Cognition and its Computation
Prof. Rajlakshmi Guha
Prof. Sharba Bandyopadhyay
Biotechnology and Bioengineering
Indian Institute of Technology, Kharagpur

Lecture - 33
Object Recognition

Welcome. We have been studying attention. So, we have discussed how attention plays a role in shaping our perception of a different items in our sensory world and that translates to what we call actually Object Recognition. Attention plays a huge role in object recognition; but object recognition in itself is something that we need to learn in order to see how attention plays a role in the overall area of object recognition.

Now, object recognition is fundamental to our survival; because, we need to identify or discriminate between objects in our visual space or auditory space or other sensory spaces that are of danger to us or are required for our survival. And, so accordingly, we have evolved in a way such that we can seamlessly recognize objects in the different sensory worlds almost immediately.

I mean, the idea of object recognition especially in the visual system is thought to be that the ability to discriminate or identify objects from other objects and maybe also to categorize them into coarse categories within less than 200 milliseconds. So, we do this in a fraction of a second and that is really important feat of the brain given the 10s of 1000s of possibilities that are there we have to stick labels on the different objects that we visualize, let us say a face from many many many faces that you know, a chair from many many objects that you know, a car of a particular kind and so on and so forth.

Similarly, in the animal world, the object recognition also plays an equally important role; especially in the visual system because it is tied to the survival of the species. And, that can be seen by the fact that more than 50 percent or almost 50 percent, around 50 percent of the cortex of non-human primates is dedicated to the visual system all ultimately trying to solve one particular problem and that is recognizing objects in a clutter or amidst of many many different objects.

And, it is key for us as I have been saying because, a lot of things depend on whether we recognize something or not. So, object recognition when we say we usually mean recognizing objects in the visual field; but it does not limit our we need not limit ourself to the visual world only. Although, a the clear distinction of objects in the auditory field are not entirely established. However, there is or there are ideas of auditory objects.

Similarly, there are definitely ideas of gustatory objects and olfactory objects and similarly somatosensory objects. So, the principles that we will talk about or the computation performed in the brain or behind or that leads to object recognition can be generalizable to the other systems although we although the definition of objects in every case is not totally clear, but, likely if the if people would agree on the definition, it would be the similar principles that would be applicable in the other sensory systems as well.

So, to begin with if we think of the visual system so, we will mainly focus on the visual system here. And, so, if you think of the visual system, you have learnt a how the visual field makes an image on the retina. And, in there is an object that you want to identify or want to discriminate from the other objects in the scene.

So, a single object may be present in the visual scene in many different positions and yet we identify the object as the same object and this is what we call as position invariance. Similarly, the objects can be of very different size or the amount of angle that it obtains at the eye based on the distance of the object from us. And, so, we still identify the object as the same object which is size invariance.

Similarly, we have orientation invariance; that is an object may be rotated a face rotated upside down would still you would still recognize the person if you know the person if in the right way or oriented properly. And, so, we are insensitive to such identity preserving transformations of objects in the visual field.

And, there are many other kinds of transformations like the emotional aspects that are observable in the face of a person does not let us stop us or does not stop us from identifying the person who is angry or who may be laughing at another point of time it is still the same person. And, so, this is also another kind of identity preserving transformation.

So, as we see at the crux or the core of recognition, is the idea of invariance; that is tolerance to identity preserving transformations of objects. So, no a car, no matter what the orientation we would identify it as a car. And, another important factor is invariance in terms of the context that is a car in a field or a car on a road or a car even in the sky an absurd thing would you would still ultimately identify it as a car.

