and kept at the transmitter side. In general, the overfitting of the first block starts from a baseline encoder. In one embodiment, the baseline encoder may be determined to be an encoder corresponding to a block of a previous image or video frame having same or nearby (e.g. adjacent) location with respect to the first block. The baseline encoder may be also determined based on similarity of the first block of the current image or video frame and a block of a previous image or video frame.

A neural network is defined by its topology or architecture (e.g., number and type of layers), and by its weights. It is assumed in this example that only the weights of the neural encoder network are changed during the optimization processes, but the topology may be changed as well, for example based on type of content in the relevant block of image. The neural network can be characterized or represented by a point in the weight space, where each dimension of this space is a weight of the neural network. The baseline network may be considered to be a point in weight space which is relatively close to the optimal points for all images, but not too close to any of those optimal points. By the optimization operation, the optimized neural network gets closer in weight space to the optimal neural network for the data on which it was optimized.

In images, video, audio, speech, text and other types of data, there is high structural correlation which means that there is lot of either spatial or temporal correlations, or both. This correlation means that two nearby blocks are likely to have similar content, and thus two neural networks optimized separately on these two blocks are likely to be close in weight space, or anyway closer than their distance from the baseline. Thus, optimizing for a second block (and subsequent blocks) by starting from the neural network optimized on the first block (or neighboring blocks) is likely to require much less training iterations to converge. Due to this the encoding speed may drastically increase.

If multiple neighboring blocks 302 (with respect to the current block 301 ) were used for overfitting encoders 305, the overfitting for the current block 301 can be started from a combination of the overfitted neural networks on those neighboring blocks 302. The combination may be an average of the weights 306, or any suitable neural network combining method.

In some situations, such as at objects boundaries, it may happen that the current block's 301 neighboring blocks are very dissimilar from the current block 301 , so the