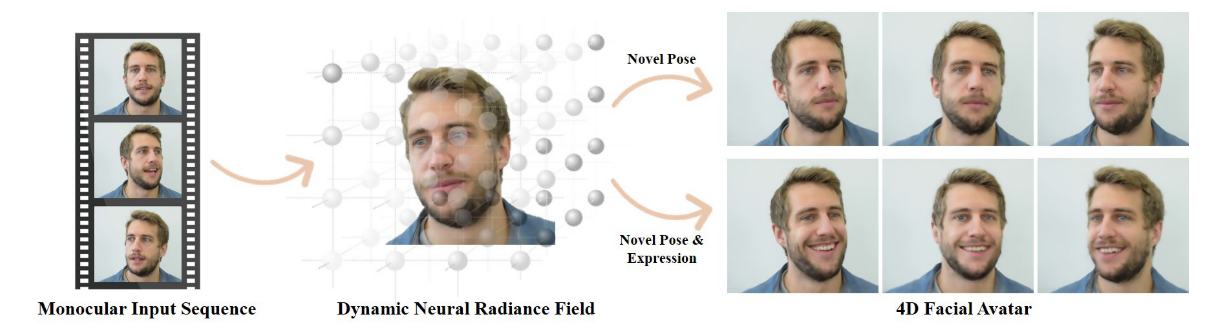
**Dynamic NeRF 2021**

****

Our approach is a neural rendering method combining classical volume rendering with a novel neural scene representation network to achieve novel head pose

and expression synthesis.

1.显示重建数字人头，高频皮肤细节需要“multi-view studio setup”，头发处理需要approximated by retrieval and refinement of hair styles [18, 50], which leads to an unrealistic visual reproduction.

2. Note that the scene representation network is not only conditioned on the sample point locations but also on the expressions of the morphable model which allows for the dynamically changing content that has to be stored in the neural network.

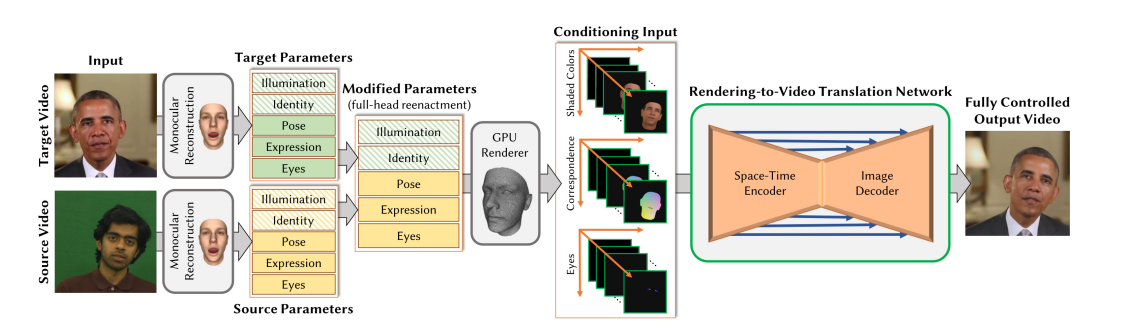
During test time, this conditioning allows us to apply novel head poses as well as expressions to synthesize a new image.

3. Hu et al. [19] combine face digitization and hair reconstruction to estimate the head geometry and appearance from a single image.

