Despite the great popularity in past, nowadays SIFT is becoming less and less popular. The reason is – CNN (convolutional neural network). This technology demonstrates incredible results in image recognition. So, lets summaries information about “ancient” algorithm and perspective technology, that could change our world at all.

We’ll start from one of the “kings” of image recognition algorithms – SIFT (scale-invariant feature transform). Most algorithms have problems with recognition and become unstable when we change image size, scale, rotate image for more than ten degrees. SIFT confidently copes with these gaps. We can highlight four basic steps which this algorithm is using:

- Building a pyramid of images, bringing images to one dimension, compute DoG;
- Search local extremes;
- Determining the neighborhood of a special point, compute gradients;
- Building a descriptor.

So, as we notice earlier, this algorithm invariant to scaling and rotations of images, and also has resistant to changes in lighting and the presence of noise. But these benefits have the other side: the computational complexity become enormous while using big codebooks. Also, if there are not enough dots or the shape is not permanent, there is a chance not to find the key points on the template. So, if we use small codebooks, we can’t guarantee the high precision of our system.

And then came CNN. We can divide them (neural networks) into 3 main classes:

- Retrieval Using Pre-trained CNN Models;
- Image Retrieval with Fine-Tuned CNN Models;
- Hybrid CNN-based Methods.

We can take pre-trained CNN models, which already have a high accuracy rate and create our own NN using fewer images from the dataset. For example AlexNet, GoogleNet, ResNet and etc. Also, there is a huge dataset called ImageNet, which we can use to train our own CNN projects and compare with global result of the correct recognition.

The next advantage of CNN is that if we want to adapt someone’s CNN, data scientists practice such a thing as fine tuning: the take off the last fully connected layers of the network, add there new ones and train the network again on new data. So, we can just tear off the last layers and use the output of the previous layers.

Also, we can use such things as convolutional auto associator, Contrastive Divergence (CD), sparse coding or deconvolutional networks (DNN) to improve the learning rate of our NN (or decrease it, if we are not enough experienced).

The next main class is fine-tuned CNN models. The outstanding representative is GoogLeNet. This DCNN won ImageNet competition in 2014 and set a record for best one-time results. This network is really enormous and very difficult. The most important thing in fine-tuned models is the strategy of transferring th training, his size and its similarity to the original dataset. If the fine tuning is successful, we can use our NN in new areas where it would be impossible to use them otherwise due to lack of data or time and cost constraints. This approach allows to achieve a significant increase in the average accuracy and efficiency of classification, for example, of medical images.

If we use hybrid CNN-based methods, we might know that they are similar to SIFT methods and often are less efficient than the single-pass methods. But if we needed to solve non-standard problems, for example, skin cancer detection, hypothetically, a hybrid model can be more successful than just a classical model. A good example of such a "hybrid" dataset is the ISIC dataset, which contains thousands of skin growth images along with information such as the patient's age and gender for each image.

To sum up, we see that convolutional neural networks dominate over SIFT methods. The accuracy is higher, tasks are wider, efficiency is higher. And it worthy adversary to SIFT models.