

NEURAL NETWORK SYSTEM 100

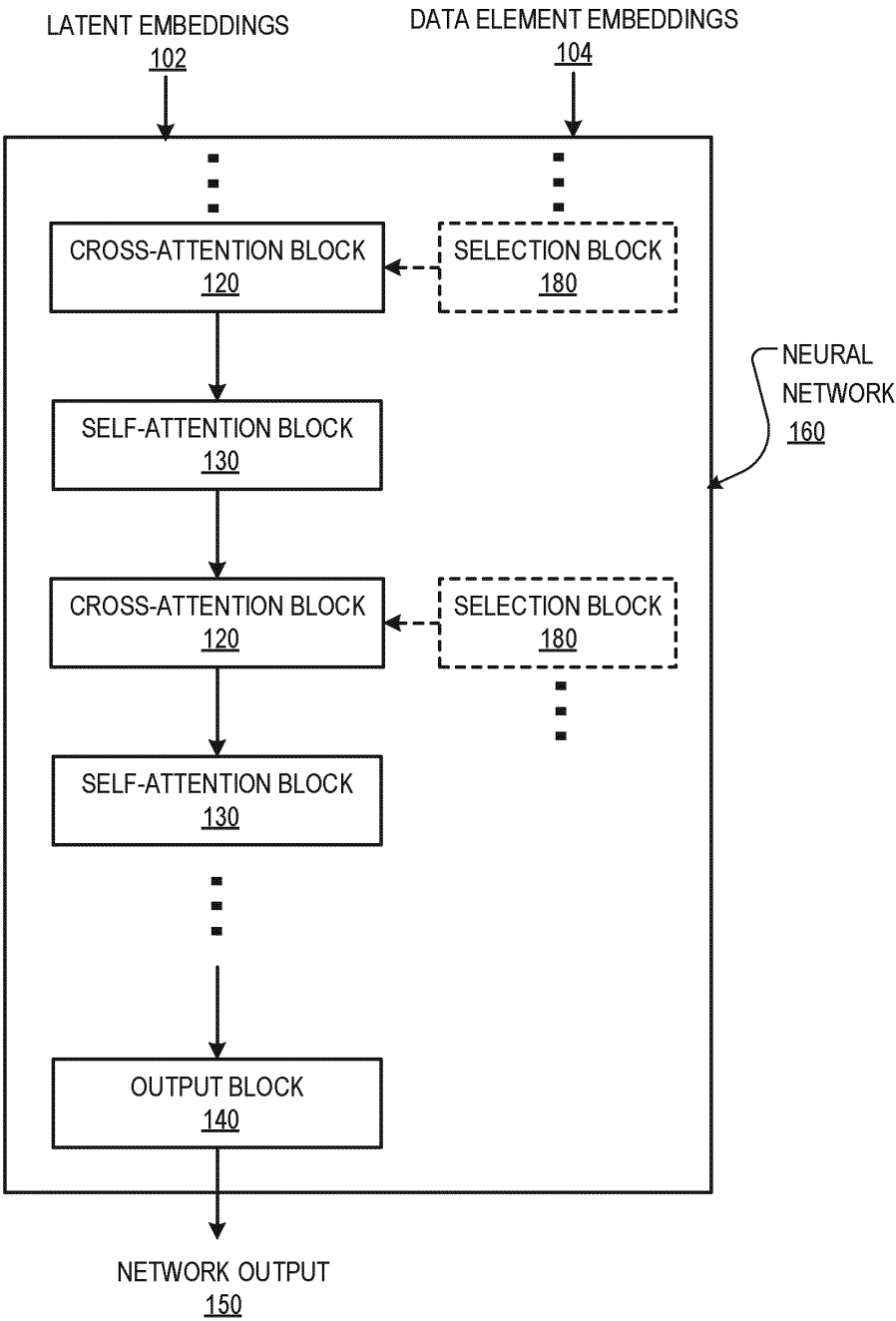


FIG. 1

NEURAL NETWORK IN MORE DETAIL 200

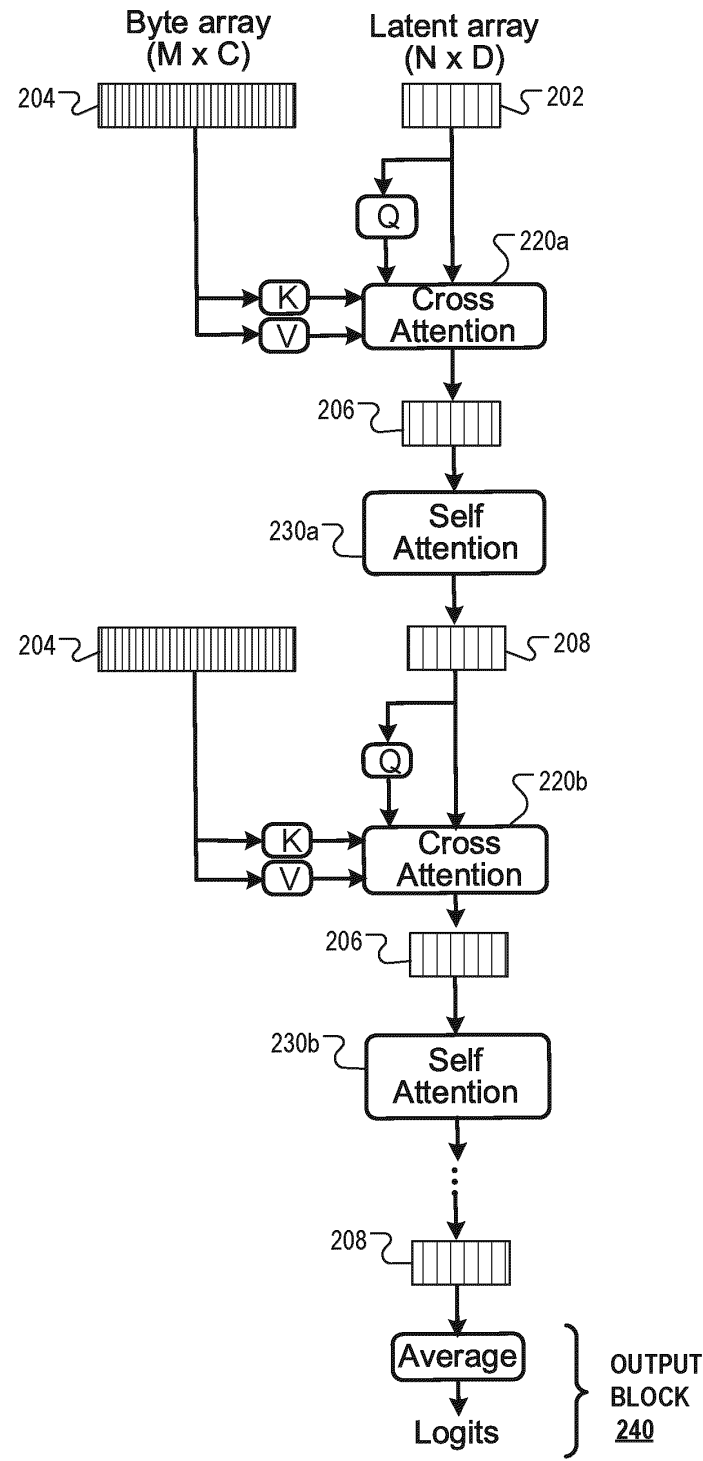


FIG. 2

SELECTION BLOCK 300

