

# HyperFields: Towards Zero-Shot Generation of NeRFs from Text



**Sudarshan  
Babu\***



**Richard  
Liu\***



**Avery  
Zhou\***



**Michael  
Maire**



**Greg  
Shakhnarovich**



**Rana  
Hanocka**

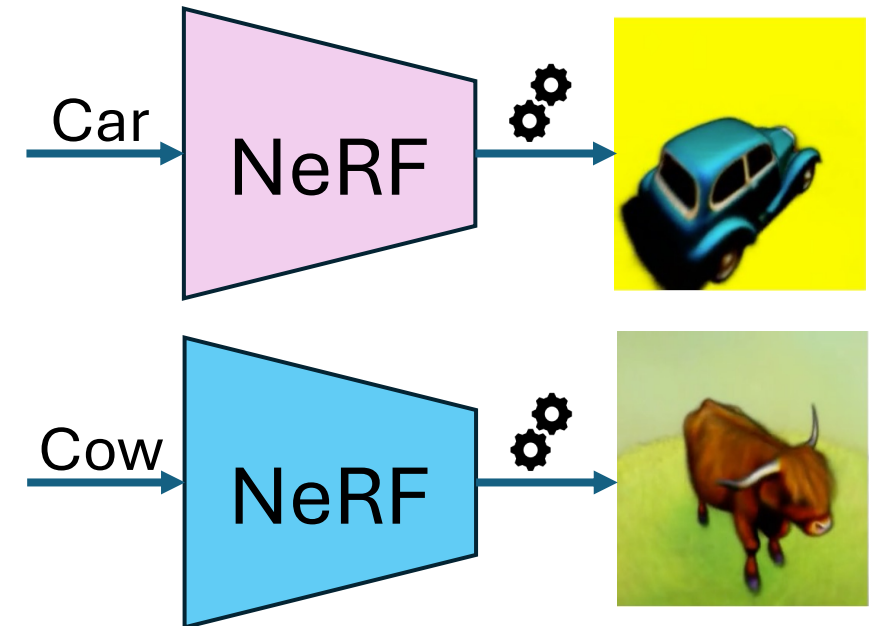
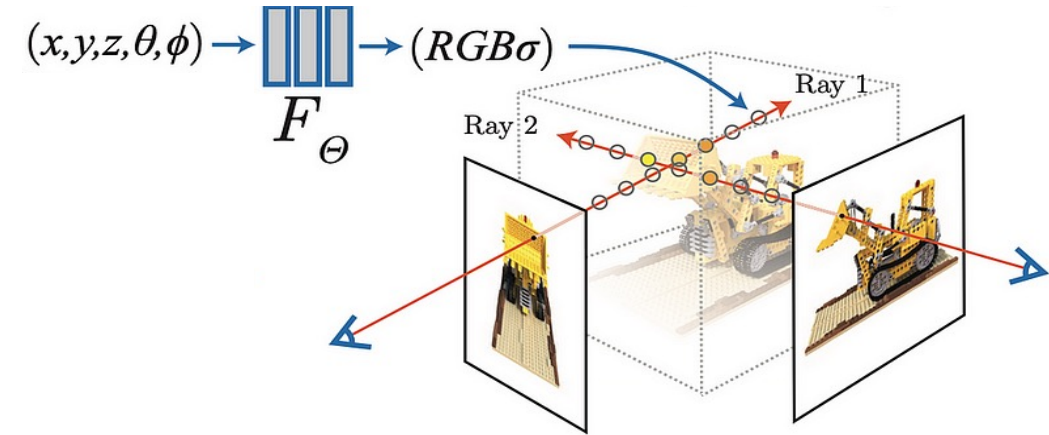


Poster Location: Hall C 4-9 #505  
Wed 24 Jul 11:30 a.m. CEST — 1 p.m. CEST  
<https://threedle.github.io/hyperfields/>



## ➤ Background

- ❑ NeRF: Implicit 3D store [1]
- ❑ Text-to-3D (TT3D) models [2] use NeRF



[1] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." ECCV 2020.

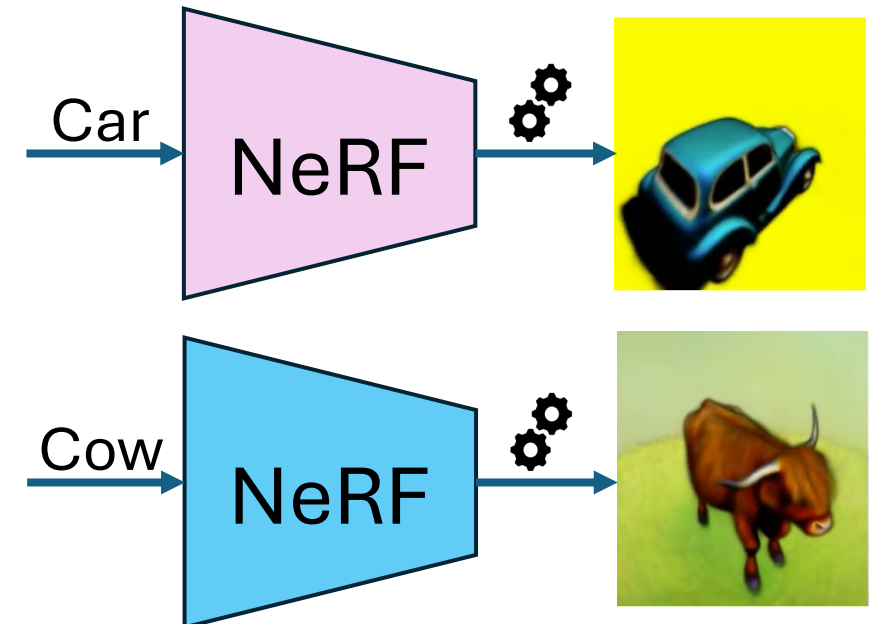
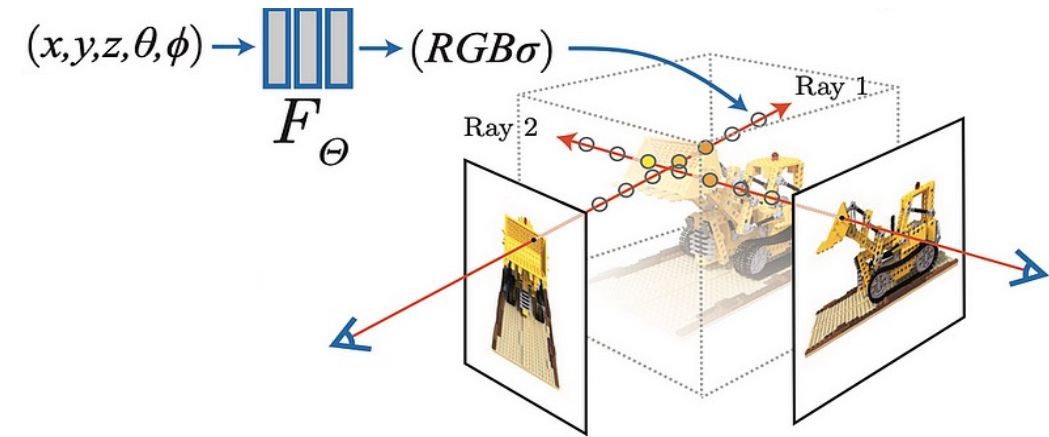
[2] B. Poole et al. "DreamFusion: Text-to-3D using 2D diffusion." ICLR 2023.

## ➤ Background

- ❑ NeRF: Implicit 3D store [1]
- ❑ Text-to-3D (TT3D) models [2] use NeRF

## ➤ Motivation

- ❑ TT3D models need per-prompt optimization
- ❑ More time, compute, storage



[1] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." ECCV 2020.

[2] B. Poole et al. "DreamFusion: Text-to-3D using 2D diffusion." ICLR 2023.

## ➤ Background

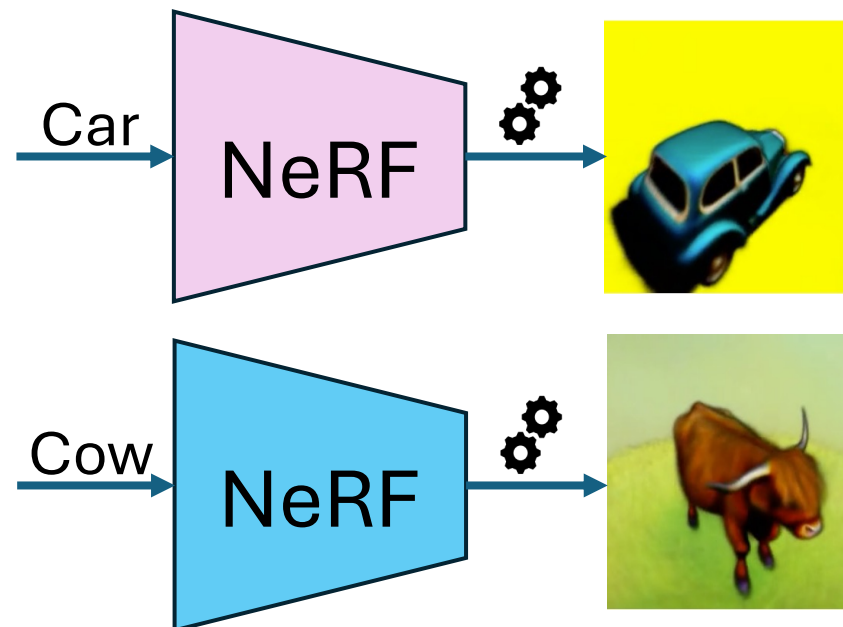
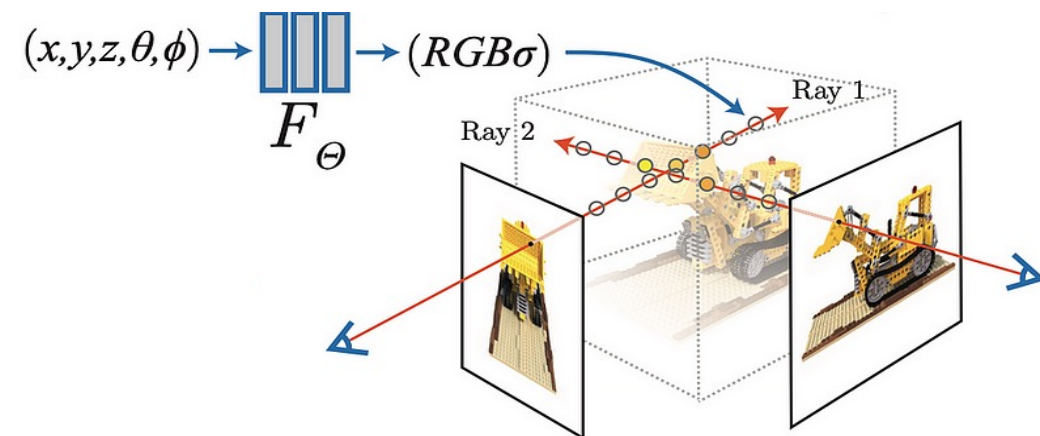
- ❑ NeRF: Implicit 3D store [1]
- ❑ Text-to-3D (TT3D) models [2] use NeRF

## ➤ Motivation

- ❑ TT3D models need per-prompt optimization
- ❑ More time, compute, storage

## ➤ Contribution

- ❑ **HyperFields**: Text-conditioned hypernetwork
- ❑ Mitigates need for per-prompt optimization



[1] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." ECCV 2020.

[2] B. Poole et al. "DreamFusion: Text-to-3D using 2D diffusion." ICLR 2023.

