### CG\_week5\_tu05face\_restruction\_1-20241012

说话人1 00:00  
You are all about the three and ten generation. Okay? First of all, we need to introduce how can we prevent the three objects. We usually know we have to do that the two d to give them these images. However, how can they represent a three d object? How many there are four representations of the first one is the breakup at least the breakout we can use the point to represent of ten. And also we can use the most. Also we can consider it as as a of aqa three q and the objective is consist of many, three qs. But we can see the also involves in high quality because of the low resolution. And also we can represent us really competitive smash. The nash is based on the way out. We can represent a for example, there is 3 points there. And we can write, we can use a triangle to represent the face of the the patch. And the sweeping can be considered consist of any three cases.

So the last one is the implicit function. This function is very popular about you can use the skf or occupancy or union to represent three d intent. And I I will give you more information about the basically function later.

Okay. First, I will discuss a recent progress in the three d kinetic generation. Have you used ai not a large language model, like a chess, ppt or wechat or doble or something? You can give the model image for a test, and you can ask the model to generate an image. For example, we can generate from a test, a strong force to get to get the image of the force. And recently writing the wave of the each generations as the three generation is forming, the target is to generate a three d shape. Typically, when you use the three mesh to represent the 37, and as soon as a rotating force, and three generation close us to view the three d content from x angle. And when we're using a three d immigration force, the left side, a color relaxed the surface in hurrying texture information. The right side are visualizes the normal methods.

And I go to lesson, i'm going to introduce the technology of generating a three d return from a task or for education as driving to the process of generating three content from a single page by sampling a distribution.

We will start with an image to three d example when we look at this 4th image. Look up. So when we look at the quality to the task of generating three match from this single movies, I am interested why is that? Because one image that is a contain enough information about the 43 structure. For example, when you look at the front of you, you cannot guess texture behind me of the bedside of my body. And this is a task is not a domestic. It's random, meaning there would be many solutions.

In this scenario, we need to we have to do this from the buyer distribution, the p distribution of this, of the three model. When given the input image, and this represents four possible three dimensions that can can be assisted with the input image and the key, then take away. There is that generating a three shift from a two image involves dealing with uncertainty. And we can approach it by assembly, possible what comes from the prior distribution.

Okay? For the generation problem, by sending from some distribution, the coding image generation has already paved the way to solve it. The key idea is that it solves the sampling problem using the duality as a proxy task. Such method is called the choosing models. And do you know any other methods to generate the two images, stable computing, endorse or computing method? Before the computer method? There there was avaevae and globalizing slow people in three methods to generate, to generate to image or some other things. Facts.

Recently, the diffusion ebola as probe has brought this a strong ability, which how do you know how to feel about the works? All the people that you don't know. So some people don't know how to, if you have more books. When I hear I think to the english, as a example, the drain model, the drain model has two processes. The first one is the whole world. And the the second one is the backward. I don't remember which one is the owner of the the land. So first, when we have an image, we need to act the gaussian, the gaussian voice. The gaussian law is to the image and a usually we use 1,000 steps, and the image will gradually become totally amounts of goods, a lots of goods.

And when we try to follow, we need to know

how can we, how can we generate. You need from the concept of the option. The customers, the solution is we can graduate minus minus the abortion employees. And finally, we can achieve the final image. And during the training process, we need to predict how much noise we did to minus in the background. The process.

In advd you can see like an image. You can ask the laws of oil and fire. It can be like this, okay? And all of us who divorce the possible ways to get the money, who gets the clean. And the model can be conditioned on additional information, such as a test or prompt for some single view images or value images. And this work and this works well.

We can try opinions capture pages from the internet and makes a model variable energy and stable, who is kind of the success of the union images to the three d format. We have a large connection of the three d of test. We we can achieve the image and the caption from the internet. And it is very difficult and very expensive to let the three objects. Last year, I used a three d scanner to collect the three d human body, 100 three d human body, and I used 3 months to finish this work. So it is a very challenging work to collect the freedom test. So we usually hope to directly use the two english generation method to generate the three d it's a regular three d mesh. Okay? So that we usually do is that the free trade models from the 20 each generation, we can copy a chinese generator and a functioning, so that instead of instead of conditioning on task, in condition, on a changing angle and interview.