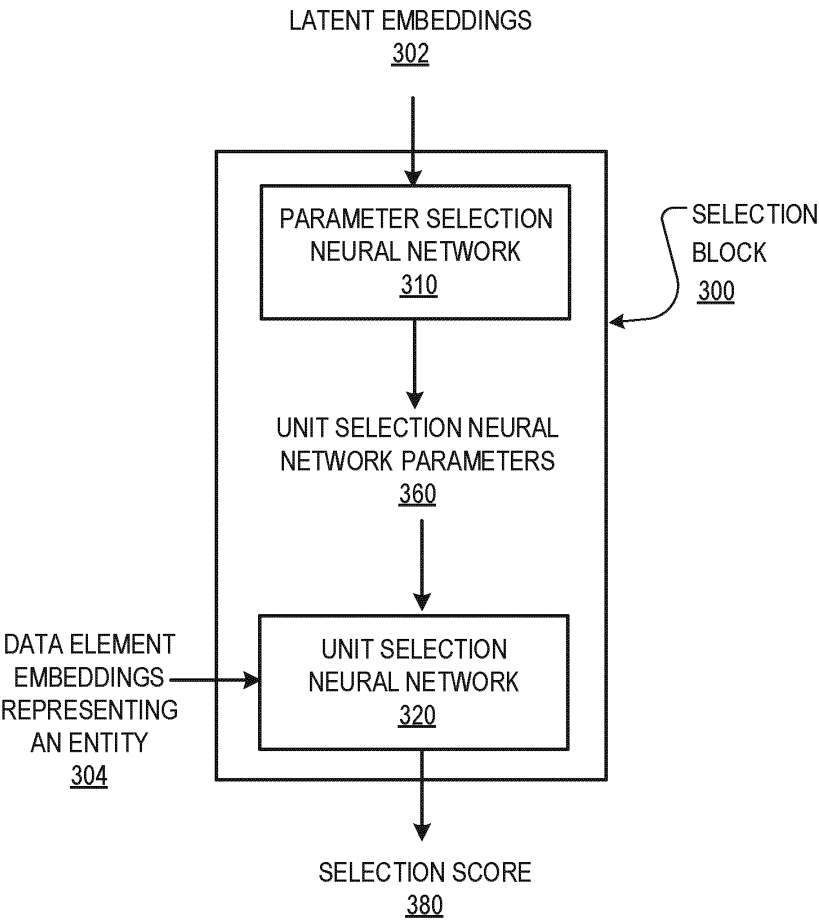


FIG. 3

USING A NEURAL NETWORK TO GENERATE A NETWORK OUTPUT CHARACTERIZING AN ENTITY

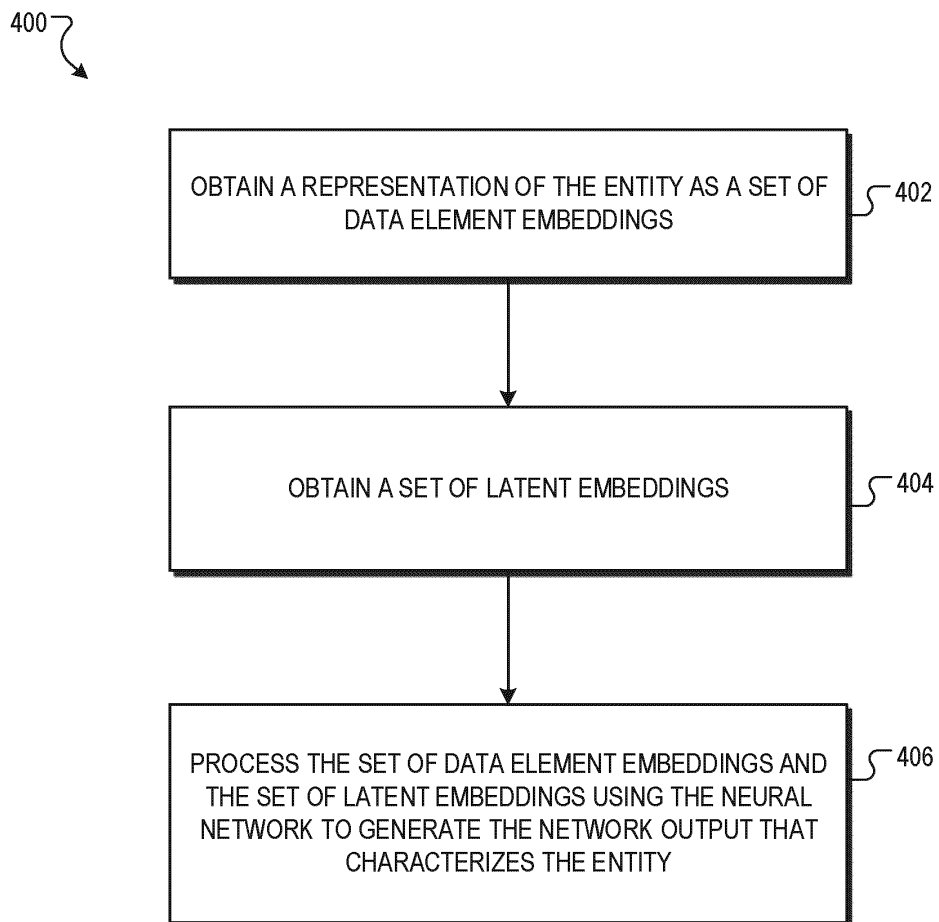


FIG. 4

EXAMPLE OF ENTITIES AND UNITS 500