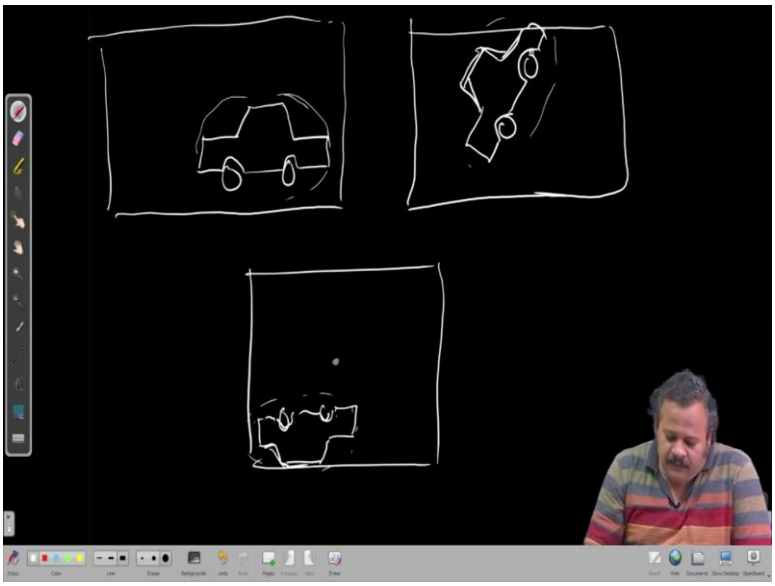
It you may be confused in the beginning, if you if a picture is portrayed that the car is in the sky, but you will ultimately recognize it as a car or we humans do recognize that as a car. And, animals trained to do so would also recognize that. So, even the context of the scene or the scenario in which we are trying to identify an object a particular object that also is can be varied, but we still identify particular objects in different kinds of contexts.

So, these invariances are at the core of the problem of object recognition. And, if you think about it if we have many different distractors in the scene that do not allow us to identify the object it makes it seemingly very very difficult; however, as we will see in the later lecture, that even in very cluttered environments this is the idea behind visual search that given familiar scenes we easily recognize objects.

And, similarly, if we think of examples if we were to say that a particular person's voice as an object, as an auditory object, which we know then if you think of hearing the persons voice in an environment where in a noisy environment you still recognize that person if you know that person well enough and that is you know the understand or know the voice a from previous occasions.

And, so, similarly if we think of a voice that is you hear it very softly or very loudly you still recognize the person who is behind the voice. And, no matter which position also the voice comes from we actually identify, the or recognize the voice of the person or identify the person whose voice that is. So, this is a similar kind of scenario even in the auditory system. So, now coming back again to the visual system so, all these invariant all these variable positions or orientations or sizes of objects, all of them have to be identified as the same object.

(Refer Slide Time: 10:27)



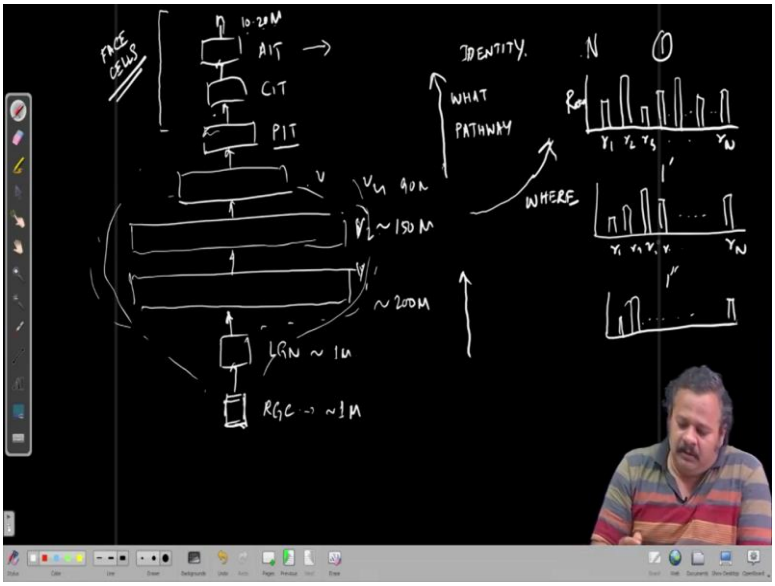
Now, let us consider identifying a car in the visual field. Let us say we have a car like this. I am drawing a cartoon of a car, that is present in the visual field at a particular location this is one object; this car could also be oriented in a different way in the same visual field like so. Assuming that it is, I am drawing the almost the same car or even in this kind of scenario where the car is let us say upside down, let us say this is the same car in 3 different position different orientations and so on.