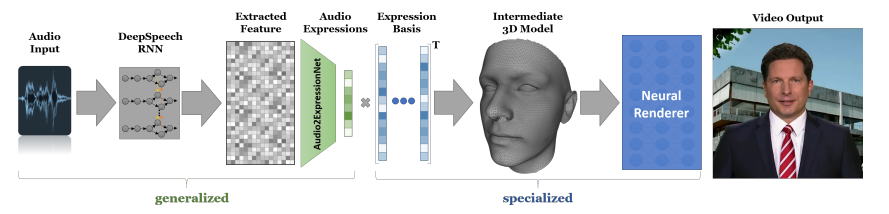
4.对于image2image的生成式方法：[3]warp；

***(1)feature map:***

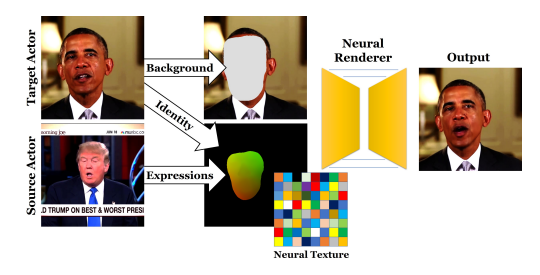
hybrids between classical rendering and learned image synthesis；Deep Video Portraits [21] is one of the **first method**s that uses rendered correspondence maps together with an **image-to-image translation network** to output photorealistic imagery. Deferred Neural Rendering [39, 38] extends this idea, by introducing neural **feature** descriptors that are embedded on the surface of a coarse reconstructed face mesh.



*[21]* *Deep Video Portraits(2018)*



*[38] Audio-driven facial reenactment(2020)*

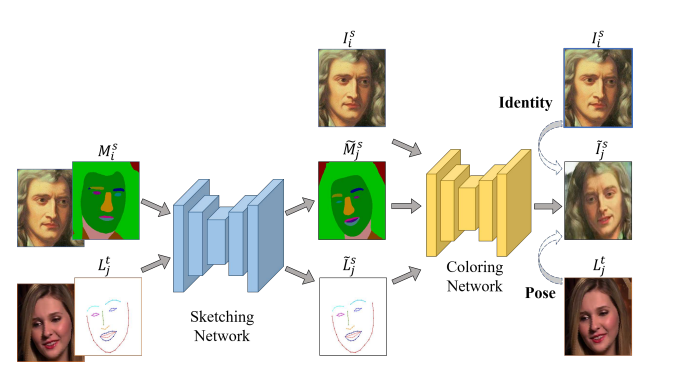
**

*[39]Deferred neural rendering: Image synthesis using neural textures.(2019)*

Instead of this dense conditioning input or rendered feature maps, there are also methods that work on rendered facial landmarks [48, 10, 47]. These approaches can also be applied to single images. First Order Motion Model [32] is a data-driven approach that decouples appearance and motion in a video of a specific class (e.g., human faces) and allows application of the motion in a source video to a target image.

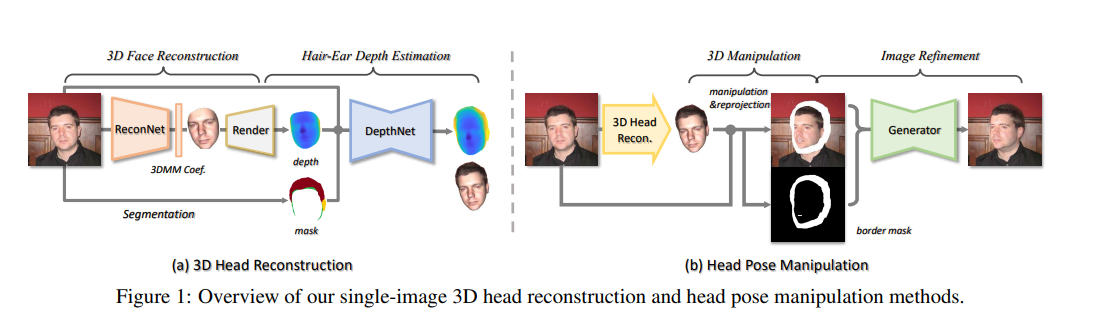
***（2）关键点/landmarks***

Puppeteergan:



