# Test Summary of Part of the Lecture

2022-01-30

Hope you guys can see this. Alright awesome I don't see the zoom window right now, so if any of you has any questions just unmute and ask the question alright. Alright, so like I said i'm Sonia and we're gonna talk about image understandings slash computer vision in the Semester. This is my email address email address if any of you need to you know, have any questions and anytime at least through January, everything is going to be online, including the office hours. You know I reserve, this time, just after class for, for you know if you guys want to stay longer on this zoom call and ask questions, otherwise just send me an email, and you can schedule something off of this hours, and we have four wonderful days. For for this class Wednesday in June lately and. they're all Grad students, so there are students that decided to do research and Grad school, so if any of you have similar plans you know, maybe you can you can talk to them about their experiences on. Wednesday and join in particular actually want to become professors of their their graduate so for this course we're going to give them an option of actually. Teaching one of the core and the classes lectures and you guys can give them some feedback, so they can learn how to do that properly and here are the email addresses. If you need to contact them. These are all the information that you guys already have class website it's actually encore quiz the link, but you know, just in case here is the link to the website or material like lectures so pdfs tutorial. Material everything is going to be put on this webpage so just keep that in opening one window and we're going to use Corporal corpus for all the announcements and any discussions of questions that you guys may have. There is no real textbook for computer vision, no one has written really a book that would sustain the test of time just because this is a field that's. Really changing rapidly, especially, especially now with you know deep learning kind of taking over there are some traditional books that are really nice. I typically you know really recommend this regularly ski kind of really core algorithms and applications mainly focus around 3D geometry. it's it's free online so it's kind of like your your your Bible of core algorithms but mind that there's really new stuff happening as we speak in this area. And you know if we're going to talk about any particular paper is or any cool code, you know we're going to pose it online. If you find some really nice links and the that you want all your class classmates to know about just pose them on quirkiness you know, this is a very empirical field and it's developing fast so let's share any nice materials you guys may found. um, what do you need to know to kind of succeed in this class right, so I really recommend you know this for. four courses, as your prerequisite so linear algebra is is a really a must, so if you don't know linear algebra this course is going to be really difficult for you, so I don't recommend you take it at this time, or you know just work really hard to catch up. All the other things are you know, very important that may be a little bit less critical critical than the Nigerian i'm programming like I said computer vision is a field where you actually do Program. In different algorithms to see how they're working on images and videos and so on, so me no, no, I had a program is important um I don't necessarily require you to know any of the specific languages, I hope you guys can see my mouse over here. But you know Python is the kind of the the programming language of choice, these days, because he had a lot of support for deep learning. And you just very simple to use so most of the stuff in this course is going to kind of assume that you know Python and examples are going to be given in Python. I personally when I did my Grad school I use matlab so it's also really nice language and has a lot of support for image based processing. So if some of you know my level one or non Muslims that are separate I would say c++ is really for those are more hardcore programmers. But we're not going to require any of this So if you hand in your assignment written any of these languages we're gonna take it all right. um you know if you know machine learning or neural nets in particular that supplies we're not going to assume you know that. This is not a requirement but we're going to actually devote at least one of the lectures to kind of have you up to speed, at least the basics in your works actually judy's gonna talk about that later in this course. was to require for this class in terms of how we're going to agree with you and we're going to have for assignments, and a project, so there is not going to be an exam or even like an oral real or oral exam we're going to be, you know doing this work as we go through through class. What our assignments So these are you know it's gonna be a mix of short theoretical questions, but mainly programming exercises are going to allow you understand the course Okay, you will be given roughly this assignments, are going to be given roughly every two weeks with maybe like. February, been exceptions, because there's reading week and we're going to start, I think the first assignment is going to be in about two weeks, maybe a little bit later. You will need to hand in a solution. So you're going to have a one week to complete your assignment Okay, and you need to work on assignment alone. We have had the algorithm before and it has not ended well, so please work on your assignment alone, this is your individual work that we're going to be grading also don't look for solutions online all right, that can also result in parentheses very easy to find if