[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: Today we're going to wrap up talking about object recognition, get into texture recognition. So remember, object recognition is a thing that we can do. You can look around scenes and name objects, OK? It's mostly effortless for people, but a pretty challenging computational problem.

In humans and other primates, object recognition is something that we typically do when we foveate objects. So it's kind of dominated by central vision. And so, oftentimes, when you're looking at a scene, you'll be making a series of saccades around things in the scene, recognizing whatever it is that your eyes focus on.

So the problem is hard, for a couple different reasons. One is that there's a large number of things that you can recognize, OK? This is just a small subset.

And then, the second issue is that the same kind of thing can produce totally different images, like depending on the viewing conditions, right? The viewpoint from which you view an object, or the illumination conditions, or what the background is that the object is on, or whether it's occluded, or whether, like if you're talking about a biological organism, whether we're non-rigidly deformed-- I can do all kinds of weird things with my body, right? Totally changes the image, OK? So you get wildly different images, depending on the viewing conditions, OK?

So we talked last time about how there's lots of evidence that object recognition is largely mediated by the ventral visual stream. Remember, dorsal ventral streams. We talked about evidence from lesion studies, where you lesion the temporal lobe, and monkeys and humans end up with recognition difficulties.

We talked about agnosia. This is the inability to recognize objects that occasionally results from brain damage following strokes. And then, we talked about the ventral visual stream and evidence that object recognition is fast. And thus, the inference from that being that, in many cases, object recognition is mediated by largely feed-forward processing, OK?

So this is in the macaque visual system. We've got latencies of about 60 milliseconds at V1. 100 milliseconds at IT. In humans, it's a little bit bigger because the head's bigger.

And then, we talked about various pieces of evidence that recognition is a quick thing, in some cases, OK? And so this is the study where people looked at series-- sequence of images. And these are event-related potentials, in response to images. And you can see that they diverge for categories of images, in this case, animals and non-animals, at a relatively short latency of 150 milliseconds, which is on the order of the latency of the early response in IT.

OK, and we talked a little bit about the differences that you see between different stages of the ventral processing stream, which much of which we've talked about at earlier stages of the class and to varying extents. So in V1, you have relatively simple tuning, orientation selectivity, and then, in simple cells and complex cells you get some tolerance to position. Question there.

AUDIENCE: For the feed-forward processing in object recognition is that-- that works for like these broad categories? If you ask for more specificity, like rather than recognizing a face, getting the person, does that require feedback? Or does it just require a longer chain of feed-forward?

JOSH MCDERMOTT: So I think that, on the whole, there's evidence that, to a large extent, even more subtle distinctions can often be made with largely feed-forward processing. I mean, there's also lots of evidence now, especially over the last five to 10 years, that images that are hard to recognize, for whatever reason-- like maybe an object is occluded pretty substantially or that are otherwise kind of unusual, in some way, where the information is degraded-- that those might require recurrent processing, OK?

But I think, in some of the things that I'll get to in just a moment, that we talked a little bit about last time, where you look at whether you can decode object information from brain responses, the brain responses that are used to do the decoding are often kind of fairly early responses that you would think would be kind of mostly feed-forward, in origin, right? And so it's not just this distinction, in other words, right? I agree, this is a pretty coarse distinction between animals and non-animals. So there are certainly things that are more fine-grained that I think you can also account for with a fairly feed-forward mode of processing, but it's clearly not the whole story, all right? Yeah, good question.

OK, so you've got selectivity for fairly simple features in primary visual cortex. Think of that as early vision. You go up to V4, things get a little bit more complicated, maybe harder to describe.

And then in IT, which we spent most of our time talking about last time, you see that it's dominated by the central visual field. So most of the projections into IT are from the central part of the visual field. And how there's lots of these examples now of pretty complicated neural tuning, often for things that are kind of behaviorally meaningful, things like hands, or faces, and so forth.

We talked about how there's some functional organization within IT and how there's this other major organizational principle that you see throughout the visual system that is also very prominent in IT, which is that the receptive field sizes increase as you go from the beginning of the visual system up to the deeper stages, OK? So even within IT, you see a gradient of receptive field sizes. Talked about how individual neurons in inferotemporal cortex show some degree of invariance to position and size.

And then, we transitioned into trying to think about what these properties mean, in computational terms, right? So these are all like different sort of clues to the basis of recognition and inferotemporal cortex, OK? But how can we actually turn this into working models and a computational understanding?

And the key thing that we talked about here was this idea of thinking about images or stimuli, in general, as being represented by populations of neurons. So you can think of the images being represented in this high-dimensional space, where every axis of the space is the response of a neuron, OK? So every image gets represented as a point in that space, OK?

And if you're dealing with different images of the same thing, you have a set of points, right, that will form what we often call a manifold. And we proposed that what a good recognition system ought to do is generate representations of images that have this particular property, whereby all of the images that correspond to one type of thing would be in one part of the space and all the other images would be in another part of the space. And the way that you quantitatively assess that is by asking whether there is a plane or a hyperplane that can separate those different sets of points. And if there is such a plane, then you can use that to build a linear classifier that can take the brain response and tell you whether you're dealing with Joe or Sam, in this particular case, OK?

So we talked about how linear classifier is based on a projection operation onto a vector, which in this case, would be perpendicular to that hyperplane. And so the crux of the recognition problem-- the reason that recognition is hard-- is that in the input representation, the classes that you care about, one kind of thing versus another kind of thing, they're all tangled up, right? So the points that correspond to one person and the points that correspond to another person, they're all sort of intermixed, in a complicated way, in the pixel space, right? That's kind of why you need a brain.

And there's this proposal that we tried to evaluate that what a sensory system might do-- in this case, the ventral visual stream-- is perform a series of transformations on the input that would end up with a representational space that would have this property, OK, where all of the representations of one kind of thing are in one part of the space and the representations of another kind of thing are in a different part. And so then we talked about evidence that this is the case. And so this is the experiment that I was alluding to in my answer to the question, just now.

So we looked at experiments where responses from a big set of neurons and inferotemporal cortex were measured to a big set of images, OK? So here, we have all these images and all these different recording sites, which you can think of as neurons or groups of neurons, OK? Some neurons respond more to some images than others.

And the question is whether or not the representation that is implied by that data would be good, in the sense that we've discussed, in making object identity explicit-- which, again, is evaluated by asking whether the two different objects are linearly separable-- which, again, you evaluate by fitting a linear classifier, and then measuring classification performance.

AUDIENCE: Also to test like different orientations and 3D rotations of the objects.

