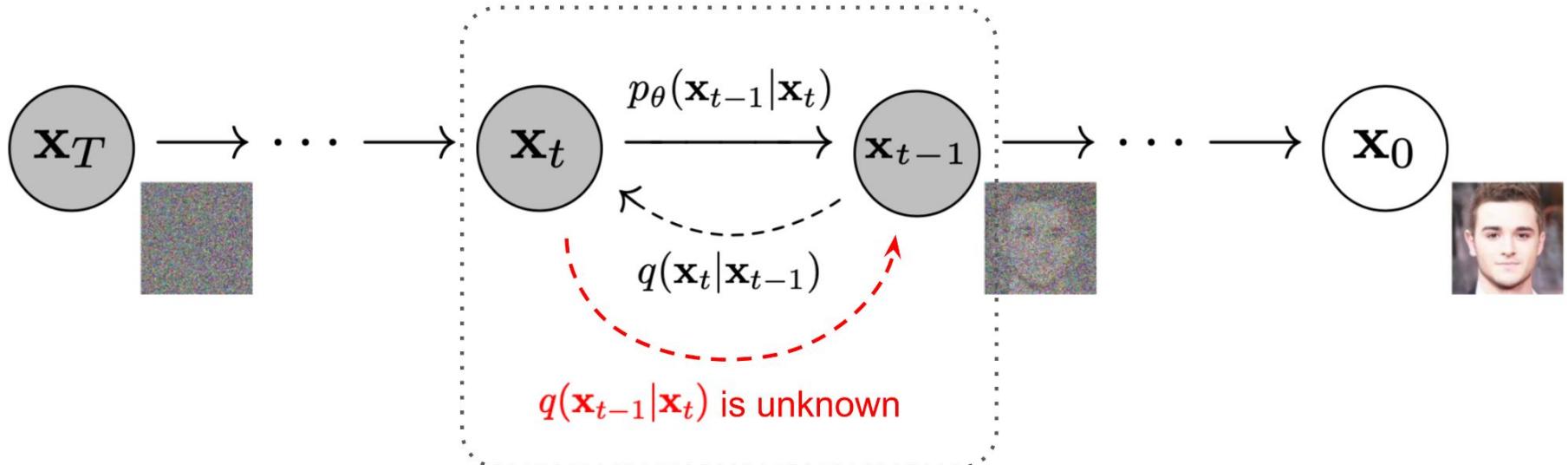
Generated image



# What is Diffusion?



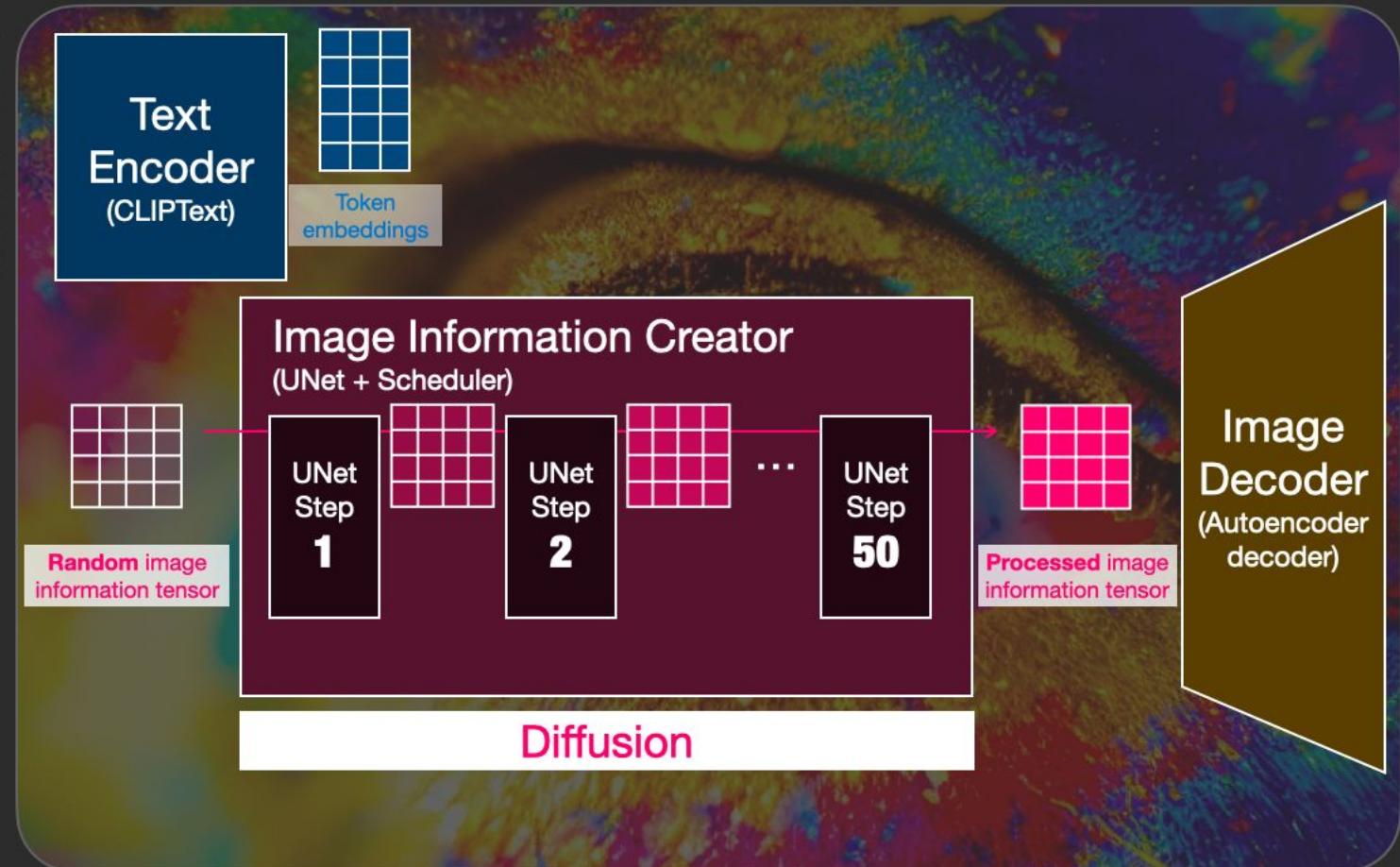
# What is Diffusion?



