

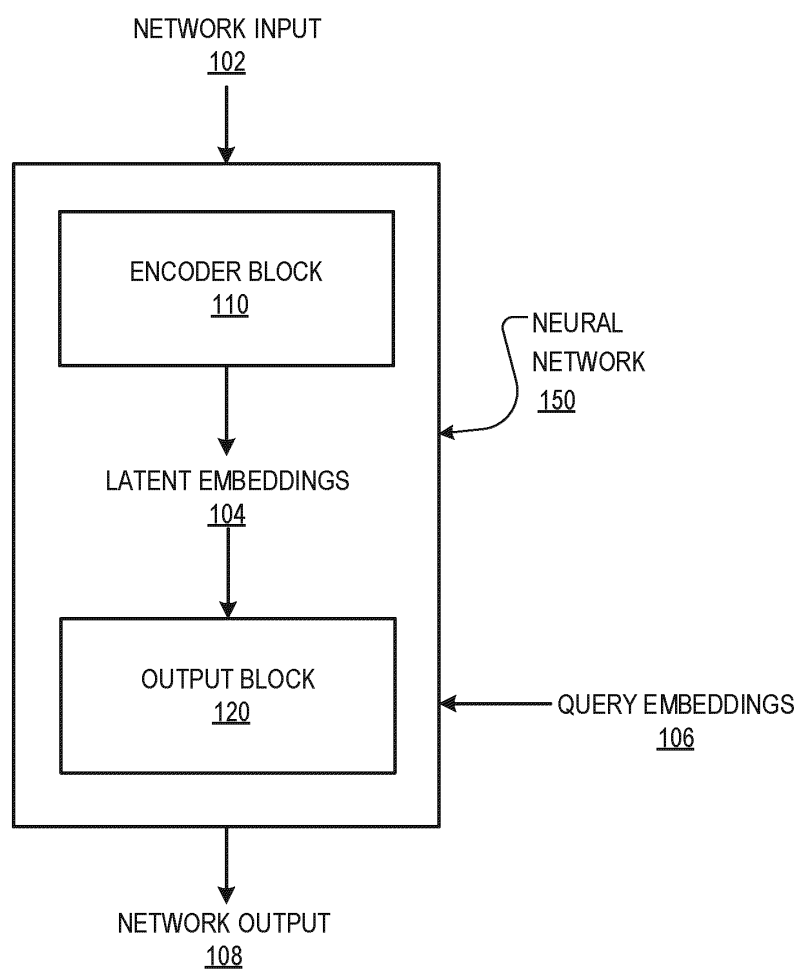
NEURAL NETWORK SYSTEM 100

FIG. 1

NEURAL NETWORK SYSTEM IN MORE DETAIL 100

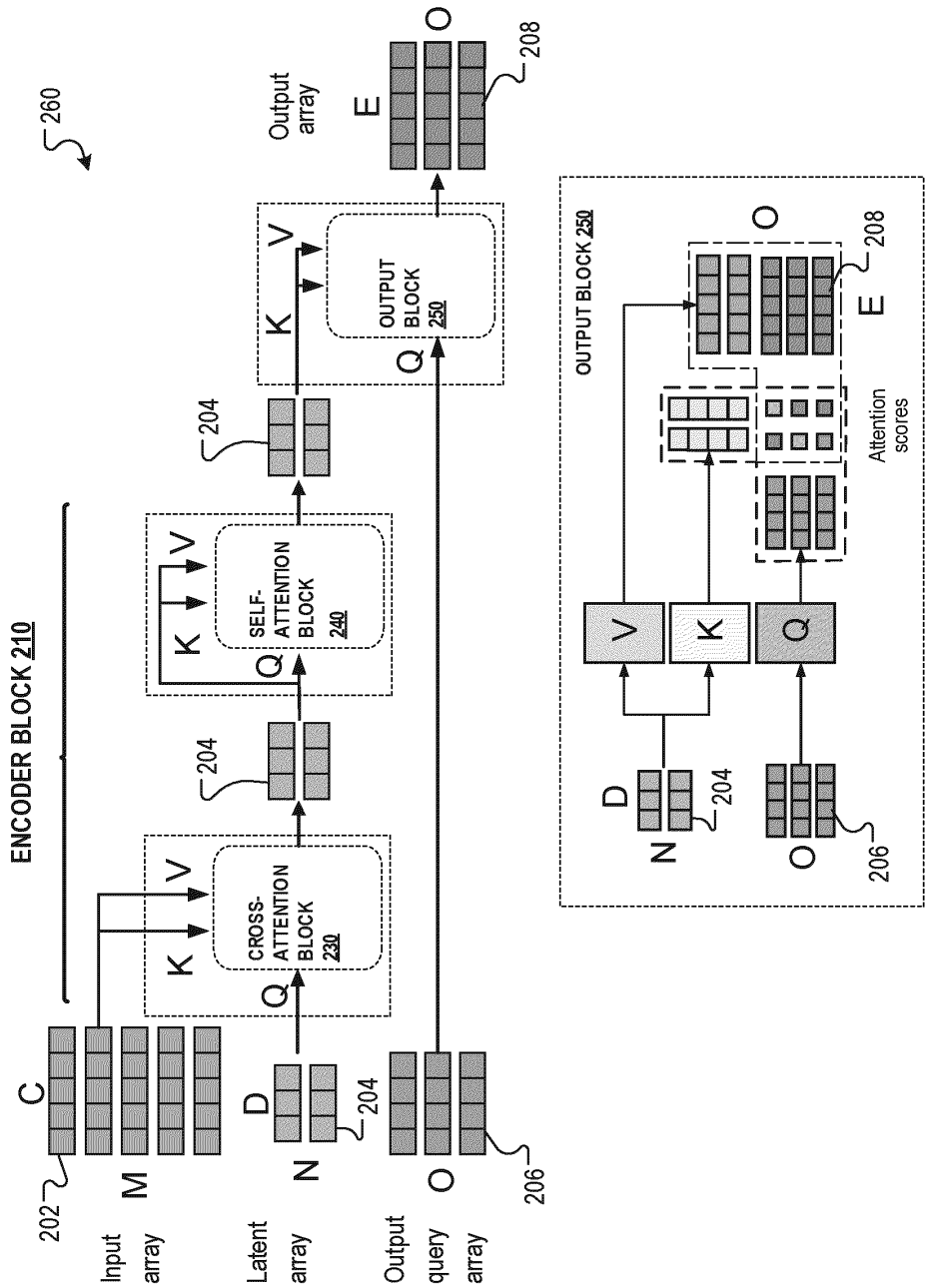


FIG. 2

USING A NEURAL NETWORK TO GENERATE EACH DIMENSION OF A NEURAL NETWORK OUTPUT

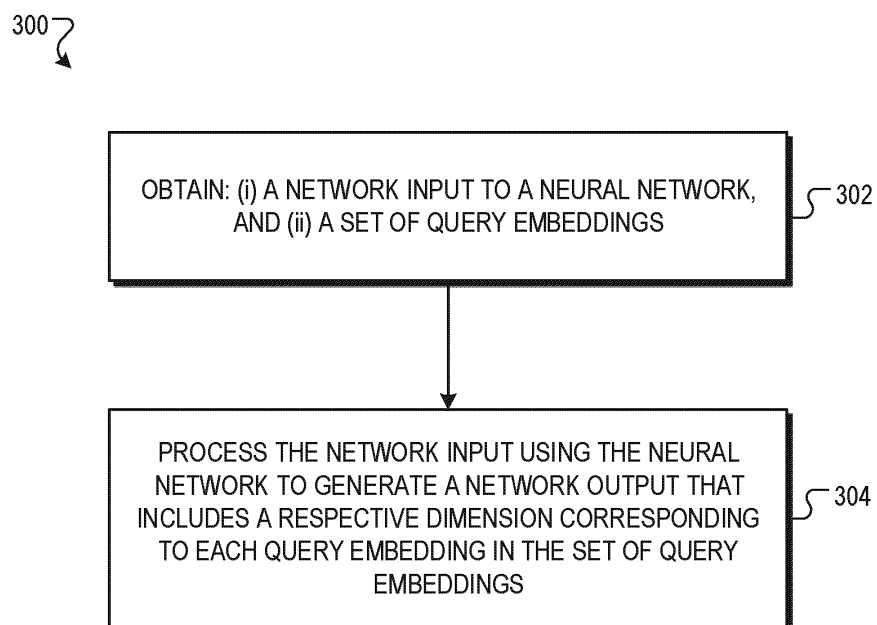


FIG. 3

EXAMPLE QUERY EMBEDDINGS 400

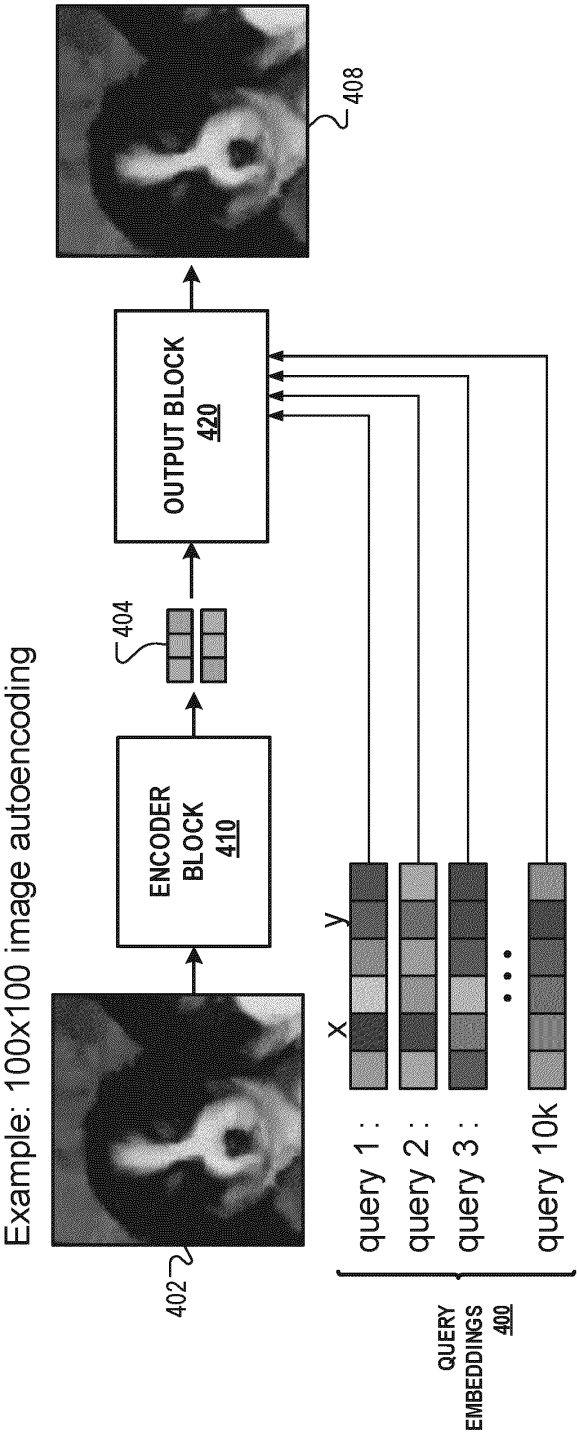


FIG. 4

EXAMPLE QUERY EMBEDDINGS 500

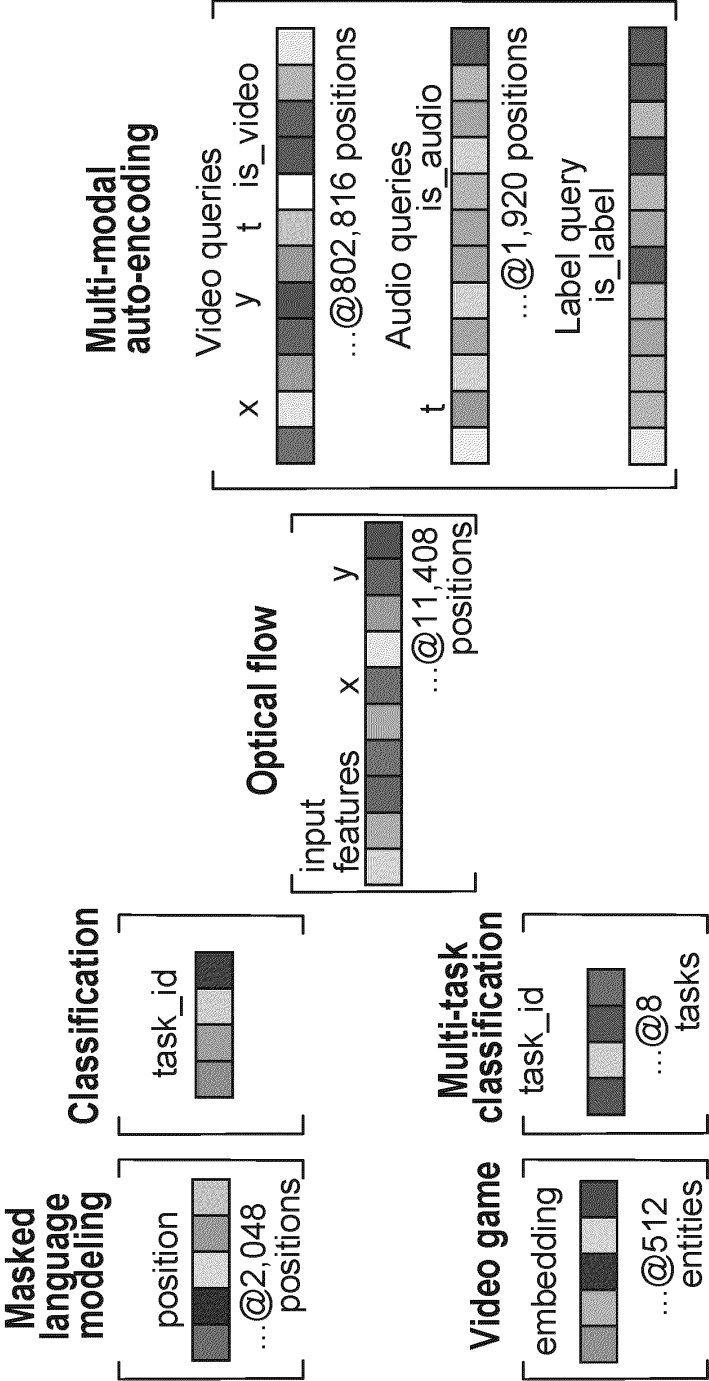