## ➤ Background

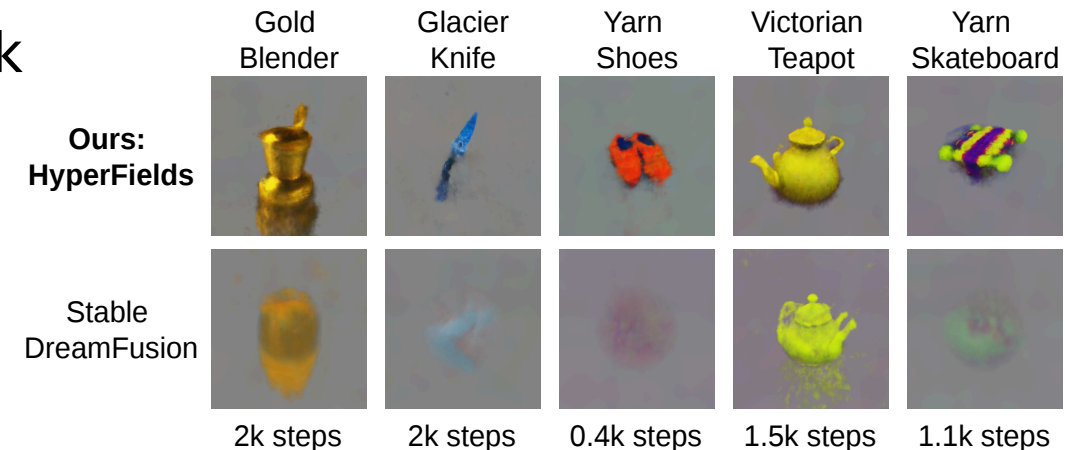
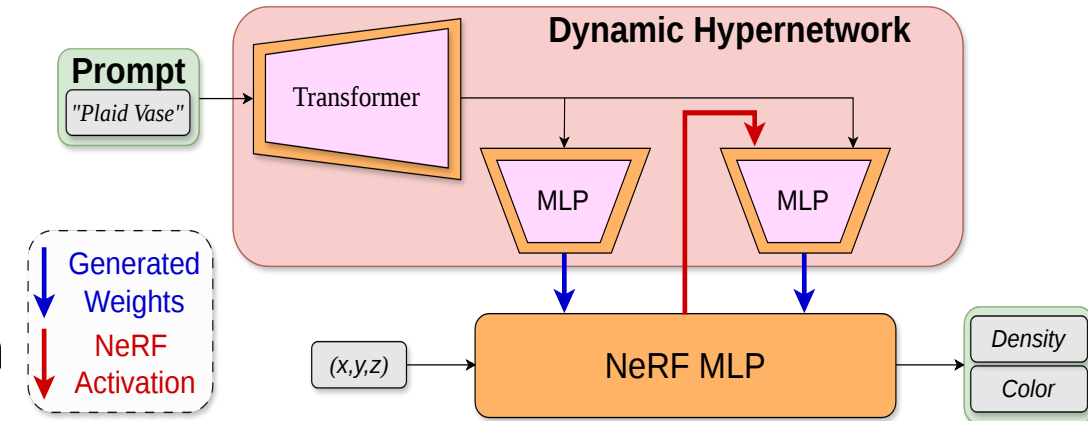
- ❑ NeRF: Implicit 3D store [1]
- ❑ Text-to-3D (TT3D) models [2] use NeRF

## ➤ Motivation

- ❑ TT3D models need per-prompt optimization
- ❑ More time, compute, storage

## ➤ Contribution

- ❑ **HyperFields**: Text-conditioned hypernetwork
- ❑ Mitigates need for per-prompt optimization



[1] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." ECCV 2020.

[2] B. Poole et al. "DreamFusion: Text-to-3D using 2D diffusion." ICLR 2023.

## ➤ Background

- ❑ NeRF: Implicit 3D store [1]
- ❑ Text-to-3D (TT3D) models [2] use NeRF

## ➤ Motivation

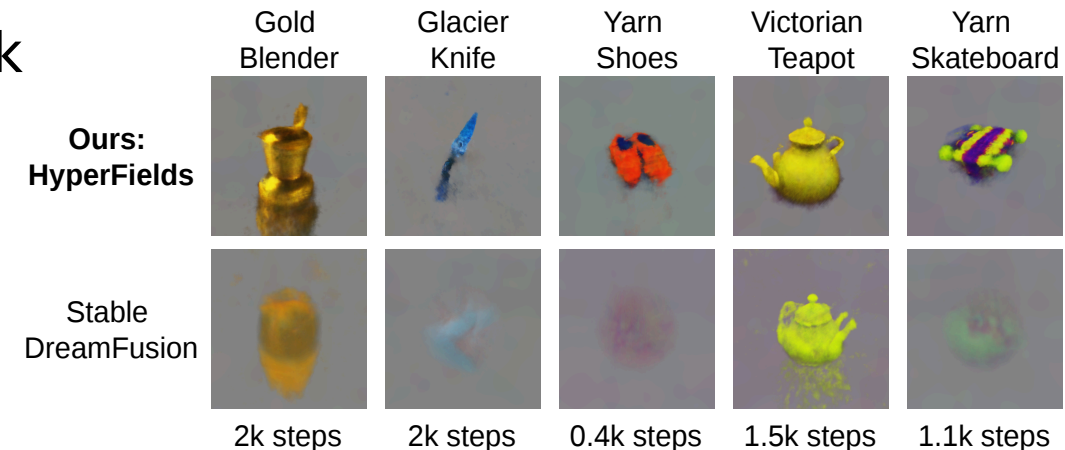
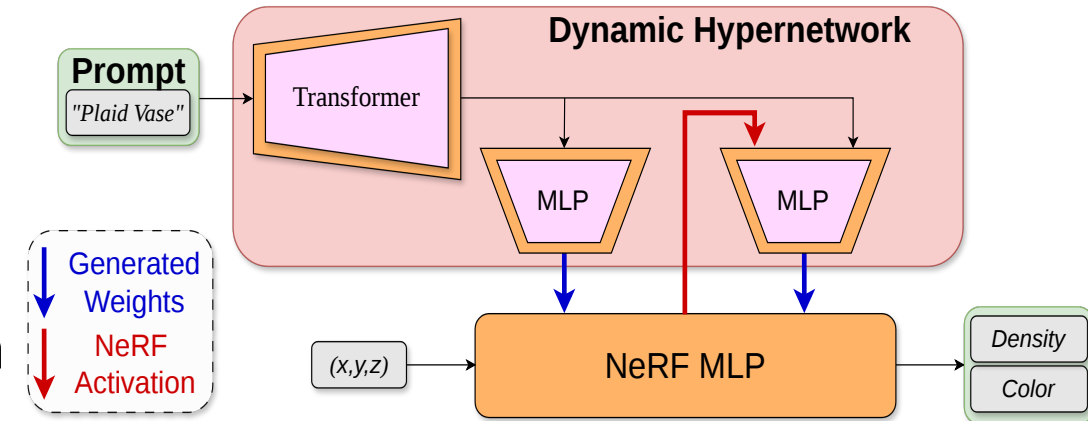
- ❑ TT3D models need per-prompt optimization
- ❑ More time, compute, storage

## ➤ Contribution

- ❑ **HyperFields**: Text-conditioned hypernetwork
- ❑ Mitigates need for per-prompt optimization

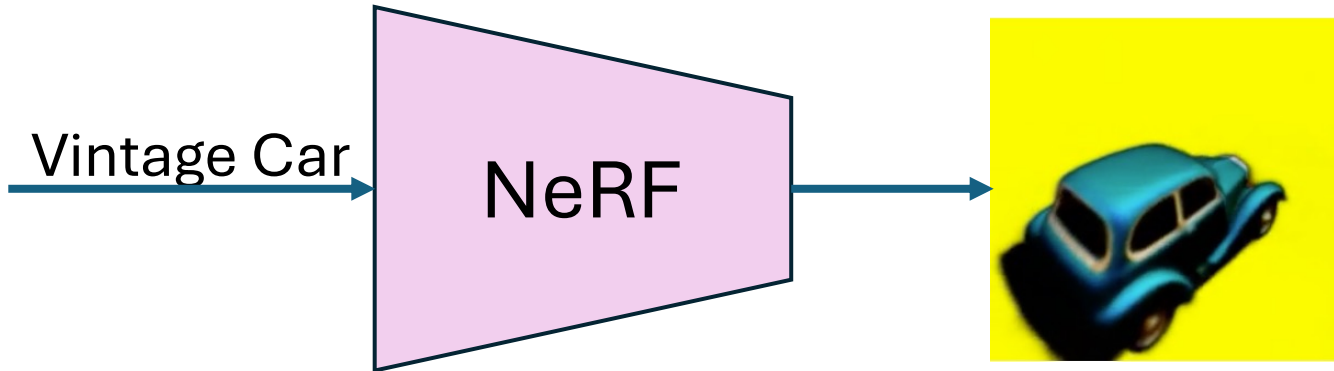
## ➤ Results

- ❑ Fast TT3D model
- ❑ Reduced compute, storage



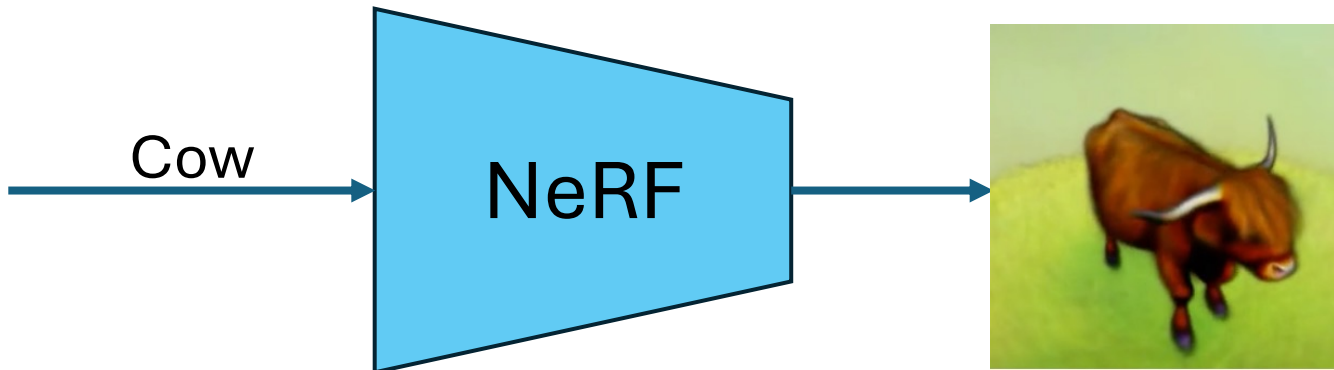
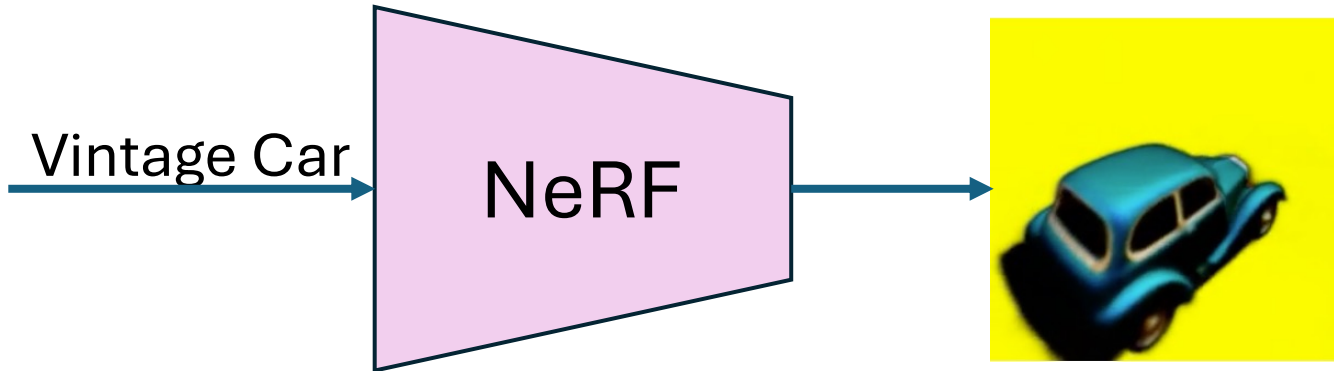
[1] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." ECCV 2020.

