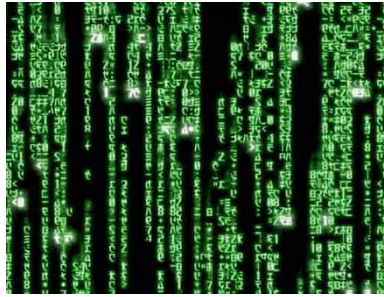


Computer Vision Notes

Ross Girshick

Why Computer Vision is Hard

We often fall prey to the *error of experience*. Vision is an automatic and unconscious process for humans; it may not be obvious that it poses great difficulty for computers. For those who have seen the film **The Matrix**, it's rather like this:



Neo: Is that...

Cypher: The Matrix? Yeah.

Neo: Do you always look at it encoded?

Cypher: Well you have to. The image translators work for the construct program. But there's way too much information to decode the Matrix. You get used to it. I...I don't even see the code. All I see is blonde, brunette, red-head. Hey, you uh... want a drink?¹

To a computer, an image is literally a matrix of numbers. The problem faced by computer vision researchers is how can we make computers *see*² what's encoded in the matrix.

¹Need to cite The Matrix screenplay.

²Here we get into the sticky business of defining what it means for a computer see. For now, let's just say that a computer can see an image if it can produce a human readable *description* of the image's contents. E.g., a bounding box labeled "cat" surrounding a cat.