# Stable Diffusion

paradise  
cosmic  
beach

77 tokens



# Diffusion



Image Information Creator

Image  
Decoder  
(Autoencoder  
decoder)



# How diffusion works

Step 1



+

Step 2



noise  
slice 1

=

Step 3



+



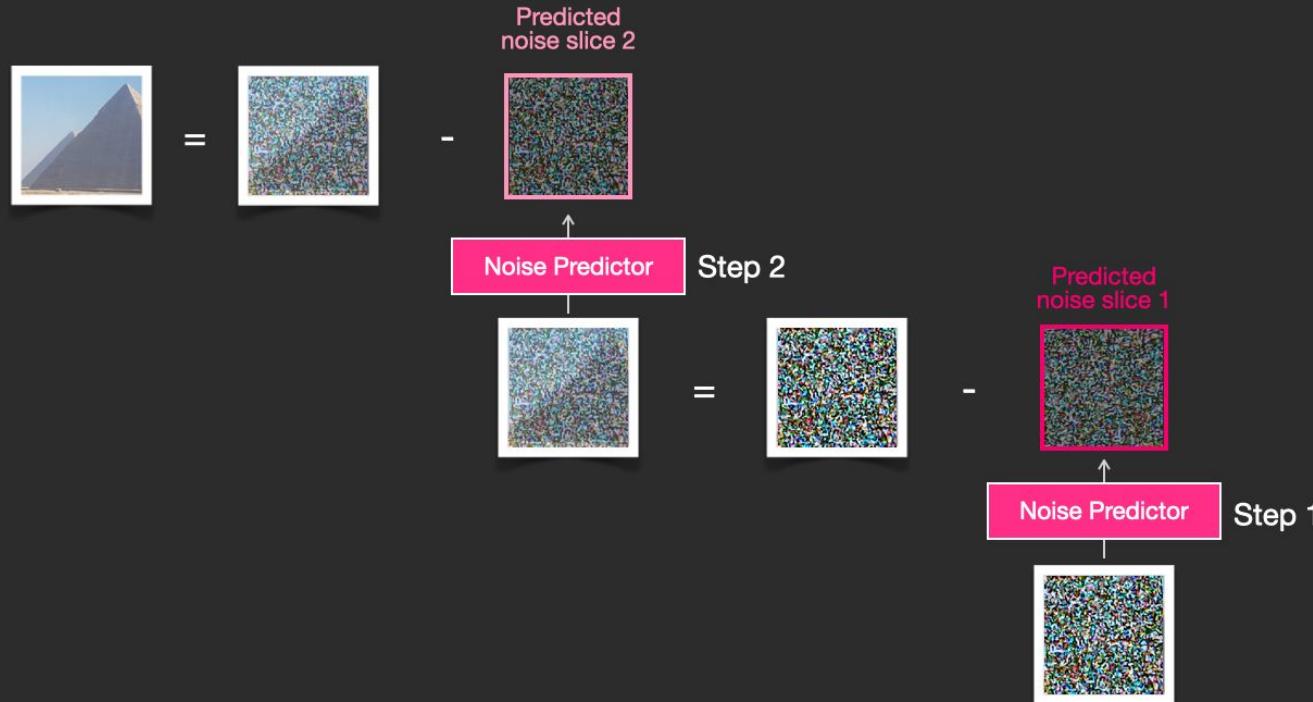
noise  
slice 2

=



# Painting images by removing noise

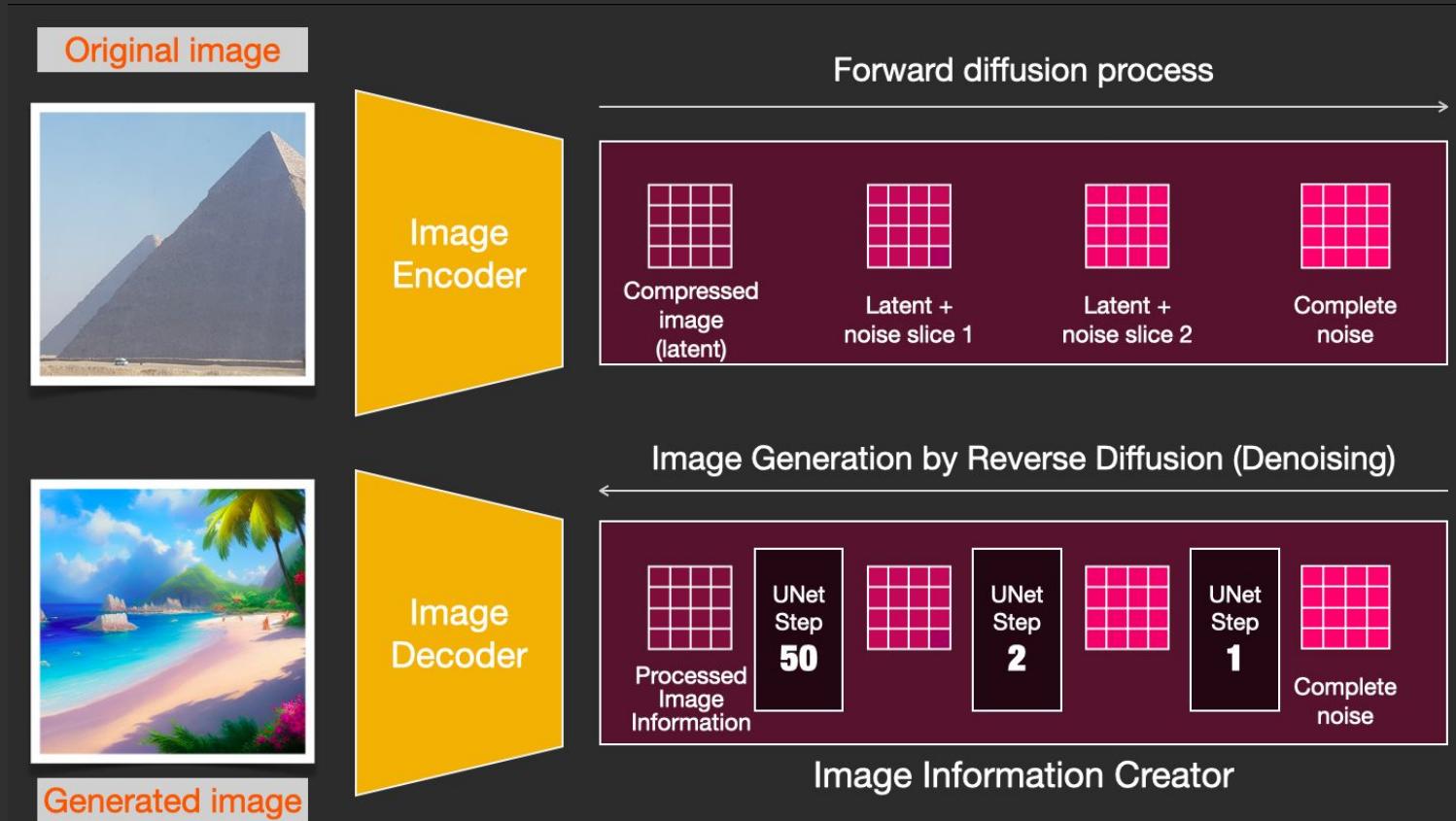
## Image Generation by Reverse Diffusion (Denoising)