In and all of these are forming very different images in the retina. So, the photoreceptors at this locations, that identify or that correlate to this particular location reserve or forms the image of the car. Similarly, in this case this particular location forms the image of the car. And, as you can see there are subtle differences in the size not that I withdrew that very carefully, but it happens to be that they are of different sizes. And, so, this is yet another representation in the retina.

So, from our understanding of the visual system, these are creating a responses along the visual pathway of what we know in the photoreceptors, in the retinal ganglion cells that convey to the lateral geniculate nucleus, in these 3 cases they are very different kinds of population activity in the lateral geniculate nucleus or in the photoreceptors or even in V 1 or even in V 2, V 4 from what we know so far. And, so, ultimately though it is we have to identify the object as a car and so; that means it is basically all the same things.

So, all these different kinds of retinal representations have to ultimately map to one particular group or category and should be different from representations or retinal representations or the representation of other objects along the hierarchy, it should be different from all of those. May be 10s of thousands of objects that we know, it should be different from all of them. So, this is the crux of the problem. So, one interesting idea or that it is actually something like that that needs to happen is what we call untangling of the visual responses.

(Refer Slide Time: 13:44)



So, if we consider that along the visual pathway. So, we have in the photoreceptors, there are actually let us say retinal ganglion cells that produce outputs they are about a 1 million in number, that project onto the lateral geniculate nucleus which is also approximately 1 million in number, that then projects and expands hugely into V 1.

And, this is approximately 200 million representation or rather 200 million units representing the 1million images or 1 million photoreceptors activity. And, this expansion in V 1 and then also into V 2 which is almost of equivalent size like around 150 million. And similarly then it get starts to get reduced from V 4. Then, it is about 90 million here which further reduces in the posterior inferotemporal cortex this is what we know as what we discussed as the what pathway in our visual system.

And remember, from the visual cortex the what and the where pathway segregate; that is the pathway that decides on the location of the object, the location specific information

and the what pathways for the identity specific information. And, it starts after the visual cortex in V 4, then the posterior inferotemporal cortex, then central inferotemporal cortex and then finally, anterior inferotemporal cortex which is thought to be the seat of object recognition.

And, so, lesions in the inferotemporal cortex actually show that they are I mean people are unable to identify objects properly and process object information properly. And, so, here ultimately it is representative a representative of 10 to 20 million outputs that go on to the prefrontal cortex. So, remember we started with 1 million units that got hugely expanded of 2 orders of magnitude higher in the primary visual cortex and secondary visual cortex in those areas.

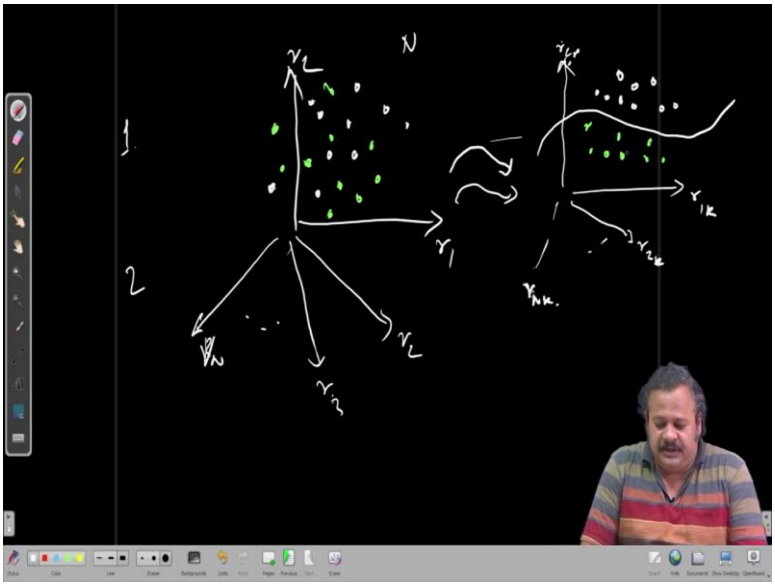
And then again, it got reduced to representing maybe 10 to 20 million units which or the some populations of those are likely to represent specific objects and they are learned as we go on. An important factor that we know about the inferotemporal cortex is that a very large class of cells represent faces or face cells which is a very important class of objects in our in human as well as non-human primates, a very important class of objects and it is important for our survival daily interactions, social communication and everything.

