This poster presents my ongoing work on developing a biologically-inspired model of visual processing in the ventral visual stream.

We know a lot about processing in the early visual cortex where oriented edges are detected, and we also know what happens at the end of the visual information processing where objects, scenes, and faces are recognized. The key question to me is what happens in between, that is, how do we get from these local very simple features to the highly complex shapes or other properties in the higher visual cortex. This work is about understanding computations in mid-level vision.   
  
In order to understand what is going on in mid-level vision, we have to ask ourselves about the outputs of this stage of processing. To begin, I am going to describe three key insights that are going to guide us in building this model. First of all, notice how the outputs of visual processing or at least mid-level processing is Isst surfaces India particular depth work. This may sound a little bit abstract, so let me illustrate that with the figure below. You are seeing now a configuration of several shapes and if I ask you to describe what you're seeing, I am pretty sure that you are not going to say that you're seeing what's shown in the middle panel, that is, four distinct disconnected regions. In fact, what you are seeing is complete shapes, so the light gray one being completed behind the black one, and also the black one extending behind the gray one, and also that the at the Dodig background extending fully behind that you shapes and not being composed of two separate regions. Let's the output that is produced in the middle figure that is actually what's usually sought after by models of vision or not interested in actually receiving the full shape robber in in just segmenting the inputs into these distinct regions and as this example of showing, this strategy might be misguided.   
  
Furthermore notice that you were able to do all of these computations or all of the segmentation without any knowledge about what's in that image. So in in this figure below to get easily describe objects of the ships of these objects, they are all 3G structure, and you could act upon them and describe the texture send Seme and so forth . And all of this without knowing or caring about what these objects are. So the point here is that many models of vision are identified this and semantic knowledge they're looking for a particular there intergender recognition tasks based on the training data off the labels on objects . But this is not what the visual system is doing bridleway . As we're seeing in these examples can do all the segmented all of the segmentation all these difficult tasks without any clue what's in BNH. I think this is a very important him remarks that on what we are trying to build Cheneric system him it doesn't it shouldn't be heavily relying on semantic knowledge although of course McHoul.  
  
Finally noticed that the task that's modelist doing him is Notz just object recognition snob just seen recognition . Julie Straighttalk look at the series of figures down here. In the last one you can see how this is an example of the typical image where you have a very clearly defined object car , and it's back. Now in the next one you may still be able to identify several objects such as buildings, but the ground truth is not so clear anymore. Now you're not sure whether the objects in the Seon are buildings or role buildings, or maybe the windows in these buildings should also be considered objects. And I'll bet it's only worse home in the next figure out where there is a mountain Cean and you may be able to point to the mountain as if it was an object called Mountain , plus crucially you no longer have the clearly defined boundaries of these objects , you will you cannot really indicates where the Mountainbiken Zelary Anse . In fact what's dominating the scene now is not so much certain objects but the spatial layout config Ración off this scene. And then the last family here you see that this is very clear now you have a clear Oración, but there are a few clearly identifiable on objects and honestly you couldn't care less about that. What your Karens here about this navigation trying to find your way and seen , understanding the spatial relations in that scene and differences between textures Rotherton trying to identify objects minutes.   
  
Again what we're looking for is Amabile that would be able to do segmentation and death for drink with Yuper with the correct filling in to Deportes on the semantically uninformative inputs and not only objects but also layouts Dexters and so forth.   
  
Next I'm going to outline the architecture goodbye believe it would be sufficiently robust to achieve these goals. At the initial stage early visual cortex on features architected so they weren't the Barsan other things just like in many models of vision. Now we have these local teachers protected with computers would perform two computations the similarity competition were we looking for him for similarities across these features followed by the cooling operation that pulls together highly similar units therefore resulting in sparse or representations. Now these computations these two operations might be performed several times over gradually larger scales such that small features are combined into more and more complex fragments finally resulting in full shape extractions . That is what's computation mid-level vision is . Now the account but this competition can be used for multiple tasks to hire visual cortex on such as objects recognition or what constructional 3-D shape him or him . I'll also noticed that the D I'll Conceptos mid-level competition the first Isaar I'll probably quite Courcier Naderi correct because this competition needs to be done fast Ossos necessarily not very accurate but with attention and Harverhill areas we can focus on the particular surface that was extracted and I'll and thus now process this particular representation of this particular surface and gain more accuracy . As you can see this architecture is very dynamic as many bypass rules this is very quick just computations and also I'll be back computations to refine representations .  
  
Call me I'm going to describing more detail how these computations on a mid-level vision to operations of similarity and pooling our unit him first you can see that a him how redundancy computation is performed first when if you just compute edges and the finding all the Melonson the figure but all over the place because we're texting that and every single axle that is of course very home very dense representation that is wasting our resources unnecessarily so. So what we can do is all for the units that have very very similar and they're the Texas orientation on and the results sparse or am representation now is against him the right hand side and that is also more accurate because were of course not seeing that many exes all over the place. The little bit larger scale now we can perform for singularity competition on the detectives and sparsest five edges . If the similarities very high as you can see you along the water off this shape these features are pulled together to form it's kind of curved fragments of sh\*t. If however we are at the place where there is clear dissimilarity across futures Tacular case edges edge orientations, then that results in hall second order edge detection as you can see indicated on the right-hand side figure. Dakkan it can be used in the later stage of processing larger scale to pool together units into longer fragments and I just like I told you before.   
  
No eventually what we are at the thing even larger scale we have a number of edges detected and or larger fragments and what we want to do is to determine now not only these more complex fragments but now the house the pool together in terms of how the closer together entrance off the whole shapes . And in order to do that we can play border ownership computation that tells us which of these features are likely to belong to the same circles because of their similar properties . How particular this is the Discret  
Enhanced of this computation I'll over all we can do the final segmentation that Wortring in a global scale .  
  
So now am I thing most of these complications are quite straightforward to match all the way they said I have not developed the algorithms that robustly be able to do that Silkers was going to be certain problems in implementing this approach anyway but am I would like to come illustrates the border ownership computation in more detail because I find it very well less intuitively clear computation and also I'll conceptually very powerful technique. If you look at this figure here I'll have to shaped you and suppose that we have the aces detectability just like to talk before and I'm indicating two of them here and rest. Have a question Azars as I said before what is the two which all surface do these edges belong. And they're only two options of what each of the these faces the Fox so either the edge is defining the borderline of the great migration here for all the white babe here . And I'm indicating these two options with the green arrows pointing into each direction. So what we need to do now is to decide somehow to wish to wish at least you are all shapes these edges block. Now of course we know that has to belong to pray one Ahela tomorrow. Here I'm going to propose that we use the complexity assumption that is that objects tend to be complex. Signed this particular case that means that's good to just have to belong to the ray Serfas because that's how they avoid form off convexed sh\*t.