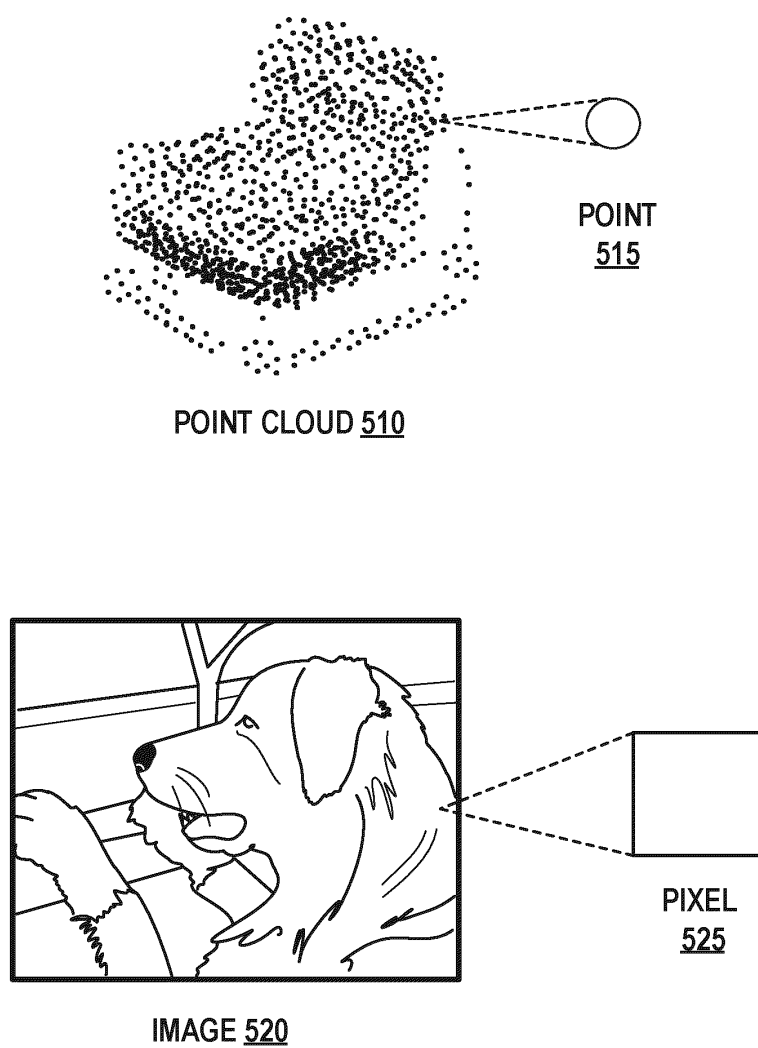


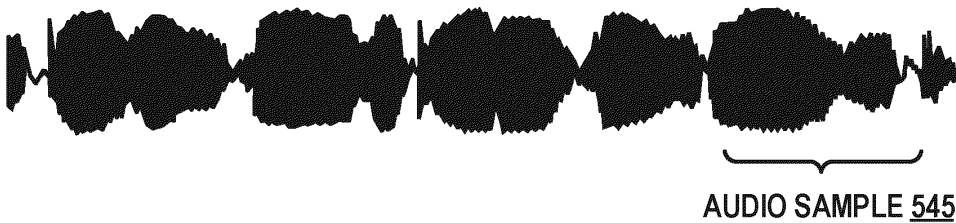
FIG. 5A

ANOTHER EXAMPLE OF ENTITIES AND UNITS 500

SEQUENCE OF WORDS 530

sapere aude