And, so, this important class of objects in A I T also shows that they can be very very specific types of faces that are encoded. And, so, the cliched term the grandmother cells in the in anterior inferotemporal cortex came about and so, there are cells there that are specifically responding to particular faces like that of maybe one's grandmother only. And, so, that is how specific they can be. So, the untangling of responses that we were talking about, that actually happens due to this expansion.

So, if you think about it ultimately, we have let us say, N for a particular objects object; let us say we have capital N neurons representing them in their firing rates. And, so, let us say object 1 provides rates that are like this is rate information; this is neuron 1, neuron 2, neuron 3 up to neuron N. And, this is the response to object 1 at a particular location, orientation and size and what have you, ok. So, this is how we know that about representation in the visual pathway; that different neurons are responding in a different ways and the entire population ultimately is going to represent the object.

And, let us say 1 prime is another representation of object 1, which provides another which has which is identified by r 1, r 2, r 3, r 4 with different firing rates and let us say 1 double prime, again the same object in some other orientation or some other transformation providing another set of rates and so on. So, all these representations have to be mapped into one particular group. And, for that we require to separate out the overlapping regions of the objects with of the of these of object 1 with the representation in the same neurons with another of another object.

(Refer Slide Time: 20:29)



And, so, if we think about the representation, if you think about the N dimensions in which let us say r 1, r 2; somehow let us say we are able to represent many dimensions r 3, r N. In this capital N dimensional space object 1, represents some set of points for the different kinds of orientations or different transformations.

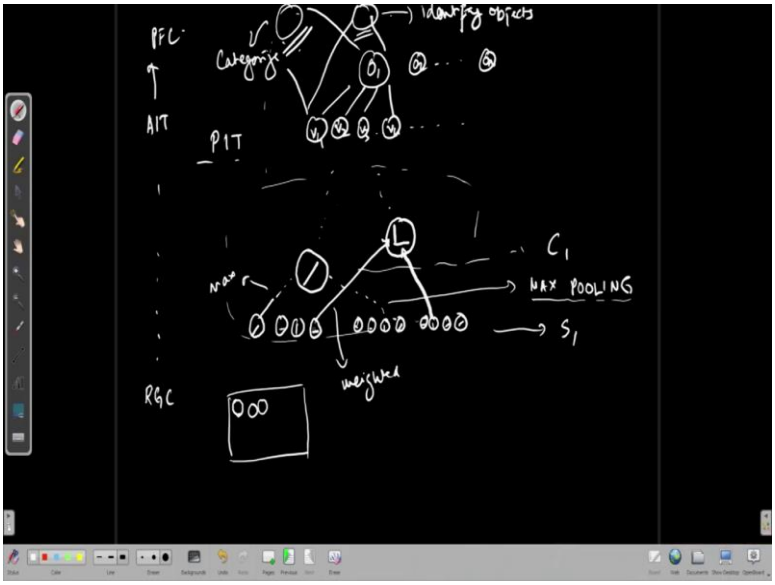
And, let us say object 2 in the same space ultimately is represented by another set of points in this. And, along the path what we actually need is that, ultimately in another set of neurons, in the final set this transformation is along the visual hierarchy we have another set of neurons let us say now r 1 at the stage k, r 2 at the stage k. Similarly, r 3 at the stage k s or and r N at the stage k.

