Fig. 2 shows an adapted transformer block of an adapted neural network system which is an embodiment of the invention;

Fig. 3 is composed of Figs. 3(a), 3(b) and 3(c) which show respectively three adapted neural network systems which are embodiments of the invention, and employ the adapted transformer block of Fig. 2;

Fig. 4 shows the steps of a method according to the invention; and

Fig. 5 shows a computer system which can be used to implement the method of Fig. 4 to generate any of the neural network systems of Fig. 3.

Equivalent elements in different ones of the figures are labelled by the same reference numerals.

Detailed description

Referring to Fig. 1(a), the architecture is shown of a known neural network which may be trained to perform a first computational task. An example network of this form is described in more detail in Alexey Dosovitskiy, et al., "An image is worth 16x16 words: Transformers for image recognition at scale". In International Conference on Learning Representations (ICLR), 2021, the disclosure of which is incorporated herein by reference.

The neural network is configured to receive a data input, and to perform the first computational task on it to generate a data output.

The neural network includes an embedder layer 11 for receiving the data input and for generating from it embedded (encoded) data referred to as "tokens". The neural network includes a sequence of transformer blocks 12. The input to a first of the transformer blocks 12 is the output of the embedder layer 11. The input to each of the other transformer blocks 12 is the output of a preceding transformer block of the sequence, and the output of each transformer block 12 (except the last transformer block of the sequence) is the input to the next transformer block 12 of the sequence. The data output of the neural network is the output of the last transformer block of the sequence.

The embedder layer 11 may have a form which depends upon the data input. For example, if the data input includes one or more images, the embedder layer 11 may include one or more convolutional layers. In some cases, the data input may be a data for each of a plurality of different times. For example, it may be a sequence of images at corresponding times (e.g. a video) or a sound signal representing sound captured by a microphone at different times. In this case, the embedder layer 11 may generate tokens, for simultaneous processing by the first transformer block 12, which represent the data input from multiple times.