[2] B. Poole et al. "DreamFusion: Text-to-3D using 2D diffusion." ICLR 2023.



➤ **Require per-prompt optimization**

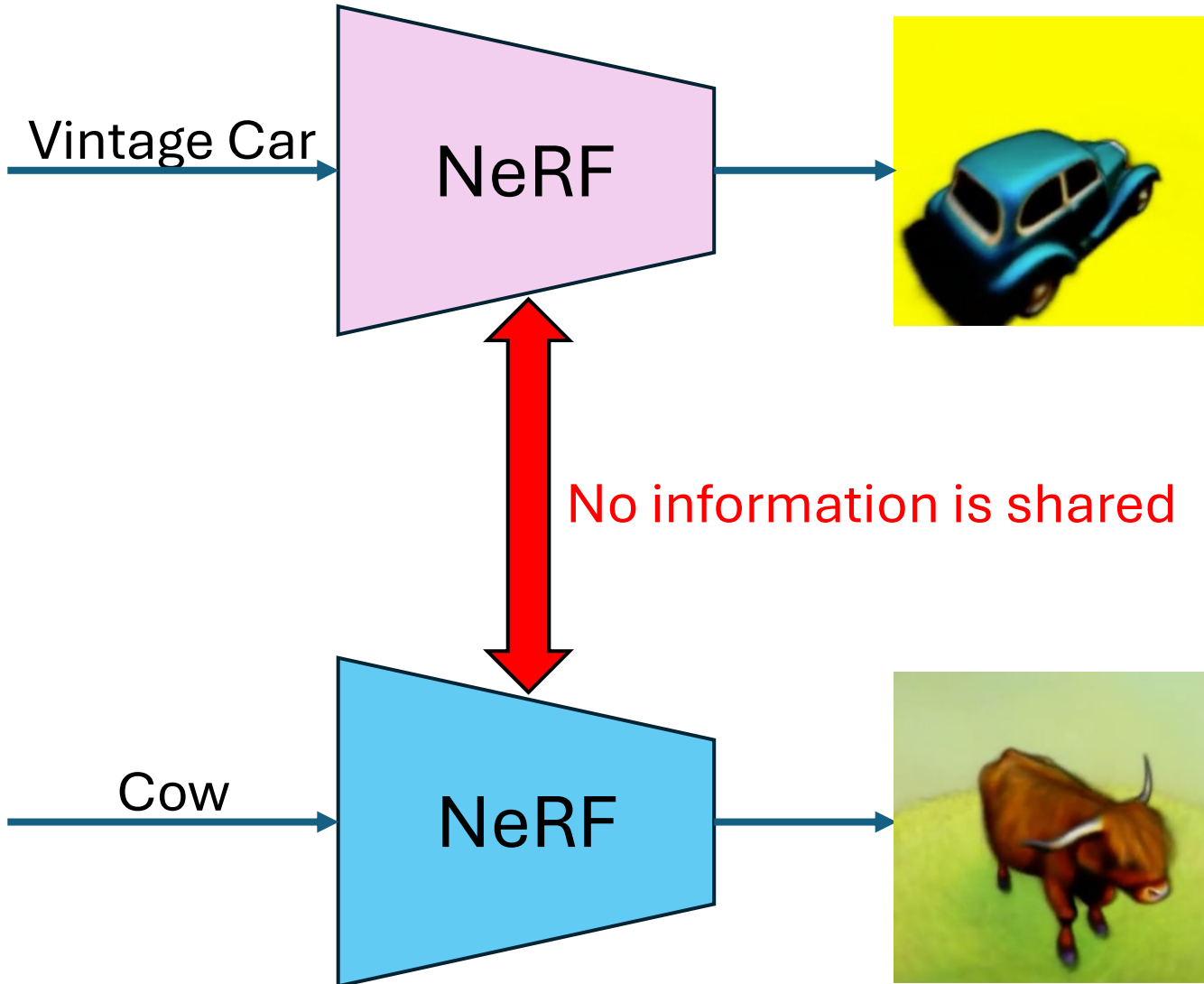
- Separate model per prompt
- No information sharing
- Time consuming



➤ **Require per-prompt optimization**

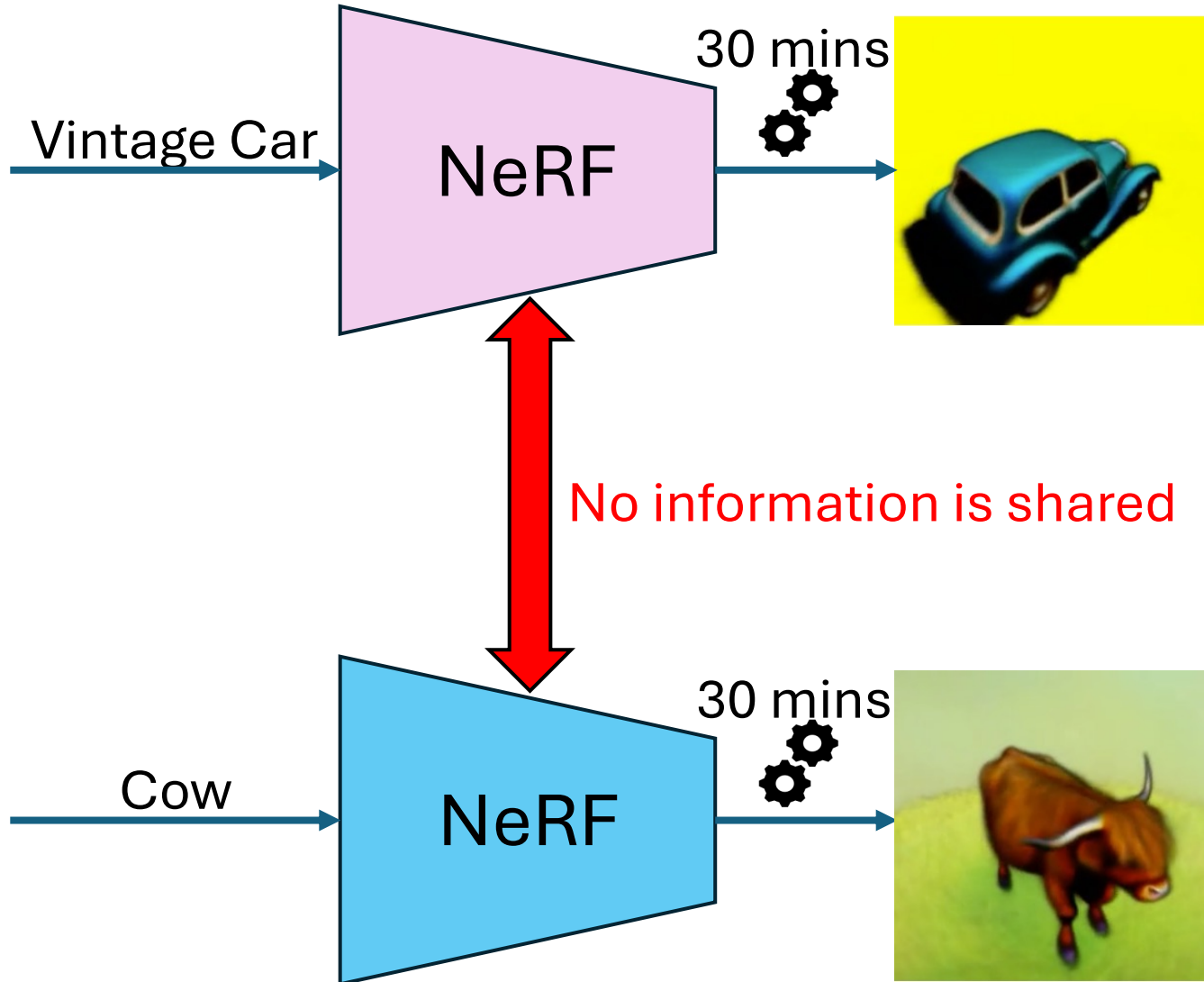
- Separate model per prompt
- No information sharing
- Time consuming