Maybe two generally a lot of you notice that the model has been tracked on millions images from the future authority already task. We can use a small amount of city there to create about you and how to use images of the 3 days. And we can use the multi view, which is to find you the new track model or the local view.

Let's start from wang one single viewing. If we can send more distribution of the we can support distribution by changing the camera angles and visions and so that we can make inspects. We can generate different angles of the image, and we can finally use a three d you can structure all with them to generate the three d match of the course.

If you are interested about this, you can google the dream fusion. This one has been published in the ipo is. And to summarize, the process occurs in two stages. First, starting with a promos, either a wish for a test, they are developing new generator produces a set of no use of the tax. That's a similar network or possibly aircraft the chinese. We construct the studio tap from those views, and these stages can be, especially for productivity, depending on the specific lead. And this organize the general master we used to tackle this problem. You might be wondering which what makes this process make more symmetry. One key factor is the consistent the consistence of the generator when generation no use of the chance. Here, the persistence means the generated views look like the same as the input view. Ideally, you want to sample healthy meals as hold of the h one h school, and that were consistent with initial view. This means the generate in global views from a joint distribution.

However, in reality, due to the mutations, in how well we can try the model. What all happens is that the model sentence use from a approximation between the joint distribution and the product of the marginal distribution. Instead of perhaps assistance across for you, you get something closer to the patent samples that not for it is a relationship between the use in a part.

And recently, in 3 generations from the use we based on assistance and construction trade off. On the x axis, we have very few persistence which measures how will different viewpoints along and on the wires. Is we have a reconstruction emperors representing how close the three d shape matches the real attempts. And a recent method for the values. Three d generation are essential. And we will push towards the upper right corner of this group means we need higher persistence and more effort to results. For lower marketing consistence, the model tesco generates lower views based on the products of the marginal distributions. From each view, this suggests that the relationship between the different views haven't been fully captured for this problem.

On the other hand, for the higher revenue consistency, the model is generating all these from a joint distribution. This means the model is better at understanding and utilizing the relationships between different views probably is more consistent results that i'm going to introduce some information about the single rule, human instruction. And we constructed a new model from the single image. A place are very important rules in various applications such as the gabby to the film production and sporting and broadcasting. And we can see for videos and where we inputs the english of the actor. We can reconstruct the new story. I know you have seen on them. This reports and music, as you see, and we can see the answer from different views, right? This is the human reconstruction.

And so, okay, the problem is, when we have a single view, arginine image, we aims to reconstruct the service of human body, the service of the service in the three nationally. As I said, there are four representations to present three objects and how to learn through information from the two images. Usually there are two channel categories. The first one is abc based method. The second one is explicit based investors. They are going to introduce to increasing investment is methods and long, explicit methods. Let me is this method means even the model wants to learn a few of the three space. And when we have a point in the three space, we can use the model to learning. It's attribute to the mean whether it is in the surface or out of the surface.

Let's first introduce the pipe. The pipe is the elastic method to reconstruct a human body from a single image. Usually we need to prepare the future. We can project the points on the only issues and achieve the peace of their officials activities. We include these images and this three d and it does three d space. We have many points. We want to project the voice on the image, such as the point x we point, we project the x on the to the plan. And we can finally achieve these rtv value, its initials features, and some other features. So the final is the input of the model, is the xyz exercises of the very . and the peaceful level features, such, as I said, the rgb that is the lack of feature and entirely utilize the occupancy that is to run against the three d object. And so the output of the pilot is likely occupancy values. And what is occupancy?

I i'm told that line here, okay? How can we represent this one?