# Components

- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

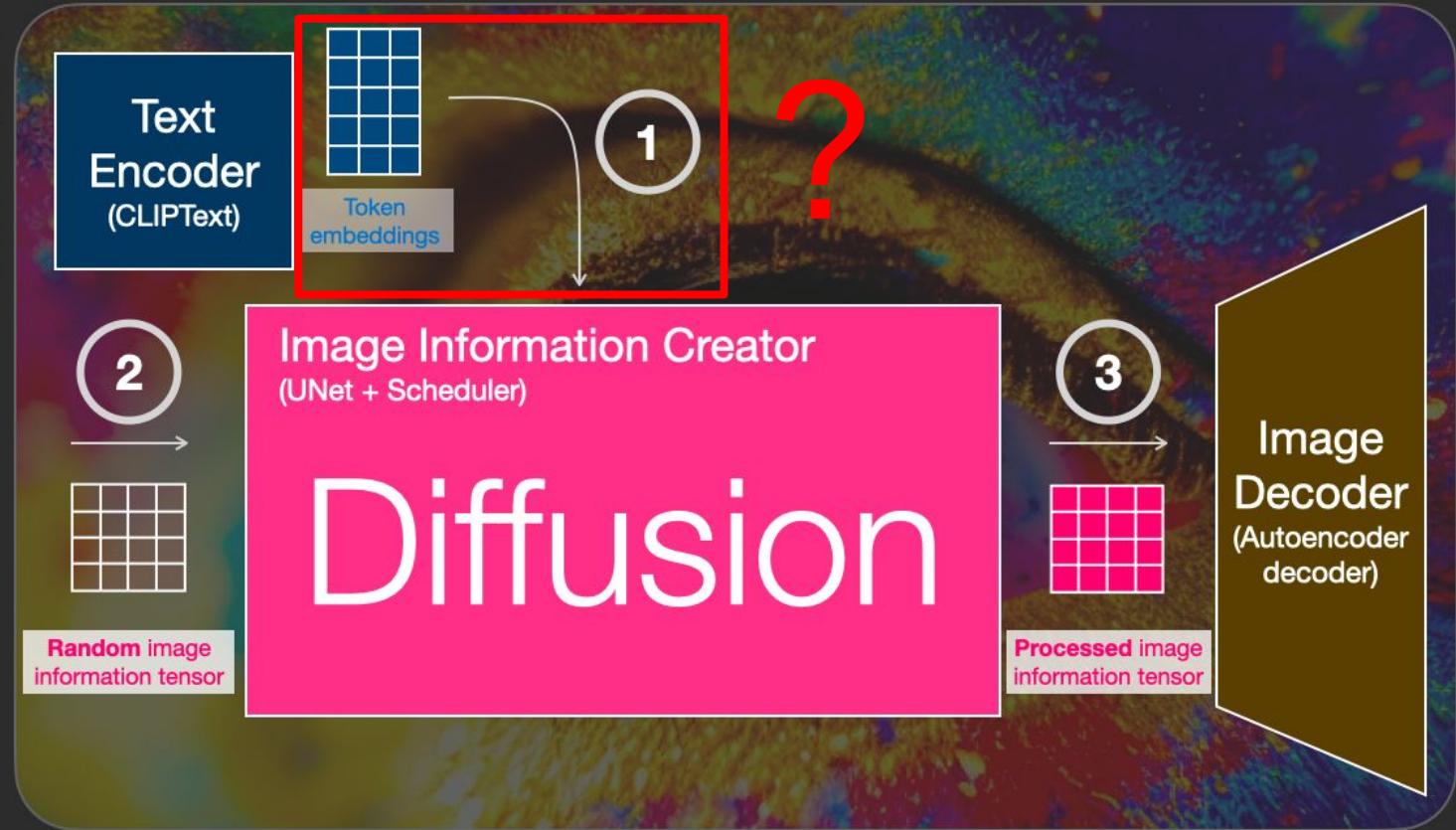
# Diffusion on compressed (latent) data instead of the pixel images