JOSH MCDERMOTT: So in this particular experiment, I'm not sure if they looked at different views. This is showing you different sizes and positions. That's certainly been done. And I think would have a similar outcome. But I don't remember about this particular study, OK.

OK, and so the outcome of this experiment was that you can classify objects pretty well using, with a linear classifier and using the responses of neurons in IT. So this graph plots the number of recording sites versus the classification performance. And once you have a couple hundred recording sites, you do fairly well, OK?

And we saw that this was at least somewhat specific to IT. So if you do the same analysis in inferotemporal cortex and in V4, which we can think of as the area of the ventral stream that precedes inferotemporal cortex, classification performance is worse. Again, these are measured as a function of the number of recording sites. And also, the consistency of the classification judgments with humans is a lot lower.

And so the conclusion from this is that IT neurons mediate object recognition by making object identity explicit, where what it means for something to be explicit is that it is easily read out. And what it means for something to be easily read out is that you can build a linear classifier that will enable you to classify the bit of information you're interested. In this case, object identity, OK? OK, so that's where we left off, last time. Are there any questions about that before we move on?

And so I'll note that in this description of the basis of object recognition, we haven't really talked about the mechanistic details of what these transformations are, at each stage, right? We've talked about the outcome of those transformations, right, in terms of a property of the representation, OK. OK, and those transformations, they might be complicated and not that easy to describe, potentially.

OK, all right. So another kind of important aspect of our recognition machinery is that there's evidence for specialization of this recognition machinery for particular classes of objects that are maybe especially distinctive or especially important, OK? So in particular, neurons that are responsive to certain special classes of objects seem to be segregated in the brain.

And we talked about our colleague, Nancy Kanwisher, who's done a lot of work on this. And so one particular class, where there's especially good evidence for this, is faces. So there are lots of patches of cortex, in both monkeys and humans, within which the neurons respond a lot more to faces than to other kinds of images, OK?

And so Nancy Kanwisher discovered one of these, the fusiform face area. There's a few others in humans. And this is the result of an fMRI experiment in macaques, where they are comparing brain responses to faces to brain responses to other types of objects. And there's a bunch of these patches that kind of emerge, OK?

So and then, in this particular study by Doris Tsao and colleagues, they then use the fMRI results to direct electrodes, OK? So they put electrodes into these face patches and now measured the responses of individual neurons. So that each row here is a neuron that they recorded from in one of these patches to a big set of images, OK?

So there's 96 different images. And the images are in different categories-- faces bodies, fruits, gadgets-- I'm not sure exactly what goes into gadgets, but I guess gadgets-- hands, and I think that's scrambled, images, OK? And so what you're supposed to take away from this-- and then, the color here represents the firing rate of the neurons, right? And so the vast majority of the neurons in these face patches are responding a lot to all of the images of faces and much less to all the other kinds of images, OK?

So here's another bit of evidence for the same of thing. This is a plot of lots and lots of recording sites. So over 1,000 in macaque inferotemporal cortex. And each recording site is a dot here. So they have lots of electrode penetrations.

And then, they're color-coding the electrode sites, as a function of the extent to which they prefer faces versus non-face objects, OK? And so the point is just that you see these regions where there's a lot of red dots, right? There's one here, and there's one here, maybe one up there.

OK, and so these various lines of evidence, they address this kind of long-standing question in the realm of visual recognition, which is like whether faces are kind of special, in different ways. And there's four pieces of evidence that I think are all kind consistent with the notion that there is some degree of specialized machinery for recognizing faces.

The first is what we've already discussed, which is evidence for a brain region that responds a lot more to faces than to non-faces. This is in humans, the fusiform face area, mostly characterized with fMRI. So it's right there. Here's another view.

Another piece of evidence that there's something special about face recognition comes from inversion effects, OK? So it turns out that face recognition is much better when the faces are upright than when they are upside down, OK? And this is not necessarily true for other objects.

And the classic piece of evidence for this is something called the Thatcher illusion, OK? And so it's called the Thatcher illusion, because the original example of this was performed on the face of Margaret Thatcher-- because it was done by somebody in Britain, and Margaret Thatcher was the prime minister of the UK, at that time, OK?

And so the illusion here-- and so here, it's done on the face of somebody else-- this is just what I happen to have handy. The illusion here results from the fact that the face is distorted, OK? So there's a few sections of the image of the face that have actually just been locally flipped upside down, OK?

But when you look at it upside down, it just looks fine. It looks like a face, right? But what it actually is-- and this becomes very obvious when it's right side up-- is this, right? So you can tell that this is not how the face is supposed to look, right?

So that and that are exactly the same, just with the exception of being rotated 180 degrees, OK? And this is not the only piece of evidence. There's lots of other evidence. You can measure the ability to recognize or discriminate faces and compare that for upright versus inverted images. And there's a huge advantage for when the faces are upright, OK? Yeah.

AUDIENCE: Do you know if this-- how this affects the facial recognition systems, possibly?

JOSH MCDERMOTT: So the question is, does this affect facial recognition systems? So are you asking whether a computer vision, face recognition system would exhibit the same phenomenon?

AUDIENCE: Yeah, so does it fall for the illusion? Or does it say, like, this is messed up?

JOSH MCDERMOTT: Yeah, so my guess is-- I actually don't know whether that's true or not. And somebody has probably looked at this. My guess is that they probably would because we think that this comes from the training data, at some level, right?

Like the naive interpretation of these inversion effects is just that most of the time faces are upright. And so your face recognition system is trained up on all this data. I mean, again, we don't know whether this is over evolution or learned via development. But either way, most of the training data, essentially, for the recognition system is in one orientation.

And a lot of objects are not like that, right? So you see chairs from all kinds of different viewpoints. And are less sensitive to that. But yeah, my guess is that probably machine facial recognition systems will show something kind of similar.

Yeah, and this is part of a long-standing debate about the extent of viewpoint invariance of human recognition. So one of the really interesting things about object recognition is that it's pretty invariant, as we've been discussing. And one of the most impressive forms of invariance is invariance to viewpoint, right? So I can look at this cup from this direction or this under this direction. I'm going to be able to tell that it's a cup, but the image that is produced by the cup is like totally different in all these different cases, right?

And so there was like a long history of people thinking about the basis of that viewpoint and variance. And so some people have postulated that you actually infer 3D models of objects. And that's the basis of recognition. Others say that you have all these stored 2D views, but they're kind of linked together.