FIG. 5

EXAMPLE MODALITIES 600

Modalities	Tasks	Preprocessing	Postprocessing	# Inputs	# Outputs
Text	Token-level pred.	Tokenization + Embed.	Linear projection	512 x 768	512 x 768
Text	Byte-level pred.	Embed.	None	2,048 x 768	2,048 x 768
Text	Multi-task (8 tasks)	Embed.	None	2,048 x 768	8 x 768
Video	Flow prediction	None	None	365,056 x 64	182,528 x 64
Video	Flow prediction	Concat	None	182,528 x 64	182,528 x 64
Video	Flow prediction	Conv+maxpool	RAFT upsampling	22,816 x 64	11,408 x 64
Video	Flow prediction	Conv+maxpool+conca	RAFT upsampling	11,408 x 64	11,408 x 64
Video+Audio+Label	Autoencoding	Patch: 1x4x4 Vid, 16 Aud	None	50,657 x 704	803,297 x 512
Image	Classification	None	None	50,176 x 3	1 x 1,000
Image	Classification	Linear projection	None	50,176 x 256	1 x 1,000
Image	Classification	Conv+maxpool	None	3,136 x 64	1 x 1,000
Video Game Unit Set	Encoding and Classification	Tokenization	Pointer network	512 x 256	512 x 128