WORD 535

AUDIO WAVEFORM 540



PROTEIN 550

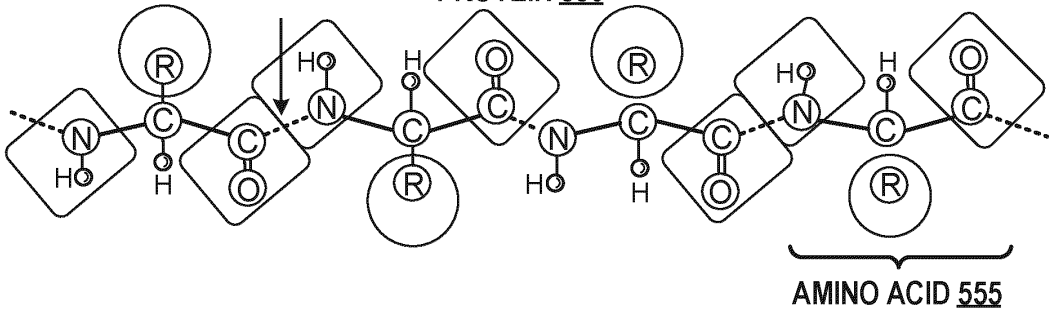


FIG. 5B

EXAMPLE ATTENTION MAPS 600

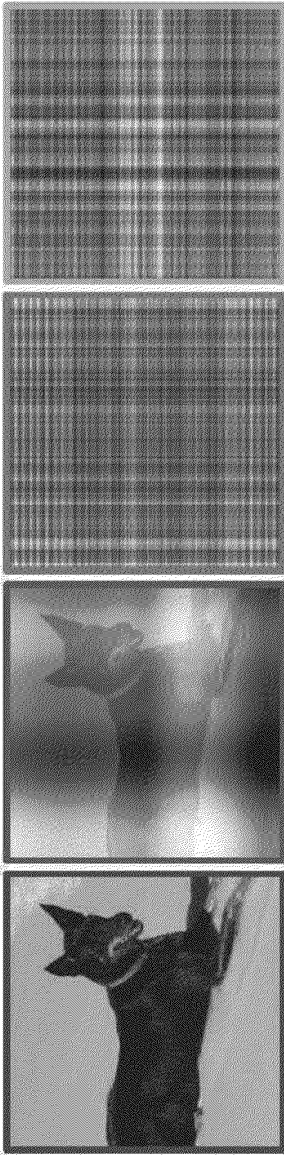


FIG. 6A

