CLAIMS

What is claimed is:

1. A method of visual data transmission for network-based visual analysis, the method comprising:

obtaining, at an imaging device, sensor data relating to a scene;

extracting an intermediate deep feature from an intermediate layer of a deep learning model based on the sensor data, wherein the intermediate deep feature comprises a plurality of feature maps;

determining whether the plurality of feature maps are in a floating point format or in an integer format;

quantizing the plurality of feature maps to obtain a plurality of quantized feature maps, respectively, if the plurality of feature maps are determined to be in the floating point format;

producing video format data based on the plurality of feature maps or the plurality of quantized feature maps, wherein the video format data is produced based on the plurality of feature maps, without said quantizing the plurality of feature maps, if the plurality of feature maps are determined to be in the integer format, or based on the plurality of quantized feature maps if the plurality of feature maps are determined to be in the floating point format;

encoding the video format data using a video codec to produce encoded video data; and

transmitting the encoded video data to a visual analysis device for performing visual analysis based on the encoded video data.

2. The method according to claim 1, wherein said producing video format data comprises repacking the plurality of feature maps based on a repacking technique to produce the video format data.

3. The method according to claim 2, wherein the repacking technique is based on channel concatenation or channel tiling.

4. The method according to claim 3, wherein

the repacking technique is based on said channel concatenation, said channel concatenation comprising determining a plurality of inter-channel distances associated with the plurality of feature maps, each inter-channel distance being associated with a pair of feature maps of the plurality of feature maps, and

said repacking the plurality of feature maps comprising forming a plurality of repacked feature maps by ordering the plurality of feature maps based on the plurality of inter-channel distances determined to produce the video format data comprising the plurality of repacked feature maps.

5. The method according to claim 3, wherein

the repacking technique is based on said channel tiling, said channel tiling comprising forming one or more repacked feature maps based on the plurality of feature maps to produce the video format data comprising the one or more repacked feature maps, each repacked feature map being an enlarged feature map.

6. The method according to claims 1 to 5, wherein the plurality of feature maps are quantized based on a uniform quantization technique, a logarithmic quantization technique or a learning-based adaptive quantization technique.

7. A method of network-based visual analysis, the method comprising:

receiving, at a visual analysis device, encoded video data from an imaging device configured to obtain sensor data relating to a scene;

producing decoded video data based on the encoded video data, wherein said producing decoded video data comprises decoding the encoded video data using a video codec to produce the decoded video data comprising video format data, wherein the video format data comprises a plurality of feature maps;

determining whether the plurality of feature maps are based on a plurality of original feature maps in a floating point format or in an integer format; and

de-quantizing the plurality of feature maps to obtain a plurality of de-quantized feature maps, respectively, if the plurality of feature maps are determined to be based on the plurality of original feature maps in the floating point format,

producing an intermediate deep feature of a deep learning model based on the plurality of feature maps or the plurality of de-quantized feature maps, wherein the intermediate deep feature is produced based on the plurality of feature maps, without

said de-quantizing the plurality of feature maps, if the plurality of feature maps are determined to be based on the plurality of original feature maps in the integer format, or based on the plurality of de-quantized feature maps if the plurality of feature maps are determined to be based on the plurality of original feature maps in the floating point format; and

performing visual analysis based on the intermediate deep feature.

8. The method according to claim 7, the video format data is de-repacked based on a de-repacking technique to produce a plurality of de-repacked feature maps.

9. The method according to claim 8, wherein the de-repacking technique is based on channel de-concatenation or channel de-tiling.

10. The method according to claim 9, wherein

the video format data comprises a plurality of repacked feature maps, and

the de-repacking technique is based on said channel de-concatenation, said channel de-concatenation comprising sorting the plurality of repacked feature maps based on repacking supplemental information to produce the plurality of de-repacked feature maps.

11. The method according to claim 9, wherein

the video format data comprises one or more repacked feature maps; and

the de-repacking technique is based on said channel de-tiling, said channel de-tiling comprising forming the plurality of de-repacked feature maps based on the one or more repacked feature maps, each de-repacked feature map being a diminished feature map.

12. The method according to claims 7 to 11, wherein the plurality of feature maps are de-quantized based on a uniform de-quantization technique, a logarithmic de-quantization technique or a learning-based adaptive de-quantization technique.

13. An imaging device for visual data transmission for network-based visual analysis, the imaging device comprising:

a memory; and

at least one processor communicatively coupled to the memory and configured to perform the method of visual data transmission for network-based visual analysis according to any one of claims 1 to 6.

14. A visual analysis device for network-based visual analysis, the visual analysis device comprising:

a memory; and

at least one processor communicatively coupled to the memory and configured to perform the method of network-based visual analysis according to any one of claims 7 to 12.

15. A network-based visual analysis system, the network-based visual analysis system comprising:

one or more imaging devices, each imaging device being configured for visual data transmission for network-based visual analysis according to claim 13; and

a visual analysis device for network-based visual analysis configured according to claim 14 and configured to receive encoded video data from the one or more imaging devices, respectively.

16. A computer program product, embodied in one or more non-transitory computer-readable storage mediums, comprising instructions executable by at least one processor to perform a method of visual data transmission for network-based visual analysis according to any one of claims 1 to 6.

17. A computer program product, embodied in one or more non-transitory computer-readable storage mediums, comprising instructions executable by at least one processor to perform a method of network-based visual analysis according to any one of claims 7 to 12.