The occupancy means is a binary valuable. It use uses the zero. And one to represent the point is in the surface or outside the surface. At least one in this circle at this point, because it's in the inside of the circle, we can assign the value of one. This is one, and also, this way is also inside so well on. And other points, they are outside the they are outside.

说话人2 21:19  
The

说话人1 21:20  
we assign of the values of zero. This way is we see, so what is five? If we increase the resolution of three, such as the I I don't know we use the besides, when the resolution is enough, we can reach the circle

more accurate like the we see the dont. You don't need? We can hurt the this is a function to service with the article $10. When we learn the when we learn, how can we secure, for example, long.

Okay. This point is zero, this point is one, the line last class, this side and this side, we can determine whether the . is inside or outside the the service. And we can finally extract the service in the market. Q and because the pipe is very simple, it only use the two images, and it has no acquire information about the human, the human body. The human body is a very strong construct. We building a local person have hands or two legs or head. But if the question cannot tackle the situation where there is a lunar situation, for example, the head of the people included its chest and we can see the results of the pilot. We cannot be construct the art of the human body.

So to solve this problem, another world icon is introduces the simple model include the model. The simple model is character can actually tumor. You can use the blender ways to control the the ship of human body. And you can also use some other parameters to control the gender, the age, the pulse of the human body. So we can use the simple model to give us the strong, higher information of the import. In icon, it utilized the sdf area instead of the opponents value. We're at the three object. I also organized a simple, normal feature and across the future to the next as the put of a of a network. And a lot of groups of island is the stf what is sdf I said, as the occupancy is, it utilizes the zero and a lot where next, whether the . is inside or outside the service.

The last step is a difference. For example, here, if a utilize a a distance value represents the three of that. Also, if point is inside of a it's not inside of a service. We can say it's asked the value. For example, this one is this one is each other reply to the distance to learn things. Instead of using a simple binary zero and one values, the problem of the icon is the simple is a very strong constraint. And when we face the images has have losing clothes, we cannot reconstruct the losing clothes likely as this one. The dress on the woman and the icon can not you can start to address.

Okay. To solve this problem, I asked to use another explicit investor as a that exclusive, she generated the body . and the phone call.

And then because we construct the human service from the going from the first, when we get argb input, we first use a depth estimation to learning. So that's and project that's into the three space. And we can achieve a partial. At this one. We put the this english and we can achieve the depths. And when we probably project the depths back into the three d space, we can achieve a a partial income of the human body. And and then then we use a simple estimation as a to ask estimate the simple model from the. Are you being put? However, as maybe the simple model, you'll be are in wrong doses, for example, axis. And this one. This is the estimating the simple model. And here is the partial point problem. We can see. The single model is a lot of work aligned with the with a partial going out. So we need to refine the single model to make its align with the partial . of.

And after the refinement we can see, the simple model has the value is a partial phone call. And now we can use diffusion according to here, although to generate the complete . human body point out, recognition with the sparse partial . out. And as far as simple . out, if you are stage a we can get a course we call of the human body. And because the company is very important of how to define a course.

For example, this is schools, this is a school smile of the body. The generally the . out of the, if you will, usually, for example, at least it's very important. It's very good to be aligned with the wrong truth service. Either we want to achieve a certainly a point is 4 points on the it is much because of the natural of the detail model is usually cannot be more possibilities.

So we need another refinement network to define a course. The first one come to a smooth surface. And you can see the because of the gpu size, we cannot use a very high resolution or protocol will be instructs the human body. So we attached the debt and estimated here and concatenate. It is the generating the point out. And then we can use a 3 . 3 pointer, 2 weeks back to the dash of the human law. When compared with online floor or the icon we can see, the sap can be construct the part, the head of the girl. And we can see the loose closing can also be constructed by our

说话人2 31:57  
method.