people actually copy. and towards the end of the course we're going to post, the project proposal or see in around March. And you know you can take our project as a bigger assignment where you're going to be able to put all the knowledge that you acquire in this course into know, maybe programming something a little bit bigger and see maybe some more interesting results in action. I want to propose a list of projects you're very, very welcome to come up with your own projects in fact i've had students come up to me in the second lecture with some ideas and you know, by the end of the course I had some really nice. projects going on, including some computer vision Apps on the phones and stuff like that which was really great to see. For the project you're going to need to hand in a report and there's going to be an oral presentation and during our presentation i'm also going to check up a little bit on your knowledge for from class. The difference between the as my my assignments and project is that you can work on projects in pairs so you guys can find. You know, someone that you want to work with and potentially work in pairs we won't allow more than two people to work on the project, but you can also work on it individually it's up to you. Great break down so each assignment is going to be 15% or so altogether altogether Center 60% of the full great. And project and oral exam are 40% splitting at around 30% for your project meaning report and then, how will you defend what you have done so on all the presentation. And then 10% of that final presentation is going to be on the oral exam we're gonna check up on your knowledge um. let's see the final presentations of projects are not going to happen in class we're going to do that. Maybe two weeks after the class is over and you're going to present to me, and maybe a ta is going to be in the call um yeah so I already talked about what you're going to be needed and. yeah that's reflected in terms of a timeline this is kind of a rough schedule when we're going to be releasing the assignments. And kind of roughly when they're going to be due why I say roughly is because you know, given how much questions you guys have or any kind of unexpected. Development you know, this might be pushed by a few days backward I like push it forward, but maybe backwards, just in case we were running behind in terms of the material Okay, but this is roughly the dates we're gonna follow. yeah like I said programming language choose your your language of choice, I put here a bullet point saying choose wisely Why did I put that. Because assignments, are going to build on top of each other so when you're going to be creating an assignment, you know, for you might want to reuse your code from assignment to. So it's actually important to write all your code all your assignments in one programming language don't use something in matlab or something as important, because you're gonna be you're gonna have a hard time reusing that code later on. Okay, so once you devote one language, I would just recommend to stick with otherwise you're going to have additional work. i'm that lane right lateness show we're going to be, you know relatively nice here so everyone is going to be given three free late days. Right, meaning that there's going to be a deadline on some day is going to be midnight and as soon as you're like one minute later 24 hours late that's going to be one day penalty Okay, but you're given three. Free ladies, meaning that you know if you are not feeling well or maybe you have other courses you're dealing with you know, I understand that I was a student myself so. You know, three free days seems fair anything beyond that means that that assignment is not going to be it's going to be as euro we're not gonna take it. Okay, now we live in hard times during the pandemic, so I understand that people may may get sick or whatever, so if that happens, let me know and we can we can discuss on on maybe extending this. Okay i'm not gonna go through the syllabus but roughly, this is what we're going to talk about these are kind of the key algorithms on that I think everyone should know for computer vision, since we don't assume that you guys know machine learning, we cannot really talk about very deep. novel approaches, but hopefully by knowing all this is going to prepare really well for more modern computer vision. Okay, any questions about just expectations or how these classes are going to proceed before we go into material. If so, just unmute and ask a question I don't see the zoom windows I can't read questions. yeah I can. yeah it's gonna happen, but maybe not all of you needed, so the tutorials gonna cover Python so basics of Python especially kind of the basic operations, you need to know for working with images. Higher lower than me image high visualize it and so on, plus maybe some very basic things about peyton So if you know that no need to come today, if you want to kind of brush on to that, then yesterday's going to be a short tutorial on on that. Any other questions. Otherwise, just you know send me an email or post it on coworkers we are going to be monitoring that alright so maybe we can get started and we're going to spend the first hour i'm just going to try to motivate computer vision, why is it such an cool fuel. That, I think, right now, especially right now it's super super important super useful. Okay, so we're gonna try to address a couple of different questions you know what is computer vision, how do we even define it. Why, you know, is really cool to actually study computer vision of what you know interesting things you can do with it and actually, why is it not still a hard problem, why is it still a hard problem was still you know at least to some extent