

CLAIMS

What is claimed is:

- 5 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;
- 10 producing encoded video data based on the intermediate deep feature; 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 the encoded video data is produced
- 15 based on a video codec.
3. The method according to claim 2, wherein
- the intermediate deep feature comprises a plurality of feature maps,
- the method further comprises producing video format data based on the plurality
- 20 of feature maps, and
- said producing encoded video data comprises encoding the video format data using the video codec to produce the encoded video data.
4. The method according to claim 3, wherein said producing video format data
- 25 comprises repacking the plurality of feature maps based on a repacking technique to produce the video format data.
5. The method according to claim 4, wherein the repacking technique is based on channel concatenation or channel tiling.
- 30 6. The method according to claim 5, 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

5 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.

7. The method according to claim 5, wherein
10 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.

8. The method according to any one of claims 3 to 7, further comprising quantizing
15 the plurality of feature maps to obtain a plurality of quantized feature maps, respectively, wherein the video format data is produced based on the plurality of quantized feature maps.

9. The method according to any one of claims 3 to 7, further comprising:
20 determining whether the plurality of feature maps are in a floating point format or in an integer format; and
 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,
25 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.

30 10. The method according to claim 8 or 9, 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.

11. 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;
5 producing an intermediate deep feature of a deep learning model based on the
decoded video data; and
performing visual analysis based on the intermediate deep feature.
12. The method according to claim 11, wherein said producing decoded video data
10 comprises decoding the encoded video data using a video codec to produce the decoded
video data comprising video format data.
13. The method according to claim 12, wherein the intermediate deep feature
comprises a plurality of feature maps.
- 15 14. The method according to claim 13, wherein
said producing an intermediate deep feature comprises de-repacking the video
format data based on a de-repacking technique to produce a plurality of de-repacked
feature maps, and
20 the intermediate deep feature is produced based on the plurality of de-repacked
feature maps.
15. The method according to claim 14, wherein the de-repacking technique is based
on channel de-concatenation or channel de-tiling.
- 25 16. The method according to claim 15, 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
30 based on repacking supplemental information to produce the plurality of de-repacked
feature maps.
17. The method according to claim 15, 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.

5

18. The method according to any one of claims 14 to 17, further comprising de-quantizing the plurality of de-repacked feature maps to obtain a plurality of de-quantized feature maps, respectively, wherein the intermediate deep feature is produced based on the plurality of de-quantized feature maps.

10

19. The method according to any one of claims 14 to 17, further comprising:
determining whether the plurality of de-repacked 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 de-repacked feature maps to obtain a plurality of
15 de-quantized feature maps, respectively, if the plurality of de-repacked feature maps are determined to be based on the plurality of original feature maps in the floating point format,

wherein the intermediate deep feature is produced based on the plurality of de-repacked feature maps, without said de-quantizing the plurality of de-repacked feature
20 maps, if the plurality of de-repacked 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 de-repacked feature maps are determined to be based on the plurality of original feature maps in the floating point format.

25 20. The method according to claim 18 or 19, wherein the plurality of de-repacked 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.

30 21. 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 10.

- 5 22. 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
10 11 to 20.

23. 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
15 data transmission for network-based visual analysis according to claim 21; and
a visual analysis device for network-based visual analysis configured according to claim 22 and configured to receive encoded video data from the one or more imaging devices, respectively.

- 20 24. 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 10.

- 25 25. 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 11 to 20.