Video+Audio	Classification	Patch: 2 x 8 x 8 Vid, 128 Aud	None	13,024 x 487	1 x 527
Video+Audio	Classification	Patch: 2 x 8 x 8 Vid, Aud → mel-spectrogram	None	17,344 x 487	1 x 527

FIG. 6

EXAMPLE MULTIMODAL RESULTS 700

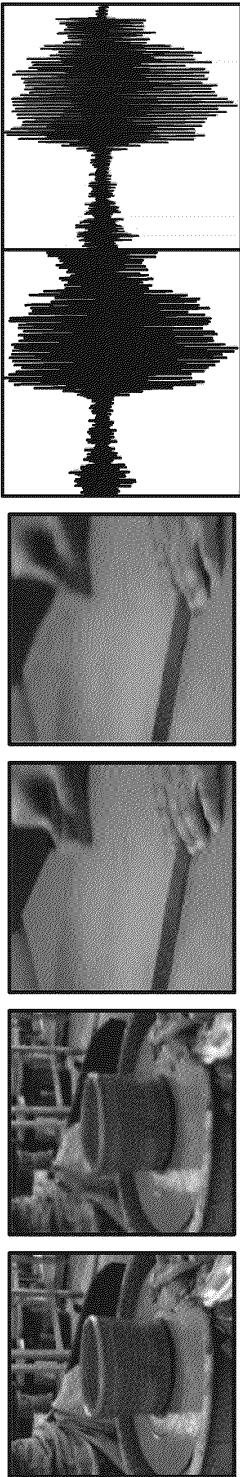


FIG. 7A

Compression Ratio	Audio PSNR	Video PSNR	Top-1 Accuracy
88x	26.97	24.37	10.2%
176x	25.33	24.27	8.6%
352x	14.15	23.21	11.5%

FIG. 7B