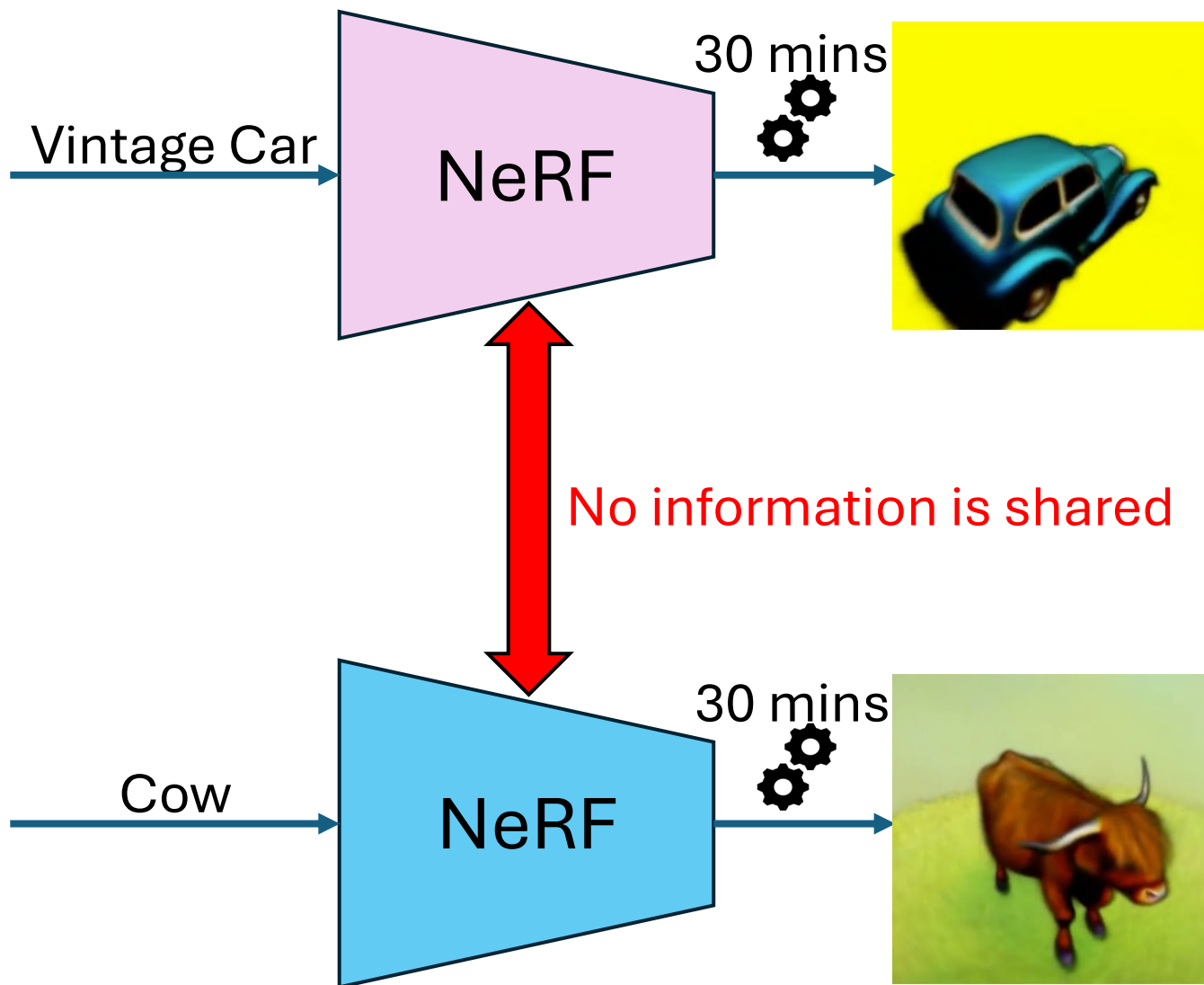
➤ **Require per-prompt optimization**

- Separate model per prompt
- No information sharing
- Time consuming



➤ **Require per-prompt optimization**

- Separate model per prompt
- No information sharing
- Time consuming



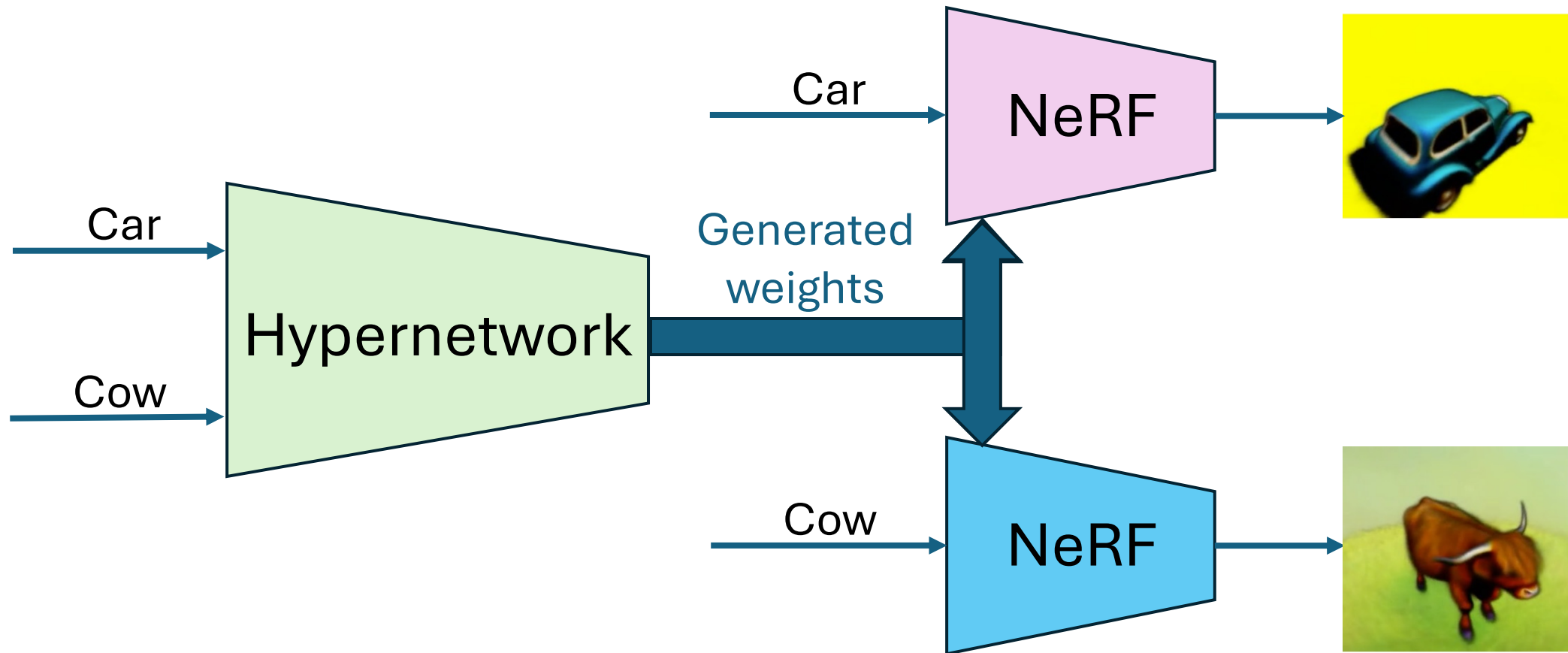
➤ **Require per-prompt optimization**

- Separate model per prompt
- No information sharing
- Time consuming

➤ **Objective:**

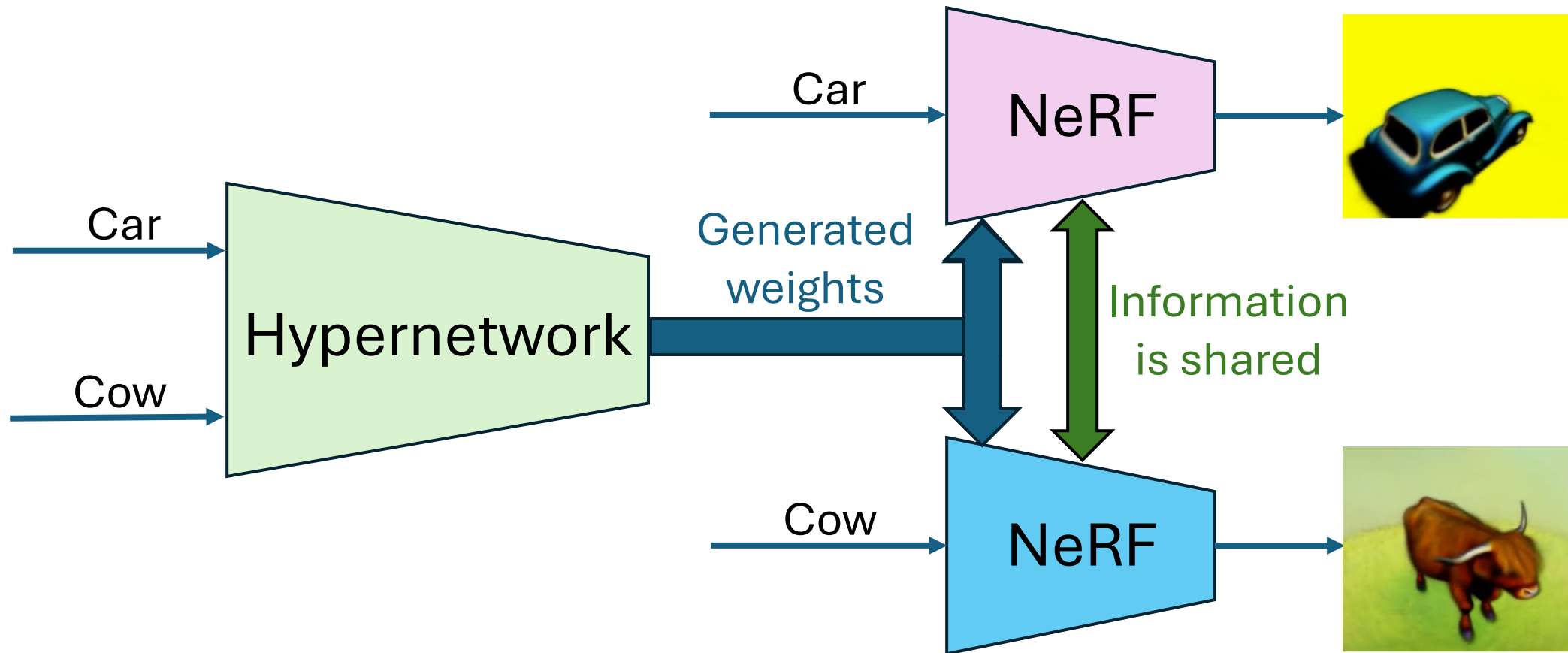
- Avoid per-prompt optimization

# Approach: HyperFields



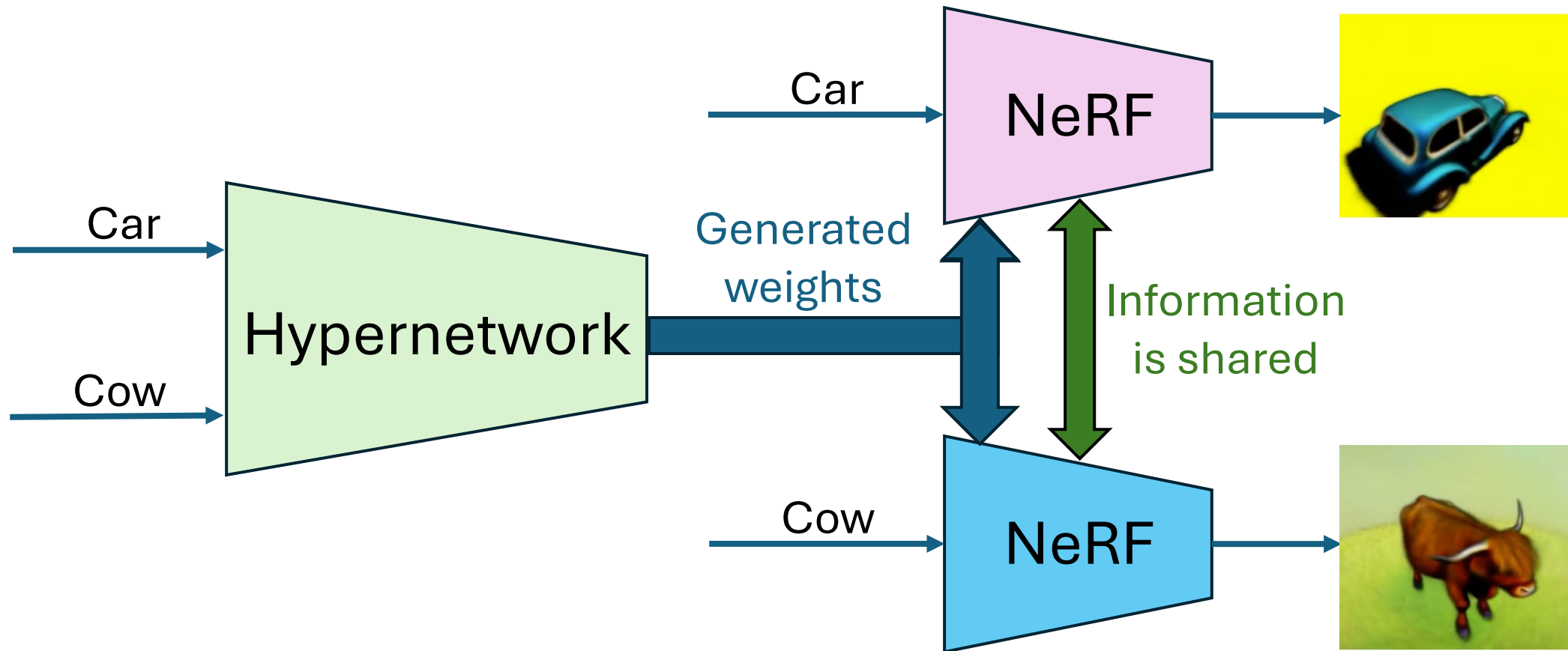
➤ **HyperFields: Text-conditioned hypernetwork** [1] shared training across prompts