an unsolved problem okay. So let's try the first talk about what what is computer vision in the first place. Okay, so the kind of the basic definition of what we're trying to accomplish here is, you know, we want to develop automatic algorithms right that would see. What does that mean right we're going to assume that you have some robot it is equipped with some sensors or maybe there is not even in the robot maybe just some sensors and you're sick. Your algorithms sitting in capturing data that comes from that sensor right that sounds are being cameras, or you know connect or you know lighter sensors right and your algorithms going to try to make. A trailer went right to see right right understand what the world is like from those images or videos from that capture data you're receiving okay now you know we're going to try to define a little bit better what it means to see. Right so imagine that, if we want to have a robot that's going to eventually go around in our House and clean and maybe you know do other things like. You know, serve your coffee or coke or something like that right So what do you need to do to actually accomplish that. right and it kind of most of it starts with computer vision Missouri service with perceiving what is actually in the scene. Right imagine just closing your eyes it's very hard to cook a meal if if you if you have closed eyes, you can go around you can touch different things, you can use your tactical. Information but it's it's it's much harder than if you actually have vision Okay, so the goal is going to be, we have cameras on this guy over here is going to move around, is going to capture. capture images and from that images will try to understand what is in the scene and try to use that information. To plan it further move right in order to move to the kitchen to find a cattle and then do something in order to to make your meal Okay, so the first and very important part is actually perceiving what is in the scene okay. So what exactly does it mean to see right imagine that they have to have taken a picture like this, like what does it mean to see what what do we even need to accomplish or here. Okay, so the first kind of definition that was given to us by by this person called mark, which is probably one of the most famous computer vision scientists. and his definition was to know what is where by looking okay What do I mean by that we want to know what every little pixel every little element of this image. Is in terms of semantics, we want to know that this is a monitor this is a chair, this is a table a floor and so once once then every single thing that is in this picture. Right now, the question is, is that enough is that enough for a robot to start let's say making your meal or you know typing your homework. Alright, and the answer is it's not enough Actually, this is a super hard problem it's a super hard problem but it's not enough right. We also need to understand where these things are in in 3D in the world was because eventually I need to grasp I need to extend my arm and pick up my cup with my tea, so I need to know exactly where it is exactly what kind of shape. Here it does exactly what kind of shape it has right because i'm going to place my hand correctly on that shape. And I want to, maybe even understand what kind of material properties right is a squishy or is it hard, is it plastic because i'm going to interact with this object in different ways if it has different materials. You know, so we also need to infer this kind of properties from from images or maybe a sequence of images okay Likewise, I need to understand that the SOFA is actually flat because maybe you want to sit on it right and again this is all very, very hard problem. On other things, that is very important right is actually understanding what kind of actions are taking place. Right, and let me give you an example, imagine that they want to create a self driving car, what does it mean right we're going to have a. Vehicle that has some sensors and cameras, maybe later some top anyone still autonomously navigate around the city and you know, ideally, not crash into anything and just be like a regular participant of traffic right and it's up to you guys to create those algorithms. Right algorithms that are going to be able to to drive that car, the same as you would, as a human. Right and here we're going underwater me our attention, mainly to perception right there's, of course. Also planning, you know how i'm going to navigate to a particular point given to understand where things are and also control how i'm actually going to actually we're going to you know press gas and rotate the wheel. But we're not going to talk about that in class maybe maybe just got the perception right now in terms of perception, what do I actually need to understand on that car. Right, yes, we need to know exactly which objects, are there in the scene exactly where they are whether something is two or five meters away from me for me. Maybe it's less important about exact material properties, but important their shape right, for example, is not just important to know that there is a car what What if it has a trunk open what it has a door open, I will understand that racer can navigate around it. and actions are important right, this is where I started talking about it, because if I actually see that there is a person on the sidewalk that is intending to cross the street is walking towards the. End of the sidewalk and tending to cross I need to react differently right my car needs to plan differently, because it's likely gonna go on the street. Right so it's actually that kind of actions it's important because, then I can compute the intent or those participants and then I can react appropriately. Right, the