And there's been lots of experiments on new kinds of objects, like where you expose people to particular views of the objects and not others. And then show that they don't generalize perfectly. And that suggests that viewpoint invariance is not automatic. So I think there's various pieces of evidence that you're only as invariant as your training data kind of forces you to be, in a lot of cases. And faces may be kind of an extreme example of that, OK? But at any rate, like this is not something that is as pronounced for most other classes of objects. So one piece of evidence that face recognition is special.

Another piece of evidence is prosopagnosia. So there are these examples of people with brain damage that end up with selective impairments in recognizing faces. And that's known as prosopagnosia.

And so the most striking cases of this come from people who've had a stroke, which is the typical cause of brain damage in people. And so here's a quote from a case study of this person, Dr. P. It says "By and large, he recognized nobody, neither his family nor his colleagues, nor his pupils, nor himself. He recognized a portrait of Einstein because he picked up the characteristic hair and mustache. In the absence of obvious markers, he was utterly lost."

Here's another report-- this is of a person describing themselves. "At the club, I saw someone strange staring at me and asked the steward who it was. You'll laugh at me. I've been looking at myself in a mirror." All right, so people can't even recognize themselves, OK?

And nowadays, so these classic examples come from people that have pretty serious brain damage. There's now, I think, growing evidence that face recognition-- probably a lot of recognition abilities-- they lie on a continuum. And some people are actually really good at recognizing faces. Others are less good.

And at the extreme end of the less good end are people who might be considered to be prosopagnosia, but just congenitally prosopagnosia, OK? And so there's thought to be a genetic basis of this. It runs in families and so forth. So but prosopagnosia is classically a neuropsychological phenomenon.

And the inference of it from this, right, is like-- and this is the power of all of these examples of brain lesions-- is that if you can selectively disrupt a particular ability, with other abilities remaining intact, that's evidence for modularity in the brain. That there is some bit of the brain that is specifically especially involved in that function. And if you're unlucky enough to have that particular bit of the brain damaged, you end up with a deficit in that particular thing, in this case, face recognition. So we saw this with achromatopsia, and here it is with face recognition.

And then we've just seen these examples of face cells, OK? So a variety of pieces of evidence that there is some degree of specialization of the recognition machinery for faces. All right. Any questions about face recognition or this particular idea? Mm-hmm?

AUDIENCE: So the time series here-- or in the responses, it like recognizes it, and then stops recognizing it? I'm a little confused as to why it's a lot more spread out, like the bottom one. Like it seems like it's continuing to fire later.

JOSH MCDERMOTT: Oh, yeah. Yeah, I mean, so these are like responses for individual trials.

AUDIENCE: Oh.

JOSH MCDERMOTT: And I mean, neurons are stochastic devices, right? So--

AUDIENCE: OK.

JOSH MCDERMOTT: Yeah, I wouldn't over-interpret this. I mean, I think if you repeated the experiment, like the exact pattern of the spikes is going to be different every time, yeah. So the main thing to take away from this is like there's a lot of spikes up here, kind of less here, some here, less there. That's sort of the point.

OK, so this large-scale organization of selectivity for different kinds of categories that shows up in these face patches and also things like the region that's selected for bodies, other things, that is evident in what we call representational dissimilarity, OK? And so this is now a pretty well established and widely used method for looking at neural representations. And so it's worth walking through.

Just curious, how many people have encountered representational dissimilarity before? OK, a few. What, in Nancy's class? Yeah.

OK, so the essential idea here-- so let me tell you how you construct this, OK? So these are what are called representational dissimilarity matrices, OK? So the matrices here consist of sets of points, OK? And each point represents the dissimilarity of the brain responds to two images, OK? And so the dissimilarity similarity is, essentially, 1 minus the correlation between the brain responses to the two images. So what is the brain response? The brain response, in this case, is measured with fMRI from a whole bunch of different voxels, OK?

So this is an experiment where you present this large set of images. Looks like there's a couple hundred here, OK, to every image. You measure the fMRI response in this big chunk of visual cortex. So that's a whole bunch of voxels. So for each Voxel you get a response to each image, OK?

So you can think of the response to each of these images as like a big vector. So let's suppose there's 1,000 voxels in the region that you measure, OK? So you get 1,000 numbers for every image, OK, image? Which is the response to that image in every voxel, OK?

And so then what you can ask is for a pair of images, how similar, overall, is the brain response, OK? And so you can measure that as the correlation. And here, this is dissimilarity. So dissimilarity, so it's just 1 minus the correlation, OK?

And so blue, here, means that things are not at all dissimilar. So they're very similar. Red, here, means that they're very dissimilar, very different, OK?

And so, in this experiment, there were like 200 different images of objects and other things, OK? And so there's lots and lots of these pairs, OK? All right.

So what do you get from this, OK? And it has to be symmetric, right? Because the rows and the columns represent the same sets of things, OK? And also, the diagonal, thus, has to be 0 because you're just comparing the response to one thing to itself.

OK, so this is what you get. And this is a particular study that compared the representational dissimilarity in monkeys and humans, OK? And we'll talk again about why you would want to do that, but let's just first get into what is actually shown here, OK?

And the other thing I should tell you is that the images that are presented in the experiment are kind of organized here, according to categories of images. So the first little bunch that you run into here are images of humans. So these are faces. Those are bodies.

These are images of animate things that are not human. Again, faces. So these would be animal faces and these would be animal bodies.

And then, down here, you have inanimate things. And that's divided up into natural and artificial. So you got a banana here and like a gun. It's just all kinds of random images, OK? All right.

And so when you look at the matrix here, there's a few things that kind of pop out. The first is that there's what we call block diagonal structure, OK? So if you step back and squint your eyes, like there's kind of a blue square here, right? And kind of a blue square here, and then red here and red here, OK?

And so that's an indication that the images of the animate objects-- so that's the first half, all right, are kind of producing similar brain responses, on average, right? And the inanimate objects are kind of producing similar brain responses, on average. But that the animate and the inanimate images tend to produce pretty different brain responses, OK?

So you can also see some really strong block diagonal structure, right here. And so what is that? Well, that's faces, right?

So you've got human faces here. So that's really solidly blue. So all of the face images are generating very similar aggregate brain responses.

And then, if you look down here, there's another kind of blue square, all right? So that's all the animal faces. OK, so those are also generating very similar brain responses.

But then, you also get a blue square here, OK? So What does that tell us? Anybody want to tell me?

AUDIENCE: Animal face, human face.

JOSH MCDERMOTT: Yeah, so the animal faces and the human faces are also producing pretty similar brain responses, OK? OK, so this is one way to characterize the brain's representation of objects, OK?