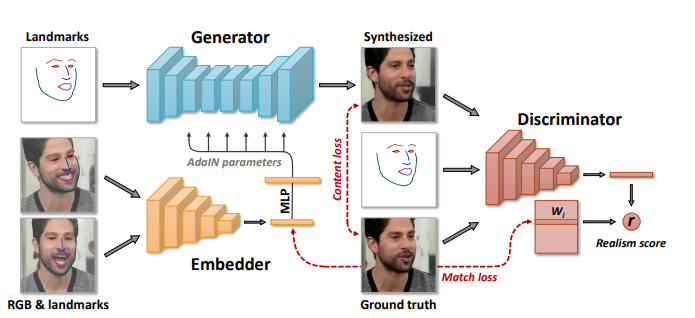
*[10] Puppeteergan(2020)*

Deep 3d portrait from a single image:

**

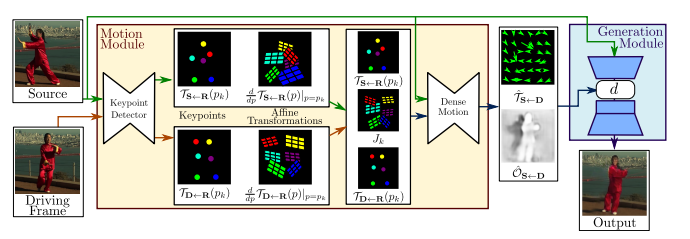
*[47] Deep 3d portrait from a single image(2020)*

Few-shot adversarial learning of realistic neural talking head model:

**

*[48](2019)*

FOMM(2019):



*[32]*

***（2）Neural Scene Representation Networks(SRN) -implicit representation***

A summary of neural rendering approaches is given in the state-ofthe-art report of Tewari et al. [37]. Sitzmann et al. [35] introduced neural scene representation networks (SRNs).

*[35]Scene representation networks: Continuous 3d-*

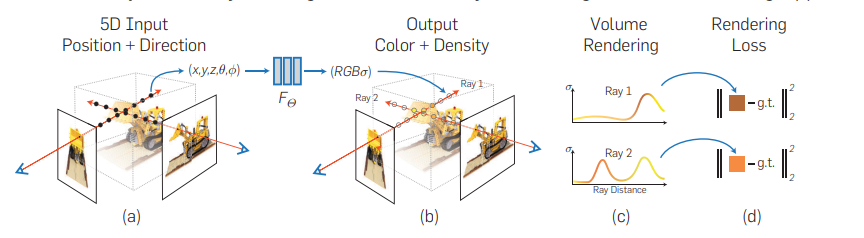
*structure-aware neural scene representations. (2019)*

*[36]State of the art on neural rendering(2020)*

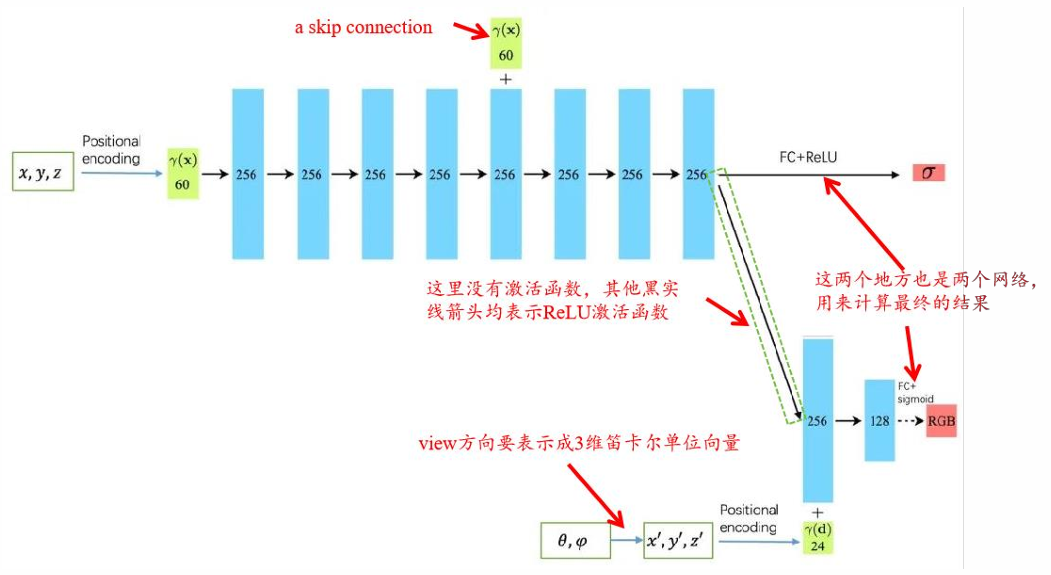
The geometry and appearance of an object is represented as a neural network that can be sampled at points in space. A ray marching approach is used to sample from the neural network to render the reconstructed surface.