so it's actually a very, very complex system. can be the original system that we need to build over here okay now there's, of course, other definitions, what does it even mean to have a full understanding of the image. And some people pitch that, as you can answer any question about image right, what does this mean if you have images, you know people asked. Humans to just assess various questions about his images questions that a human would easily understand and questions that are important to you know, maybe create do certain tasks like. You know cook cook a meal or something like that. Right and what are the different questions that people ask right I talk about shape they talk about where is a particular object and talk about size which which object is the largest or something like that. They talk about colors are what your subject is dress is red talk about counting humans are particularly counting how many drawers are there, how many onions, you need to create this particular meal, how many doors are open. And then there is also more high level questions you know, can you make a pizza in this room, what does this mean right, we need to have at least an hour one or something like that, so that actually require some reasoning. Okay, so imagine if you need to design an algorithm is going to answer these questions right, you can imagine that this is actually really hard. Alright, so maybe we can talk about you know why should we even start the computer vision, what are the really cool things we can do with it once we we master it. Right, so one, one thing is, you know, maybe you want to have your robot do stuff for you like this is what I mentioned already before tried to play this video I. Hope you see this video. shape. Right even really old movies from the 80s already imagining. A little bit about what these algorithms should be doing Polish be recognizing objects. Okay, so when I was a student, you know this kind of Ai robot and it's kind of always inspired me to actually study computer vision, I thought it would be really cool. i'm driving right autonomous driving cannot work cannot work if you do not have really amazing computer vision system on top okay so here i'm going to show you mobile I everybody's familiar with this is. You know one company was created this Israeli Professor. And was probably one of the first kind of systems determines driving system for driving assistance. So it basically has some algorithms on on on the scarlet detecting lanes, so the warning for the for the lane it's also detecting pedestrian so you know it can do automatic breaking. in Canada let's say similar things like that, but it's not able to navigate autonomously, let me show you this real quick. And this is an old video sensors mobilized has been leading an ambitious and transformative effort to develop the most effective accident prevention system available using artificial vision. Through significant research and development, moving a sense of the kind of solution vision passively when we just saw the word winning technology. This system is contracting cost effective bundling multiple applications into one single package. Much like the human eye the mobilized solution is capable of performing driver seen interpretation detecting and classic now that card is driven by human goals what you're saying here is just configuration options traffic signs and more. This and some of these algorithms are implemented on most of the cars teachers. are very successful company was recently acquired by Intel revolutionized alright, I wanted to show you actually. something nice let's see if I can do it was looking at this cruise be there are you able to see the YouTube. All right, I was just saying this was just recently released, this is November so crazy another autonomous driving company very large company. And this is their system in action, actually, this is actually driven by the car so when you see the camera moving that's a car basically moving and that car is true and automatically. so high, the entire system computer vision prediction planning for what you're seeing the top, is what the car is actually understanding from the environment. So the texting other crazy little people here walking across the street. You also see trajectories here like you see this Green lights. Right, so this is this is your car over here, so this is showing you where the car is planning to go there, so here it's properly understanding or something in front of it, so no its trajectory trajectory is becoming smaller so it's waiting. And what you see around here is also the map, so it has the map of the environment, probably created offline before I know he's just registering to that map it looks too good otherwise. But you know you it's pretty impressive how stable this predictions are over here. And how well this car is driving, so this is basically 26 minutes of or driving in this pretty hard you know challenging environments, you have a lot of course it's kind of like maybe not Manhattan which is super urban with a lot of traffic, but it seems pretty hard. All right, so there's a couple of southern companies way more cruise zoox video is also creating their solution, there is a startup that was just recently created. Almost a year ago here in Toronto, the Professor or color the software as a professor University of Toronto it's called lobby. So if you guys are interested in self driving should check that out all right i'll stop here and go back over here okay. And there's a lot of competition in this space, you know it's a trillion dollar market whoever solves computer vision for driving and puts cars on the road. To control, you know tracks to control cars that's trillions of dollars right, so you have so many competitive