Now, one of the things that this was used for-- and this was really the point of this particular paper, which introduced this method-- but then what they were really using it for was to try to compare the monkey visual system and the human visual system, all right? And so the monkey and the human visual system have-- there's certain homologies, in terms of the areas.

But really, they're kind of different, right? So like, in humans, like the visual system is kind of scrunched into the back of the brain because you use your temporal lobes for language and stuff like that. Whereas in the monkey, the visual system kind of extends into the temporal lobe, all right?

So the actual sort of physical substrate is kind of different, OK? So it's not really totally obvious, like how you would actually make a comparison of the human and the monkey visual system at some of really fine-grained quantitative level, OK? But this particular analysis method gives you a way to actually make comparisons across brains, even if it's like different species, right? So you can ask is the structure of the representational space, as measured in this particular way, of like the similarity of the responses to different images, is that's shared between monkeys and humans?

And what you're supposed to take away from this is that the thing on the left looks a lot like the thing on the right, OK? So in both species, you see this coarse scale structure of different responses to animate and inanimate objects. Faces seem to really stand out as being kind of special. Animal and human faces seem to be represented similarly, both in monkeys and humans. Yeah.

So the take home message here is that this is one way to analyze brain representations. It's called representational dissimilarity. It's in pretty widespread use now. It's based on population level responses. So in this case, it's like the fMRI responses from a big chunk of brain, OK, to like a pretty large set of stimuli, OK?

And we believe that the structure here is mostly driven by the large-scale organization of category selectivity. So the fact that there's this fusiform face area that responds a lot more to faces than to other stuff. And so what that means is that if you look at the response of the entire visual system to objects, faces are going to tend to all produce pretty similar responses-- because there's this one set of voxels that generates a huge response to faces and not much of a response to anything else, all right?

And you see kind of similar stuff for bodies. So this large scale category selectivity that we certainly see, at the level of fMRI, contributes a lot to this kind of thing. Ask me questions. Yeah.

AUDIENCE: What's the benefit of using dissimilarity for second level correlation?

JOSH MCDERMOTT: I'm not sure that there is much of one. I think there's some reason for it, and I forget what it is, yeah. I mean, it's got the same information in it, yeah.

AUDIENCE: Yeah, so it just seems like one extra step.

JOSH MCDERMOTT: Well, yeah, I guess the idea is that it's more like a distance, all right? And so people often think of this as like capturing geometry. And so just dealing with a distance-based measure is appealing, from that standpoint, right? But I think you could draw a lot of the same inferences without doing that, yeah. Yeah?

AUDIENCE: Can you use this to draw any information on differences between monkeys and humans? Like, for example, like the human one looks to be like a lot more well defined, in terms of borders. And also, like for example, in the monkey one, high dissimilarity between human face and not human body. But you don't really see that in humans. Like, I don't know. Is there some explanation for these kinds of things?

JOSH MCDERMOTT: Yeah, I mean, I think your observation is correct. There are some subtle differences. I mean, exactly what that means and why there is not totally clear.

But yes, you can conclude that the representations are not identical. And without something like this, it's not even clear how would you go about evaluating that, right? So yeah, there's some similarities, but they're also not exactly the same, and. Yeah.

AUDIENCE: Has this been done for animals other than monkeys? Like, is there any comparison between the different types of animals and humans?

JOSH MCDERMOTT: I'm not aware of this having been done with any animal other than monkeys. Yeah, I mean, it's interesting that the animal models that are popular in neuroscience has changed over the years. So the early days, like the '60s and '70s, like there were lots of experiments done on cats and some on monkeys. And then, but really, hardly anybody like, does anything with cats anymore.

And then, nowadays, really because of the molecular genetic tools that are available in rodents, there's lots of people who work on mice and rats. But their visual system is very different from humans. So I don't think anybody has done an experiment like this with them. That would presumably would look totally different. Big area for cheese or something, right? So yeah.

AUDIENCE: I'm curious if a similar type of representation of the similarity matrix has been done at different levels of development. So like if these associations are more ingrained or they're learned through time.

JOSH MCDERMOTT: Yeah, so I don't believe it has. And the reason for this is that doing an experiment like this in children is really hard. Because really, the number one thing that determines the success of an fMRI experiment is whether the participants hold still. And kids are just not as good at staying still.

And so it's really hard to get the kind of power that you need to do an experiment where you have like 200 stimuli. So I mean, there are people in this building, in like Rebecca Saxe's lab that are working on doing infant fMRI to look at questions like this. But it's because it's so much harder to get good data, like typically the experiments will have maybe four kinds of conditions, rather than 200.

So I mean, you could look at a very, very coarse representational dissimilarity matrix. In fact, I think they may have done that, actually. But nothing like this, with like 200, yeah. So yeah. So it's a really interesting question.

But they do find that you kind of see evidence for face selectivity in really young infants. And I think, also, maybe body selectivity, yeah. So Heather Kosakowski is a grad student who did a lot of that work. You should check out her papers. Yeah.

OK, so we talked about this idea that we can kind of think of recognition systems as having a computational goal, right? Which is that you take this image and you want to transform the representation of that image into a representation that makes object categories explicit, right, by kind of putting them in different parts of the representational space, right? And so we talked about evidence that if you look at the representations in inferotemporal cortex, they have that property. But we didn't really talk at all about how you would go about achieving that property, OK?

And so one kind of interesting development in this field is the advent of artificial neural networks that can recognize objects from images pretty well, OK? And this is something that it just wasn't the case even 10 years ago, right? So this is a pretty recent kind of development.

And so this has led to a lot of interesting analogies between, in this case, the ventral visual stream and a neural network, an artificial neural network that recognizes objects, OK? So this is a diagram of one common type of artificial neural network called a convolutional neural network. And so these are machine systems that consist of cascades of simple operations. The operations that are in these kinds of systems are, arguably, loosely inspired from things that people initially observed or theorized about in the brain. So they perform filtering operations, thresholding operations, pooling, normalization.

And these are each individually simple operations, but then when you stack them, OK, you can get pretty complicated behavior out, OK? And the critical thing that kind of makes these architectures work and makes them useful is that they're pretty easy to optimize, all right? So people, for a long time, we're not actually able to successfully optimize big systems that looked like this. But nowadays, we just know how to do it, OK?

And the essential way by which this works is really pretty simple. And I'll tell you about it. But as I said, the most common architecture for sensory tasks, in the one that's really useful for making comparisons to biological sensory systems, is what's called a convolutional neural network.