Mildenhall et al. [28] extend this idea to store radiance fields in a neural network.**(NeRF)**

*(They assume a static object and multi-view data. A key contribution is the volumetric in tegration and the usage of positional encoding for higher detailed reconstructions.)*



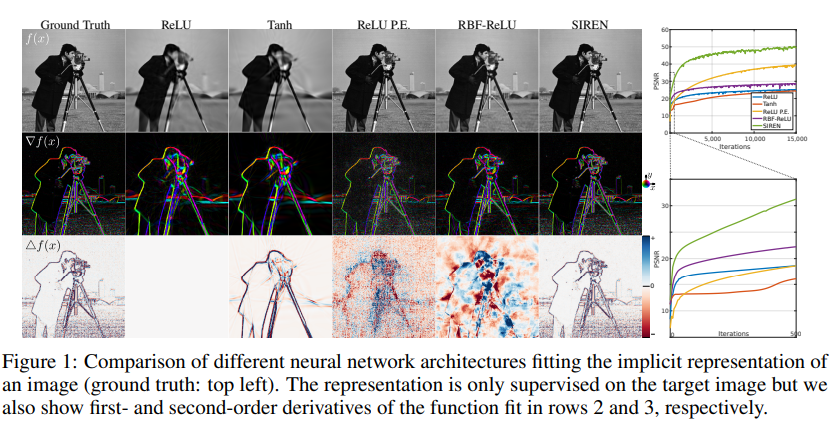
*NeRF(2020)*



*NeRF的神经辐射场MLP网路*

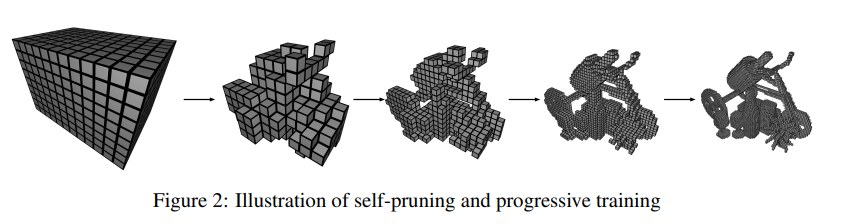
Follow-up work extends this idea by using different positional encodings and in-the-wild training data including appearance interpolation.

Concurrent work of Sitzmann et al. [33] proposes the usage of sinusoidal activation functions for the scene representation network.(改进NeRF高频细节)



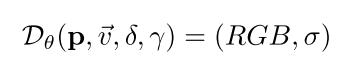
*[33]SIREN(2020)*

Neural Sparse Voxel Fields [25] employ an Octree to cull empty space and speed up rendering.



*[25]NSVF(2020)*

We use a similar volumetric integration scheme to [28] with an additional layer for the static background. The dynamic neural scene representation is not only conditioned on the sample position and view direction, but also on the facial deformations.



