Exposure range in images

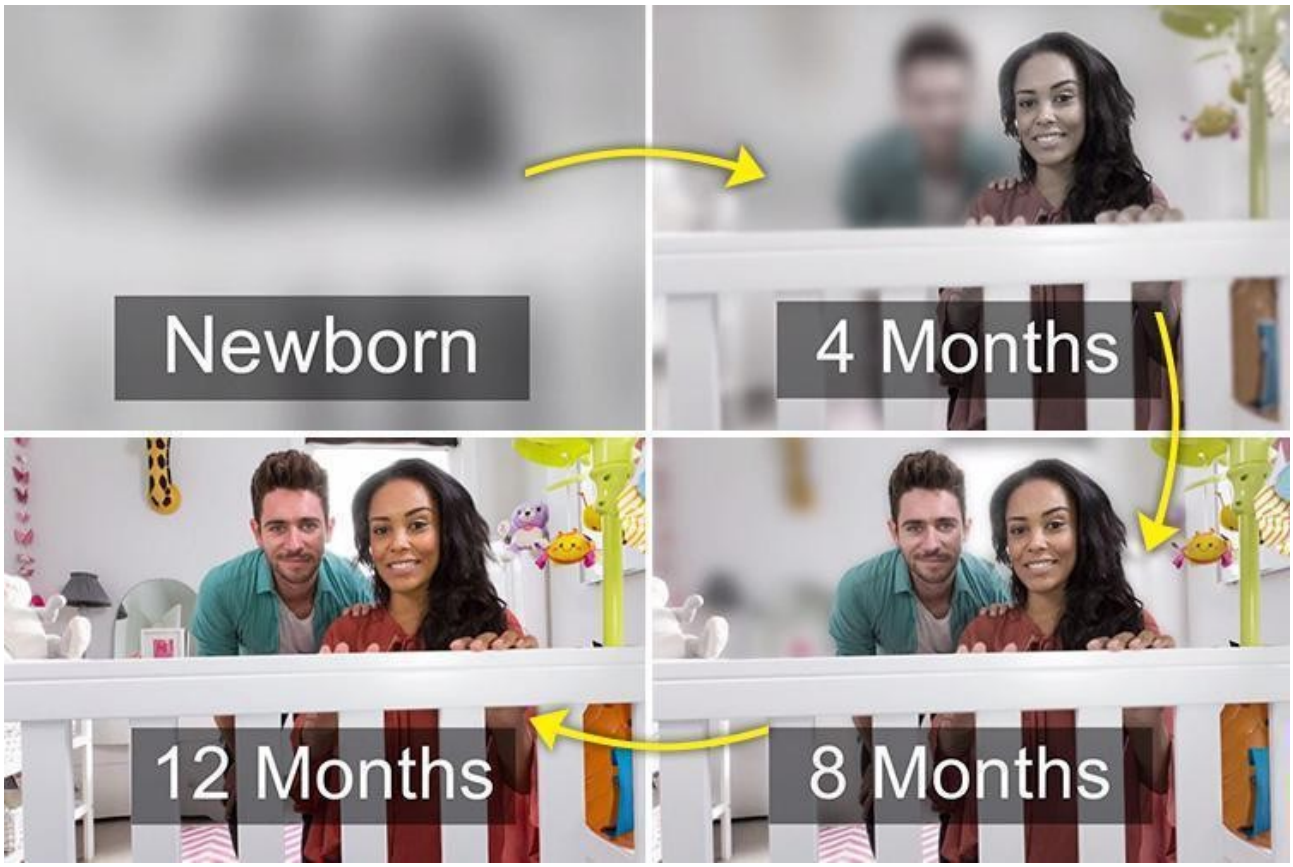
“The dynamic range (i.e. the differences between areas of shadow and light) of the visible is considerably higher than that which perceives the human eye and that of the digitally recordered images. However, while the human eye is able to adapt to very different brightness levels, usually cameras and monitors can only reproduce a fixed dynamic range. Photographers, film artists and generally those who work with digital images must be very selective in determining what is important in a scene, because they work with a limited dynamic range.”

To be able to cover with a conventional good quality camera the entire adaptable dynamic range of the human eye, it takes at least 5 shots with different exposure times.

Anyway there are many aspects to consider related the ability of human eye to adapt the dynamic range as needed. And they don’t touch only the light sensibility. For example the ability to focus an image at a determined distance has a great influence, as well as the ability to de-focus an image.

Spatial frequancy range and Chrominance

Following an image on how a baby transform and adapt his watching ability over the first year of life.



Has to be clearly understood that the baby has the ability since the 3rd month to build images like the one shown in the above picture at 12 months, but he doesn't acquire them intentionally, because he is structuring the vision biologic system. In fact the baby is not able to understand until the 6th month the hierarchy of object and the meaning of spatial position, especially distributed along the optical ax. This has several implication: inability to de-focus images to simplify the semantic abstraction of the scene, inability to understand that an object before was visible and now is hidden by an obstacle, inability do detect correctly and rapidly the image regions in YOLO conditions.
Of course all these consideration refer mainly to 'feature engineering analysis', instead of deep learning and convolution networks.
But, apart the incredible advantage to implement some feature recognition engineering in the hardware, the considerations exposed have a dramatic impact both on comprehension of how a baby watch a scene and on how a network behave in real conditions.
Let's start from the situation occurring when the baby reaches the 3rd-4th month of age: the spatial frequency dynamic range is very limited, so what mainly happen is that only the low spatial frequency features on the image contribute to the process of vision.
I often try to think about this, like the process of starting a network training only with very low resolution images. This is useful both to simplify the regions detection and to not overfit: the attempt of the baby is to progressively increase the number of classes meanwhile the brain cortex simplify the semantic representation. Only after the baby reaches some confidence about the world he begin increasing the spatial resolution and the chrominance dynamic.
As a robotic engineer, my personal experience is that usually biology makes things better than us. So it is natural for me to start from the position that, if I want a network model to work well, I have to refer to the Nature first. Only after, I try to organize my simple available tools to emulate Nature as much as possible.

Dynamic Range Propagation

A vision dataset is composed usually from color images, with resolution hopefully greater than some hundreds of pixels cross some hundred of pixels.
When we use those images we already apply a first spatial filter, resizing the image to match our first input layer. Of course we don’t modify the original dataset, to be able later to rethink about our first layer dimensioning.
Another consideration regards the exposure range of the images: we calculate mean and standard deviation of each image and we use those values as a criteria to normalize the dataset. Depending on normalization algorithm this can, or can not, have an impact again on the spatial frequency bandwidth: only if the normalization is performed before the resizing, the impact on spatial frequency bandwidth will be very limited, leaving the information content of the image quite the same.
This said, depending on the pre-processing of the images, we can calculate, or at least have an idea, of the dynamic qualities of the image we input in the network.
After been acquired as input by the first layer, the ‘features’ of the image will be dissected by the first layer so that each relevant base features stick to a particular node.
As the dynamic range of the image (both spatial and color chrominance, i.e. the information content) we could say to come from the sum of those features, we can imagine as if the dynamic of the image was spreaded on all those nodes. Going trough the other inner layers, walking to the output, progressively the sub features will match in combinations, so that this dynamic range is again aggregated, going to be focused on the output layer in the corresponding class.
For this reason doesn’t seem reasonable that the Learning Rate LR is the same along all the network, moving from the first layer to the last.

Roberto Moretti
3DSF LTD