So you remember we talked about the operation of a convolution, right? So what a convolution is, is you've got a filter. And you apply that filter at all the locations in your signal, right? So in an image that would be all the x and the y locations, right? In a sound, this would be all the temporal positions, OK?

And so we talked about this idea that, in the context of vision, you can think of the convolution operation as representing what a population of neurons that have the same kind of receptive field, just at different spatial positions-- so think of center surround receptive fields, in the retina or orientation selective receptive fields in V1, what they would essentially do to an image when they're all acting in concert, OK?

And so one of these neural networks is performing that operation. So each one of these sheets kind of represents the response of a filter at all the different possible locations in an image, OK? And the different sheets kind of represent different filters. So this one could be horizontal, this one could be vertical, this one could be diagonal, and so forth, OK? All right, so you have multiple layers of these filters separated by non-linearities and pooling operations.

And what makes them useful, what makes them really work is that you can learn the filter parameters using gradient descent, in order to maximize the performance of a task, right? And really a very popular task and one that's very widely used is object classification, OK? And so these systems are trained using these big data sets with lots and lots of images, where people have gone through, and with every image, they label the image, and say, that's a cat, that's a dog, that's a chair, that's a building, and so on, and so forth.

So you have this huge set of images. You can present the images to the model. The model will give you an answer for what class it thinks it is. Initially, like the weight the filters will be randomly set. And so the answers will mostly be wrong.

But the gradient tells you how should I change the parameters of the filters, in order to have the model make fewer errors, OK? So you do gradient descent over many, many iterations. And the model moves into a place where it can perform well at the task, OK? And so this is now kind of become one of the dominant methods in engineering. And object recognition was like a key in early success story, OK.

So in some sense, this is the realm of computer vision. But these systems, because they're solving a problem that we think is also important for biological vision, they give you a useful computational model that you can propose as a model of biological vision. And you can ask, well, does this model actually do some of the same kinds of things that humans do, OK? So this is now sort of a cottage industry of people building these models, and analyzing them, and comparing them to human sensory systems, in different ways.

And so one kind of thing that you can do is measure representational similarity, OK? So these are representational dissimilarity matrices. This is the kind of thing that we just looked at. This is from human IT using fMRI.

So again, you get this block diagonal structure. And this is that same quantity measured in a-- HCNN is Hierarchical Convolutional Neural Network that's trained to recognize objects. And you can see that they're not exactly the same. But there are some common features.

So again, you see this pretty clear distinction between animate and inanimate objects. You see that faces are kind of special. It's interesting that, in the neural network, maybe you don't see this kind of quite as strongly as you do in human IT.

So there's some differences too. They're not exactly the same, but there's some similarities. So this is kind of one way that you can make these model brain comparisons.

Another very popular way to compare brains to models is to use the features in the model, as what's called an encoding model, OK? And so the way that this works is that you try to model each neuron or each unit of measurement in a brain as a weighted sum of unit responses in some stage of your model, OK?

So here's just a schematic to illustrate this. So imagine you do the kind of experiment that we've been talking about, where there's a big set of images that gets presented to a person. And you're measuring their brain responses, OK? So let's say with fMRI, OK?

So here's a voxel. And the color here is supposed to represent the response of the voxel to each image. So every voxel gives you a vector, which is the response to each image, OK?

So here's our neural network. Consists of a bunch of stages. You can think of each stage of the model as consisting of a bunch of units, all right? You could think of that as a filter at a particular point in the image, for instance.

And so each one of those units will give you a numerical response to any image that you present to it. So you can take all of the images that you used in your fMRI experiment. And you can present those to the model, all right? And you'll get some big complicated matrix of responses.

All right, and so one way to ask whether the model is representing the same information that is represented by the brain is to try to predict the brain responses, using the model responses, OK? And the standard way to do this is to take the model responses and to fit a linear mapping-- here, that's f-- between the model responses and the brain responses, OK?

So you take the model responses to some subset of images and you optimize this linear mapping. So it's a weighted combination of the units in the model that allow you to predict the responses to some subset, usually half of the images. And then what you do is you take that matrix, that linear mapping, and you evaluate the predictions on another set of images, OK? And so you ask, does the predicted response that you're getting from the model features actually match the measured response from that little bit of the brain, OK?

And so you can think of what the linear mapping, it's kind of helping you align the coordinate systems, right? So it's telling you how to rotate and scale the model responses to get them to match up with the brain, all right? But that doesn't guarantee you that you'll be able to make a good prediction. So if the features that are in the model are kind of totally unrelated to the kinds of things the brain is measuring, the predictions might be terrible, all right? Sometimes they are.

All right, so here's just an example of what would come out of this. So these are example IT neural responses to a big set of images. So every point here is the response to one image. And these shaded regions are kind of categories of images.

So the black line is the actual measured response of a particular neuron in IT. And then the red line is the predicted response using the artificial neural networks features, OK? And you can see that there is not like a perfect match. But the prediction here is good, in the sense that it's kind of high, for the images in this particular region, which happened to be images of faces and low for everything else, OK?

OK, so the outcome of this exercise is typically expressed as variance explained, OK? So there are some variants in the brain responds. And then you ask whether that variance is matched by the predictions, OK? And that gives you a number. And that's a quantitative metric of the extent to which the model is replicating the brain's representation, OK?

All right, so here's one finding that kind comes out of this exercise. And that is that particular stages of artificial neural networks tend to be best at predicting particular stages of the visual system, OK? And so this is one early example of this. So the graph on the left is showing the predictions of V4 responses by different sets of models and different stages of this artificial neural network. And the graph on the right is showing the same thing, but for inferotemporal cortex. So remember, V4, we kind of think of as an earlier stage of the visual system than IT, OK?

So the y-axis here is variance explained. So bigger numbers mean better predictions, OK? 0 is like about the worst that you could do.

And the gray bars here are a bunch of old school models. So some of this is a model of V1 that's based on Gabor filters. Remember Gabor? It's one model of V1 simple cells. SIFT is like old school computer vision features. HMAX is a early model from this building from 20 years ago, so on and so forth, OK?

And then the red bars, here, are the different layers of this artificial neural network that's been trained to recognize objects. And so what you're supposed to take away from this is two things. One is that, in the best cases, the red bars are pretty far above the gray bars, OK? So the neural networks are giving us better predictions of the brain responses. And in V4, the best predictions kind come from the middle layers here, layer 2 and layer 3, OK? If you look at IT, the red bars are still doing the best. But now, the red bar that produces the best predictions is actually the deeper layer, layer 4, OK?