The we require the white dots, a to be on one side and the green representations or those activities representing another objects to be separated by some manifold in this space. And, that will let us so, this transformation is what we call untangling of the response's

in the into another set of in into another space and that is essentially also the idea behind what we know as support vector machines and other many other machine learning tools.

However, the way the brain solves this problem is although the idea is similar, we will look into what we know as the standard model and discuss how we achieve this kind of representation this invariance. So, and able to discriminate the objects. So, the standard model that has been propounded by a Tomaso Poggio and others for a very long time in that case what we know is that, we have this is happening along the pathway from the retinal ganglion cells up to the AIT.

(Refer Slide Time: 23:25)



And then that goes on to the prefrontal cortex, as we know to the more executive regions of the brain. So, if we start with V 1, we know that the simple cells are there which are orientation selective. So, if we think of the image on the retina, if we have a particular object, then for as we know in the primary visual cortex for a particular retinotopic location in a hyper column, we have a whole set of orientation tuned units and for another position also we have a whole set of orientation tuning units so on.

And, for another position also we have such orientation tuned units. And, similarly, we have a whole set of color tuning, we have a whole set of depth based tuning and so on. So, what happens is that, if we are able to combine different positions with the same orientation let us say this orientation, with a max tuning curve, then we get an orientation invariant representation of that particular tuning.

So, this is what we call as max pooling. So, that is the strongest afferent drives the output neuron so; that means, no matter where the input is coming from in the retinal image, if whichever one is the strongest that is what is driving the output of the second layer of the complex cell C 1 and this is let us say the S 1. So, this provides us with position invariance. And, similarly, if we have a max pooling for different sizes, we will get size invariance and so on.

So, this kind of max pooling followed by integration of 2 different kinds of orientations which is so, these dot dotted lines are max pooling. And, this solid lines are weighted averaging. Then, this gives us an this gives us selectivity to different kind of integration. And, now, from these can be combined at different positions and that will give us position invariance of this particular feature shape.

And, so, this is how it is organized in the hierarchy to finally, produce what is known as the view tuned units, which is by the time it reaches the posterior inferotemporal cortex. So, this is the same objects in different views and so on. And, now combining all the different views we can get what we know as the object tuned units. And, these along with the objective units in the PFC in both ways can give us the ability to categorize objects and identify objects.

So, the biological mechanisms behind the max pooling there is a lot of evidence suggesting that this can happen. And, so, this particular level here is known as the view module. The view-based module and then higher up these weightings have to be learnt in order to be able to categorize units.

And, so, there have been experiments that have shown that we the PFC units can actually be tuned to different categories based on learning. This was done by a famous experiment from Carl Meyers group, where they train monkeys to identify cats versus dogs of 4 different species of cats and 4 different not species breeds of cats and breeds of dogs and they were morphed at 2 different degrees.

And, the animal clearly learnt to identify the majority cat versus majority dog or identify the majority cat as cat and the majority dog as dog in the morphed images. And, it was found that the PFC neurons were category selective that they responded highly to only the cat or only the dog. And, so, that is what we mean the categorizability from the PFC

and this information again. So, this is all feed forward. We have not considered the feedback part of it, which actually enhances all this.

And, the particular role the prefrontal cortex plays is that this category information and the identity information and attentional mechanisms actually feedback to enhance the representations here. And, that actually what we have discussed earlier that actually results in easier identification of objects or easier object recognition.

So, these go hand in hand in terms of what role attention or the prefrontal cortex has to play and the object task of object recognition. So, with this we will close our discussion on object recognition. And, later in the next lecture we will consider a visual search or visual search and pattern recognition that is basically, how we are able to rapidly identify objects by searching through a visual scene.

Thank you.