Slide 2.
3D reconstruction, also commonly addressed as the inverse graphic problem, is about recovery of a 3D object from 2D or partial 3D observations. The research interest in this field has been rising rapidly due to its wide application in robotics, simulation and content creation.

In contrast to the learning in 2D domain, there is an ongoing dispute on which 3D representation is the best for learning. The desired representation should be compact, expressive and efficient.

Slide 3.
Before the deep SDF and other implicit representations came out the dominant representations could be categorized into three categories: their voxels, point cloud and mesh representations.

voxel can represent arbitrary typology but it has a cubically growing compute and memory requirements.

more compact representation **point cloud** but do not describe surface

methods that produce triangular **mesh** directly at that time, are limited to the topology of the template that they are using. Also they often exhibit artifacts from inaccurate deformation as shown in the example

Recently there are a number of works proposed to overcome the issue of fixed templates; these methods can output mesh with arbitrary topology such as in MeshRCNN and DefTet.

Slide 4.
prior work on 3D representation lacks the ability of representing fine surface details.
The key inside of this work is to directly regress the continuous sign distance field using a neural network. a sine distance function is a continuous function that for a given point in 3D space it output the points distance to the closest surface with the sign which encodes whether the point is inside or outside the watertight mesh. The underlying surface boundary is inconsistently represented by the iso-surface in which case SDF value equals zero.

Slide 5.
This is more informative than just having occupancy value and conceptually the surface can be extracted with less query points than the concurrent work occupancy network.

To compare occupancy function with sine distance function, let's look at a quick example here, we are running a marching tetrahedron with a grid size of 100. if we only have occupancy value which is decided at each grid position, we have to place the surface at the center of each edge and the resulting shape is very wavy.

However if we have the ground truth signed distance value,
We can place faces by interpolating the distance and the result is much closer to the ground truth while we query the same number of points.

Slide 6.
Contributions
With that being said the sign distance function is a very expressive 3d representation.
in this paper the authors proposed

A formulation of generative shape modeling with SDF in continuous domain,
notice that they are not the first ones who use SDF in the learning setup. A previous work from Geiger's group proposed a model that predicts SDF at discrete locations. It is similar to the last example I showed you.
However deep SDF is the first work that's that use neural network to approximate SDF in continuous domain

Moreover, they propose a novel learning method based on a probabilistic auto decoder.

Further they demonstrate the application of their formulation by obtaining state-of-the-art results on shape reconstruction and completion next i will dive into the method details

Slide 7.
Problem Setting
Deep-SDF regresses the sign distance function using a fully connected neural network and this function takes input as xyz coordinate And shape code z and outputs the predicted SDF of that particular position conditioned on the shape code.

In addition for shape reconstruction we only care about the surface boundary,
so we want the network to focus its prediction near the zero level set therefore they further clip the SDF value with a small threshold.

Slide 8.
The author empirically showed that with a smaller truncation distance they achieve better reconstruction quality.

Slide 9.
Learning
Next, let's look at how to obtain meaningful codes for a batch of shapes. We want the shape code z to encode the information of a single shape that can be interpreted by the decoder.

One approach is to train an auto encoder such that the latent space of the bottleneck layer will automatically be the shape code.
However the authors claim that it is not straightforward to design an encoder on SDF, also since only decoder is retained for inference it is unclear whether using an encoder is the most effective use of computation resources during training.
This drives them to use a framework as they called auto decoder for learning a shape embedding without the encoder.

They first randomly initialize the code for each training shape, then the code is attached to xyz coordinate and with the ground truth SDF values they can jointly optimize the shape code as well as the decoder weights to minimize the prediction error with simple back propagation.

Slide 10.
Inference
At inference time the optimal code that best explains the input shape can be obtained using gradient descent minimizing the same objective function.

Notice that during optimization the trained decoder weights are kept frozen and only the code is optimized.
The full reconstruction can be achieved by running inference with the optimized latent code in full 3D space.

This example shows shape completion from a single depth map.

Slide 11.
Surface Extraction
There are two ways to use deep SDF representation.

In real applications the deep SDF can be used in the rendering process with ray casting where the user should raise from each pixel until it meets a zero crossing to obtain the depth map.

The surface normal can be calculated by taking gradients through back propagation.

Furthermore a marching cube algorithm can be applied to extract the mesh

While both methods can be used for iso surface extraction in practice marching cubes runs faster but it also imposes quantization error due to fixed screen size

Slide 12.
Results and Discussion
Now let's look at some results
Although deep SDF extracts the shape code by optimization at inference time it can effectively generalize to unseen shapes.

Their method significantly outperforms Atlas-net on a wide variety of shape classes and matrix as shown in the table.

Next they demonstrate their method on the shape completion task from partial scans, basically a single depth maps.
Deep SDF produces more visually pleasing and accurate shape reconstruction than the state-of-the-art approach at that time.

Notice that one key advantage of deep SDF pipeline is that the inference can be performed from an arbitrary number of SDF samples with the same trained model,
For other models like occupancy networks typically it needs a separate trained model to reconstruct shape from complete point cloud and partial point cloud.

In addition they test the robustness of their shape completion method by using the noisy depth map as input as shown in the figure,
As the noise level increases the reconstruction error increases much slower implying that the shape prior encoded in the network plays an important role regularizing the shape reconstruction.

This table is showing that deep SDF has a much smaller network size compared to all other methods using different Representations, however the inference time is also much higher than other methods because of the iterative optimization at the inference time.

They show an interesting demo video of feature space interpolation as you can see the transition between shapes is smooth as the parts emerging and changing size demonstrates that the learned feature embedding space is complete and continuous.

This explains the capability of reconstructing novel shapes which could be seen as the interpolation of two scene shapes

Slide 13.
Limitation
One limitation of this work is that inference needs optimizing latent code with SDF, this is not applicable to 2D observations.
A later word called disney addressed this problem with a novel image encoder.

Another limitation is that decoding with mlp taking a global vector as input leads to over smooth results, one approach to this problem is to use a 3D feature volume along with 2D feature map in orthogonal views. This one is proposed in the paper called convolutional occupancy network.
This work is capable of representing scene level details.

Slide 14.
In **conclusion, in the** deep SDF paper the authors proposed the formulation of generated shape modeling with sine distance function in continuous space that is efficient, expressive and continuous.

Also they propose the learning method for 3D shape based on a probabilistic auto decoder, further they demonstrate the application of their formulation by obtaining state-of-the-art results on shape reconstruction and completion.