So it's consistent with this idea that the hierarchy that we believe exists in the visual ventral stream, where it consists of the sequence of regions, and as you go from one region to the next things change, right? The receptive fields get bigger. The responses is get more complicated, in sort of an intuitive sense, OK? Object information becomes more explicit.

All right, so remember, we talked about how you can read out object information pretty well from IT less well from V4, right? So things change. We think that there's this series of transformations going from region to region that culminates in representations that can mediate object recognition.

And in this artificial machine system, the artificial neural network that is also optimized to recognize objects, there's some evidence of an analogous hierarchy, right? In particular, the middle layers are giving you the best predictions of a middle layer of the ventral stream. And the deep layer is giving you the best predictions of the deep stage of the ventral stream, OK? Questions about that? Yeah.

AUDIENCE: Any like current research being done on CNN's or artificial networks to model or like to perform object recognition, is like the focus more so on editing the actual models to make them more replicated of the brain or changing the data that we're giving them, at least?

JOSH MCDERMOTT: Yeah, it's a great question. So the question is like if we want to make these models better, do we change the training data or do we introduce architectural constraints that would make them more brain-like.

And I think there's a lot of interest in both. And I think it's an open question, as to what the relative importance of those things will be. There's one perspective that really-- like the task and the diet kind of predominantly determine the characteristics of the representations. There's another perspective that anatomical constraints really kind of matter a lot.

So I mean, and the other thing I should say is that when we talk about training a neural network, right? So training often kind of means adapting the weights via gradient descent. But it can also mean optimizing the architecture.

So any time you build one of these models, there's all these choices that you make about the number of stages, and like the operations that are in each stage, and the sequence of the operations, and how big the filters are, and how many there are. And there's a million choices that you make, right? And those choices can matter.

And so that sometimes that can actually be the subject of an optimization process. So you search over like a big set of architectures to find something that kind of works well. And that may, also, be one way to potentially find architectures that are computationally promising and maybe that also end up kind of resembling the visual system in some abstract way, yeah. So these are kind of, I think, these are open questions that lots of people are interested in, right now, yeah.

OK, so this is one kind of early piece of evidence that there is some tendency for artificial networks to replicate what we think of as the hierarchy in the ventral visual stream. And this is from recordings in monkey V4 and IT. This is another more recent piece of evidence using fMRI responses in humans.

So this is a comparison between a visual convolutional neural network and human fMRI responses. And so this is a big flat map of the visual system in humans. So we've got V1, V2, V3, V4, the extrastriate body area, LO, the fusiform face area, parahippocampal place area, OK. It's the human visual system.

And then the color denotes the stage of this neural network model that produced the best predictions of the voxels in a particular bit of the brain, OK? And these are different hemispheres, like left and right, OK? And so the take home message here is that the early visual areas, like V1 and V2, are best predicted by lower levels of the model.

And you then kind of have places like V4, which we think of as intermediate visual areas that are kind of green. Those are in the middle. And then some other regions here that are red. So they're best predicted by kind of deeper stages of this model, OK? So again, another piece of evidence that the hierarchy that we see in the visual system can, in some cases, kind be mirrored by the differences in representations across stages of this artificial neural network that's been trained to recognize objects.

So these findings are not unique to the visual system. So this is some work from my lab doing something kind of analogous in the auditory system. So this is a similar kind of comparison between fMRI responses in humans to a big set of natural sounds and deep neural networks that are trained to recognize words.

So same kind of analysis, where the voxels here are color-coded based on the stage of the model that produces the best response. So pink is kind of an intermediate stage of the model. So here's early to late. So pink is kind of in the middle. And the cyan is deeper stages.

And so you can see that there's this region here where you get the voxels colored pink. And that kind of loosely corresponds to what we think of as primary auditory cortex. So these black outlines, here, are three anatomical subdivisions of Heschl's gyrus, the site of primary auditory cortex. So things around primary auditory cortex are tweaks are pink. And then, beyond that, you see a lot of blue, OK?

OK, any questions about that? OK, so this is, as I said, a very active area of research. Lots and lots of interest, right now, in the extent to which artificial neural networks will mirror the structure of sensory systems, and under what conditions, and how you can use them to do interesting things.

So overall summary of what we talked about, about object recognition. So the ventral stream is believed to mediate object recognition. Overall, the complexity of the neuronal stimulus tuning seems to increase from V1 to IT.

The key computational challenge of object recognition can be thought of as untangling object representations. We talked about evidence that basic object recognition is pretty quick. And we believe, likely, mostly feedforward.

We discussed evidence that IT representations make object identity explicit. Where again, explicit means that they enable linear readout of the object identity. We discussed evidence that artificial neural networks replicate aspects of visual ventral stream processing. And we talked about evidence for object class-related brain specialization. The poster child being face recognition, OK?

We just talked about object recognition. It's tempting to conclude that object recognition is kind of the point of vision, right? I showed you of simple schematic, where we got early vision and mid-level vision. And then it culminates in object recognition, OK?

But the point of this lecture is that perception is not just about objects, all right? So on the left, we have an object. What is that object?

AUDIENCE: A leaf.

JOSH MCDERMOTT: A leaf. Yes, it's a leaf, all right? On the right, we have lots of those objects, all right?

And when you look at that thing, if you look at it closely, you can probably pick out individual leaves, right? But at a glance, you can tell that that's leaves, all right? So there's some collective properties, when you have a lot of leaves that create what we call a texture, OK?

So textures are everywhere in the visual world. They tell us about materials. So you can glance at these things and quickly tell that you're dealing with grass or gravel.

We've already talked about how they're an important depth cue. So they give us information about shape. So just via the texture gradient here, you get a sense of the three-dimensionality here.

Texture also helps us to segment scenes. So you can see how there are different image regions here that have different textural properties. You can distinguish the lawn from different parts of the house. The roof that's kind of covered in ivy. OK, all right.

So texture is an important part of the visual world and vision. And humans are pretty good at detecting some texture boundaries, OK? So if you look at the image on the left, you can very clearly see this distinction between the left and the right side, OK?

But some kinds of texture boundaries are a lot easier to see than others. So in the image on the right, it's harder to tell that there's one kind of thing on the left and another kind of thing on the right, OK? You have to inspect it a little bit more. And there's lots of examples of this. So all of these images contain texture boundaries. Some of them are very salient, and others, like this one, are really kind of hard to see, OK?

OK, so one question that people have historically been interested in is what determines the salience of these boundaries, right? So here's just a bunch more examples where, in some cases, it's super obvious that you've got two different types of texture. In other cases, it's not so obvious.