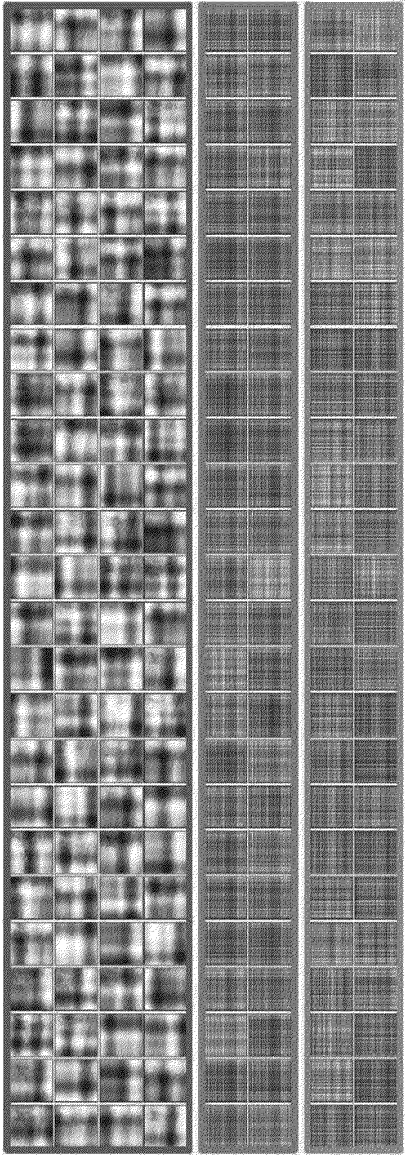


FIG. 6B

**EXAMPLE CONFIGURATIONS OF A NEURAL
NETWORK SYSTEM 700**

# cross-attends	Acc.	FLOPs	Params
1 (at start)	76.7	404.3B	41.1M
1 (interleaved)	76.7	404.3B	42.1M
2 (at start)	76.7	447.6B	44.9M
2 (interleaved)	76.5	447.6B	44.9M
4 (at start)	75.9	534.1B	44.9M
4 (interleaved)	76.5	534.1B	44.9M
8 (at start)	73.7	707.2B	44.9M
8 (interleaved)	78.0	707.2B	44.9M