# Stable Diffusion

paradise  
cosmic beach

77 tokens



Generated image



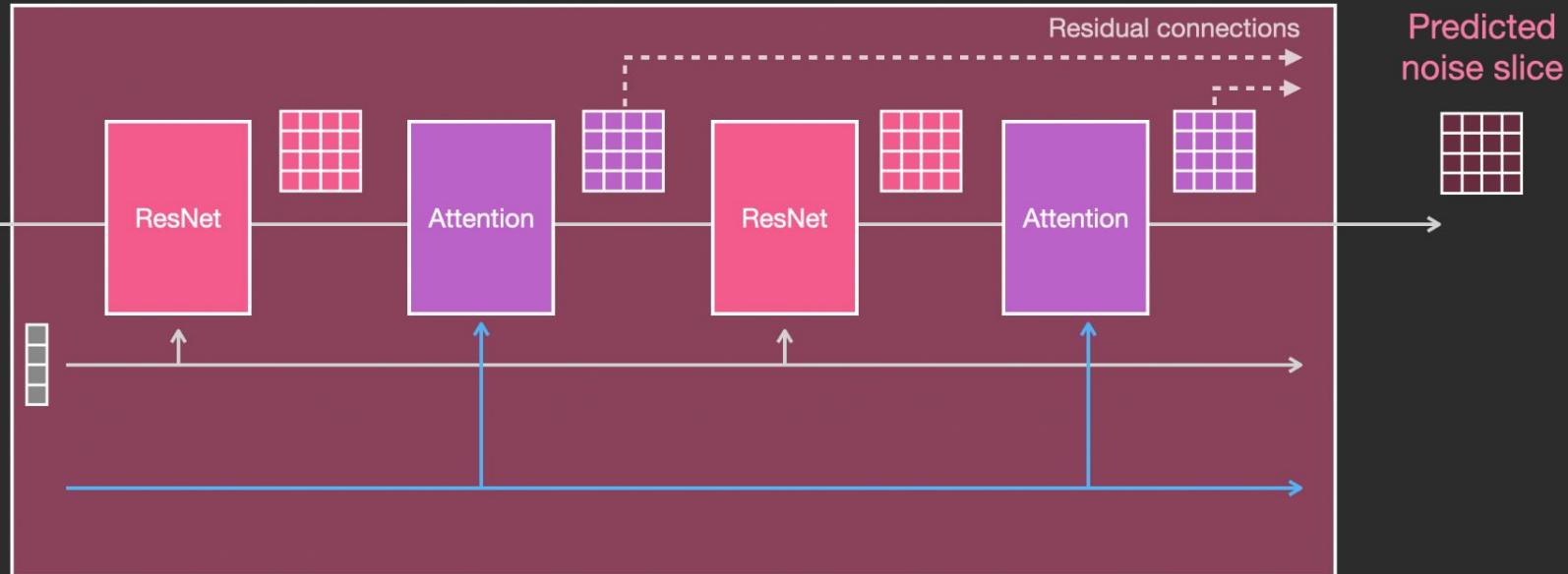
# Feeding text information into the image generation process

## Noise Predictor with **Text Conditioning** (UNet with **attention**)

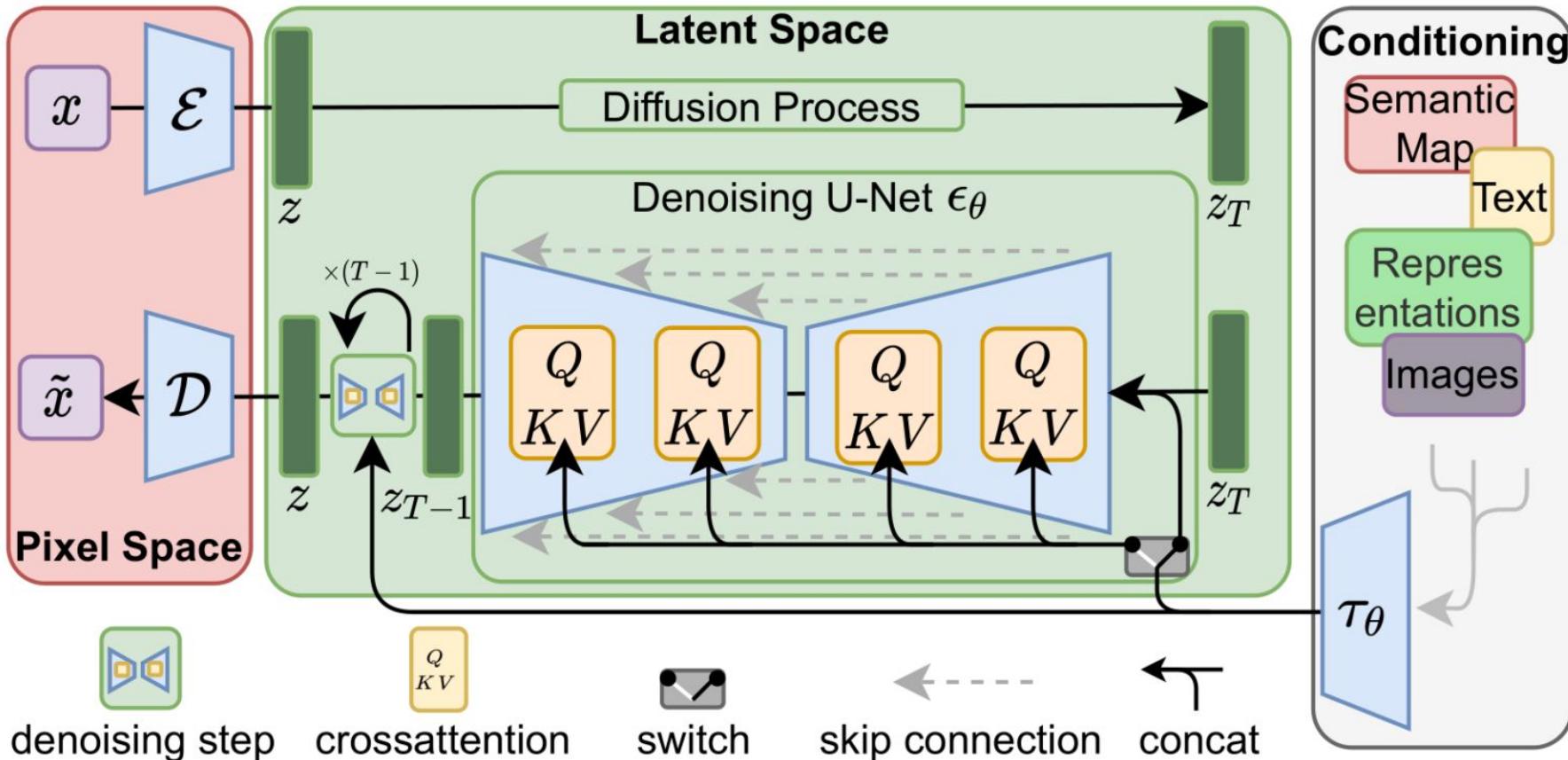
Noisy image  
information  
(latents)