And so just informal observation kind of suggests that the texture differences that are salient involve differences in what we would think of as basic features-- things like color, shape, orientation, size-- that we often think of as being kind of detected in early vision. And so one kind of obvious question is whether we can predict the salience of these texture boundaries using the models of early vision that we've learned about and grown to love, right?

So here, we have an image. It's convolved with two Gabor filters or approximately Gabor filters. And this is the output. So what you can see is that when you convolve it with a filter that's tuned to horizontal orientations, the regions that have more horizontal energy in the middle give you larger responses. The reasons that are more vertical give you smaller responses. And the vertical filter, it's the opposite, OK?

All right, so how can we actually take these responses, and then and turn this into something quantitative that we can compare with our perception of boundaries? So here, we have one of those convolutions that we just looked at, right? And we're going to take a 1D slice through the convolution, and we plot it, all right?

So this is the filter response as a function of space, all right? The y-axis is the response. The x-axis is space.

And so you can see that this response wiggles around, from positive to negative. But then you get here, in the middle. And the amplitude of the wiggles is kind of smaller, OK?

All right, so we can make that explicit. So this is like a simple cell response, more or less, right? So we can make that explicit by squaring the filter response. So that makes everything positive.

So now, you've got responses here that get big and go down to 0. And then, here, they just don't get that big, OK? And then to summarize that, we can average over local regions with some kind of smoothing operation. And then we get this thing that's shown here, all right?

So this, now, has a big response over here and a small response here. And then another big response. So sort of distinguishing between these regions that we see.

All right, so what we just did here, which is squaring and averaging, gives us a quantity that is a lot like an energy measure or an envelope that we've talked about in other parts of the class. Remember, when we were talking about simple cells and complex cells, we talked about this fact that complex cell responses can be modeled by combining even and odd simple cell responses, right?

So remember, you've got an even in an odd simple cell response. These are Gabor functions with just different phases. So then, you square them, and you add them, and you get this energy quantity. And so the energy quantity is always positive and smooths out these effects of phase that you see with the simple cell receptive fields, OK?

And so this is one way of computing energy that is especially simple to think about. But you can get a similar measure by squaring and just averaging over space, right? So if you think of the sine and the cosine as having similar receptive fields, just at slightly different spatial phases-- if you took one of them and you averaged over space, you'd get something pretty similar. And so that's what we're looking at here.

So this is an energy measure that is derived from a simple cell-like receptive field that's been squared and averaged, OK? And in some cases, these energy measures can predict texture boundaries pretty well. So we've got different texture here, in the middle. And that kind of shows up as a difference in the amount of energy, OK?

All right, so the key idea that I want you to take away from this is that texture is defined by the average properties of an image region. OK, so in this case, the texture kind of is generated by all these elements, OK? But what kind characterizes the look of the texture-- or so that our hypothesis for what characterizes the look of the texture is its average properties, OK? And so this is one way to turn that into a quantifiable, measurable thing by taking some response, and then kind of locally averaging it over space, OK?

All right, so this is a really simple idea. So what about seemingly more complicated textures? Can we account for the boundaries that we see in something like this with such a simple idea, OK?

And the answer is that these very simple energy-based measures do a pretty good job, OK. So here's a bunch of examples where we've got an image on the left and the output of an energy-tuned measure on the right. And you get differences in the responses that co-vary with your subjective sense of the salience of the texture region, right?

So this one is pretty obvious visually, and you get a pretty big difference, OK? Here's another one that is even more obvious. And you get a really big difference.

Here's one where it's really actually kind of hard to see the difference between the center region and the surround region. So the center region has got these little Xs. The surround has got the Ls. But the sizes of them have been adjusted. And it's a lot harder to see. And lo and behold, the energy measure also doesn't give a very big difference in the response, OK? All right, so this is kind of one early development in the world of texture. So this is an example paper by Bergen and Adelson that espoused this sort of idea.

OK, so energy measures can predict some texture boundaries. But are they a complete description of what we actually see when we look at this texture, all right? And I'm going to make the claim that if we have correctly identified the brain's representation of texture, then if we measure that representation for an example texture, like this, OK? And then we generate synthetic textures that produce the same values of that candidate representation, well, then the synthetic textures, they ought to look like the original example, OK? So that would be like the purest test of our theory of representation, OK? And so that's what we're going to try to do. And this is a pretty influential method for evaluating texture models known as texture synthesis, OK?

All right, so the idea is that, for any given type of texture, There's like lots of example images that are the same texture. So if you just look down-- everybody look down-- all right, there's a carpet, right? And that carpet has a texture, OK? There's a microtexture and a macro texture.

But you can imagine taking lots of images of that carpet. They would all be the same texture, all right? Different images, but same texture. They have the same properties.

So we would call those equivalent, right? So in this case, those different images of the same texture, they would be generated by exactly the same physical process, right, which is like somebody sewing together the carpet, OK? Now we think that when you look at the textures and you have the sense that they're the same thing, well, your brain is doing something, right? There's some representation, in your brain, that is being produced that is the same for all of those different images, OK?

All right, and so those brain representations will cause there to be a set of perceptually equivalent textures for any example image that we might come up with, OK? And that might not be exactly the same as the set of textures that is generated by the same physical process, all right? But these are things that would look the same to you, OK?

OK, so the synthesis concept is that we're going to use image synthesis to test whether a representation captures texture perception, all right? And so the way that this is going to work is we'll have some example texture image, like this, right? It's just got a certain look to it, OK? We're going to start out with a noise image. And we're going to try to change that noise image to cause its representation, in our texture representation, to enter this green set, OK?

And so here, here we've moved it, to this point. It's not all the way in the green set. You can see that actually doesn't look exactly like this. Although, it's maybe part of the way there, OK? All right. And the idea is that if we have the right representation of texture, right, that when we do this, when we take that noise image, and we cause it to have the same representation as some candidate image, they should look like the same type of thing, OK, in all respects.

OK, so the first example that I'm going to give you of this was really the first attempt that people made at instantiating this general kind of scientific philosophy, OK? And they did it in a particular way that was simple and elegant. And so I'm going to walk you through how that worked, all right?

And so the essential idea is that the representation of a texture consists of the distribution of filter responses, OK? So that our energy measure is the average squared filter response. So we measure the filter response. And then we square it and we average it, all right? And you can think of that as kind of matching one statistic. So the average of the square is more or less like the variance, if the thing has 0 mean, OK?