FIG. 7

EXAMPLE PARAMETERS OF A NEURAL NETWORK SYSTEM 800

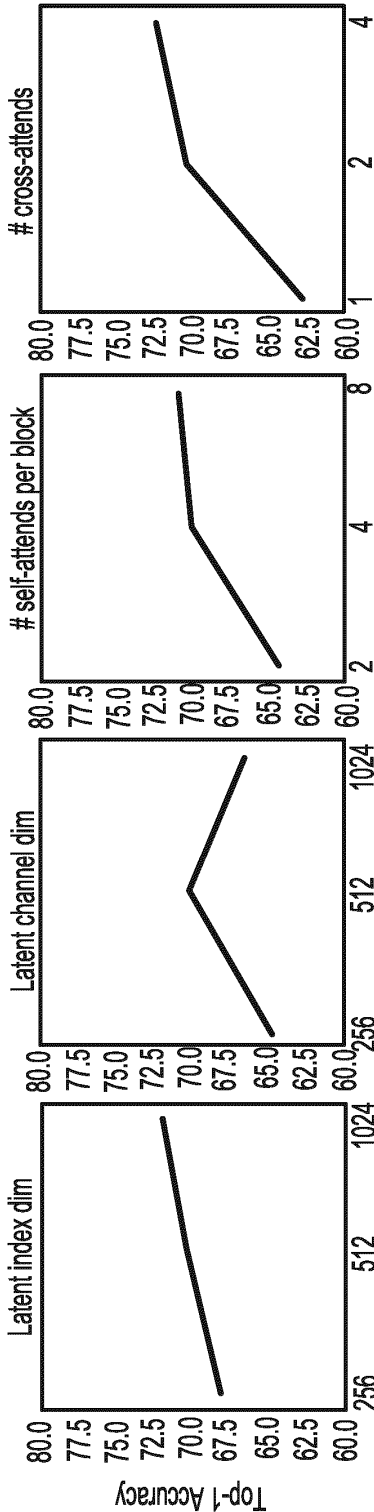


FIG. 8A

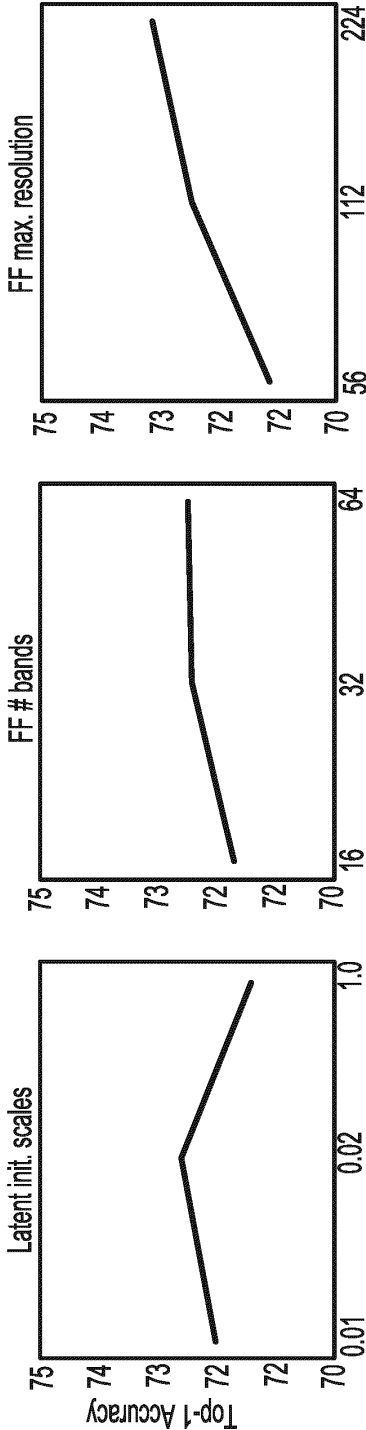


FIG. 8B

910 ↘

ResNet-50 (He et al., 2016)	77.6
ViT-B-16 (Dosovitskiy et al., 2021)	77.9
ResNet-50 (FF)	73.5
ViT-B-16 (FF)	76.7
Transformer (64x64, FF)	57.0
Perceiver (FF)	78.0

920 ↘

	Raw	Perm.	Input RF
ResNet-50 (FF)	73.5	39.4	49
ViT-B-16 (FF)	76.7	61.7	256
Transformer (64x64) (FF)	57.0	57.0	4,096
Perceiver:			
(FF)	78.0	78.0	50,176
(Learned pos.)	70.9	70.9	50,176

FIG. 9

1010

	Accuracy
PointNet++ (Qi et al., 2017)	91.9
ResNet-50 (FF)	66.3
ViT-B-2 (FF)	78.9
ViT-B-4 (FF)	73.4
ViT-B-8 (FF)	65.3
ViT-B-16 (FF)	59.6
Transformer (44x44)	82.1
Perceiver	85.7

1020

	Audio	Video	A+V
Benchmark (Gemmeke et al., 2017)	3.14	-	-
Attention (Kong et al., 2018)	32.7	-	-
Multi-level Attention (Yu et al., 2018)	36.0	-	-
ResNet-50 (Ford et al., 2019)	38.0	-	-
CNN-14 (Kong et al., 2020)	43.1	-	-
CNN-14 (no balancing & no mixup) (Kong et al., 2020)	37.5	-	-
G-blend (Wang et al., 2020c)	32.4	18.8	41.8
Attention AV-fusion (Fayek & Kunar, 2020)	38.4	25.7	46.2
Perceiver (raw audio)	38.3	25.8	43.5
Perceiver (mel spectrogram)	38.4	25.8	43.2
Perceiver (mel spectrogram - tuned)	-	-	44.2

FIG. 10