companies right now trying to get to that point. All right, so it's a good space to work on very good space to work on. Okay, other other applications of computer vision right, so you have an image or you've been a tourist you went to Prague, or something. and take a picture and suddenly, you will notice there is this crane over here right so it's a bit annoying to send a picture to your friends. But you can maybe use, you know computer vision for his attack that or maybe when you just market and remove it so maybe make it look like it was a normal day in the city. You know you're going to say that you want to have a different sky, so you can actually mark the sky over here and replace it with you know, a different kind of sky right so image manipulation is one really nice application computer vision as well. um I don't know whether i've seen this, but again it's it's a really amazing technology by I mean you from nvidia. and basically what it allows you to do is to sketch so over here on the top left. You sketch what object, you want to have in where we're in the image, so in this case, one would draw this should be a mountain. They should over here should be a sky and maybe this is a lake so it's up to you you're a painter you're the Creator. Okay, and once your pain, this is going to convert into a photograph automatically okay so it's called image synthesis or conditional image sentences, because the conditions on the semantic user created. notation so let me play this is conceit. This is an old version so go again to was just recently released an idea have time to replace the the video, but basically here on the left the user draw this. On the right you see one click of a button you get out a photo. So there is no more need for you to guys actually trevor's travel somewhere to take this kind of beautiful for taken droid and turn them into components and it's just quite revolutionary technology. There you go, this is a user drying if you wanted to come on the right side. You have an image shapes of the objects that's been created and the neural network and then fill in all the details. gargan to is even more amazing it can go up to 4k if we had a water feature the network is able to add reflections, not because we call it that, because yeah like. Brian you saying you know you do, you understand those reflections the shadows, so you don't know what has learned something about how these images are formed which is pretty great architects designers people making virtual worlds, to train robots and self driving cars. The input to this model is something we call a segmentation map and you only even be an amazing painter right even I could draw this and i'm not particular talent. And this gives you a really cool images out shadows and the colors based on things that it's learned that from a large database of real world images. All right, you guys can watch you some nights called galligan go again and check o'callaghan to it's actually in line, you can play with it, if you want. Alright um other type of work, you can also change style of images imagine, I have captured this with my camera right and I want to make it look artistic. I want to make it look like one gold right Okay, so what I can do, I can give my neural network here in just a picture of you know, some some picture of an artist that I like like here would be painting of one go. And now can stylized this particular photograph to look more like that Van Gogh painting Okay, or some other artists, you know i'm not an expert right, but you can have other artists here the paintings and then you can get you know kind of stylized images out she's put the code. And you can also do that for videos right imagine to when I actually. create a full video in one style, let me play this one. We can skip the model explanation. Here is the original image here is a picture i'm giving it and then this is a stylized video. And this is a few years ago I didn't put a better example, so the it's a little tiny bit. On flickering if you look at details. really works pretty well and last year actually someone implemented this algorithm so they were showing me some video and this kind of style as results, and it was really nice to see okay someone some of you went to the. Alright, so. And i'm sure many of you have watched twilight and you know Kristen Stewart so actually they follow up on this work and i've taken this. neural style and shot a movie and completely style is, and will read this Ai technology to look like more scary and so actually this technology is being used by you know really users real artists already she's awesome. Okay, other applications, you can also do really nice things from just a single image imagine, I have taken this particular photo I can do things like this. Right, this is an active area of research, this was many years ago, it was very manual at the time. A lot of people are working now on making this a little bit more streamlined and more automatic in fact Wednesday in June, are both working in his domain, so if you have any questions you can ask them. i'm sure you guys worse and worse watch sports, you know there's Olympics coming up. All these things you know, like when you, you see, who won whatever with this overlay of text overlay of some size on the image. That is very simple computer vision that you can use over here okay and we're going to do that, for one of the assignments in class you guys are going to be pasting some texts like this onto images. much harder is, if you guys watch, you know American football where where there's also our tracking all the players and understanding how much they run how many passes they do that's much harder but that's also done pretty well these days it's a good technology. Also for special effects