So what we're going to try to do here is to match the entire distribution of the filter's response, OK? So this has the consequence of matching all possible statistics. And whereas, statistic is capturing some average tendency of a signal, OK? All right. So this may not make a whole lot of sense. So let's walk through an example, OK?

All right, and so we're going to do this in a filter bank representation of an image. So let's review filter bank. So here's an image, OK? This is a set of filters, this row. They vary in spatial scale. So we have things over here that are small at high spatial frequencies and things over here that are bigger and at lower spatial frequencies.

And they vary in orientation, all right? So we've got vertical, horizontal, and the two diagonals. All right.

The next row plots the convolution of that image at the top with each one of these filters, OK? So remember, the convolution gives you a filter response at every point in the original image. So you can display it like an image.

And what do you see? Well, the low spatial frequency filters, their response varies kind of slowly over space. They're picking out low frequency structure at different orientations. So the horizontal filter responds a lot, whenever there's horizontal edges. The vertical filter responds a lot, whenever there's vertical edges, OK? And then the high spatial frequency filters do the same thing, but at a higher spatial frequency scale, OK?

All right, so the technical term for each one of these things is a subband, OK? So that's what people refer to the output of a bandpass filter as in the engineering world. So but for us, it's just the convolution of the image with the filters. And from the prism of neuroscience, you can think of each one of these images as kind of a simulation of a population of neural responses, all right? So it's like all of the neurons in the visual cortex that have receptive fields that look like that, where every neuron has the receptive field at a different spatial position, OK? All right, so the model of texture perception that we're going to test here is one in which textures are represented with the distribution of activity in these subbands. Again, simulating the response of oriented spatial frequency-tuned neurons.

All right, so this image that we're looking at here is not a texture. It's like just something else. Here, we have a texture.

And this is the subband representation of the texture. So this is like the exactly the same thing that we saw here. So you're taking the image and convolving it with all of the filters-- with one wrinkle, which is that when the filters are coarser, you can downsample the image, OK?

And so one kind of common type of representation in image processing is called a pyramid representation, where the low spatial frequencies get downsampled, so that they can be represented with fewer numbers, OK? So here, we have the high spatial frequencies. Here's the middle. Here's the lower. But they've been downsampled, and so the actual image representation of this is smaller. It's got the same information content.

And so it's called a pyramid representation because you have these subbands at the base that are big, all right? And then it progresses to subbands that can be represented as smaller things. OK, so pyramid representation.

Now what we're going to do is take each of these images and turn it into a histogram, OK? So each one of these things is an image. So at every point, you get a number, which is the response of the filter, all right?

And this is just a histogram of all of those numbers. It's like you throw them all into a bag, and you just measure the distribution, OK? So that's the histogram.

This here is a histogram of the pixel intensities, all right? So what does this show? Well, obviously, when you look at this image, it's mostly white. And then, there's some little bits that are black.

So the histogram here has a small little peak down here at 0. That's all the bits that are black. And then a big peak up here at 255 or 1-- I forget what the scale is that they're using. So that's all the white pixels, OK?

So this here is the histogram for one of these particular subbands. And so you can see that this has a big peak at 0. So that's what gray is, right?

So there's all these places where the filter is not giving you any response. And that's because the image is uniform, OK? And then, these places, where the response is either lower or higher than 0. All right, so that's what you get. And so the proposal is that the representation of texture is going to be a whole bunch of these histograms, one for each one of these filters.

OK, so what does this mean to actually have this as your representation. Well, OK, the histogram is throwing out all information about space, all right? So you're taking all of these values. You put them in a bag, and then you just measure how frequently you get one value or another, OK?

So that's a pretty big change from this as a representation, right? This has spatial structure, OK? OK, so instead what we're just what this is capturing is kind of the average properties of this thing when you pull it over space.

All right, so this will capture the energy. So the energy would be captured by the variance of that distribution. But it'll also capture all kinds of higher order statistics, which might be important.

OK, so now, what we're going to try to do to test this theory of texture representation is to measure these histograms and some example images, like that one there. And now synthesize an image that we're going to force to match all of the histogram, all of the subband histograms, OK, the histograms of all of these filter responses, OK?

So we're going to start with noise and equate the histograms, all right? So you'll start with some kind of distribution. And then, you're going to mess with the image to change that distribution. So it turns out that this process of histogram equalization is super simple. So it's a kind of classic trick that is used in image processing.

So the way that this works is you have the image that you want to match to. You generate its histogram. So this is now just done with the pixel intensities to convey the idea, all right?

So remember, there's a little bit of black and a lot of white. You generate the cumulative histogram, OK? So that's just kind of integrating this thing over the intensity axis. And this is the same thing for the noise image that we start with.

So this is a noise image. And so it's got a pretty uniform distribution. And so its cumulative histogram looks like this, OK?

And so, now, all that we're going to do is generate a lookup table, all right, that basically says that you take a pixel here. And you find what its value is. And then you just map that onto the cumulative histogram that you get for the other image, OK? And so, in this particular case, something that has a value of 0.8 gets turned into having a value of 1, all right?

So this image, here, is that noise image that's now been histogram-equalized to that texture, all right? But just in the pixel domain, OK? So now we end up with a noise image that's mostly white and a little bit of black, all right, but otherwise quite random.

All right, now, we're going to try to do the same thing, but now in the domain of the subbands, the filter responses, OK? So here, we have the filter responses of our texture that we're interested in. And now, we have the filter responses of our noise that's just been histogram-equalized in the pixel domain to the texture, all right? And so those look very different.

And lo and behold, the histograms look totally different, OK? So this one is very peaked at 0 and has these long tails. This looks a lot more Gaussian, OK?

So we're going to do the exact same thing, but on these images, all right, these subbands. So we'll equate their histograms. And so here we have the new subband that has now been histogram-equalized to the original texture. All right, they've got the same distribution.

So we're just going to do this for every subband, OK? This actually kind of has to be done iteratively. But it's super simple.

So you just take your image. You split it up into sub bands. Each one of those gets histogram-equalized to the original. And then you can add them back up to get an original image and do this iteratively.

And so what you can see is the red and the blue plots here. One of them is the target histogram. The other is the histogram of the synthetic image. Each one of these is for a different subband, so a different filter.

All right, so then we're going to add these things up, and we'll get an image, all right? And so the key question here is whether these things look like the original, right? So remember, the idea is that if this candidate texture representation captures what is in our brain's representation of texture, then the synthetic examples, they should look like the original, OK? And when we come back on Tuesday, I will tell you whether they do.

[LAUGHTER]

OK.