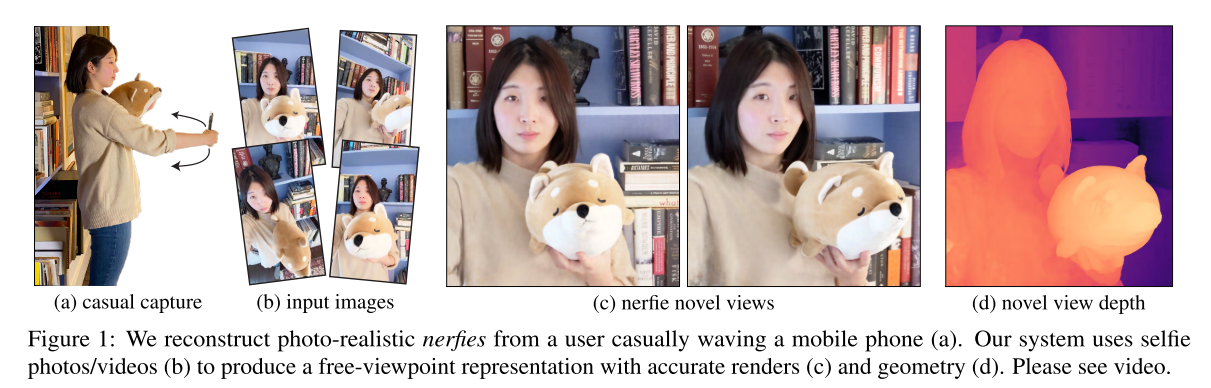
Pose参数4\*4，让人脸采样点从观测空间转换到规范空间（正面）为公式里的P，表情参数76维条件控制神经辐射场；添加额外的隐编码γ（32）维，目的是弥补损失的信息。

**Nerfies: Deformable Neural Radiance Fields-2021**

**3D 任意场景重建**

Our approach augments neural radiance fields (NeRF) by optimizing an additional continuous volumetric deformation field that warps each observed point into a canonical 5D NeRF.**（deformation field 到底解决了什么问题？）**

We observe that these NeRF-like deformation fields are prone to local minima, and propose a coarse-to-fine optimization method for coordinate-based models that allows for more robust optimization.（By adapting principles from geometry processing and physical simulation to NeRF-like models, we propose an elastic regularization of the deformation field that further improves robustness.）



We show that our method can turn **casually captured selfie photos/video**s into deformable NeRF models that allow for photorealistic renderings ofthe subject from arbitrary viewpoints, which we dub “nerfies.

**先前工作**

1. Non-Rigid Reconstruction:[9; 10; 14; 15; 22; 25; 33; 39; 48] [20; 34; 52]

Earlier works focused on sparse representations such as keypoints projected onto 2D images [10, 48], making the problem highly ambiguous. Multi-view captures [14, 15] simplify the problem to one of registering and fusing 3D scans [22].

**DynamicFusion** [33] uses a single RGBD camera moving in space, solving jointly for a canonical model, a deformation, and camera pose.

More recently, learning based methods have been used to find correspondences useful for non-rigid reconstruction [9, 39].

Unlike prior work, our method does not require depth nor multi-view capture systems and works on monocular RGB inputs.

Neural Volumes [25] learns a 3D representation of a deformable scene using a voxel grid and warp field regressed from a 3D CNN.

(2) Domain-Specific Modeling:

1）人脸领域[4; 6; 8]

In contrast, our work does not rely on domain specific knowledge, enabling us to model the whole scene, including eyeglasses and hair for human subjects.

(3)NeRF发展[3; 24; 29; 41; 53]

These methods have been used to represent shapes [13, 31, 35] and scenes [32, 44]. Of particular interest are NeRFs [32], that use periodic positional encoding layers [43, 47] to increase resolution, and whose formulation has be extended to handle different lighting conditions [3, 29], transient objects [29], large scenes [24, 53] and to model object categories [41].

Our work extends NeRFs to handle non-rigid scenes.

（4）形变场[23; 37; 49; 50]

Two concurrent works [37, 49] propose to represent deformable scenes using a translation field in conjunction with a template.