Step 3



Text information  
(token embeddings)



# Stable Diffusion

- How does it work?
- Applications

# Text to image



Prompt: "Vienna deep learning meetup in October"

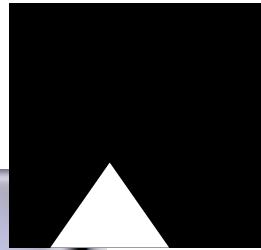
# Image to image



<https://huggingface.co/spaces/huggingface-projects/diffuse-the-rest/discussions/204>

+ Mask

# Inpainting



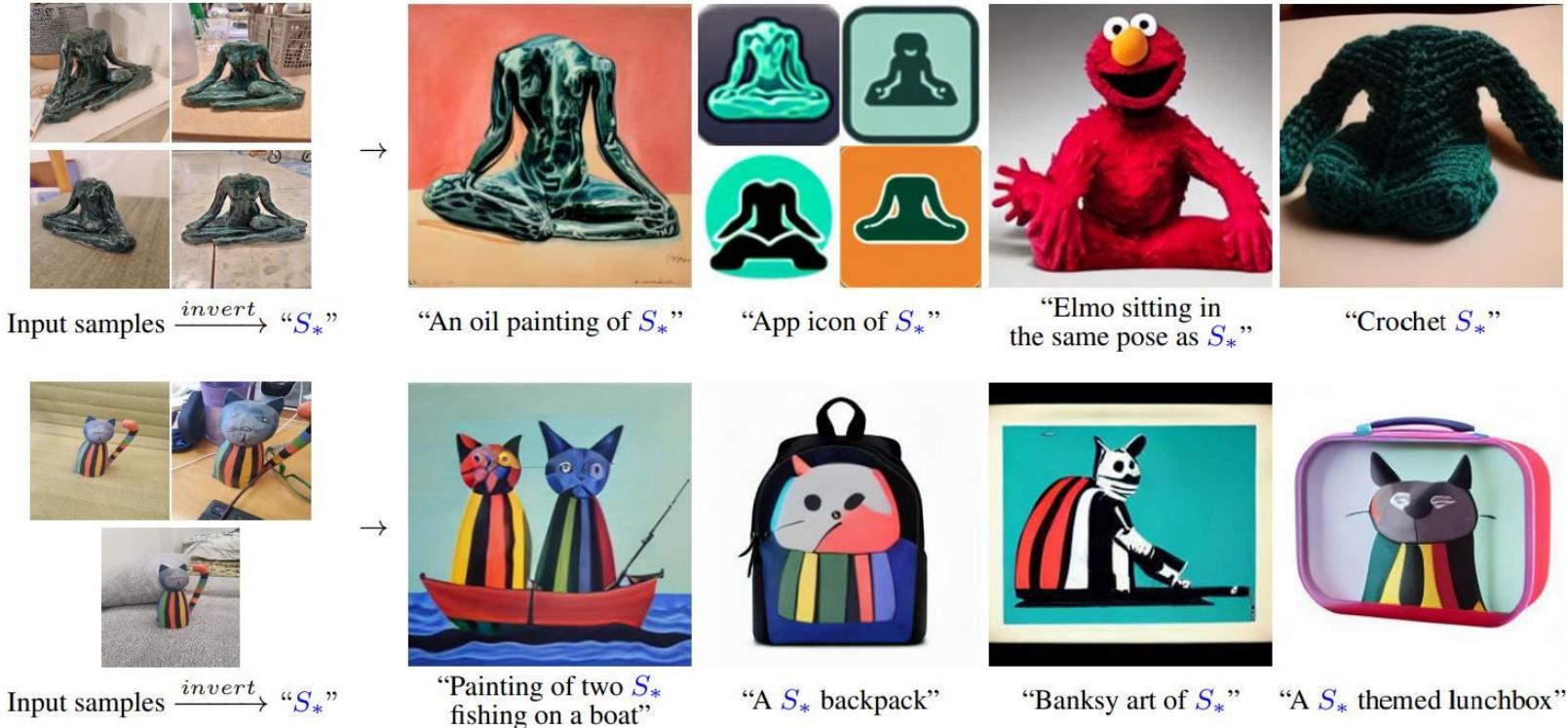
+ Prompt: "Cinematic movie still of  
Elon Musk in 2001 space Odyssey"