# Approach: HyperFields



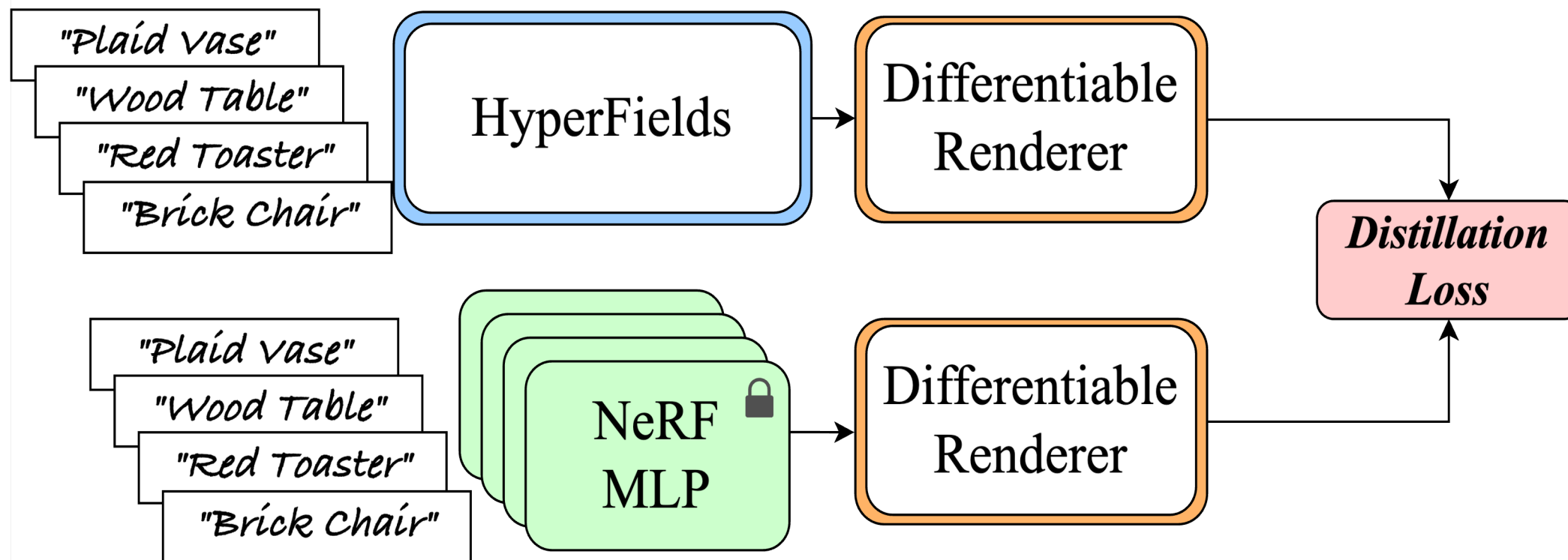
➤ **HyperFields: Text-conditioned hypernetwork** [1] shared training across prompts

# Approach: HyperFields



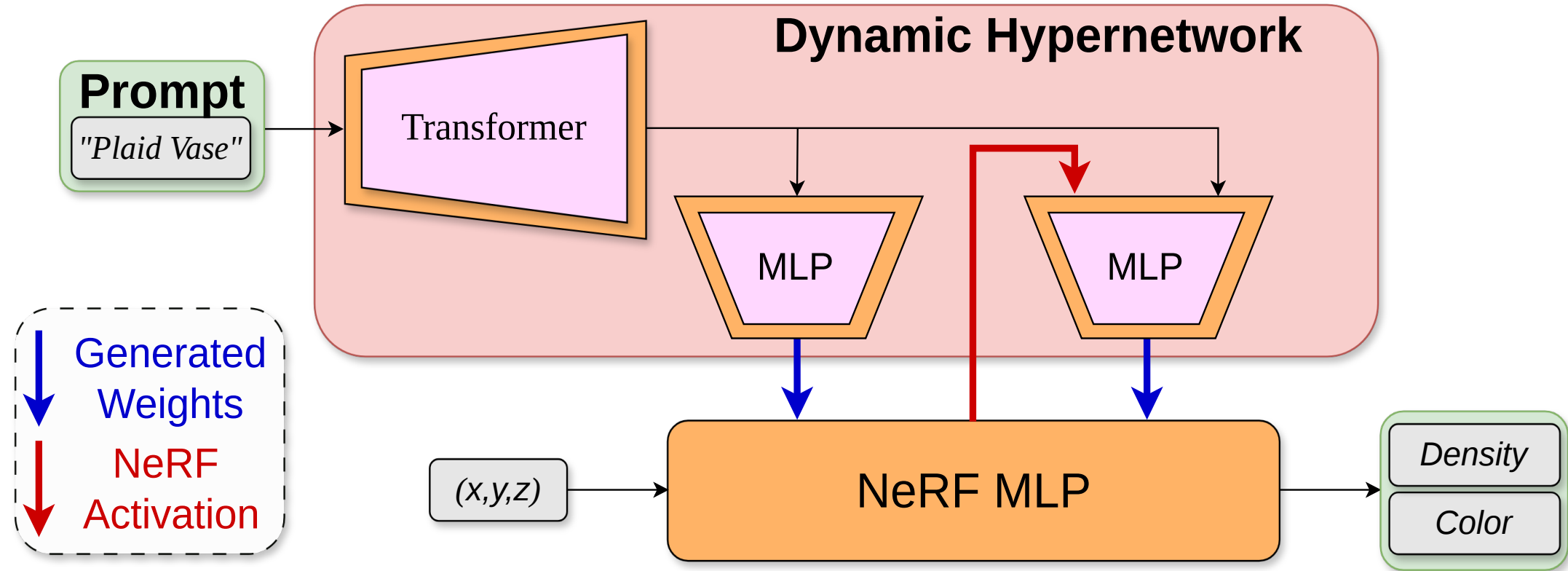
- **HyperFields: Text-conditioned hypernetwork** [1] shared training across prompts
- **Outcome:** Mitigation of per-prompt optimization and consequently faster generation

# Approach: HyperFields Training



- **Shared Training:** Distill multiple single scene NeRFs to a single HyperFields model
- **Distillation Loss: MSE loss** between the renders of HyperFields and teacher NeRFs

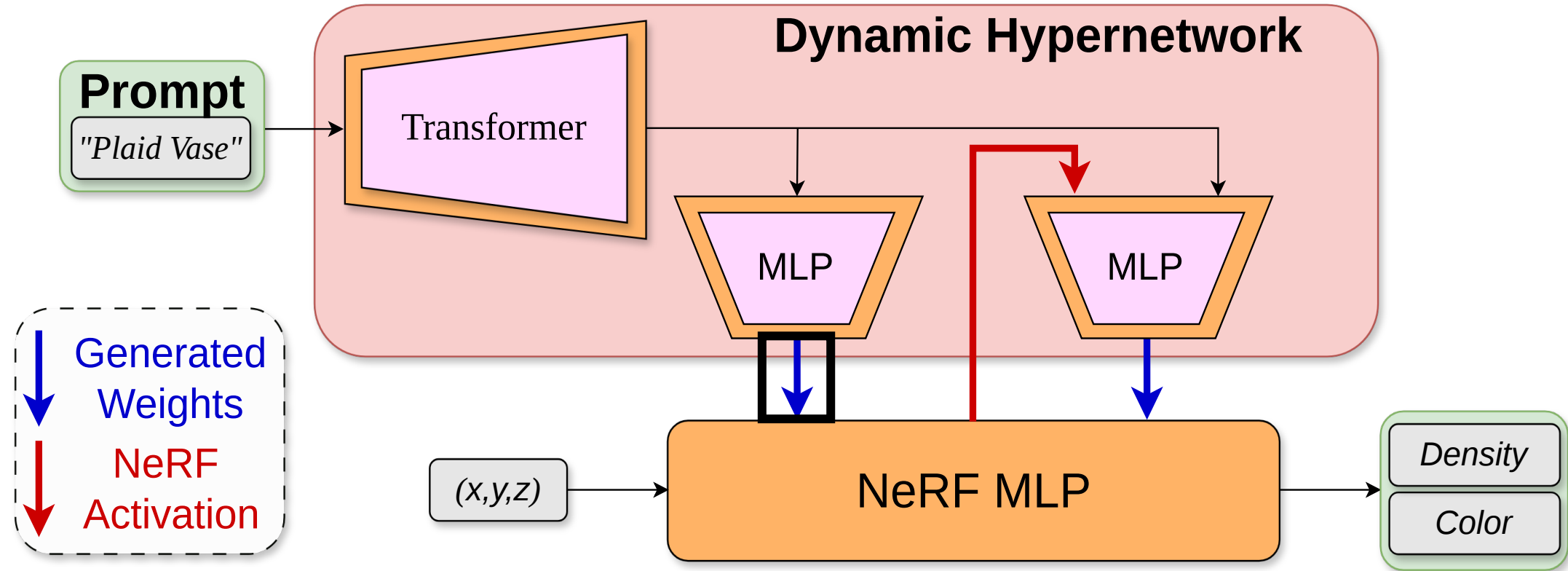
# Approach: HyperFields Architecture



- **Forward Pass:** Text prompt is mapped to the weights of NeRF MLP
- **Dynamic Hypernetwork:** **Generated weights** are function of **activations** and prompt