说话人1 31:58  
And that's i'm going to show you some code of the occupancy network, the occupancy. The occupancy network is based on the iphone and icon method. So there are 3 minutes in training a new network, I thought. We use package for some other furthermore, and has a goal or mssok there are three elements in training. You are right from the first one is the data order to prepare the date for the process. The second one is the apple architecture. And we use the metaphor to produce the optimus value, the asset value, or such as the hap and we can use the effort to graphically generate to the . of account.

The last one is the foundation objective. This one is used to supervise the training process of the platform. And when we preparing the training data, we can send for 3 . + in a long boss. And the strategy of the title and icon is several more points, and there is a surface and combine them as the integrity for the chinese. Here is a big order. The big order also have some. How to do is the first one is a neat function. The second one is the next function to return the number of the samples. And the fi is a function to get the data label. Here there is a code of the occupancy network. We can see the device it needs. But here we initialize the solution is that is to define the axis of the ai and the category of the data and some other use of the data and the lens. The last option is used to cars get older. I thought how much model of how much samples I have.

And during the operating the training process, we use again, either function to get the trying to get an actor each chinese relation and there is a network architecture. It also has three. I have it. It has two articles. The first in each function is defined the network architecture. And the forward function is used to work input into the network and achieve the prediction. There is aa code of the occupancy network. It defines the network. Here. We can see it has a indian area, and some reset arrests have gone. Finally, it use aa project to gets the final output of the network. And here is the in australia occupancy effort, first computer, the value of the input, the input can be a mash, a partial going out or partial english, and the occupancy. That is the input points through the variable, pulling back to the rest and loss.

And in this west side blocks, we can use the conditional professional to condition the network on the inventing see and finally from china to the output, on the network to long financial use, a single connected there and find the single function to look at the occupancy, the occupancy value.

And finally, we can learn occupants pure of the cd of that. And the super media cost is, it's like this. We can directly compute the error loss between the predicts options value and around tuesday.

And this is over. Do you have any questions in

说话人2 37:47  
chapter

说话人1 37:50  
for sure? 3 hours. You are a ppt. I know the child

说话人2 38:50  
going to go to

说话人1 38:51  
the same. True. We have to do it. Nana. No. I don't know something. You can show it. There are kind of a you create a hardcore anyway, you want to do what is you? I got one of the same. I absolutely. How do you see what I

说话人2 40:35  
mean?

说话人1 40:36  
Time to pass only people? That's all right. So you have to actually higher. I you. I'm gonna show you. Are the human body.

And so we if you are doing kind of some of these play with to driving a you are. How are you? A absolutely. It is.

说话人2 43:07  
For someone

说话人1 43:12  
sure you are right. Nana. Higher. You don't want I need an hour there for our ways of work. So on, he went on towards for the first years. So I have a member get over there.

说话人2 43:48  
She is

说话人1 43:49  
different. We so which we are done. If you I a if I have we have the people into the right and we have how do you know? Do you have any more

说话人2 45:12  
about shanghai to

说话人1 45:17  
the

说话人2 45:18  
producer shanghai.

说话人1 45:21  
I I we can. I a tau tau tau tattoo.

I in my case, you as I that is

t I don't know. So you can if you divorce, if you

说话人2 47:37  
you run away.

说话人1 47:44  
Area. Ok. I awesome. So you can understand. All right. Good.

说话人2 48:20  
Is

说话人1 48:25  
it has, if I ever changed it that

说话人2 48:29  
I was talking about, I don't have to do it. Nothing.

说话人1 48:38  
N I think it is also ahead of that.

Take a look at

说话人2 49:10  
what are you doing in china jail

说话人1 49:14  
country. This is I think that it is killing children in china n

说话人2 49:30  
however, we could give us their chair.

For water.

说话人1 49:57  
We're not insanity.

说话人2 50:03  
So how do you

说话人1 50:47  
are you

说话人2 50:49  
working

说话人1 50:50  
on? What are you doing?

Options are so human problem and then iPhone dong I

说话人2 51:21  
am doing more I choose no

说话人1 51:32  
the donor. It is. You can. Nana? No. We are now. How could you do it?