in movies okay let's match these particular there's a lot of computer vision nothing it for that to. hear is like the making of game of thrones. This is what you see right, this was the serious. But you can see that the quickie painting castles. With something that's easy to capture something is with you guys seen the serious. And this is again a combination of computer vision and graphics. what's computer vision has to do with here you have original footage you need to infer lighting this lighting music, you can simulate adding new object inside this capture videos. need to do some sort of segmentation. area if I need to insert my object somewhere, I need to know there's something wherever which pixels belong to a particular object. So almost nothing you see actually this series is real, a lot of it is kind of compositing synthesize. And a lot of that has to do with algorithms from computer vision. Alright awesome. And you know you can also do 3D reconstruction, maybe a skip this one. And this is another startup from Toronto by a professor at university of Toronto actually there was recently acquired for but with by which company, maybe we or something, and they were doing this really cool application on the phone. And what they've done is basically annotated. Much 200,000 images with detailed hair every single hair. Was entertaining every single here, not just the full hair. Was annotated and they train and you're on the deck here and then. allow users to modify. Their hair color maybe when. You can see some our artifacts actually use it loosely here in the middle, this is the founder, and this is a professor of status guy. Right, as you can see this as you can imagine, this is actually a pretty popular technology and useful. Okay now these guys spend six years annotating each individual hair all right, maybe the future of this is making it a lot more automatic maybe we don't need so much Labor leader is to train successful machine learning algorithms to do exactly that that could be your syrup. There is also a lot of kind of work on faces rather know whether you guys, who played with face APP it allows you to kind of modify modify your face on here is icon and are the copier you can kind of turn him into a woman or a young boy or an old older version of himself. Here are some more examples. here's my example put in trump kind of funny you can also upload your photograph okay. What you see here. it's not real none of these people exist, none of these people exist, this is all synthesized it's a neural network that you know you give it some noise you give it some random draw rector of numbers. And it's going to produce a face and it's going to look like it's a real face. Okay, this is this technology is called generative adversarial networks gans or maybe now there are also other types of generative models. In particular, this guy this this results here by style again you guys can actually download it and play with it okay. And it's super cool there's so many applications, you can do, once you actually can synthesize. Data here that you were only showing faces, of course, it also works and other types of objects for cars bedrooms you know it's kind of up to you on what kind of imagery you train it on okay. But it's able to produce really, really realistic output as point and it allows you to edit if you look at actually our own work on edit again allows you to edit images. With language or we just modifying you know I can draw different iowa here is going to produce a nice image so it's a really powerful technology. and other stuff that is super cool to do, maybe you want to capture images okay when take an image and actually write in language, what do you see in the image. Okay, why is this super important well it's important for many things, one of them is you know, imagine you want to build a system for blind people they don't see what's in front of them. Right, so you maybe want to have a camera and algorithm is going to narrate you know, be careful, this is, you know, this is what on the road so that's one use case the other use cases retrieval right like what Google is doing. But maybe even better right maybe you want to write a sentence and you want to retrieve relevant image right. and stuff like that okay. This is a, this is a reverse alright so here the user is going to write a sentence, this is a text that they're writing. And this is synthesize with a neural network, none of it is real, none of this is real all right, this is this is generated based on the language description. Okay, this created a huge boom last year I was done by openai end and the method it's called Dolly now they already have better versions of that. But it's quite incredible what they've done is they, they basically crawl the web, for I don't know hundred 200 million images and text right, because when you guys posting what pages are posting. stuff on the web, you typically write something about image right okay i've been in Paris, look at look at this cool whatever. Right, so they actually crawl that tax and that images and they train this this this network, why it takes a sentence and synthesize images. And once you go to that scale once you have available hundred or maybe even more data everything kind of starts to work like it's just quite incredible replica. The the, the number of probably never have seen you know, an arm chair chair, in the shape of avocado, but it has seen on cherries and consumed avocado and amazing thing is because it comes to generalize it can actually create an armchair in the shape of avocado in economic sense. All right, anything like this really complicated images which are called composited rubber baby daikon radish even know what that means is walking a dog it kind of looks good.