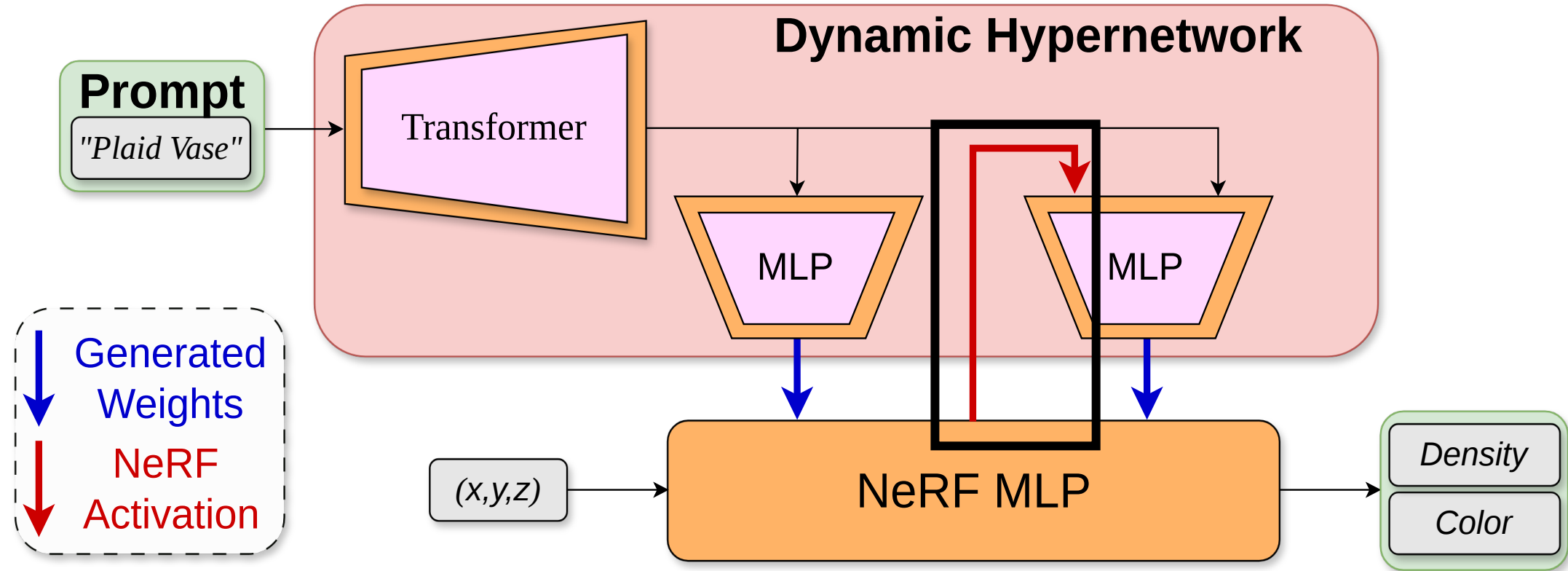


# Approach: HyperFields Architecture



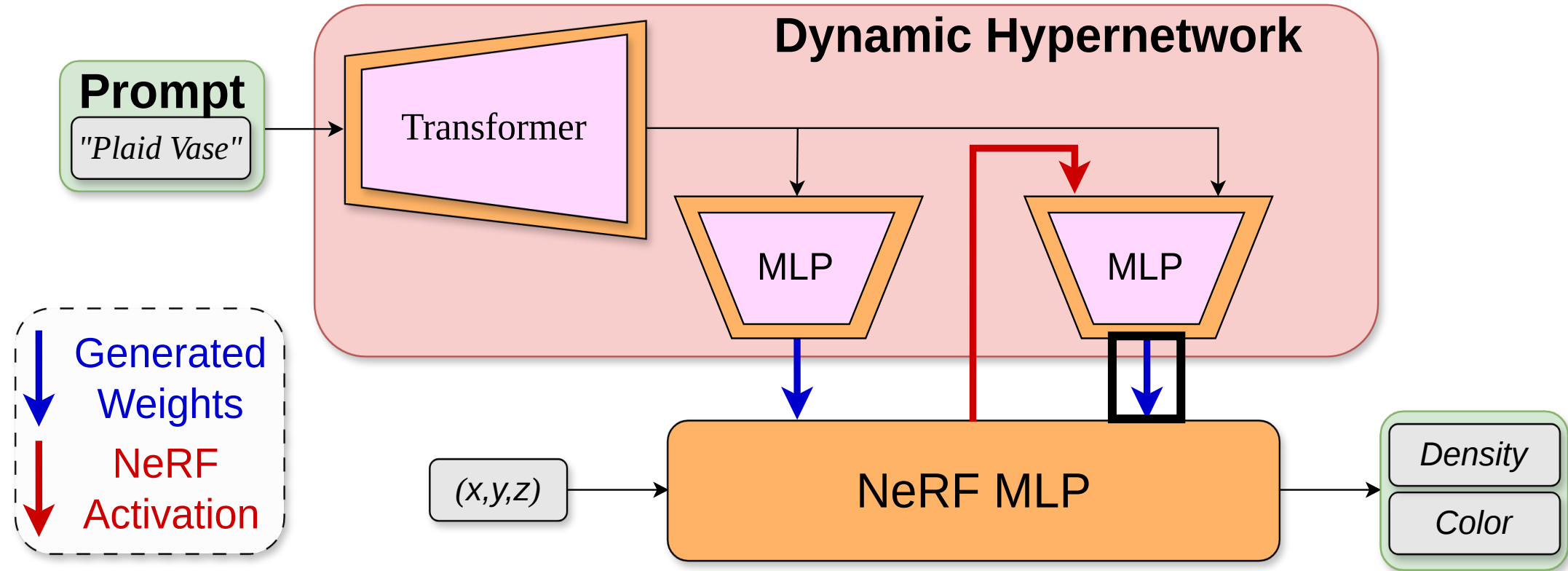
- **Forward Pass:** Text prompt is mapped to the weights of NeRF MLP
- **Dynamic Hypernetwork:** **Generated weights** are function of **activations** and prompt

# Approach: HyperFields Architecture



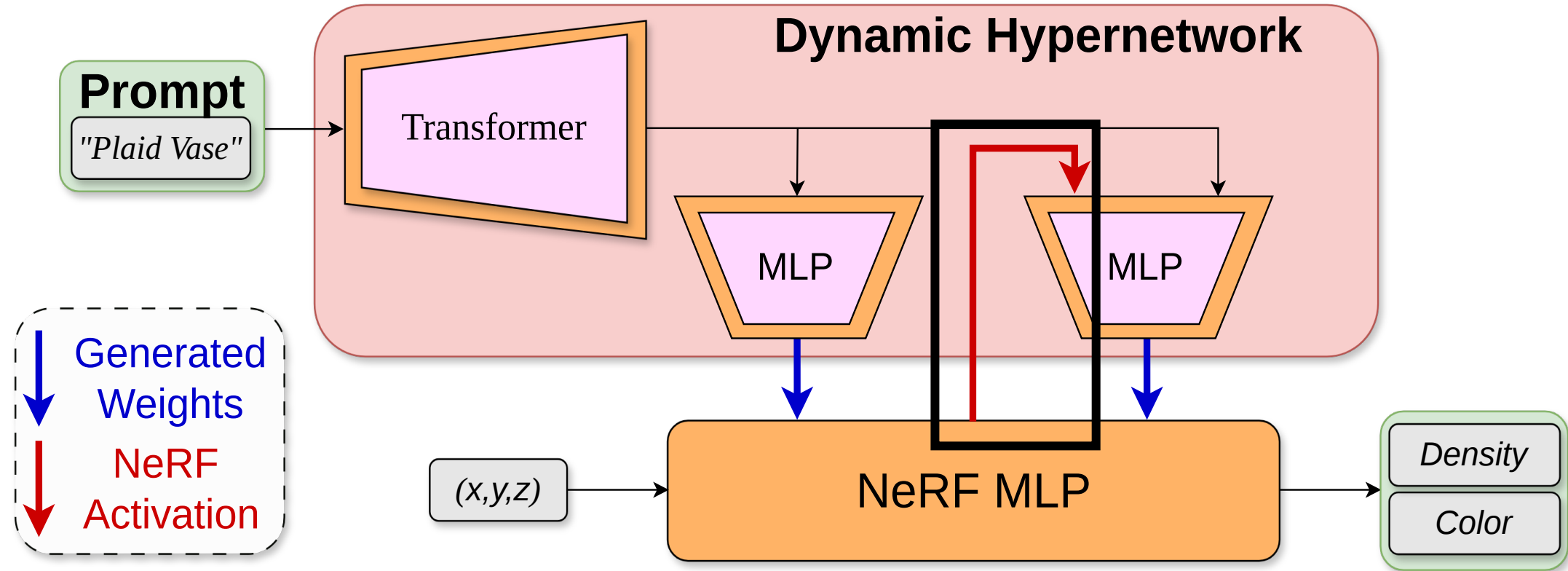
- **Forward Pass:** Text prompt is mapped to the weights of NeRF MLP
- **Dynamic Hypernetwork:** **Generated weights** are function of **activations** and prompt

# Approach: HyperFields Architecture



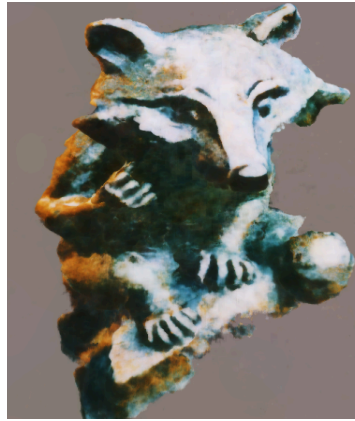
- **Forward Pass:** Text prompt is mapped to the weights of NeRF MLP
- **Dynamic Hypernetwork:** **Generated weights** are function of **activations** and prompt

# Approach: HyperFields Architecture



- **Forward Pass:** Text prompt is mapped to the weights of NeRF MLP
- **Dynamic Hypernetwork:** **Generated weights** are function of **activations** and prompt