# Fine tune Stable Diffusion

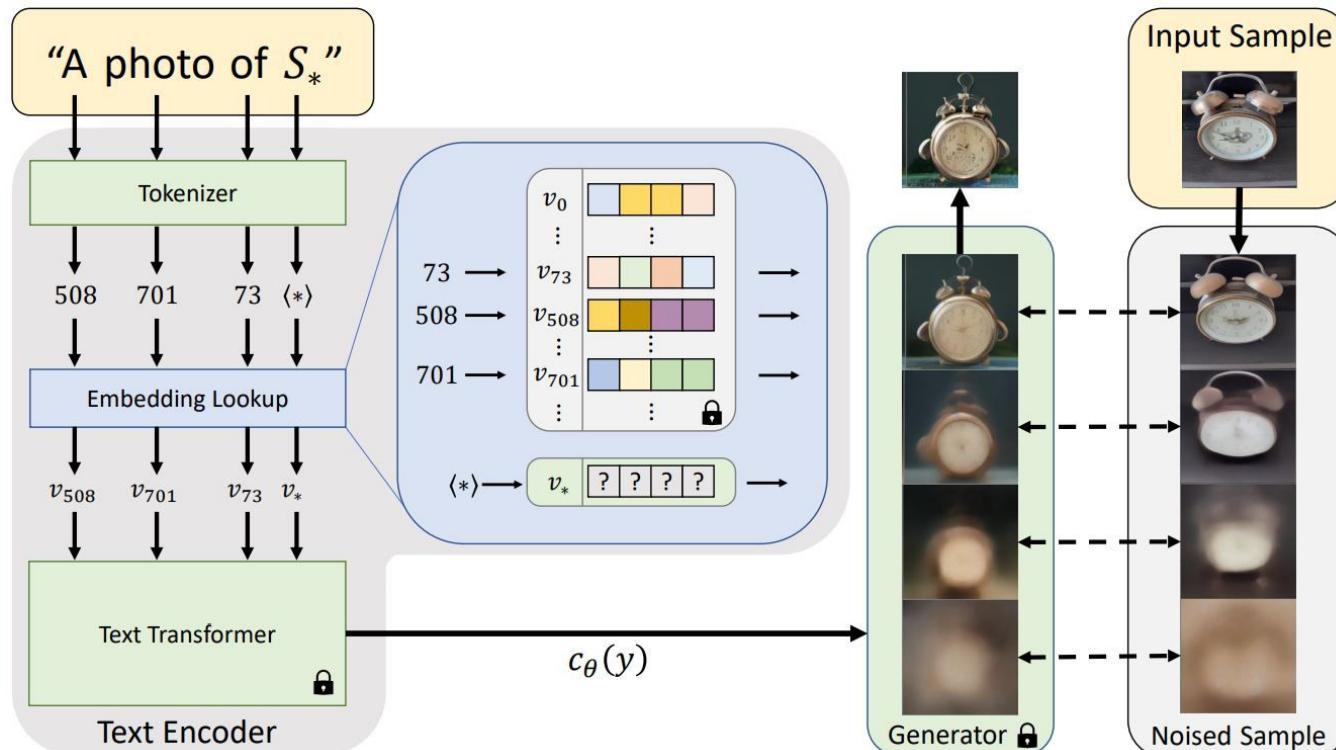


<https://lambdalabs.com/blog/how-to-fine-tune-stable-diffusion-how-we-made-the-text-to-pokemon-model-at-lambda/>