# Results: Packing Multiple Scenes



Michelangelo statue of a racoon



An elephant with a shovel



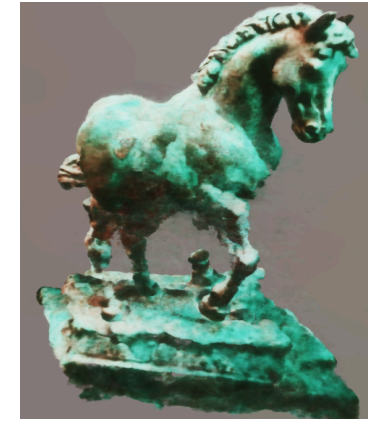
A bear playing chess



A rabbit playing chess



A frog playing chess



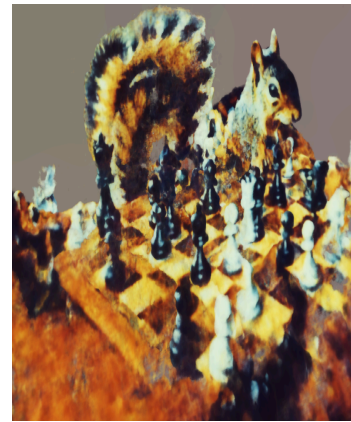
Michelangelo statue of a horse



A monkey reading a newspaper



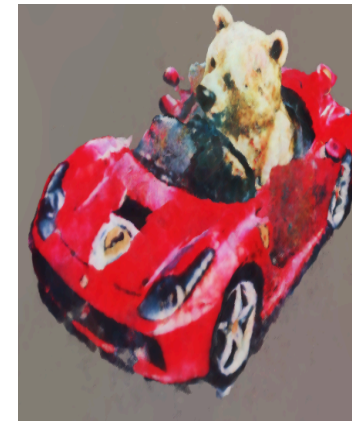
A rabbit wearing a sweater



A squirrel playing chess



A cat reading a newspaper

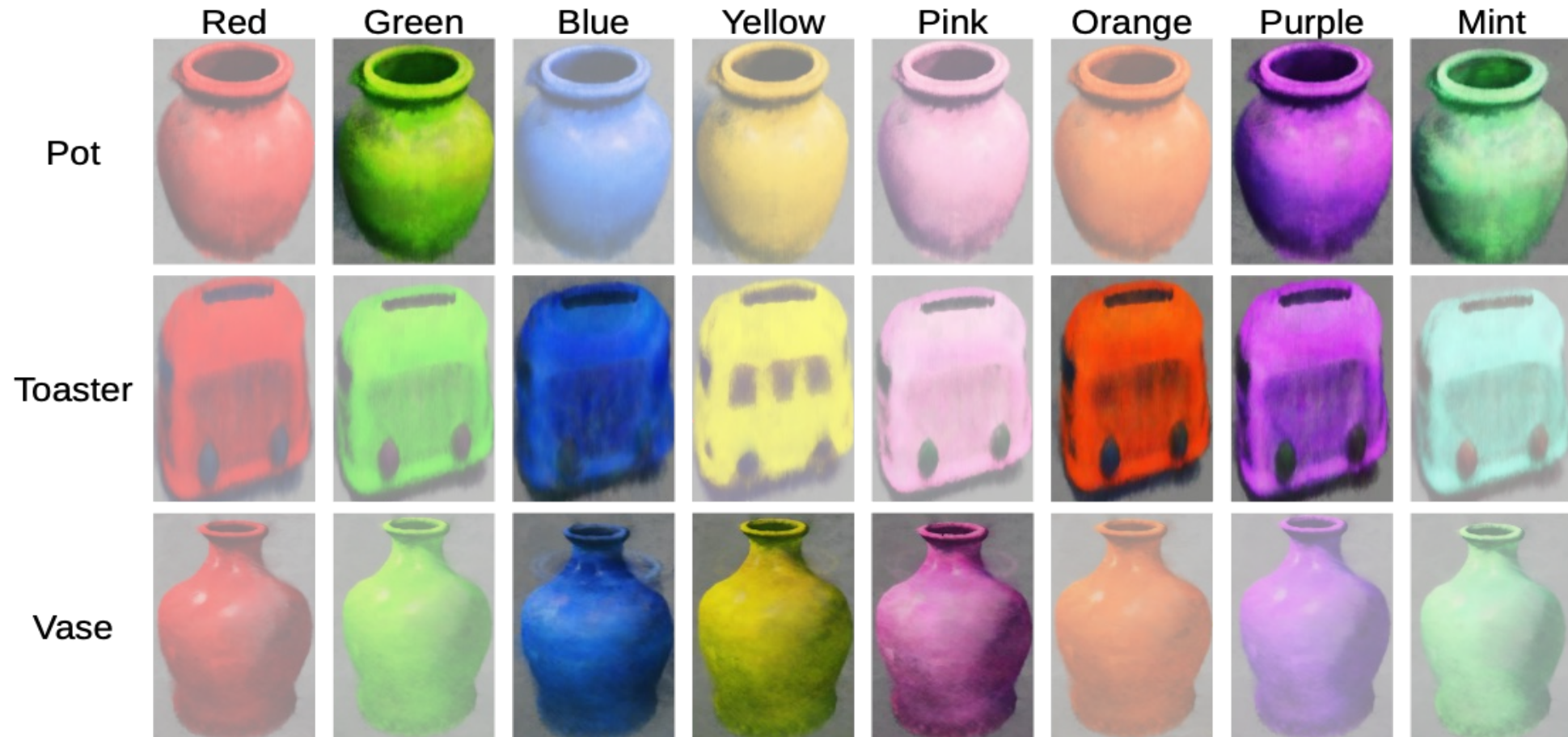


A bear driving a Ferrari



A rabbit playing a guitar

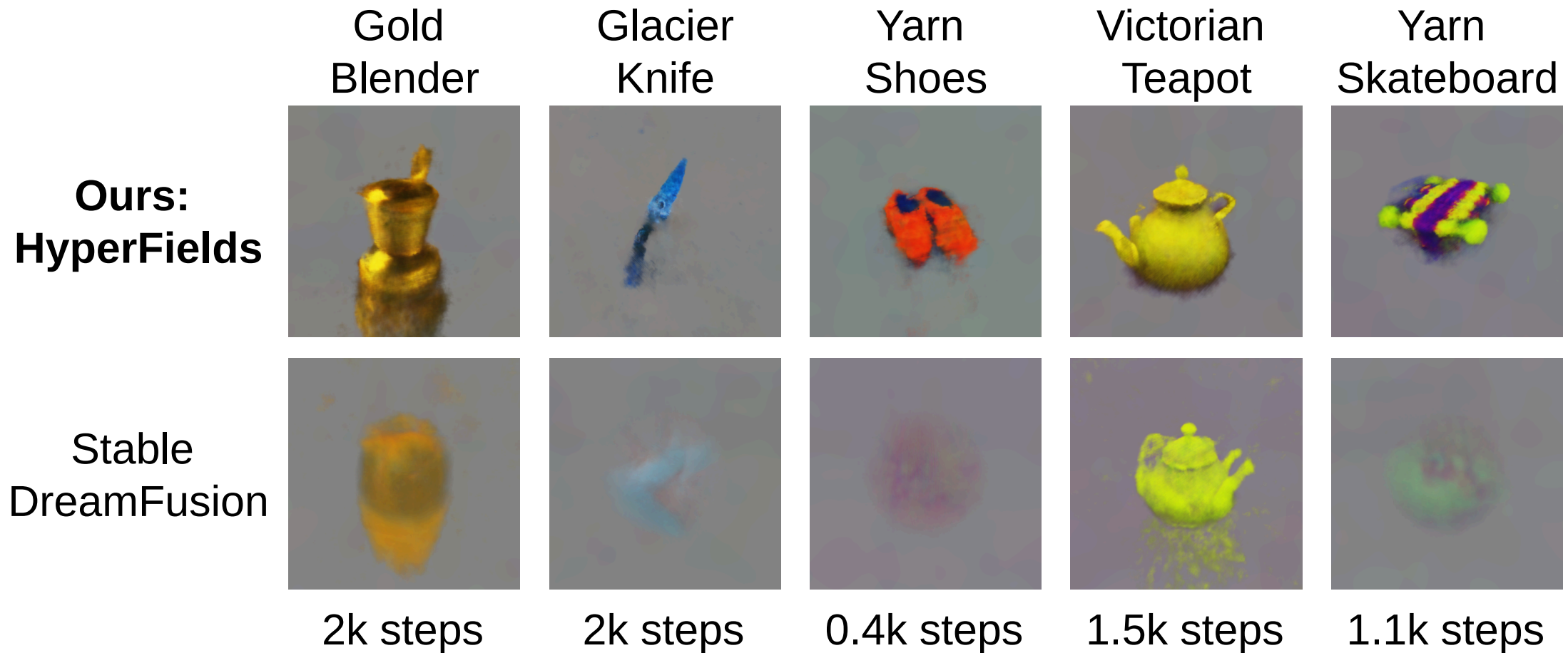
# Results: Zero-shot Generation



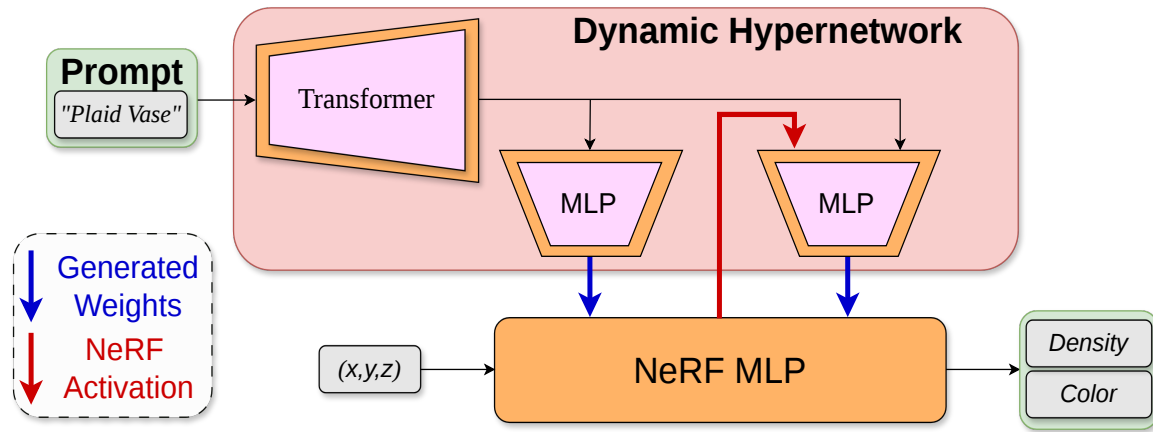
➤ Train: Faded, e.g., **Pot**, **Toaster**, **Vase**

➤ Test: Unfaded, e.g., **Pot**, **Toaster**, **Vase**

# Results: Generalization to OOD Scenes



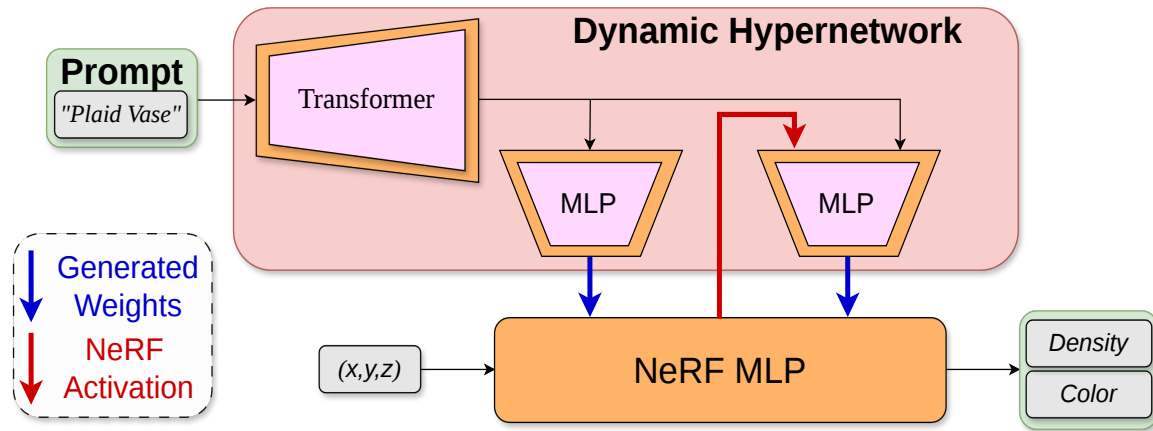
➤ Improved quality for the same steps



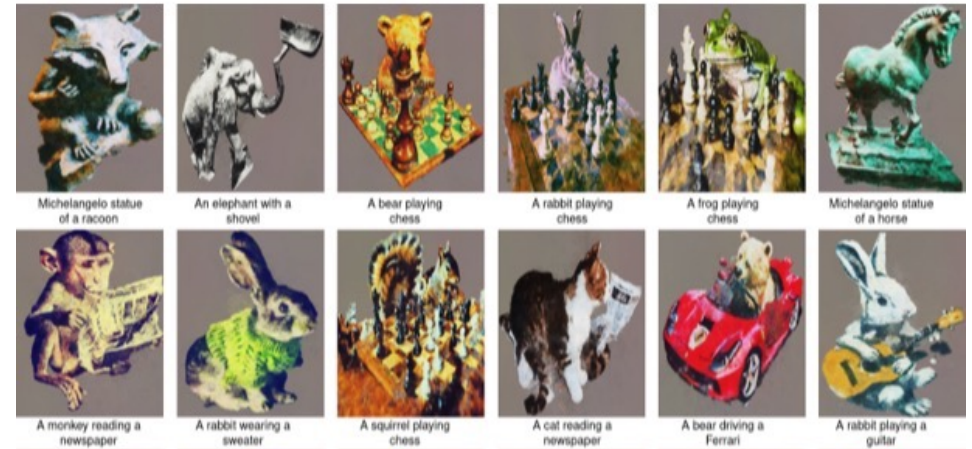
## HyperFields: Dynamic Hypernetwork



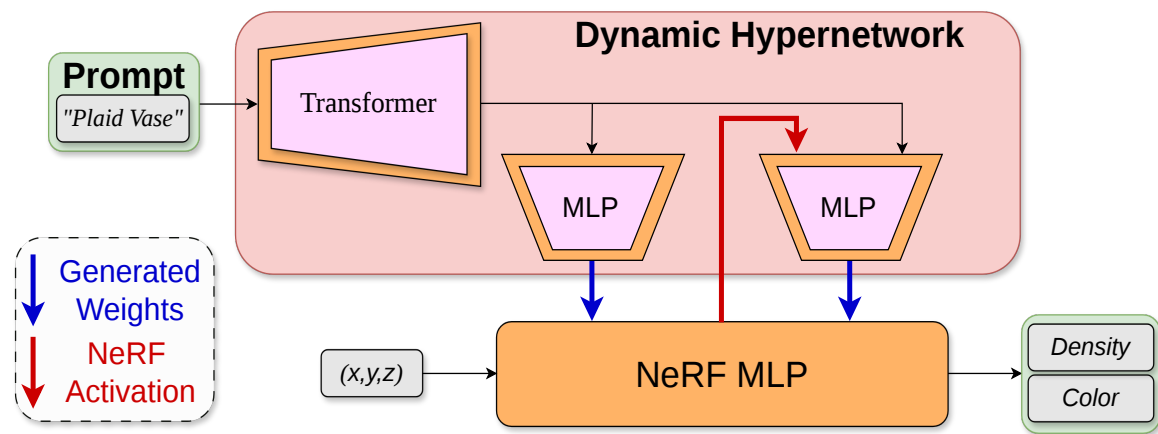
# Summary



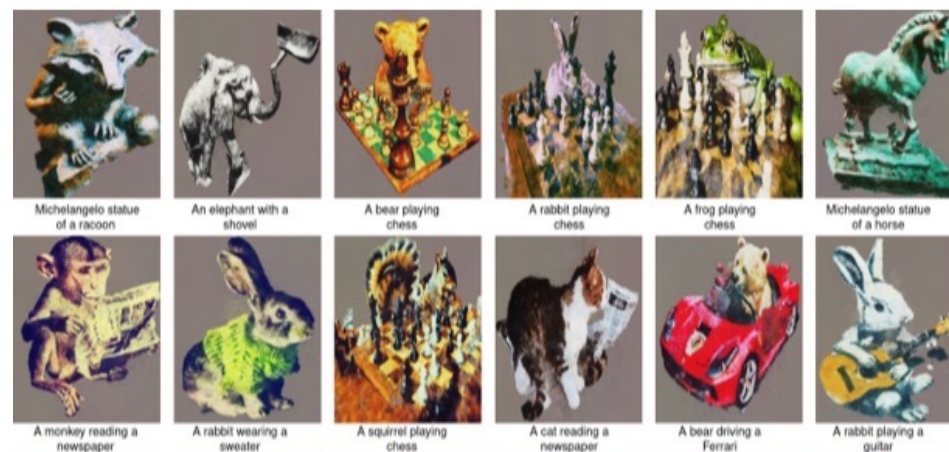
**HyperFields: Dynamic Hypernetwork**



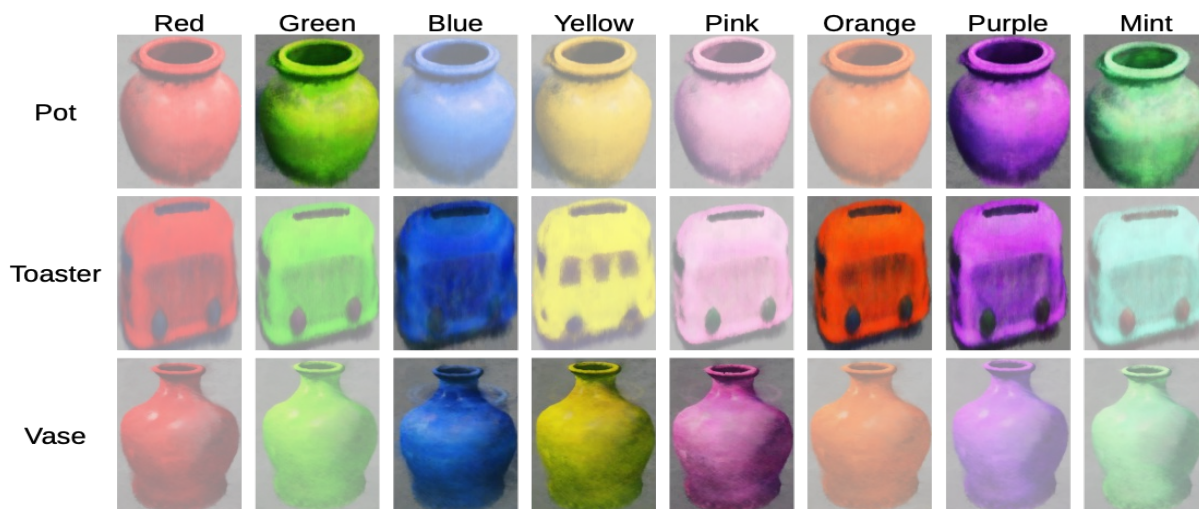
**Multiple Scenes Packed in Single Model**



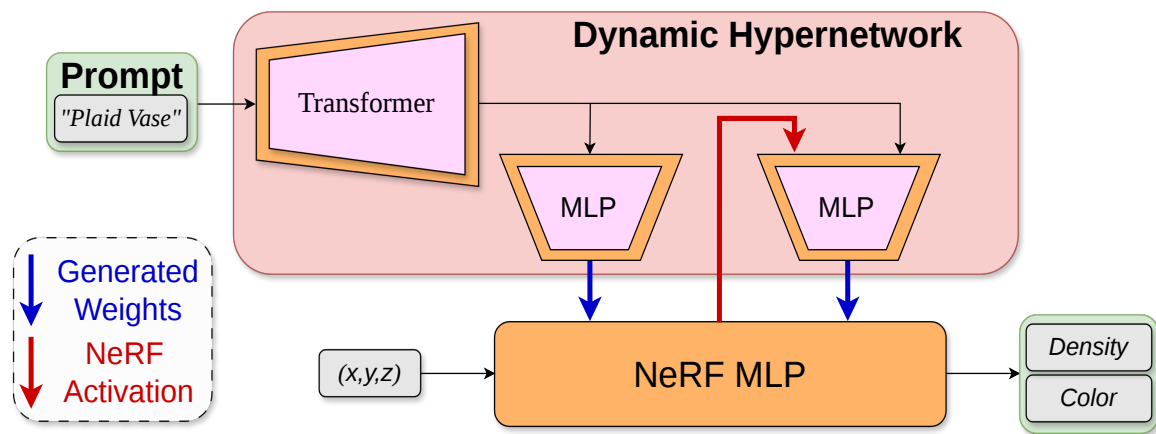
## HyperFields: Dynamic Hypernetwork



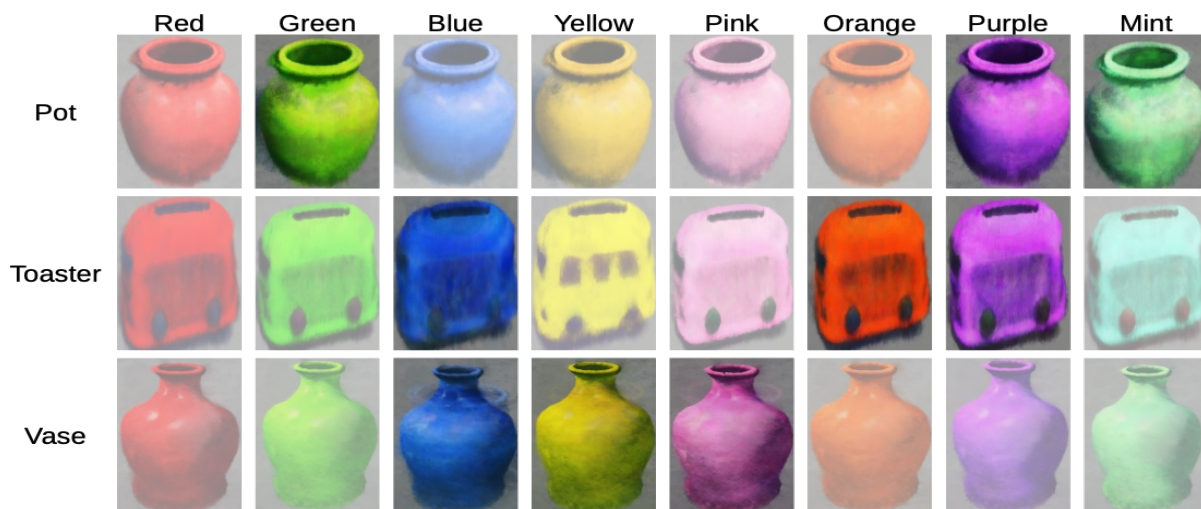
## Multiple Scenes Packed in Single Model



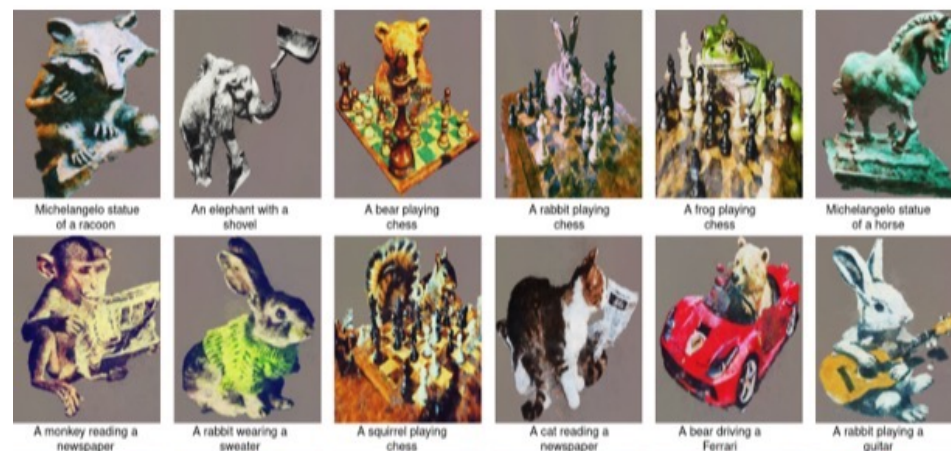
## Zero-shot Generation of 3D scenes



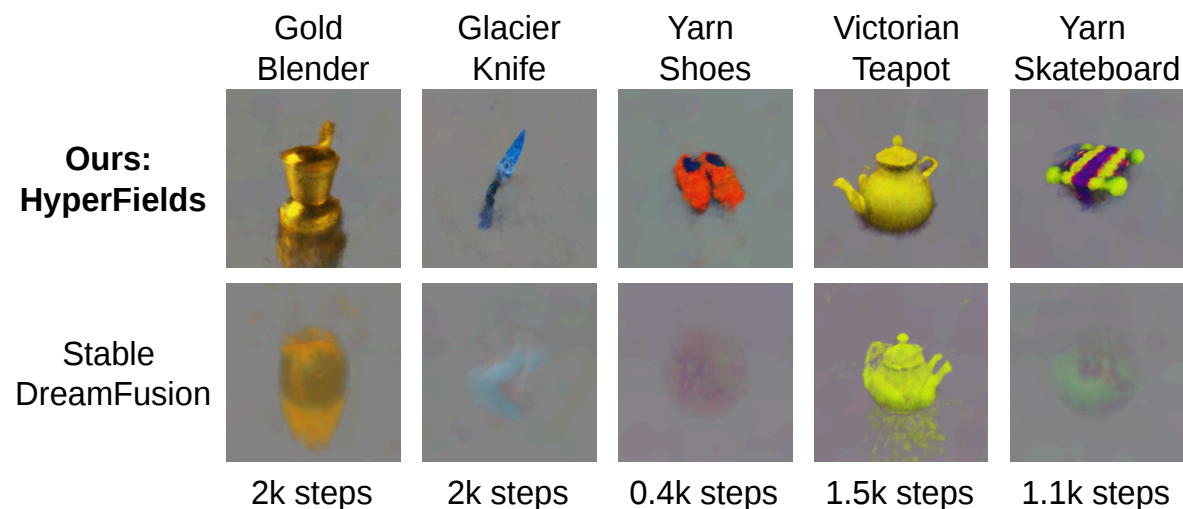
## HyperFields: Dynamic Hypernetwork



## Zero-shot Generation of 3D scenes



## Multiple Scenes Packed in Single Model



## Generalization to OOD Scenes