# Textual inversion



# Textual inversion setup



# Dreambooth



Input images



in the Acropolis



swimming

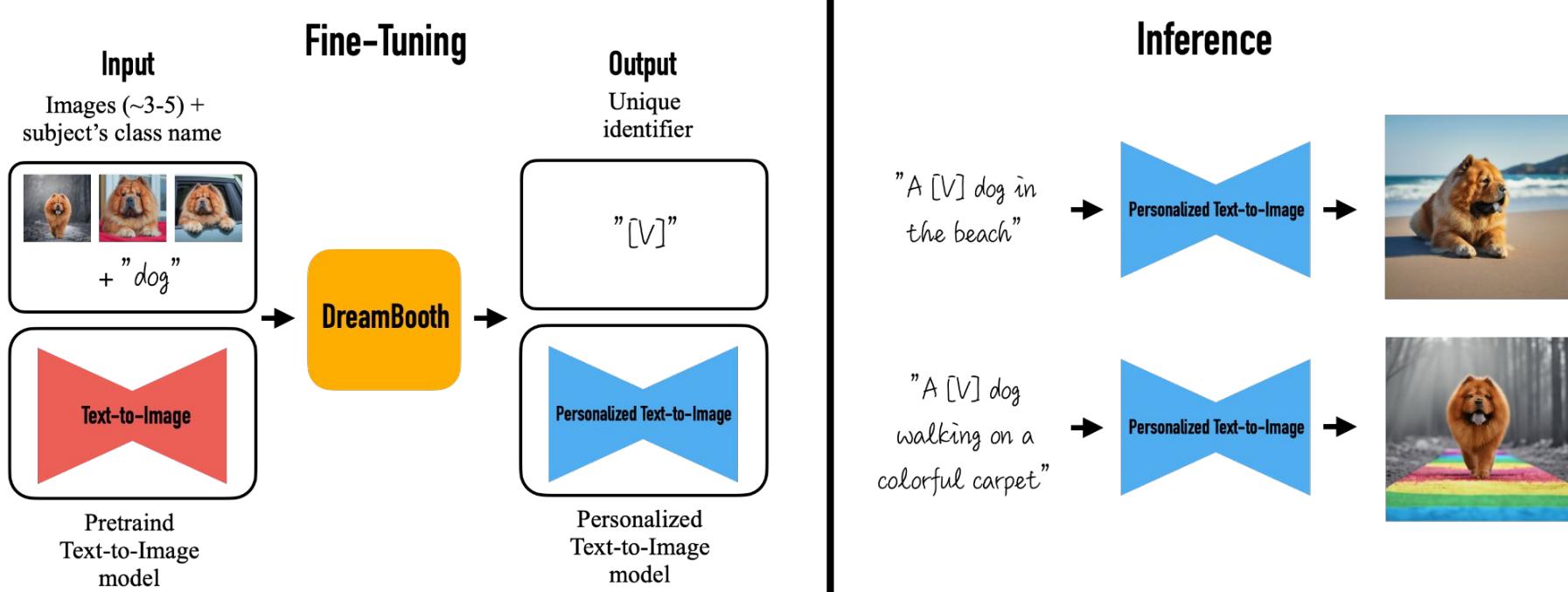


sleeping

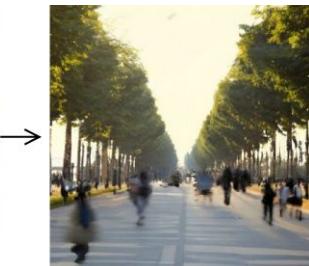


getting a haircut

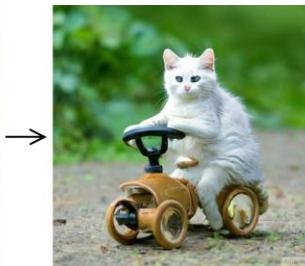
# Dreambooth setup



# Prompt-to-prompt image editing



"The boulevards are crowded today."  

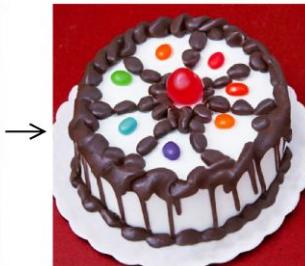
"Photo of a cat riding on a ~~bicycle~~  
 car."



"Landscape with a house near a river  
and a rainbow in the background!"  




"My  fluffy bunny doll."



"a cake with decorations."  




"Children drawing of a castle next to a river!"  


# Thank you for your attention!

## Additional resources

- [Stable Diffusion Public Release - Stability.Ai](#)
- [The Annotated Diffusion Model](#)
- [Stable Diffusion with 🎨 Diffusers, Diffusers notebooks](#)
- [How diffusion models work: the math from scratch | AI Summer](#)
- [What are Diffusion Models? | Lil'Log](#)
- [Generative Modeling by Estimating Gradients of the Data Distribution | Yang Song](#)
- [Diffusion Models: A Comprehensive Survey of Methods and Applications](#)
- [How DALL-E 2, Imagen and Parti Architectures Differ](#)
- <https://github.com/CompVis/stable-diffusion>
- <https://github.com/lucidrains/denoising-diffusion-pytorch>
- New fastai part 2 course 2022 will be released soon and has a focus on Stable Diffusion.

**Next meetup:**

**November 17 @ Technikum Wien**

**Real Time Video Analytics //  
Embedded DL Hardware**

[emotion3d.ai](http://emotion3d.ai)